Early detection of physical fatigue in industry using wearable sensors and contextual modeling

Safety Science 194 (2026) 107041

Contents lists available at ScienceDirect

Safety Science

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

Early detection of physical fatigue in industry using wearable sensors and contextual modeling

Carlos Albarrán Morillo [a,*], Huxiao Shi [a], John Suárez-Pérez [a], Micaela Demichela [b]

[a] Applied Science and Technology Department, Politecnico di Torino, Turin, Italy
[b] Escuela de Ingenieria y Ciencias, Tecnologico de Monterrey, Mexico

A R T I C L E I N F O

Keywords:
Occupational safety
Workplace physical fatigue
Wearable sensors
Physiological monitoring
Machine learning
Fuzzy logic
Industry 5.0

A B S T R A C T

Physical fatigue in repetitive production lines contributes to musculoskeletal disorders and absenteeism. This study investigates a pharmaceutical packaging environment in Colombia with 43 operators (42 female; 19–53 years) performing repetitive inspection and packing. Smartwatches captured pulse rate, electrodermal activity, skin temperature, and motion, complemented by demographic (age, experience) and occupational factors (task load, line, shift, timing). Principal Component Analysis (PCA) reduced dimensionality, and a fuzzy logic–based labeling method—adapted from prior controlled experiments—generated binary and four-class fatigue labels without mid-shift self-reports. These labeled datasets were used to train multiple machine-learning classifiers. Integrating contextual features with biometrics substantially improved performance: in binary classification, F1 increased from 0.8848 (biometrics only) to 0.9375; in four-level classification, F1 rose from 0.8232 to 0.8793. Motion-related metrics emerged as the most informative predictors. Critically, feature integration improved reliability: accuracy for intermediate states (Higher Non-Fatigue and Higher Fatigue) rose by ~10 percentage points, while false negatives in the Pure Fatigue class were eliminated—3% of cases previously misclassified as Higher Non-Fatigue were instead correctly mapped within the fatigue spectrum. This shift strengthens the system's effectiveness for real-time safety interventions. The novelty of this work lies in combining biometric and contextual modeling to reduce false negatives in critical fatigue states, providing a scalable, non-intrusive, and human-centered early-warning system. By aligning with Industry 5.0, this approach demonstrates how wearable and contextual data can jointly support proactive and trustworthy safety interventions while maintaining operational flow.

1. Introduction

Industry 4.0 integrates advanced technologies such as the Internet of Things (IoT), artificial intelligence (AI), and robotics to enhance industrial automation and efficiency (Ahmmed et al., 2024). These technologies have significantly improved production processes, optimizing workflows and reducing operational costs. However, despite these advancements, human workers remain a critical component of industrial systems (Mital and Pennathur, 2004). Traditional quality-focused approaches often overlook human factors, such as worker fatigue and ergonomics, which are essential for maintaining safety and productivity in industrial environments (Trstenjak et al., 2025). Neglecting these factors has been linked to increased accidents, a higher risk of injuries, and potential quality issues in production (Breese, 2024). In response, the emerging Industry 5.0 framework emphasizes a human-centered approach, blending human creativity with smart automation to

Abbreviations: AI, Artificial Intelligence; BR, Breathing Rate; CDFs, Cumulative Distribution Functions; CFS, Chronic Fatigue Syndrome; CTS, Carpal Tunnel Syndrome; EDA, Electrodermal Activity; EEG, Electroencephalography; EMG, Electromyography; EPC1, External Principal Component 1; EPC2, External Principal Component 2; EPC3, External Principal Component 3; EPC4, External Principal Component 4; HR, Heart Rate; HRV, Heart Rate Variability; IMU, Inertial Measurement Unit; IoT, Internet of Things; IPC1, Internal Principal Component 1; IPC2, Internal Principal Component 2; KNN, K-Nearest Neighbors; ML, Machine Learning; MSDs, Musculoskeletal Disorders; NN, Neural Network; PCA, Principal Component Analysis; PDFs, Probability Density Functions; PCs, Principal Components; RF, Random Forest; RPE, Rating of Perceived Exertion; SVM, Support Vector Machine; SMOTE, Synthetic Minority Over-sampling Technique; SNS, Sympathetic Nervous System; Temp., Temperature; XGBoost, eXtreme Gradient Boosting; 3D, Three-Dimensional; LightGBM, Light Gradient Boosting Machine.

* Corresponding author.
E-mail addresses: carlos.albarran@polito.it (C.A. Morillo), huxiao.shi@polito.it (H. Shi), jf.suarez@tec.mx (J. Suárez-Pérez), micaela.demichela@polito.it (M. Demichela).

https://doi.org/10.1016/j.ssci.2025.107041
Received 2 September 2025; Received in revised form 13 October 2025; Accepted 26 October 2025
Available online 1 November 2025

prioritize worker well-being, operational performance, and sustainability (Verma, 2024; Ruiz-de-la-Torre et al., 2023).

Physical fatigue is a widespread concern in industries involving repetitive tasks such as packaging, lifting, pushing, and assembly-line work (Panneerselvam et al., 2023). Over time, fatigue contributes to exhaustion, increased human errors, and workplace accidents (Sieber et al., 2022). It also impairs cognitive and motor functions, elevating the risk of incidents such as slips, falls, and injuries (Parijat and Lockhart, 2009). Prolonged exposure to fatigue not only affects worker performance but also leads to long-term health issues, including chronic fatigue syndrome (CFS), musculoskeletal disorders (MSDs), and weakened immune function (Fukuda, 1994; Andersen et al., 2007).

The global prevalence of MSDs underscores the critical need for effective workplace prevention strategies. According to the World Health Organization, approximately 1.71 billion people worldwide are affected by MSDs, making them the leading contributor to disability globally (World Health Organization, 2023). Among these disorders, carpal tunnel syndrome (CTS), linked to repetitive hand movements, has been identified as a major occupational health concern (Szabo, 1998). In Colombia, MSDs have been a significant concern in occupational health. According to a study analyzing data from 2009 to 2013, 88 % of reported occupational diseases were MSDs (Guzmán-Velasco and Diago-Franco, 2019). Notably, women are disproportionately affected, with a prevalence ratio of eight-to-one compared to men (Quiroga et al., 2022). The most impacted age group is 41–60 years, but an increasing number of cases have been reported among younger workers aged 16–25 in recent years (Quiroga et al., 2022). Studies have also identified that workers in the packaging area of pharmaceutical production lines are the most affected by ergonomic risk factors, with 69.8 % of operators in the hand/wrist conditioning area reporting pain-related symptoms (Ferrerosa et al., 2015). These findings underscore the urgent need for improved fatigue monitoring and prevention strategies.

Despite its impact, fatigue assessment remains challenging due to its subjective nature, as it is influenced by individual health conditions, job demands, and personal circumstances (Bazazan et al., 2023). Traditional subjective fatigue evaluation methods, such as self-reported questionnaires, are widely used but often suffer from bias and variability (Gawron, 2016). A more objective approach involves physiological measurements linked to the sympathetic nervous system (SNS), such as heart activity, blood parameters, and electrodermal responses, which provide quantifiable fatigue indicators (Valenza et al., 2018). When a person engages in physically demanding tasks, the SNS responds by increasing heart rate, elevating sweat gland activity (reflected in electrodermal activity), and altering thermoregulation, among other responses. These physiological markers are not only objective indicators of fatigue accumulation but also align with well-established theoretical models linking fatigue to homeostatic imbalances and autonomic stress regulation (Behrens et al., 2023; Fernández et al., 2025; Raizen et al., 2023), positioning them as reliable alternatives to subjective self-assessment. Recent advances in wearable and wireless sensor technologies now make it possible to monitor these signals in real-time, non-intrusively, and continuously, making them particularly well-suited for dynamic industrial environments where early fatigue detection is critical for safety and performance (Moon and Ju, 2024).

However, many existing fatigue classification models primarily rely on biometric data or subjective assessments, overlooking critical occupational and demographic factors that significantly influence physical fatigue. This gap can lead to less accurate and less generalizable models, limiting their effectiveness in real-world industrial settings. To address this, our study proposes an integrated approach that combines biometric data with demographic and occupational factors. This comprehensive model aims to enhance fatigue detection accuracy, reduce bias, and improve the robustness of physical fatigue classification in industrial contexts.

This study focuses on a pharmaceutical packaging area in a company in Colombia, where workers perform repetitive upper limb movements as part of their daily tasks. These movements involve short work cycles that require repeated muscle exertion, engaging muscle groups, bones, joints, tendons, ligaments, and nerves. Over time, this repetitive strain can lead to MSD-related complications. In this context, we analyzed biometric data collected via smartwatches, including electrodermal activity (EDA), internal temperature, and pulse rate, to monitor physiological responses associated with physical fatigue and motion data. Additionally, demographic factors such as age and work experience, task-related factors like physical load (measured by the number of drugs packed per box), and operational variables such as production line location, day of the week, work shift, and shift timing were considered. Ambient temperature, noise, and humidity were monitored with an IoT sensor system installed in the packaging area. Readings remained within a very narrow range and fully complied with Colombian standards throughout data collection: noise levels were consistently below 85 dB (aligned with OSHA action levels and Colombian limits), ambient temperature remained between 20–25 °C (in line with USP controlled room temperature standards), and relative humidity stayed within 40–55 % (within the ASHRAE 55 recommended 30–60 % range). We considered including these variables as predictors; however, they showed little to no within-shift variation over the 20-minute sessions and were measured at the area level rather than per worker—providing insufficient discriminative signal and risking overfitting. Accordingly, we used them solely as quality-control checks. In environments with greater fluctuation—or when personal/workstation-level environmental sensors are available and time-aligned with the wearable windows—we would incorporate these factors and assess their incremental value via nested model comparisons.

The remainder of this paper is structured as follows. Section 2 identifies critical gaps in the literature, emphasizing the common reliance on biometric signals without considering demographic and occupational factors, and highlights the limitations of subjective labeling methods, which can introduce bias and reduce model generalizability. Section 3 describes the data collection campaign and the data analysis process, including principal component analysis (PCA) and fuzzy logic techniques before the machine learning (ML) classification algorithms. Section 4 presents the results for binary and four-level physical fatigue classification, followed by Section 5, which discusses these findings. Finally, Section 6 provides the conclusion of this study and outlines potential directions for future work.

2. State of the art in physical fatigue detection

Physical fatigue detection has become a critical area of research in recent years, driven by the need to improve worker safety. Numerous studies have investigated various methods for detecting physical fatigue, with a particular emphasis on biometric data. While some approaches rely on more intrusive techniques—such as electromyography (EMG), blood biomarkers, or electroencephalography (EEG)—these methods typically require electrodes, skin preparation, or laboratory settings, making them less practical for continuous use in production environments. In this work, we chose to focus on non-intrusive, real-world applicable wearable methods, such as heart rate, heart rate variability, EDA, motion, and temperature monitoring. These modalities can be embedded in smartwatches or lightweight sensors that operators can realistically use during long shifts without interfering with their tasks. Table 1 summarizes some of the most influential studies in this field, including their participants, metrics, labeling methods, fatigue levels, best algorithms, and performance outcomes.

Several key approaches have emerged in recent years, each focusing on different biometric signals and machine learning algorithms for fatigue classification. For example, HRV-based methods have achieved strong performance in construction (Anwer et al., 2023) and in larger experimental cohorts using subjective Borg-scale labeling (Ni et al., 2022), while combinations of heart rate, breathing rate, temperature, and limited personal characteristics have also been tested in firefighters

Table 1
Overview of fatigue detection methods, algorithms, and performance.

Study	Participants (n)	Metrics	Labeling Method	Fatigue levels	Best Algorithm	Performance
(Anwer et al., 2023)	15	HRV	Borg-20 scale	4	RF	0.935 F1
(Bustos et al., 2022)	24	HR, BR, temp., personal characteristics	Borg-20 scale	4	XGBoost	0.88 F1
(Antwi-Afari et al., 2022)	10	Plantar pressure, 3-axis acceleration	Borg-20 scale	4	RF	0.8346 F1
(Ni et al., 2022)	80	HRV	Borg-20 scale	3	LightGBM	0.801 F1
(Marotta et al., 2021)	8	Motion capture	Borg-20 scale	3	RF	0.905 F1
(Nagahanumaiah et al., 2022)	22	HRV	Borg-20 scale	2	SVM	0.96 F1
(Lambay et al., 2022)	14	IMU and HR	Self-reported	2	RF	95.7 % accuracy
(Nasirzadeh et al., 2020)	8	HR	Borg-20 scale	2	NN	90.36 % accuracy
(Sedighi Maman et al., 2020)	24	IMUs and HR	Borg-20 scale	2	RF	87.9 % accuracy
(Narteni et al., 2022)	15	IMUs and HR	Borg-20 scale	2	SVM NN	0.90 F1

HRV – Heart Rate Variability, HR – Heart Rate, BR – Breathing Rate, Temp. – Temperature, IMU – Inertial Measurement Unit, 3D – Three-Dimensional, RF – Random Forest, NN – Neural Network, LightGBM – Light Gradient Boosting Machine, XGBoost – eXtreme Gradient Boosting, SVM- Support Vector Machine.

(Bustos et al., 2022). Other research has explored wearable insole sensors with plantar pressure and acceleration (Antwi-Afari et al., 2022), motion capture (Marotta et al., 2021), and HR- or IMU-based approaches (Nagahanumaiah et al., 2022; Lambay et al., 2022; Nasirzadeh et al., 2020; Sedighi Maman et al., 2020; Narteni et al., 2022). Together, these studies demonstrate the potential of physiological and motion data to detect fatigue with high accuracy (F1 scores between 0.80 and 0.96). However, they rely heavily on biometric signals and, like the rest of the reviewed literature, rarely integrate demographic or occupational factors that are known to shape fatigue outcomes.

Taken together, this body of research highlights two key gaps. First, most current approaches focus almost exclusively on biometric signals while neglecting contextual elements such as demographic and occupational data that strongly influence fatigue levels. Second, many methods employ supervised learning techniques that require labeled data for training, often relying on subjective self-reports such as the Borg-20 scale (Borg, 1982) (ranging from 6, "no exertion at all," to 20, "maximal exertion"). While widely used, these assessments are intrusive, requiring participants to interrupt their tasks to provide ratings, which can disrupt workflows and introduce variability. This reliance on subjective reporting also introduces bias and limits the generalizability of the resulting models, reducing their practical effectiveness in real-world industrial settings.

Our work addresses these gaps by proposing a more all-encompassing strategy that integrates biometric, demographic, and occupational data to improve fatigue classification accuracy. The approach seeks to reduce the bias of purely biometric methods and provide a stronger, context-aware fatigue detection model by including factors such as age, work experience, task type, production line location, and shift timing. This integration is expected to improve the generalizability of fatigue classification systems, thereby increasing their effectiveness in real industrial environments where individual worker traits and task variability strongly influence fatigue development. Furthermore, instead of relying on subjective self-reports, we employ a fuzzy logic-based data labeling method offering a non-intrusive, consistent, and objective framework for fatigue classification. Conventional safety programs often rely on lagging indicators (e.g., injury or incident rates), which detect problems only after harm occurs. By contrast, the proposed wearable-plus-context model functions as a leading indicator of elevated risk that can be embedded in routine operations (e.g., brief recovery micro-pauses or task rotation) without disrupting workflow. Design choices—non-intrusive sensing, avoidance of mid-shift self-reports, and thresholds tuned to minimize false negatives for high-fatigue states—align with risk-based safety management (e.g., ISO 45001 PDCA). In this way, early detection directly supports injury prevention and safer work pacing on packaging lines.

In the following section, the methodology used for data collection and analysis is explained in detail.

3. Materials and methods

This section is divided into two main parts. The first part describes the data collection campaign, including the types of data collected and the conditions under which they were obtained. The second part outlines the methodology used to analyze the collected data, detailing the steps followed to develop a predictive model for physical fatigue.

3.1. Data collection campaign

The data collection campaign was conducted in a pharmaceutical company, specifically in the packaging area, where operators perform highly repetitive tasks, including inspecting, packaging, and sealing drugs ranging from 1, 10, to 100 items, as well as handling larger trays. These tasks require sustained fine motor control, high precision, and continuous visual focus, often involving rapid hand movements and repetitive gripping, making this an ideal environment for evaluating physical fatigue.

To illustrate the experimental context, Fig. 1 presents the workstation layout with a heat map generated by the Tobii Pro Glasses 3 eye-tracking system, highlighting areas of operator visual attention. Red zones indicate the highest focus, followed by yellow and green for moderate and low focus, respectively. The strongest focus is concentrated on the drugs and the packaging boxes positioned directly in front of the seated operator on the worktable. The task flow follows a highly repetitive cycle: (i) grasping drug ampoules from the worktable, (ii) performing a rapid visual inspection for integrity and labeling, (iii) placing each ampoule into the designated slot of the packaging box, and (iv) closing and stacking the box before starting a new cycle. Table 2 summarizes the typical work cycles across different packaging categories, detailing load sizes and average cycle times.

With 43 participants (42 females, aged 19–53 years, mean age 32.3 ± 7.5 years), this study surpasses the sample sizes of many previous

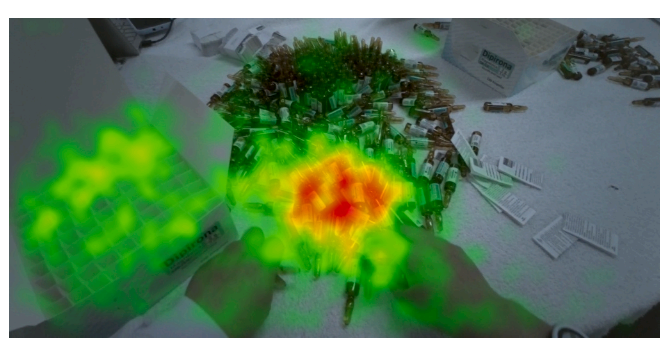

Fig. 1. Workstation layout and operator visual attention heatmap from Tobii Pro Glasses 3. Red areas show highest fixation density, followed by yellow (moderate) and green (low) focus zones.

Table 2
Packing load and average cycle times.

Load / Units per Box	Avg. Cycle Time (s)
1	~7
10	~27
100	~94
Large Trays (>1000)	~780

works, including (Anwer et al., 2023) (15), (Bustos et al., 2022) (24), (Antwi-Afari et al., 2022) (10), (Marotta et al., 2021) (8), (Nagahanumaiah et al., 2022) (22), (Lambay et al., 2022) (14), (Nasirzadeh et al., 2020) (8), (Sedighi Maman et al., 2020) (24) and (Narteni et al., 2022) (15), which achieved high performance despite smaller samples. This gender distribution also reflects the actual workforce composition, where approximately 80 operators include only 3–4 men, making this sample contextually appropriate and representative sample. While this enhances ecological validity, it may limit generalisability if gender-related differences in fatigue physiology or motion patterns exist. To validate transportability, future work will use gender-balanced sampling across multiple sites and job roles. Participants' work experience in the packaging area ranged from less than one month to approximately 20 years, reflecting a realistic cross-section of both newly hired and long-tenured operators within the company.

Prior to participation, all 43 operators provided informed consent, confirming that they did not have any injuries or medical conditions that would prevent them from carrying out their usual work. Participants were informed that they could withdraw from the study at any time without any consequences for their employment. No financial compensation was provided. The study was approved by the Ethics Committee of the Universidad de América (Colombia) and adhered to the ethical principles of the Declaration of Helsinki (Protocol No. 002-–2024).

Each participant contributed data twice per shift — once during the first part of the shift (before the scheduled 40-minute break) and once during the second part (after the break) — to capture fatigue progression within a typical work cycle. This timing reflects operational constraints and was designed to detect variations in physiological responses due to accumulated task load, while minimizing interference with production routines. During each session, participants wore an Empatica EmbracePlus smartwatch for 20 min to capture biometric data relevant to physical fatigue analysis (Albarrán Morillo and Demichela, 2023; Albarrán Morillo and Demichela, 2023). The device continuously recorded pulse rate (mean ± SD: 82.32 ± 10.25 bpm) as an indicator of cardiovascular activity, skin temperature (31.35 ± 1.22 °C) to monitor peripheral thermoregulation and metabolic changes, and electrodermal activity (EDA) (0.423 ± 1.223 μS) to assess autonomic nervous system activation and arousal levels. Additionally, movement-related data included accelerometer readings (standard deviation of G-forces; 0.0667 ± 0.0310 g) representing movement intensity and variability, step count, and activity count, which quantified overall motion throughout the recording period.

In addition to biometric data, demographic and occupational factors were incorporated into the dataset to enhance the predictive model for physical fatigue. The demographic factors included age and work experience in the packaging area. The occupational factors encompassed physical load, referring to the number of pharmaceutical products packed per box (1, 10, 100, or long trays), and operational conditions, which considered the moment or timing of data collection specifically distinguishing between measurements taken at the start (first two hours) and end (last two hours) of the work shift, the day of the week (Monday to Friday), the work shift (three 8-hour shifts: 06:30–14:30, 14:30–22:30, and 22:30–06:30), and the production line location, where data were collected from four different scenarios: Plant 4, Line 2; Plant 4, Line 3; Plant 4, Line 4; and Plant 8, all of which involved the same packaging activity.

Once the dataset was collected, an initial baseline period at the start of each recording session was removed to eliminate artifacts related to device setup. The cleaned dataset, consisting of 2306 samples with 13 features, was then prepared for further analysis, following the methodology outlined in the next section.

3.2. Data analysis techniques

This subsection describes the data analysis methods used to pre-process and analyze the collected dataset. The overall end-to-end workflow is summarized in Fig. 2, which presents the data-processing pipeline from raw smartwatch and contextual inputs through pre-processing/normalization, feature extraction with PCA, fuzzy logic-based labeling, and machine learning classification, up to model evaluation and deployment. Each of these steps is described in detail in the following subsections.

The analysis framework consisted of three key steps: reducing dimensionality with PCA, labeling the data using fuzzy logic based on a previous study (Shi et al., 2024), and applying ML classification algorithms to develop a predictive model for physical fatigue. All physiological signals were collected using the Empatica EmbracePlus device, which applies proprietary algorithms to raw sensor streams and transforms them into validated digital biomarkers (pulse rate, electrodermal activity, skin temperature, and movement indices). These biomarkers were aggregated on a per-minute basis, providing a stable representation of participants' physiological states while reducing high-frequency noise. PCA was applied directly to these preprocessed biomarker datasets rather than to raw data. By avoiding heavy manual preprocessing, the framework supports more efficient and practical deployment of the fatigue detection model in real-world industrial environments.

3.2.1. Principal Component analysis for feature extraction

PCA was used in this study not just to reduce dimensionality, but to address multicollinearity and redundancy among variables. Preliminary analysis showed inter-correlations that could affect model robustness and interpretability. PCA helped transform correlated variables into orthogonal components, improving model clarity and stability. To

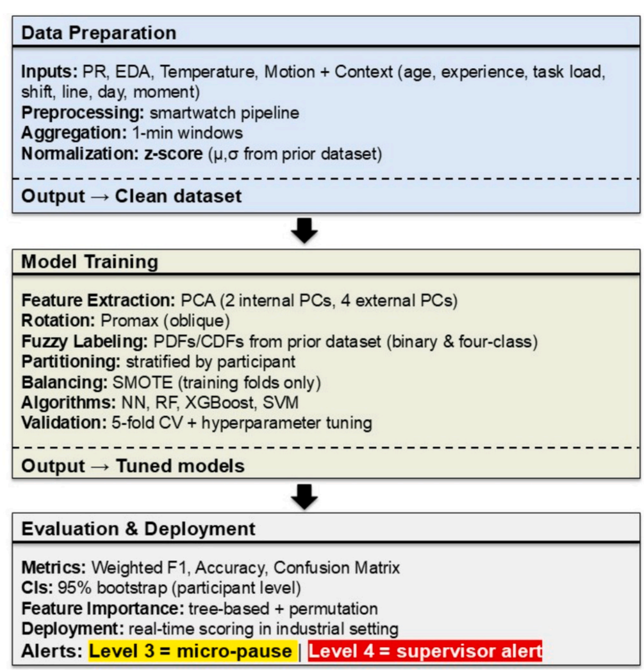

Fig. 2. Overview of the data processing and modeling pipeline for physical fatigue detection, including data preparation, feature derivation, model training, and real-time deployment.

ensure methodological soundness, only components with eigenvalues > 1 were retained, and the cumulative explained variance was guaranteed at an acceptable level.

After extracting the Principal Components (PCs) from both internal (biometric data) and external (demographic and occupational factors) datasets, a Promax rotation with Kaiser Normalization was applied (Shi et al., 2024; Albarrán Morillo et al., 2024). This approach was adopted to improve physical interpretability by ensuring that the extracted components correspond to meaningful physiological and work-related variables while simplifying the data structure, reducing redundancy, and enhancing the clarity of factor loadings (Shi et al., 2024; Albarrán Morillo et al., 2024). Promax, an oblique rotation technique, allows for correlations between factors – an appropriate choice given that biometric responses and occupational characteristics likely interact in influencing fatigue. The transformed components were then used in the subsequent fuzzy logic classification step to assign fatigue levels before training the ML classification models.

3.2.2. Fuzzy Logic-Based data labeling for physical fatigue classification

To address the research gap related to the subjectivity and intrusiveness of self-assessment methods for fatigue labeling, this study uses a fuzzy logic-based data labeling approach. This method, based on earlier work (Shi et al., 2024), assigns each sample a degree of membership to either the fatigue or non-fatigue group, rather than making a strict binary classification. This provides a more flexible and realistic representation of physical fatigue, capturing gradual transitions between fatigue states that conventional approaches often miss.

The fuzzy logic approach (Mendel, 1995) relies on probability density functions (PDFs) and cumulative distribution functions (CDFs) derived from a previous dataset (Albarrán Morillo and Demichela, 2023; Albarrán Morillo and Demichela, 2023; Shi et al., 2024; Albarrán Morillo et al., 2024) collected in a controlled fitness setup designed to simulate industrial tasks. In this earlier study, participants performed physically demanding activities such as pushing, pulling, lifting, picking up, tugging, and bending, selected to replicate the physical workload typically encountered in industrial environments (Albarrán Morillo and Demichela, 2023; Albarrán Morillo and Demichela, 2023; Shi et al., 2024; Albarrán Morillo et al., 2024). These tasks were conducted in a controlled fitness setting designed to simulate real-world conditions in a safe and repeatable manner, emphasizing repetitive and monotonous task execution. The nature of these activities closely resembles those observed in pharmaceutical packing lines, particularly in the case study involving manual operations under time and precision constraints. Physiological data were continuously recorded using Empatica EmbracePlus smartwatches, which monitored four key indicators linked to physical fatigue: pulse rate (PR), electrodermal activity (EDA), skin temperature, and movement data. These parameters were chosen due to their strong association with activation of the sympathetic nervous system, providing objective insights into levels of physical exertion. To complement the physiological measurements, participants also self-reported their perceived exertion using the Borg Rating of Perceived Exertion (RPE) scale (ranging from 6 to 20). RPE scores were collected every one minute to capture the progression of fatigue with high temporal resolution. These values were synchronized with the physiological data and used as subjective ground truth labels in our previous PCA-based fuzzy logic framework (Shi et al., 2024), where they defined fatigue and non-fatigue states. The fuzzy logic method transformed these subjective ratings into probability-based membership functions, reducing inter-individual variability and creating a more objective labeling scheme. Importantly, these subjective ratings were not used in the labeling of the present case study; instead, the fuzzy logic framework established in the previous work was directly applied here to generate fatigue labels in a non-intrusive manner.

By applying these pre-established PDFs and CDFs, this approach calculates fuzzy probability coefficients for each sample in the current dataset, creating a more objective and consistent framework for fatigue classification. This method significantly reduces the subjectivity and potential bias introduced by traditional self-reports, providing a more accurate representation of physical fatigue in real-world industrial contexts.

To ensure consistency, the same normalization method used in the previous study was applied before implementing the fuzzy logic classifier. We adopted the parameters of the normal distribution (μ, σ), where μ denotes the mean and σ denotes the standard deviation, from (Shi et al., 2024) for four reasons, including a key design nuance of the "fitness" protocol. First, the controlled tasks were not heavy-weight, maximal lifts; they intentionally used light-to-moderate loads, defined as approximately < 30 % of the participant's estimated maximum capacity for light and 30–60 % for moderate (Schoenfeld et al., 2021). These loads were applied with high repetition and short rests (pushing, pulling, lifting, carrying, bending) to accumulate fatigue over time, rather than provoke brief, high-intensity spikes.". This load profile mirrors the low-to-moderate, repetitive effort of the packaging line and was chosen specifically to emulate cumulative fatigue rather than strength peaks. Second, biometric signals were collected with the same smartwatch (Empatica EmbracePlus) and identical preprocessing, ensuring measurement invariance. Third, in both datasets, Principal Component Analysis yielded an internal principal component (IPC1) with the same physiological structure—high loadings on pulse rate and electrodermal activity and a moderate loading on skin temperature—indicating a stable autonomic-arousal axis across contexts. Fourth, after Z-score normalization using (μ,σ) from (Shi et al., 2024), the IPC1 distributions aligned and behaved as theory predicts in the real plant (higher later in the shift and under greater task load, lower after breaks), supporting construct and scale equivalence. This correspondence justifies applying the original Gaussian PDFs/CDFs to the real-world dataset. Nonetheless, for new sites or climates that exhibit distributional drift, we recommend lightweight recalibration (e.g., re-estimating (μ,σ) on a small anchor sample or quantile mapping) to refine thresholds without altering the overall pipeline. Specifically, Z-score normalization was used:

$$z = \frac{x - \mu}{\sigma} = 1, \tag{1}$$

where x represents the raw data value, μ is the mean, and σ the standard deviation of the original dataset. The same μ and σ values from the previous dataset were applied to the new dataset to ensure scale consistency and allow the fuzzy classifier to correctly map the new data onto the pre-established fatigue and non-fatigue distributions.

Once the data was normalized, we compute the fuzzy membership values for each sample xi using the pre-fitted PDF and CDF functions for the fatigue and non-fatigue states:

$$Fatigue = \int_{x_i}^{x} PDF_{fatigue}(x)dx = 1 - CDF_{fatigue}(x_i)$$

$$Non - Fatigue = \int_{x_i}^{x} PDF_{non-fatigue}(x)dx = 1 - CDF_{non-fatigue}(x_i) \tag{2}$$

These membership values indicate how strongly each data point is associated with either the fatigue or non-fatigue group. For the final labeling, two classification approaches were implemented: binary classification and multiclass classification with four levels. In the binary approach, samples were assigned to fatigue or non-fatigue groups based on the crossover point of the membership functions, providing a straightforward division between safe and risky conditions. In contrast, the four-level classification introduced intermediate fatigue states, categorizing samples into pure non-fatigue, higher non-fatigue, higher fatigue, and pure fatigue states, based on their degree of belonging to each group. Specifically, the *pure non-fatigue state* represents low exertion and stable physiological responses, where workers are not at risk. The *higher* *non-fatigue state* indicates the early accumulation of strain, where performance remains unaffected but exertion is building. The *higher fatigue state* reflects significant exertion that can impair performance and requires recovery measures. Finally, the *pure fatigue state* corresponds to critical exertion levels, where workers are fully fatigued, and task continuation poses high risks of error or injury. This classification provides a nuanced view of fatigue progression, supporting more accurate monitoring and safer application in industrial environments

3.2.3. Machine learning classification algorithms for physical fatigue prediction

After applying the preliminary analysis, we obtained a reduced dataset consisting of principal components derived from biometric data (internal factors) and demographic and occupational factors (external factors), along with the fatigue labels generated through fuzzy logic analysis. We subsequently applied various machine learning classification algorithms to evaluate their performance in both binary and four-level fatigue classification tasks. The decision to implement a four-level fatigue classification was driven by the need for a more granular and proactive assessment of physical fatigue, enabling earlier identification of fatigue symptoms before reaching critical levels – an essential consideration for maintaining safety in industrial environments. This multi-level classification approach is supported by previous studies (Anwer et al., 2023; Bustos et al., 2022; Antwi-Afari et al., 2022), which have shown that distinguishing between multiple fatigue levels not only captures a broader range of fatigue intensities but also contributes to more actionable insights for fatigue management.

The analysis evaluated the weighted F1 score for each classifier and feature set – considering both binary and four-class fatigue classification – using a 5-fold cross-validation process. The weighted F1 score was chosen to address the challenges posed by class imbalance, as it provides a more balanced evaluation by accounting for both precision and recall across all classes while adjusting for their relative frequencies. Unlike accuracy, which can appear deceptively high when the model favors the majority class, the weighted F1 score offers a more reliable measure of performance, especially in imbalanced scenarios where correctly identifying minority class instances is critical. To prevent data leakage between training and testing sets, cross-validation was stratified by individual participants, ensuring that all data from a given participant appeared exclusively in either the training or testing set within each fold. This approach preserved the independence of the evaluation process and prevented the model from learning participant-specific patterns that could bias the results. Overfitting was further monitored and controlled through several strategies. Hyperparameter tuning for each classifier was conducted within the cross-validation framework, using either grid search or manual tuning based on the average cross-validation performance. For models prone to overfitting, such as Neural Networks and tree-based ensembles, regularization techniques were applied: L2 regularization for Neural Networks, and parameter constraints such as maximum tree depth, reduced learning rates, and early stopping (where applicable) for XGBoost and Gradient Boosting. Additionally, to further mitigate imbalance effects, SMOTE (Synthetic Minority Over-sampling Technique) was applied to generate synthetic samples for underrepresented classes, enhancing the robustness of the training data. The analysis also compared two feature configurations: one using only internal factors (biometric data), and another combining both external (demographic and occupational) and internal factors, to evaluate the potential performance improvement gained by incorporating contextual information.

Based on insights from the literature review on physical fatigue prediction using wearable devices in industrial settings (Anwer et al., 2023; Bustos et al., 2022; Antwi-Afari et al., 2022; Ni et al., 2022; Marotta et al., 2021; Nagahanumaiah et al., 2022; Lambay et al., 2022; Nasirzadeh et al., 2020; Sedighi Maman et al., 2020; Narteni et al., 2022), a set of machine learning algorithms was selected for evaluation. These included Neural Networks (NN), Random Forest, XGBoost, Gradient Boosting, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression – all of which have demonstrated competitive performance in prior fatigue detection studies. For each classification task, the weighted F1 score was used as the primary performance metric. The results present the F1 scores of the four best-performing models, along with the best hyperparameter configurations identified through tuning. To assess the statistical significance and robustness of the F1 scores, a bootstrapping technique was applied without setting a fixed seed during resampling. This involved resampling the training data with replacement, retraining the model on each resampled dataset, and evaluating it on the original test set. The process was repeated 1,000 times, producing a distribution of F1 scores from which the 95 % confidence interval was computed.

4. Results

This section presents the main findings, following the same sequence of the data analysis techniques described earlier. First, we outline the results of PCA, highlighting the extracted components from biometric, demographic, and occupational factors. Next, we present the fuzzy logic classification results, detailing the assigned fatigue levels based on the membership functions. Finally, we evaluate the performance of ML classification algorithms, comparing their effectiveness in predicting binary (two-level) and multiclass (four-level) physical fatigue states. The evaluation is conducted in two stages: first, using only internal data, and then with a combined approach that integrates both external data (demographic and occupational factors) and internal biometric signals to assess their impact on classification performance.

4.1. PCA results

After extracting the PCs from both internal (biometric data) and external (demographic and occupational factors) datasets, the Promax rotation with Kaiser Normalization was applied to enhance interpretability. For biometric data (internal factors), two principal components were retained based on the Kaiser criterion (eigenvalues > 1), explaining a total of 62.81 % of the variance:

IPC1 (internal principal component 1) had an eigenvalue of 2.566, explained 42.77 % of the variance, and was primarily associated with physiological responses, including EDA (0.703), pulse rate (0.711), and skin temperature (0.505) (Table 3).

IPC2 (internal principal component 2) had an eigenvalue of 1.202, contributed an additional 20.03 % of the variance, and was associated with motion-related metrics such as accelerometer data (0.867), step count (0.918), and activity count (0.863) (Table 2).

This analysis resulted in distinct components that grouped variables based on their underlying relationships, aligning with previously published findings (Shi et al., 2024; Albarrán Morillo et al., 2024).

For external factors (demographic and occupational data), four principal components were retained based on the Kaiser criterion (eigenvalues > 1), together explaining a cumulative 85.21 % of the total variance:

EPC1 (external principal component 1) had an eigenvalue of 2.489

Table 3
Principal component matrix with factor loadings for biometric data (internal factors).

Variable	Principal	Component
	1	2
EDA	0.703	
Pulse rate	0.711	
Temperature	0.505	
Accelerometer		0.867
Step count		0.918
Activity count		0.863

Factor loadings with an absolute value < 0.5 are not indicated in the table.

and explained 35.56 % of the total variance. It reflects work-related process factors, including shift (0.734), production line (0.929), and number of products (0.811) (Table 4) which are likely to contribute to workplace fatigue through operational demands and task complexity.

EPC2 (external principal component 2) had an eigenvalue of 1.356 and explained 19.37 % of the variance. It captures personal attributes, specifically age (0.955) and experience (0.845) (Table 4) forming a demographic dimension linked to worker characteristics and potentially influencing fatigue susceptibility.

EPC3 (external principal component 3) had an eigenvalue of 1.113 and explained 15.89 % of the variance. It is dominated by the variable day of the week (0.968) (Table 4) and while it represents a single-variable component, it was retained due to its contribution to the cumulative variance.

EPC4 (external principal component 4) had an eigenvalue of 1.007 and explained 14.39 % of the variance. It is associated primarily with the moment of measurement (0.997) (Table 4). Despite being driven by a single variable, it was also retained to preserve the comprehensive variance structure.

Promax has been adopted here with the capability of involving correlations between factors. Compared with orthogonal methods like Varimax that may oversimplify the relationship, it seems to be more reasonable to apply oblique techniques given that biometric responses and occupational characteristics likely interact in influencing fatigue. To verify robustness, Oblimin rotation was also tested as another oblique rotation technique and yielded highly consistent results with Promax. Although minor differences in loading magnitudes were observed, the overall factor structure and model performance remained stable across both methods.

According to Tables 3 and 4, the original 13-dimensional feature set was effectively reduced to a 6-dimensional feature set through Principal Component Analysis (PCA), consisting of 2 components derived from biometric variables (explaining 62.81 % of the variance) and 4 components from external variables (explaining 85.21 %). This transformation addressed multicollinearity and produced structured, interpretable inputs for the fatigue classification model (Vaillant and Sagrilo, 2024). The difference between the internal and external variable reduction shows a stronger multicollinearity within biometric data and more independence exhibited by external variables. Moreover, the factor loading explored through the Promax rotation technique enhances the interpretability of extracted components, where significant variables were highly focused. This helps to interpret the physical meaning of each PC, like physiological response-related PC (IPC1), motion metric-related PC (IPC2), work process-related PC (EPC1), etc. Such information is meaningful to reflect the hidden latent variables of physical fatigue for a comprehensive understanding. Notably, EPC3 and EPC4 are effectively single-variable components—Day and Moment, respectively—with loadings ≈ 1 and negligible cross-loadings. This indicates that each contributes unique, largely orthogonal variance not shared with other external variables. Because EPC3 and EPC4 are essentially centered-and-scaled versions of Day and Moment, we report their effects individually while retaining PCA for the remaining external variables to preserve an orthogonal, standardized predictor set (and maintain symmetry with the internal PCs).

Table 4
Principal component matrix with factor loadings for demographic and occupational data (external factors).

Variable		**Principal**	**Component**	
	1	**2**	**3**	**4**
Moment				0.997
Shift	0.734			
Production line	0.929			
Day			0.968	
Number of products	0.811			
Age		0.955		
Experience		0.845		

Factor loadings with an absolute value < 0.5 are not indicated in the table.

This finding is consistent with our initial hypothesis: in pharmaceutical industry contexts, external factors – due to the complexity of tasks, demographic and occupational characteristics – may provide essential and distinct information that complements biometric signals. Therefore, incorporating external variables helps to construct a more reliable physical fatigue classification model, rather than relying solely on biometric data.

4.2. Fuzzy Logic-Based label classifier results

To label the fatigue level of monitored data at each time, the established fuzzy logic classifier from our previous work has been adopted (Shi et al., 2024). Specifically, all internal factors contribute to calculating the membership degree through their explored factor loadings (Table 5). These loadings reflect the relative contribution of each variable to the classifier, with higher absolute values indicating stronger influence on fatigue labeling. As shown, activity-related measures such as activity counts (0.892), step counts (0.852), and accelerometer data (0.833) dominate, underscoring the importance of movement intensity and frequency in fatigue detection. Pulse rate (0.657) also plays a substantial role, while electrodermal activity provides a moderate contribution (0.456). Skin temperature, by contrast, shows a negative loading (–0.212), suggesting an inverse relationship with fatigue where peripheral cooling may coincide with exertion. The same normalization method from the previous study was applied to ensure consistency in data scaling (equation (1)). Being aligned to the equation (2), the fuzzy logic approach computed membership degrees for each sample with the detailed parameters provided by equation (3), determining its probability of belonging to either the fatigue or non-fatigue category. Based on these probabilities, labels were assigned using binary classification (fatigue vs. non-fatigue) and four-level classification, where fatigue was categorized into progressive intensity levels.

$$PDF_{fatigue}(x) = \frac{1}{0.7021\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x+0.7079}{0.7201}\right)^2}$$

$$PDF_{non-fatigue}(x) = \frac{1}{0.7805\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-0.5491}{0.7805}\right)^2} \tag{3}$$

The first membership function plot (Fig. 3a) illustrates the probability distribution for fatigue classification based on IPC1. A threshold was set at the intersection of the Non-Fatigue and Fatigue states, enabling binary classification. In this case, samples to the left of the crossover are labeled as Non-Fatigue, and those to the right as Fatigue. The corresponding histogram (Fig. 3b) shows the distribution of samples, with a higher frequency of Non-Fatigue cases compared to Fatigue cases.

To enhance classification granularity, the second membership function plot (Fig. 4a) introduces a four-level fatigue classification, subdividing the space around the crossover to capture transitional states:

Pure Non-Fatigue: high Non-Fatigue membership, negligible Fatigue membership (far left).

Table 5
Factor loading of each variable in the established fuzzy logic classifier.

Variable	Factor loading
Electrodermal activity	0.456
Pulse rate	0.657
Skin temperature	−0.212
Accelerometer data	0.833
Step counts	0.852
Activity counts	0.892

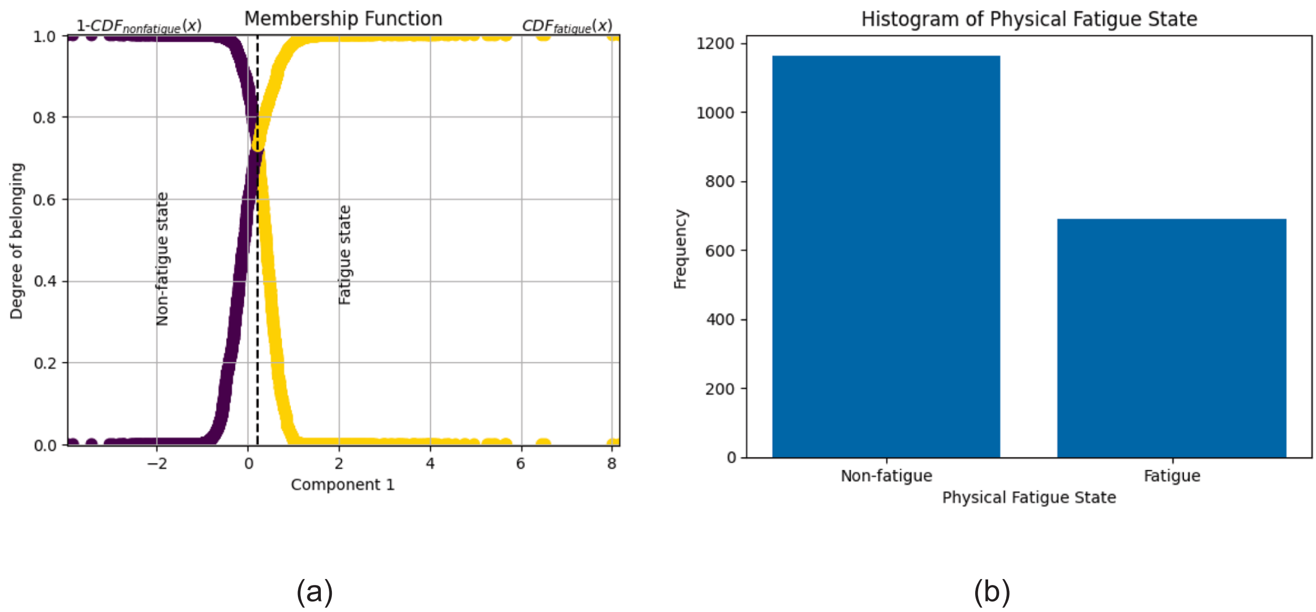

Fig. 3. Fuzzy logic-based labeling for binary classification. (a) Membership function for binary fatigue classification based on IPC1. (b) Histogram of physical fatigue label distribution for binary classification.

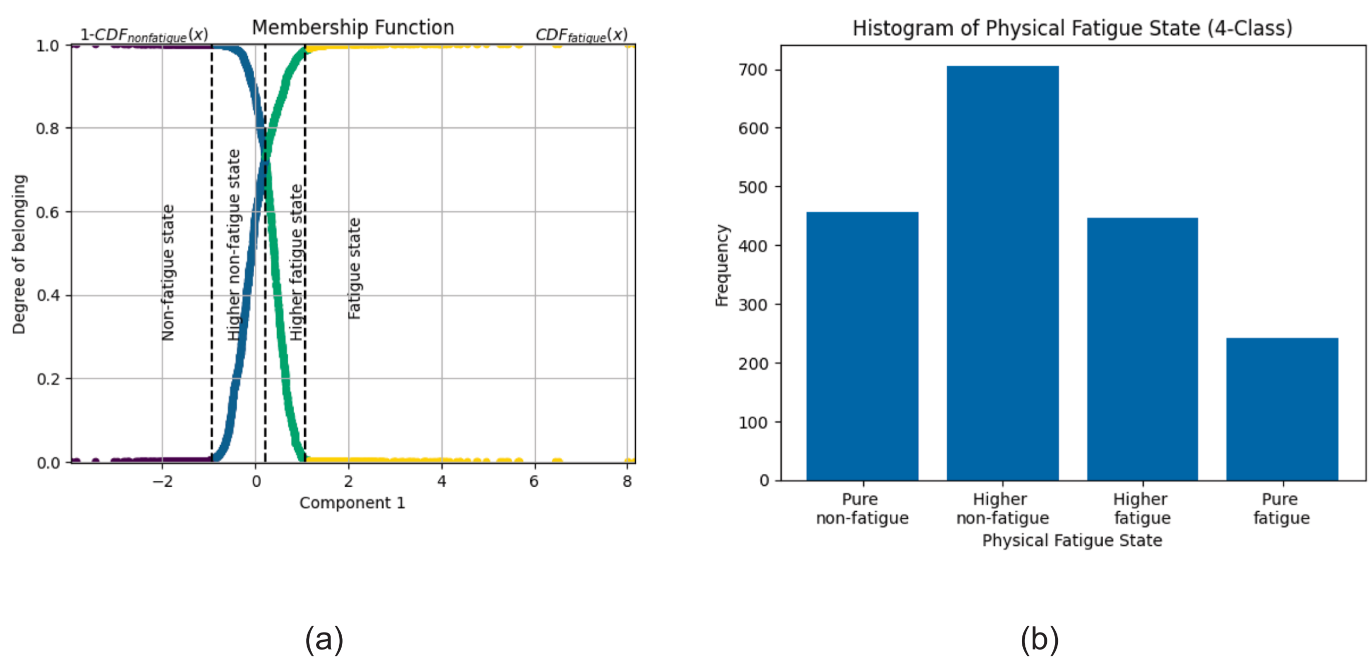

Fig. 4. Fuzzy logic-based labeling for four-level classification. (a) Membership function for four-level fatigue classification based on IPC1. (b) Histogram of physical fatigue label distribution for four-level classification.

Higher Non-Fatigue: samples near the crossover, with Non-Fatigue membership greater than Fatigue membership.

Higher Fatigue: samples near the crossover, with Fatigue membership greater than Non-Fatigue membership.

Pure Fatigue: high Fatigue membership, negligible Non-Fatigue membership (far right).

These decision rules are derived directly from the fuzzy probability density and cumulative distribution functions adapted from our prior controlled dataset and consistently applied in this industrial dataset.

The histogram of the four-class labels (Fig. 4(b)) shows an imbalanced distribution of fatigue states, with a higher frequency of Higher Non-Fatigue cases compared to the other categories. This indicates that while the classification differentiates between subtle fatigue levels, the distribution of samples across the four categories is not uniform.

These labeled fatigue states serve as training targets for the next stage, where ML classification models will be implemented to predict physical fatigue. As previously explained, SMOTE was used to address class imbalance, and F1-weighted scoring was employed to ensure fair performance comparison across both binary and multi-level classification settings.

4.3. Machine learning classification algorithms results

In this section, we present the results of ML classification algorithms applied to predict binary (fatigue vs. non-fatigue) and four-level physical fatigue states, defined as pure non-fatigue, higher non-fatigue, higher fatigue, and pure fatigue states. The analysis was conducted using two feature configurations: one considering only internal data, consisting of the principal components IPC1 and IPC2, and another integrating both external and internal data, incorporating EPC1, EPC2, EPC3, and EPC4 along with the internal components. The results are structured into two subsections, each evaluating the impact of these feature sets on classification performance. To assess the model effectiveness, the F1-weighted score is reported, as it is particularly well-suited for this study given the imbalanced nature of the dataset, especially in the multi-class configuration. For brevity, “F1 score” hereafter refers to the F1-weighted score. The evaluation focused on the four top-performing classification algorithms.

4.3.1. *Internal features*

For binary classification, among the evaluated classifiers, the top four models based on F1-weighted score were the NN, SVM, XGBoost, and Logistic Regression (Table 6). The NN achieved the highest test F1 score of 0.8848, with a 95 % confidence interval ranging from 0.904 to 0.919. Its best performance was obtained using a tanh activation function, a low regularization strength (alpha = 0.0001), and a single hidden layer with 100 neurons. The SVM followed with an F1 score of 0.8798 and a confidence interval of 0.895–0.911, performing best with a radial basis function kernel and a regularization parameter of 1. XGBoost achieved an F1 score of 0.8785, with a confidence interval of 0.878–0.910, and was optimized with a learning rate of 0.05, a maximum tree depth of 3, and 200 boosting rounds. Lastly, Logistic Regression performed reliably with a test F1 score of 0.8782 and a confidence interval of 0.904–0.914.

Feature importance is reported exclusively for XGBoost, as it consistently ranked among the top-performing models and, thanks to its tree-based structure, provides reliable and easily interpretable measures of feature relevance. In contrast, feature importance was not reported for NN due to their complex and non-linear architecture, which does not offer direct or transparent feature contribution metrics. Similarly, SVMs rely on kernel transformations that obscure the direct influence of individual features, making interpretation difficult. Although Logistic Regression provides model coefficients, these are not directly comparable to the feature importance scores generated by tree-based models. The XGBoost analysis showed that IPC2 had a significantly higher importance (0.7331) than IPC1 (0.2669), highlighting IPC2′s stronger influence on the fatigue classification task.

For the four-level physical fatigue classification the four best-performing models in terms of F1-weighted score were XGBoost, NN, RF, and Logistic Regression (Table 7). XGBoost achieved the highest F1 score of 0.8232, with a 95 % confidence interval ranging from 0.773 to 0.815. This model performed best using a learning rate of 0.1, a maximum tree depth of 3, and 200 boosting rounds. The NN followed closely with an F1 score of 0.8200 and a confidence interval of 0.811–0.832, using the ReLU activation function, a regularization strength of 0.0001, and a two-layer architecture with 100 and 50 neurons, respectively. Random Forest achieved a comparable F1 score of 0.8195, with a confidence interval between 0.788 and 0.826, tuned with a maximum depth of 10 and 200 trees. Logistic Regression performed reliably with an F1 score of 0.8179 and a confidence interval from 0.805 to 0.826. Feature importance was examined for the tree-based models, XGBoost and RF, as both provide interpretable measures of feature relevance. For XGBoost, IPC2 had a higher contribution (0.7718) compared to IPC1 (0.2282), while RF showed a similar trend with IPC2 (0.7366), contributing more than IPC1 (0.2634).

Table 6
F1-weighted scores for binary fatigue classification using internal features.

Algorithm	F1-weighted
NN	**0.8848**
SVM	0.8798
XGBoost	0.8785
Logistic Regression	0.8782

Table 7
F1-weighted scores for four-level fatigue classification using internal features.

Algorithm	F1-weighted
XGBoost	**0.8232**
NN	0.8200
RF	0.8195
Logistic Regression	0.8179

Fig. 5 presents the normalized confusion matrix for the four-level physical fatigue classification using just the internal features as predictors and the XGBoost algorithm. The matrix illustrates the model’s ability to correctly classify instances across the four fatigue states: Pure Non-Fatigue, Higher Non-Fatigue, Higher Fatigue, and Pure Fatigue. The diagonal elements represent the true positive rates for each class, indicating classification accuracy per category. The model achieved the highest accuracy in identifying the Pure Non-Fatigue state with 94 %, followed by Pure Fatigue at 85 %. Both intermediate classes, Higher Non-Fatigue and Higher Fatigue, showed balanced but lower accuracies of 77 % each. These results highlight the model’s strong performance in distinguishing between distinct fatigue levels, particularly at the extreme ends of the fatigue spectrum.

4.3.2. *Internal + External features*

The inclusion of external features derived from demographic and occupational factors (EPC1, EPC2, EPC3, EPC4) alongside internal factors significantly improved classification performance across all models.

For binary classification, the four top-performing classifiers based on F1-weighted score were the NN, RF, XGBoost, and Gradient Boosting (Table 8). The NN achieved the highest test F1 score of 0.9375, with a 95 % confidence interval between 0.930 and 0.953, using the ReLU activation function, a regularization strength of 0.0001, and two hidden layers with 100 and 50 neurons. The RF followed closely with an F1 score of 0.9294 and a 95 % confidence interval of 0.920–0.950, tuned with a maximum tree depth of 10 and 100 estimators. XGBoost reached an F1 score of 0.9242, with a confidence interval from 0.921 to 0.948, using a learning rate of 0.1, a maximum depth of 5, and 200 estimators. Gradient Boosting also performed strongly, with an F1 score of 0.9163 and a confidence interval between 0.911 and 0.937, optimized with a learning rate of 0.1 and 200 boosting rounds.

Feature importance was examined for the tree-based models, Random Forest (RF). As shown in Table 9, IPC2 (motion-related metrics) was by far the most influential predictor (0.6431), followed by IPC1 (physiological responses, 0.1945). Among the external features, EPC2 (work-related process, 0.0626) and EPC1 (demographic data, 0.0513) contributed more than temporal variables such as Day (0.0313) and Moment (0.0171). This indicates that integrating contextual drivers, especially work-related and demographic factors, enhances the reliability of binary fatigue classification.

For four-level classification, the top four models based on weighted F1 score were the NN, RF, XGBoost, and Gradient Boosting (Table 10). The NN achieved the highest F1 score of 0.8793, with a 95 % confidence interval ranging from 0.847 to 0.880. This model was configured with a tanh activation function, a low regularization strength (alpha = 0.0001), and two hidden layers with 100 and 50 neurons. The Random Forest followed closely, with an F1 score of 0.8693 and the confidence interval ranging from 0.842 to 0.878, using a maximum depth of 10 and 200 estimators. XGBoost performed strongly as well, achieving an F1 score of 0.8620 and a confidence interval of 0.825–0.864, with a learning rate of 0.1, maximum depth of 5, and 200 boosting rounds. Gradient Boosting delivered an F1 score of 0.8530, with a confidence interval from 0.817 to 0.857, using the same learning rate and number of estimators as

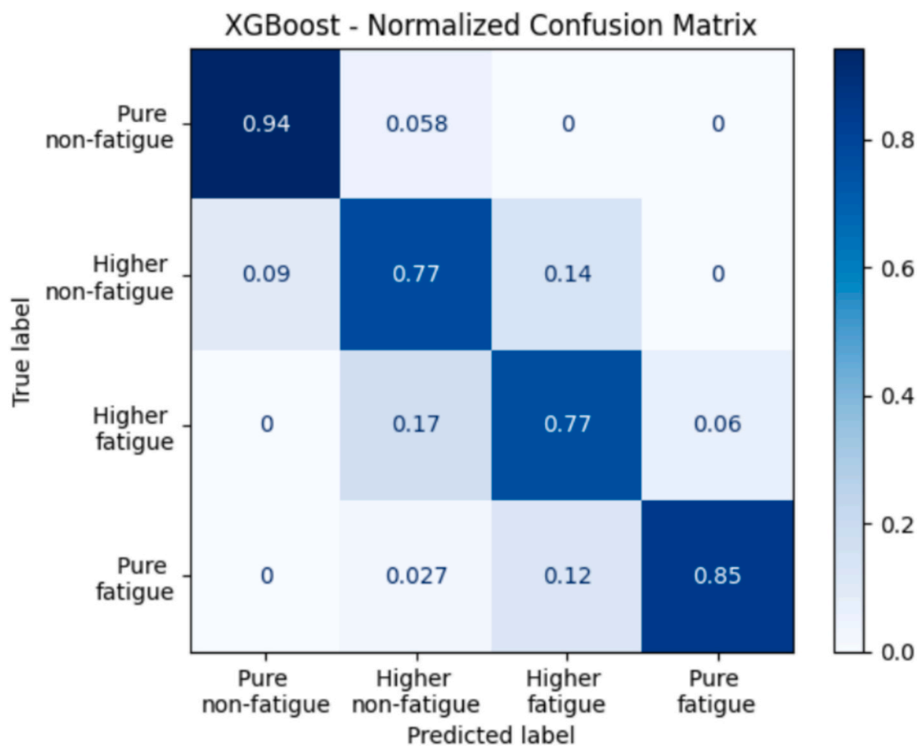

Fig. 5. Normalized confusion matrix for the four-level physical fatigue classification obtained for the XGBoost algorithm. Diagonal values indicate the proportion of correctly classified instances per fatigue state: Pure Non-Fatigue (0.94), Higher Non-Fatigue (0.77), Higher Fatigue (0.77), and Pure Fatigue (0.85).

Table 8
F1-weighted scores for binary fatigue classification using external and internal features.

Algorithm	F1-weighted
NN	**0.9375**
RF	0.9294
XGBoost	0.9242
Gradient Boosting	0.9163

Table 9
Feature importances for binary fatigue classification using the Random Forest algorithm.

Feauture	Importance
IPC2	**0.6431**
IPC1	0.1945
EPC2	0.0626
EPC1	0.0513
Day	0.0313
Moment	0.0171

Table 10
F1-weighted scores for four-level fatigue classification using external and internal features.

Algorithm	F1-weighted
NN	**0.8793**
RF	0.8693
XGBoost	0.8620
Gradient Boosting	0.8530

XGBoost.

Feature importance analysis for RF provided insights into the relative contribution of internal and external features in this multi-class task. Table 11 shows that IPC2 again dominated (0.5928), with IPC1 also playing a substantial role (0.2151). Among external features, EPC2 (0.0783) and EPC1 (0.0514) were the most relevant, followed by Day (0.0410) and Moment (0.0215). Compared to binary classification, the external drivers contributed more strongly here, suggesting that demographic and work-related variability—together with temporal context—helped the model better distinguish intermediate fatigue states.

Table 11
Feature importances for four-class fatigue classification using the Random Forest algorithm.

Feauture	Importance
IPC2	**0.5928**
IPC1	0.2151
EPC2	0.0783
EPC1	0.0514
Day	0.0410
Moment	0.0215

Fig. 6 shows the normalized confusion matrix for the four-level physical fatigue classification obtained using the NN model with both internal and external features. The matrix demonstrates improved classification accuracy across all fatigue states compared to the previous model that used only internal features. The model achieved its highest accuracy in identifying Pure Non-Fatigue instances at 96 %, followed by Higher Non-Fatigue at 87 %, Higher Fatigue at 82 %, and Pure Fatigue at 88 %. These results represent a notable improvement, especially for the intermediate fatigue states, where accuracy increased by approximately 10 percentage points compared to the internal-features-only confusion matrix. Another important improvement related to the real-time physical fatigue alert system is the enhanced reliability in detecting Pure Fatigue states. With internal features only, the model correctly identified Pure Fatigue 85 % of the time; however, it misclassified 12 % of instances as Higher Fatigue and 3 % as Higher Non-Fatigue—the latter being a non-fatigue status, which could compromise alert effectiveness. In contrast, when external contextual information was included, Pure Fatigue was predicted correctly 88 % of the time, with the remaining 12

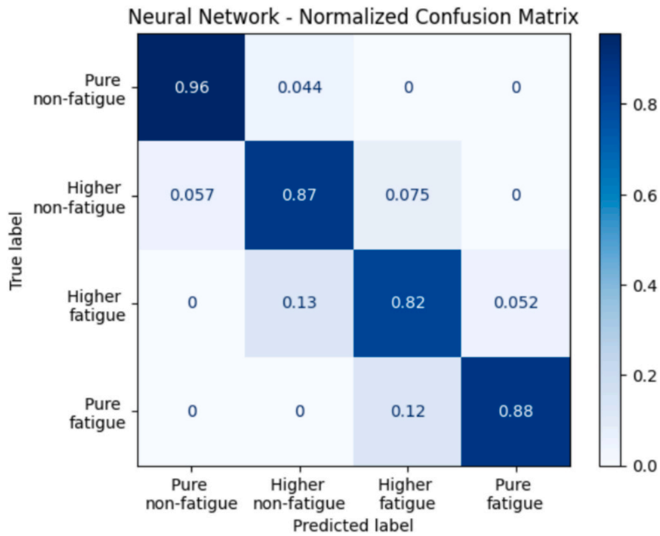

Fig. 6. Normalized confusion matrix for the four-level physical fatigue classification using a NN with combined internal and external features. Diagonal values indicate the proportion of correctly classified instances per fatigue state: Pure Non-Fatigue (0.96), Higher Non-Fatigue (0.87), Higher Fatigue (0.82), and Pure Fatigue (0.88). Classification accuracy improved across all classes compared to the internal-features-only model.

% predicted as Higher Fatigue—a misclassification still within the fatigue spectrum.

While accuracy is reported for the best-performing algorithm in each task, it should be interpreted with caution due to the imbalanced nature of the original dataset. In the binary classification task, the Neural Network achieved an accuracy of 91 % using only internal features, which increased to 95.3 % when both internal and external features were included. For the four-level classification task, overall accuracy reached 82.4 % using XGBoost with internal features alone and improved to 87.9 % when the Neural Network was used with the combined internal and external feature set.

5. Discussion

The integration of external features—specifically demographic and occupational factors (EPC1, EPC2, EPC3, EPC4)—alongside internal components significantly enhanced model performance in both binary and multi-class (four-level) fatigue classification tasks. These external variables provided additional contextual information that helped the models better capture individual differences in fatigue expression, ultimately leading to improved predictive performance and greater model stability, as evidenced by the narrower 95 % confidence intervals. In almost all cases, the test F1 scores fell within or very close to the center of their respective intervals, indicating consistent and reliable model performance across resampled training sets. For instance, in the binary classification using only internal features, the Neural Network had an F1 score of 0.8848, but an interval ranging from 0.904 to 0.919, which did not fully contain the score. However, with the addition of external features, the F1 improved to 0.9375 and the interval to 0.930–0.953, showing both improved accuracy and reduced variability. In the binary classification task, the best-performing model, the Neural Network, improved its F1 score from 0.8848 (using only internal features) to 0.9375 when external features were included—an increase of over 5 percentage points. For the four-level classification task, the best model using only internal features was XGBoost, which achieved an F1 score of 0.8232. However, when external features were added, the Neural Network became the top performer with an F1 score of 0.8793, demonstrating a substantial performance gain of more than 5.6 percentage points. These results highlight the importance of integrating external factors into fatigue prediction models, as they not only significantly boost classification effectiveness—particularly in more complex, multiclass scenarios—but also enrich the data representation, improving generalization and interpretability in real-world industrial settings. Age and work experience capture differences in physical endurance and adaptation to repetitive tasks, enabling the system to distinguish between novice and experienced workers under similar workloads. Task type reflects the varying physical demands of specific operations, production line location accounts for environmental conditions such as ergonomics, space constraints, or workflow pace, and shift timing incorporates circadian and workload-related effects, acknowledging that fatigue is more likely at certain points of the shift.

Feature importance analysis revealed that among the internal components, IPC2—associated with motion-related metrics—was the most influential predictor across all models and classification tasks, followed by IPC1, which reflected physiological responses. Regarding external features, EPC2 (work-related process factors, including task load, production line, and shift context) and EPC1 (demographic variables, such as age and experience) consistently emerged as the most influential external features across both classification tasks, though with relatively modest importance values. Their contribution highlights how operational conditions and worker characteristics shape fatigue outcomes, complementing biometric signals and enhancing the model's real-world applicability, above all in the four-level classification, where their relative importance was higher. These findings indicate that while the inclusion of external features enhanced overall classification performance, the internal components—particularly IPC2—remained the dominant factors guiding the models' decisions. This improved feature representation translated directly into better classification outcomes, as evidenced by the confusion matrix results. Notably, classification accuracy for intermediate fatigue states (e.g., Higher Non-Fatigue and Higher Fatigue) increased by approximately 10 percentage points when external features were included. Furthermore, the system's reliability in detecting Pure Fatigue states improved significantly: with internal features alone, 3 % of Pure Fatigue instances were misclassified as Higher Non-Fatigue—a non-fatigue state that could result in a missed alert. In contrast, the model with combined features eliminated such misclassifications, classifying the remaining 12 % as Higher Fatigue—still within the fatigue spectrum. This shift strengthens the effectiveness of the real-time fatigue alert system, reducing false negatives (fewer missed high-risk episodes) and improving the precision of safety interventions.

When compared to previous studies, our models achieve performance that is comparable to, and in some cases exceeds, existing fatigue classification approaches. Reported F1 scores in the literature vary depending on the number of fatigue levels modeled. For two-level (binary) classification, F1 scores typically range from 0.90 to 0.96 (Nagahanumaiah et al., 2022; Narteni et al., 2022), with reported accuracies between 87.9 % and 95.7 % (Lambay et al., 2022; Nasirzadeh et al., 2020; Sedighi Maman et al., 2020). In our study, the Neural Network achieved an F1-weighted score of 0.9375 and an accuracy of 95.3 %, placing it at the upper end of this range, and outperforming several methods that used only physiological or motion data. For three- and four-level classification, F1 scores in previous works range more widely, from 0.801 (Ni et al., 2022) to 0.935 (Anwer et al., 2023), though the highest scores typically come from models using HRV without contextual information. Notably, our model achieved an F1-weighted score of 0.8793 in the four-level classification task, outperforming (Antwi-Afari et al., 2022) (F1 = 0.8346), which also targeted four fatigue levels but relied only on plantar pressure and acceleration data. These results demonstrate that our integration of both internal physiological measures and external contextual features—including demographic and occupational data—not only achieves state-of-the-art performance but also improves interpretability and relevance for real-world industrial applications. In particular, demographic variables (EPC2: age and work experience) capture inter-individual variability in autonomic responses and physical endurance, shaping how quickly fatigue develops under repetitive tasks—for example, younger or less experienced workers may exhibit sharper physiological changes compared to more experienced operators performing the same activity. Work-related process factors (EPC1: task load, production line, and shift context) capture operational demands and circadian influences that directly affect physiology and motion metrics, with late-shift work or repetitive high-load tasks amplifying autonomic strain and motion variability in ways biometric-only models might overlook. Additionally, Day and Moment capture temporal dynamics across the workweek and within shifts, highlighting patterns such as cumulative fatigue toward the end of a shift or greater strain on certain days. Taken together, these contextual variables enhance model accuracy, support generalizability across diverse settings, and improve fairness by adapting to both worker characteristics and operational conditions.

An additional strength of this study lies in the use of a fuzzy logic-based labeling method, originally developed in a controlled fitness setup (Albarrán Morillo and Demichela, 2023; Albarrán Morillo and Demichela, 2023), where monotonous and repetitive tasks were simulated—closely resembling those in the present case study. Although widely used, subjective tools such as the Borg-20 scale come with important limitations—especially in real-world industrial settings. Workers may unintentionally underreport or exaggerate their fatigue due to social pressures, fear of judgment, or simply not fully understanding the scale (Bazazan et al., 2023). What's more, how a person rates their effort can vary from one day to the next, even if their physical state hasn't changed much (Parijat and Lockhart, 2009). These methods also require workers to pause and report how they feel, which can disrupt their flow and affect the accuracy of the data. Another challenge is the lack of standardization—what feels like a "15 – hard" effort to one person might only be a "12 – moderate" to someone else (Valenza et al., 2018). By contrast, the fuzzy logic approach used in this study relies on physiological markers objectively linked to sympathetic nervous system activation and avoids these issues by enabling continuous, non-intrusive, and more replicable labeling.

Taken together, these findings carry important implications for both practitioners and researchers. For practitioners, the proposed system can function as a leading indicator of fatigue risk, enabling proactive safety measures such as micro-pauses or task rotation without disrupting workflow and reducing the likelihood of missed high-risk episodes. For researchers, the study demonstrates the value of integrating physiological signals with demographic and occupational factors, advancing fatigue detection beyond biometrics alone and offering a scalable, reproducible framework adaptable to diverse industrial settings. The key takeaway is that fatigue monitoring can evolve into a proactive, human-centered tool that aligns with Industry 5.0 principles—supporting both safety and productivity in complex operational environments.

6. Conclusions

In the context of Industry 5.0, where humans remain central and human–machine collaboration is emphasized, this study offers a foundational step toward a real-time physical fatigue monitoring and alert system (Fig. 7). From a safety-management perspective, the primary contribution is a practical leading indicator of fatigue risk that can be integrated with existing Safety Management Systems (e.g., ISO 45001) to support prevention rather than post-event reaction. By combining wearable biometric signals (pulse rate, EDA, skin temperature, motion) with contextual workplace data (task load, line, shift, moment, day, age, experience), the model continuously estimates fatigue and assigns one of four states: Pure Non-Fatigue, Higher Non-Fatigue, Higher Fatigue, and Pure Fatigue. For illustrative deployment, per-window class probabilities can be aggregated into a fatigue-risk score and evaluated over short rolling windows with simple persistence/hysteresis rules to avoid spurious alerts. A tiered policy then supports immediate decisions: workers in Higher Fatigue (level 3) receive a discreet haptic/visual prompt suggesting a brief recovery micro-pause, while Pure Fatigue (level 4) triggers supervisor notification for actions such as task rotation or a short, assisted break. Thresholds can be tuned to prioritize safety (minimize false negatives) and optionally adapted to context (e.g., slightly lower thresholds late in the shift). Privacy-by-design principles (local processing/pseudonymized dashboards) can be applied. Detailed alert-policy evaluation and interface design are beyond the scope of this paper and are reserved for future work; our goal here is to demonstrate the feasibility of a non-intrusive, context-aware pipeline that can enable such real-time support. Additionally, the fuzzy logic-based labeling method, adapted from a controlled setup, removes the need for subjective self-reports, offering a more objective and scalable solution. Crucially, adding contextual features not only improved classification performance but also reduced false negatives, especially in the critical Pure Fatigue state. With biometric data alone, some cases were misclassified as Higher Non-Fatigue, risking missed alerts. Including demographic and task-related variables ensured all misclassifications stayed within the fatigue spectrum, enhancing the system's reliability and safety.

Future applications of this approach include extending its use to other sectors, such as heavy industry and high-demand operational settings, where physical fatigue is influenced not only by physiological and task-related factors but also by environmental conditions such as

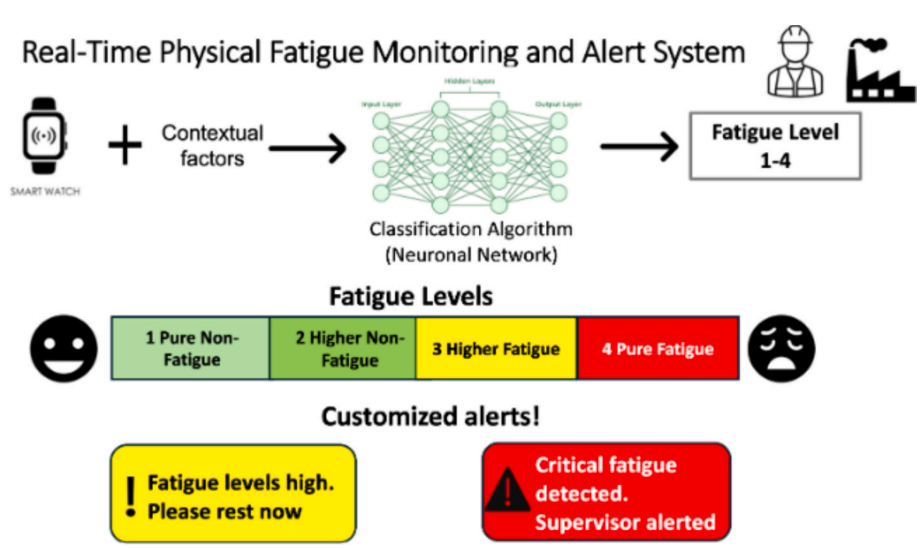

Fig. 7. Real-time physical fatigue monitoring and alert system.

noise, temperature, or lighting. In the present study we did not use ambient variables as predictors: temperature, noise, and humidity were continuously monitored via IoT sensors, remained within narrow, stable ranges, and complied with Colombian standards, providing insufficient within-shift variability for modeling. Still, in more variable or less controlled work environments, these factors could interact with both biometric signals and work conditions in meaningful ways—so they're important to consider in future studies looking to expand the model's reach. The model is well-suited to incorporate these additional variables, thanks to the use of Principal Component Analysis (PCA) in the preprocessing stage. PCA facilitates the integration of new features by identifying inter-correlations and minimizing redundancy, ensuring that the model remains efficient and interpretable even as the input space grows. This makes the approach highly adaptable for future, more comprehensive fatigue monitoring systems across diverse industrial environments. As future implementations generate larger and more diverse datasets across different industries, the opportunity to re-estimate and adapt the fuzzy logic-based labeling model will become more feasible. We acknowledge a current limitation in the direct transfer of fuzzy logic parameters, which were originally developed in a controlled fitness setting. Due to potential differences in data scale and distribution, these parameters may not perfectly align with those of new industrial datasets. To address this, we recommend that future applications include parameter re-calibration based on task-specific or sector-specific data. This will improve the adaptability and precision of the fuzzy classification approach, ensuring more accurate fatigue detection across diverse industrial domains. Additionally, participants were predominantly women, reflecting the actual demographic profile of the packaging area. While this strengthens external validity, gender-related differences in fatigue responses could introduce bias. Subsequent research should aim for gender-balanced samples and multiple workplaces to validate robustness.

7. Institutional review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Universidad de América (Colombia) (protocol code 002-2024 June 2024).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

CRediT authorship contribution statement

Carlos Albarrán Morillo: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Huxiao Shi:** Writing – review & editing, Validation, Software, Methodology, Formal analysis, Data curation. **John Suárez-Pérez:** Writing – review & editing, Visualization, Validation, Supervision, Software, Formal analysis, Data curation. **Micaela Demichela:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Funding

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

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.

Acknowledgments

The authors express their gratitude to Anyela Jineth Mateus Olarte and Diana Cathalina Muñoz Giraldo from the pharmaceutical company and Mónica Andrea Camargo Salinas and Nasli Yuceti Miranda Arandia from the Universidad de América for their invaluable support and collaboration throughout this research.

Data availability

The anonymized data files from the previous study conducted in a fitness environment to simulate industrial tasks—used for fuzzy logic labeling—are available on Zenodo at https://doi.org/10.5281/zenodo.13906740. In addition, the anonymized dataset collected from the pharmaceutical case study, used for the classification models presented in this work, can be accessed at https://doi.org/10.5281/zenodo.14891916.

References

Ahmmed, M.S., Isanaka, S.P., Liou, F., 2024. Promoting Synergies to Improve Manufacturing Efficiency in Industrial Material Processing: a Systematic Review of Industry 4.0 and AI. Machines 12, 681. https://doi.org/10.3390/machines12100681.

Mital, A., Pennathur, A., 2004. Advanced Technologies and Humans in Manufacturing Workplaces: an Interdependent Relationship. Int. J. Ind. Ergon. 33, 295–313. https://doi.org/10.1016/j.ergon.2003.10.002.

Trstenjak, M., Benešova, A., Opetuk, T., Cajner, H., 2025. Human Factors and Ergonomics in Industry 5.0—A Systematic Literature Review. Appl. Sci. 15, 2123. https://doi.org/10.3390/app15042123.

Breese, M., 2024. The effect of Human Factors and leadership on Safety. Chem. Eng. Prog. 120, 44–49. https://www.aiche.org/resources/publications/cep/2024/may/effect-human-factors-and-leadership-on-safety.

Verma, D., 2024. Industry 5.0: a Human-Centric and Sustainable Approach to Industrial Development. Int. J. Soc. Relevance Concern 12, 17–21. https://doi.org/10.26821/IJSRC.12.5.2024.120507.

Ruiz-de-la-Torre, A.; Guevara, W.; Río-Belver, R.M.; Borregan-Alvarado, J. Industry 5.0: The Road to Sustainability. In Towards a Smart, Resilient and Sustainable Industry; Springer: Cham, Switzerland, 2023; pp. 247–257. DOI: 10.1007/978-3-031-38274-1_21.

Panneerselvam, S., Kumar, A.S., Subramanian, C., 2023. Physical challenges in Assembly Line Production Systems from Ergonomics Point of View – Review. J. Ind. Mech. 8, 24–43.

Sieber, W., Chen, G., Krueger, G., Lincoln, J., Menéndez, C., O'Connor, M., 2022. Research Gaps and needs for Preventing Worker Fatigue in the Transportation and Utilities Industries. Am. J. Ind. Med. 65 (11), 857–866. https://doi.org/10.1002/ajim.23346.

Parijat, P., Lockhart, T.E., 2009. Effects of lower Extremity Muscle Fatigue on the Outcomes of Slip-Induced Falls. Ergonomics 51, 1873–1884. https://doi.org/10.1080/00140130802567087.

Fukuda, K., 1994. The Chronic Fatigue Syndrome: a Comprehensive Approach to its Definition and Study. Ann. Intern. Med. 121 (12), 953. https://doi.org/10.7326/0003-4819-121-12-199412150-00009.

Andersen, J.H., Haahr, J.P., Frost, P., 2007. Risk Factors for more Severe Regional Musculoskeletal Symptoms: a Two-Year prospective Study of a General Population. Arthritis Rheum. 56 (4), 1355–1364. https://doi.org/10.1002/art.22513.

World Health Organization (WHO). Musculoskeletal Conditions. WHO Fact Sheets, 2023. Available online: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions (accessed on 14 July 2022).

Szabo, R.M., 1998. Carpal Tunnel Syndrome as a Repetitive Motion Disorder. Clin. Orthop. Relat. Res. 351, 78–89. https://doi.org/10.1097/00003086-199806000-00011.

Guzmán-Velasco, A., Diago-Franco, J.L., 2019. Coexistence of Musculoskeletal Disorders in the Upper Body of Labor Origin. Duazary 16 (2), 193–203. https://doi.org/10.21676/2389783X.2749.

Quiroga, S., Largo, J.F., Rodriguez, F., Sánchez, V., 2022. Carpal Tunnel Syndrome Diagnosis and Prevention System. Universidad de los Andes, Bogotá, Colombia. Available online: https://repositorio.uniandes.edu.co/server/api/core/bitstreams/1f81ae44-9cf2-4954-9089-cff1fee3ff4d/content.

Ferrerosa, B., López, J., Reyes, E.G., Bravo, M., 2015. Painful Musculoskeletal Symptoms and Ergonomic Risk in Upper Limbs in Workers of a Cosmetics Company. Rev. Colomb. Salud Ocup. 5 (3), 26–30. https://doi.org/10.18041/2322-634X/rcso.3.2015.4912.

Bazazan, A., Noman, Y., Norouzi, H., Maleki-Ghahfarokhi, A., Sarbakhsh, P., Dianat, I., 2023. Physical and Psychological Job Demands and Fatigue Experience among Offshore Workers. Heliyon 9 (6), e16441. https://doi.org/10.1016/j.heliyon.2023.e16441.

Gawron, V., 2016. Overview of Self-Reported measures of Fatigue. Int. J. Aviat. Psychol. 26 (3–4), 120–131. https://doi.org/10.1080/10508414.2017.1329627.

Valenza, G., Citi, L., Saul, J.P., Barbieri, R., 2018. Measures of Sympathetic and Parasympathetic Autonomic Outflow from Heartbeat Dynamics. J. Appl. Physiol. 125 (1), 19–39. https://doi.org/10.1152/japplphysiol.00842.2017.

Behrens, M., Gube, M., Chaabene, H., Prieske, O., Zenon, A., Broscheid, K.C., et al., 2023. Fatigue and Human Performance: an Updated Framework. Sports Med. 53, 7–31. https://doi.org/10.1007/s40279-022-01748-2.

Fernández, F.-X., Raizen, D.M., Mullington, J., Anaclet, C., Clarke, G., Critchley, H., et al., 2025. A Comprehensive Model for the Converging Biologies that Underpin the Homeostatic sleep Signal. Sleep Med. 134, 106723. https://doi.org/10.1016/j.sleep.2025.106723.

Raizen, D.M., Mullington, J., Anaclet, C., Clarke, G., Critchley, H., Dantzer, R., et al., 2023. Beyond the Symptom: the Biology of Fatigue. Sleep 46, 9, zsad069. https://doi.org/10.1093/sleep/zsad069.

Moon, J., Ju, B.-K., 2024. Wearable Sensors for Healthcare of Industrial Workers: a Scoping Review. Electronics 13, 3849. https://doi.org/10.3390/electronics13193849.

Anwer, S., Li, H., Umer, W., Antwi-Afari, M.F., Mehmood, I., Yu, Y., Haas, C., Wong, A.Y. L., 2023. Identification and Classification of Physical Fatigue in Construction Workers using Linear and Nonlinear Heart Rate Variability Measurements. J. Constr. Eng. Manag. 149 (7). https://doi.org/10.1061/jcemd4.coeng-13100.

Bustos, D., Cardoso, F., Rios, M., Vaz, M., Guedes, J., Costa, J.T., Baptista, J.S., Fernandes, R.J., 2022. Machine Learning Approach to Model Physical Fatigue during Incremental Exercise among Firefighters. Sensors 23 (1), 194. https://doi.org/10.3390/s23010194.

Antwi-Afari, M.F., Anwer, S., Umer, W., Mi, H.-Y., Yu, Y., Moon, S., Hossain, M.U., 2022. Machine Learning-based Identification and Classification of Physical Fatigue Levels: a Novel Method based on a Wearable Insole Device. Int. J. Ind. Ergon. 93, 103404. https://doi.org/10.1016/j.ergon.2022.103404.

Ni, Z., Sun, F., Li, Y., 2022. Heart Rate Variability-based Subjective Physical Fatigue Assessment. Sensors 22 (9), 3199. https://doi.org/10.3390/s22093199.

Marotta, L., Buurke, J.H., Van Beijnum, B.-J.-F., Reenalda, J., 2021. Towards Machine Learning-based Detection of Running-Induced Fatigue in Real-World Scenarios: Evaluation of IMU Sensor Configurations to Reduce Intrusiveness. Sensors 21 (10), 3451. https://doi.org/10.3390/s21103451.

Nagahanumaiah, L.; Singh, S.; Heard, J. Diagnostic Human Fatigue Classification Using Wearable Sensors for Intelligent Systems. In Proceedings of the 2022 17th Annual System of Systems Engineering Conference (SOSE), Rochester, NY, USA, 6–9 June 2022; pp. 424–429. DOI: 10.1109/SOSE55472.2022.9812694.

Lambay, A., Liu, Y., Ji, Z., Morgan, P., 2022. Effects of Demographic Factors for Fatigue Detection in Manufacturing. IFAC-PapersOnLine 55 (2), 528–533. https://doi.org/10.1016/j.ifacol.2022.04.248.

Nasirzadeh, F., Mir, M., Hussain, S., Darbandy, M.T., Khosravi, A., Nahavandi, S., Aisbett, B., 2020. Physical Fatigue Detection using Entropy Analysis of Heart Rate Signals. Sustainability 12 (7), 2714. https://doi.org/10.3390/su12072714.

Sedighi Maman, Z., Chen, Y.J., Baghdadi, A., Lombardo, S., Cavuoto, L.A., Megahed, F. M., 2020. A Data Analytic Framework for Physical Fatigue Management using Wearable Sensors. Expert Syst. Appl. 155, 113405. https://doi.org/10.1016/j.eswa.2020.113405.

Narteni, S., Orani, V., Cambiaso, E., Rucco, M., Mongelli, M., 2022. On the Intersection of Explainable and Reliable AI for Physical Fatigue Prediction. IEEE Access 10, 1–11. https://doi.org/10.1109/ACCESS.2022.3191907.

Borg, G.A.V., 1982. Psychophysical bases of perceived exertion. Med. Sci. Sports Exerc. 14 (5), 377–381. https://doi.org/10.1249/00005768-198205000-00012.

Albarrán Morillo, C., Demichela, M., 2023. A Data-Driven Framework to Model Physical Fatigue in Industrial Environments using Wearable Technologies. Neuroergonom. Cognit. Eng. 160–168. https://doi.org/10.54941/ahfe1003016.

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. Chem. Eng. Trans. 99, 167–172. https://doi.org/10.3303/CET2399028.

Shi, H., Albarrán Morillo, C., Baldissone, G., Demichela, M., 2024. The Design of the Principal Component Analysis (PCA)-based Fuzzy Logic Classifier on Physical Fatigue in Process Industries. Chem. Eng. Trans. 111, 189–194. https://doi.org/10.3303/CET24111036.

Albarrán Morillo, C., Shi, H., Baldissone, G., Demichela, M., 2024. Customizing a Weighted Scale for Precision in Fatigue Assessment within the Process Industry. Chem. Eng. Trans. 111, 193–198. https://doi.org/10.3303/CET24111033.

Mendel, J.M., 1995. Fuzzy Logic Systems for Engineering: a Tutorial. Proc. IEEE 83 (3), 345–377.

Schoenfeld, B.J., Grgic, J., Van Every, D.W., Plotkin, D.L., 2021. Loading Recommendations for Muscle Strength, Hypertrophy, and Local Endurance: a Re-Examination of the Repetition Continuum. Sports 9 (2), 32. https://doi.org/10.3390/sports9020032.

Vaillant, A.G., Sagrilo, L., 2024. On the use of Dimension-Reduction Methods in Fatigue Analysis of Flexible Risers Subjected to Bimodal Seas. J. Offshore Mech. Arct. Eng. 147 (3), 1–61. https://doi.org/10.1115/1.4065773.