



OPEN ACCESS

EDITED BY

Stawomir Kujawski,
Ludwik Rydygier Collegium Medicum in
Bydgoszcz Nicolaus Copernicus University in
Toruń, Poland

REVIEWED BY

Antonio Esposito,
Kore University of Enna, Italy
Swagata Barik,
University of Calcutta, India

*CORRESPONDENCE

Yubin Zhou
✉ yubinzhou@163.com
Guangyun Wang
✉ gfkdwgy@163.com

RECEIVED 01 May 2025

ACCEPTED 13 June 2025

PUBLISHED 02 July 2025

CITATION

Guo D, Wang C, Qin Y, Shang L, Gao A,
Tan B, Zhou Y and Wang G (2025)
Assessment of flight fatigue using heart rate
variability and machine learning
approaches.
Front. Neurosci. 19:1621638.
doi: 10.3389/fnins.2025.1621638

COPYRIGHT

© 2025 Guo, Wang, Qin, Shang, Gao, Tan,
Zhou and Wang. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Assessment of flight fatigue using heart rate variability and machine learning approaches

Dalong Guo, Cong Wang, Yufei Qin, Lamei Shang, Aijing Gao,
Baosen Tan, Yubin Zhou* and Guangyun Wang*

Air Force Medical Center, Air Force Medical University, Beijing, China

The accurate identification of flight fatigue is crucial for managing pilot training intensity and preventing aviation accidents. However, as a subjective perception, flight fatigue is often difficult to evaluate objectively. Heart rate variability (HRV), derived from electrocardiogram signals and regulated by the autonomic nervous system, is recognized as an effective biomarker for assessing fatigue status. This study proposes a novel HRV-based method for the automatic and objective classification of flight fatigue. This study involved an experimental investigation conducted with a cohort of 90 pilots. First, we conducted statistical analyses to investigate whether HRV features and respiratory rate indicators significantly differed across various fatigue levels. A subset of HRV features and the respiratory metric were used as input variables for four machine learning algorithms: decision tree, support vector machine, K-nearest neighbor, and light gradient-boosting machine (LightGBM). These models were applied to perform a three-level classification of flight fatigue. Finally, classification performance was evaluated using average accuracy, precision, recall, and F1 score. Among these models, LightGBM demonstrated the best performance, achieving an accuracy of 0.886 ± 0.057 , precision of 0.837 ± 0.064 , recall of 0.861 ± 0.086 , and F1 score of 0.849 ± 0.067 . These findings indicate that a LightGBM model trained on 12 selected HRV features and one respiratory indicator can accurately categorize flight fatigue into three levels. Fatigue can be detected even when mild, enabling real-time monitoring and early warning of flight fatigue. This approach holds potential for reducing fatigue-related flight accidents.

KEYWORDS

flight fatigue, machine learning, heart rate variability, light gradient-boosting machine, electrocardiogram

1 Introduction

Flight fatigue refers to the cumulative physical and mental exhaustion experienced by pilots during flight operations, which is primarily attributed to factors such as extended flight durations, circadian rhythm disruptions due to jet lag, and heightened psychological stress (Wingelaar-Jagt et al., 2021). Flight fatigue can lead to a decline in both psychological and physiological functioning in pilots, manifesting as slower reaction times, impaired judgment, and reduced motor control precision, posing serious threats to flight safety

(Gaines et al., 2020; Wingelaar-Jagt et al., 2021; Jun-Ya and Rui-Shan, 2023). Historically, fatigue-related incidents have been common among military and commercial pilots (Gaines et al., 2020; Alzhairi et al., 2021; Wingelaar-Jagt et al., 2021). Such accidents are often attributed to decreased concentration and impaired performance caused by fatigue. Research has suggested that about 20% of aviation accidents are closely linked to flight fatigue (Boksem et al., 2005; Faber et al., 2012). Therefore, real-time assessment of fatigue states holds significant practical importance, as it allows for timely intervention measures to prevent major flight safety incidents (Wilson et al., 2020).

Based on the aforementioned considerations, numerous researchers have employed a variety of physiological detection techniques to assess flight fatigue. Of these, the detection of flight fatigue using an electrocardiogram (ECG) is regarded as the most promising method (Abe et al., 2014; Hu and Lodewijks, 2020; Pan et al., 2021; Li et al., 2022). Heart rate variability (HRV), obtained from processed ECG signals, has been shown to reflect autonomic nervous system activity and is widely recognized as an effective indicator for assessing drowsiness and fatigue levels in the human body (Abe et al., 2014). Furthermore, HRV monitoring serves as a non-invasive detection method that poses no risk to the pilot's physical wellbeing and does not elicit any obvious discomfort; it can also be reliably collected through various lightweight wearable devices (Hu and Lodewijks, 2020; Chen et al., 2024). In comparison with other fatigue detection technologies such as an electroencephalogram (EEG), HRV demonstrates superior stability in-flight environments and is less susceptible to factors, including head movement, noise, light, temperature, and electromagnetic interference (Chen et al., 2024). Consequently, it provides a more precise reflection of the pilot's actual fatigue status. In recent years, machine learning algorithms utilizing HRV have become a major focus in fatigue detection, with significant breakthroughs in the automatic analysis of mental fatigue. One study used key HRV features related to fatigue as inputs and successfully distinguished between normal and fatigued status (Wang et al., 2023). Another study used neural network algorithms to develop a driver fatigue model using the power spectral density of HRV features as input, achieving a fatigue detection accuracy of up to 90% (Patel et al., 2011). Additionally, models that employed a support vector machine (SVM) and LightGBM to classify physical fatigue also achieved promising results (Ramos et al., 2020; Cosoli et al., 2021). Research has also shown that combining ECG and respiratory signals can improve the accuracy of fatigue classification, possibly because respiratory rate (Rsp), like HRV, effectively reflects fatigue status (Chen et al., 2016; Xiang et al., 2022).

Using a range of physiological data, four supervised machine learning algorithms commonly used to classify fatigue levels are: decision tree (DT), SVM, K-nearest neighbor (KNN), and LightGBM (Kirodiwal et al., 2022; Ramkumar et al., 2023). The DT model presents its decision-making process in a tree-like structure, which facilitates an intuitive understanding of the model's reasoning. This is particularly valuable for identifying key fatigue-related features and interpreting classification outcomes. In addition, DT models make minimal assumptions about data distribution, allowing them to be adapted to various types and distributions of fatigue-related data (Lin et al., 2010). SVM models have strong generalization ability, even when trained on limited

sample sizes, which enhances their accuracy on unseen data, making it suitable for studies such as this one, where data collection can be challenging (Cervantes et al., 2020). The KNN algorithm can classify data based on the similarity of instances. It is relatively easy to implement, which allows for rapid responses to new data samples in fatigue classification tasks. It is also adaptable to various types of fatigue data and exhibits robustness against noise and outliers in the dataset (Boateng et al., 2020). LightGBM utilizes a histogram-based algorithm that enables the rapid processing of fatigue datasets, improving both training and prediction efficiency. Furthermore, LightGBM can effectively address class imbalance issues by fine-tuning its parameters, thereby enhancing the model's classification performance (Zeng et al., 2019). From the above analysis, we conclude that all four machine learning algorithms discussed demonstrated the capability to classify fatigue states, each exhibiting distinct advantages. Consequently, this study employed these four algorithms for the automatic classification of HRV data collected during flight.

In prior studies that utilized HRV data to train machine learning models for fatigue classification, the standard 12-lead ECG acquisition equipment was predominantly employed (Rudroff, 2024). In practice, this approach is not feasible for pilots during flight as they are required to wear safety harnesses, and fighter pilots must wear anti-G suits and helmets, making the acquisition of such physiological signals impractical. Previous studies have generally classified fatigue into only two levels: fatigue and non-fatigue states (Hu and Lodewijks, 2020; Xiang et al., 2022; Chen et al., 2024). This is typically effective only for identifying severe fatigue, limiting its effectiveness for timely warning and intervention. To address these limitations, we propose an automatic classification method for flight fatigue based on wearable ECG and Rsp monitoring devices. This method involves training the four aforementioned machine learning algorithms to perform a three-level classification of flight fatigue using selected physiological features. This study aimed to evaluate the potential of wearable ECG devices in conjunction with machine learning algorithms for accurate classification and early detection of flight-related fatigue.

2 Materials and methods

2.1 Participants

We recruited 90 male Chinese volunteer pilots via volunteer programs with an average cumulative flight time of 896 ± 131 h. The mean age of the participants was 31.6 ± 6.8 years, with an average height of 173.5 ± 3.6 cm and an average weight of 73.8 ± 7.9 kg. All pilots were prohibited from consuming any medication, alcohol, coffee, tea, or other beverages that may influence their nervous systems' excitability within 1 week prior to the flight. Additionally, all volunteers underwent rigorous physical examinations within 1 month of the experiment to ensure that they had no mental illness or sleep-related disorders. All volunteers were thoroughly informed of the potential risks prior to participation, including but not limited to drowsiness and fatigue. Written informed consent was obtained from each

participant. The study protocol and measurement procedures were approved by the Ethics Committee of the Air Force Medical Center. Prior to formal testing, all participants received training to ensure proper operation of the measurement devices used in the study.

2.2 Data acquisition method

As shown in Figure 1, each participant was required to wear an ECG chest strap (SensEcho, Sensecho-1A, Beijing, China) prior to the experiment. This device was specifically developed for ECG monitoring during physical activity. It uses three adhesive electrodes to record single-lead ECG signals at a sampling rate of 200 Hz. Wires embedded in the elastic chest strap also permitted respiratory signal monitoring at a sampling rate of 25 Hz. The device incorporated hardware-based noise reduction and software filtering algorithms to ensure high-quality signal acquisition. Details about the hardware design and software algorithms may be obtained from (Zhang et al., 2019).

2.3 Experimental procedures

Prior to data collection, all participants were introduced to the Rating of Perceived Exertion (RPE) scale and received training on its use. This scale ranges from 6 to 20, and participants were instructed to select any integer within this range that best represented their perceived levels of fatigue. Based on the RPE scores, fatigue was classified into three levels: (1) “non-fatigue,” with RPE scores of 6–10; (2) “mild fatigue,” with scores of 11–16; and (3) “severe fatigue,” with scores of 17–20 (Boksem et al., 2005; Jiang et al., 2024; Rudroff, 2024). This refined classification of fatigue levels enabled the precise evaluation and analysis of the participants’ fatigue levels, providing strong support for the accurate interpretation of the experimental data.

Participants were asked to wear the ECG monitoring device 30 min prior to flight. Baseline data were collected in a seated position, during which participants also completed the RPE scale

assessment. ECG data were continuously monitored throughout the flight, and data collected within the first 5 min after flight completion were used for fatigue evaluation. Participants were again asked to complete the RPE scale immediately after data collection.

Each participant took part in at least one and up to seven data collection sessions. Flight durations varied depending on the nature of the mission, with a maximum duration of 120 min. In total, data from 392 sessions were collected. All aircraft operated by the pilots were jet fighters, and flight tasks comprised a variety of routine flight operation training exercises, including takeoff and landing procedures, instrument flight training, and aerobatic maneuver training. Table 1 describes the distribution of fatigue states across the sessions collected during the experiment. The results suggest clear individual differences in fatigue perception. For the same duration of flight, pilots reported varying levels of fatigue; however, level of fatigue may have also been influenced by the complexity of the flight tasks performed.

2.4 Data processing method

ECG signals were sampled at 200 Hz. To eliminate noise and motion artifacts, a low-pass filter with a cutoff frequency of 50 Hz was applied during signal preprocessing. This was conducted using the PhysioNet Cardiovascular Signal Toolbox (Clifford Lab, Emory University, Atlanta, GA, United States) in MATLAB R2018a (Vest et al., 2018). The selection of HRV features played a critical role in developing a model with high sensitivity and specificity for fatigue detection. Including irrelevant or redundant information related to fatigue can increase computational cost and the risk of overfitting during model training. Based on the preprocessed ECG and respiratory signals and previous findings (Zeng et al., 2019; Wang et al., 2023; Rudroff, 2024), 16 ECG-derived HRV features and one respiratory feature with demonstrated sensitivity to fatigue were chosen. Table 2 presents the names and descriptions of these features.

Existing research indicates that, during states of fatigue, the sympathetic nervous system (SNS) becomes suppressed, leading to slower energy metabolism, delayed responses, and reduced mental alertness. Conversely, the parasympathetic nervous system (PNS) becomes more active, inhibiting SNS activity, reducing heart rate, and promoting rest and recovery. In fatigued states, time-domain HRV features—such as mean heart rate (HR), mean R-R



TABLE 1 Distribution of pilot flight times and fatigue states.

Flight table fatigue state	Before flight	Within 30 min	30–60 min	60–90 min	90–120 min
No. of participants in non-fatigue state	392	52	17	2	0
No. of participants in mild fatigue state	0	31	66	27	25
No. of participants in severe fatigue state	0	2	5	81	84

TABLE 2 Extracted ECG and respiratory features.

Type of feature	Feature	Description
ECG features	Mean HR	Mean heart rate
	Mean RR	Mean R–R interval (the time between successive heartbeats)
	SDNN	Standard deviation of all RR intervals
	SDSD	Standard deviation of differences between adjacent RR intervals, reflecting short-term variability
	RMSSD	Root mean square of successive RR interval differences
	pNN20	Percentage of successive RR intervals that differ by more than 20 ms
	pNN50	Percentage of successive RR intervals that differ by more than 50 ms
	LF	Absolute power of low-frequency band (0.04–0.15 Hz)
	LFn	Normalized low-frequency (LF) power (i.e., LF divided by total power)
	HF	Absolute power of high-frequency band (0.15–0.4 Hz)
	HFn	Normalized high-frequency (HF) power (HF divided by total power)
	LF/HF	Ratio of low-frequency to high-frequency power
	S	Total HRV ellipse area, proportional to SD1 and SD2
	SD1	Standard deviation of instantaneous beat-to-beat variability (short-term HRV, perpendicular to line-of-identity in Poincaré plot)
SD2	Standard deviation of continuous long-term RR variability (along the line-of-identity in Poincaré plot)	
SD1/SD2	Ratio of short-term to long-term HRV	
Respiratory indicator	MeanRsp	Mean respiratory rate

interval (RR), root mean square of successive differences (RMSSD), standard deviation of normal-to-normal intervals (SDNN), and the percentage of successive RR intervals that differ by more than 50 ms (pNN50)—typically show an increase. These changes reflect enhanced parasympathetic regulation, as the body attempts to conserve energy and promote recovery by reducing heart rate and increasing HRV. In terms of frequency-domain indicators, low-frequency (LF) power increases, high-frequency (HF) power decreases, and the LF/HF ratio rises. These changes suggest a relative dominance of sympathetic activity and a reduction in parasympathetic regulation, indicating that the autonomic nervous system is in a state of stress during fatigue (Ni et al., 2022).

2.5 Machine learning methods

The performance of the classification models was evaluated using the following metrics: accuracy, defined as the proportion

of correctly predicted samples out of the total number of samples; precision, the proportion of true-positive predictions among all samples predicted as positive; recall, the proportion of true-positive predictions among all actual positive samples; and F1 score, the harmonic mean of precision and recall (Dauner et al., 2023). The mathematical formulas for these four evaluation metrics are defined as follows (Dauner et al., 2023; Zahid et al., 2023):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{1}$$

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

$$F1 = \frac{2 \text{ Precision Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

In this context, true positive (TP) refers to the number of samples correctly classified into a given category. False positive (FP) represents the number of samples incorrectly classified into a category to which they do not belong. True negative (TN) denotes the number of correctly classified samples that belong to another category, while false negative (FN) refers to the number of samples that belong to a given category but are misclassified into another category. The average values of these metrics were calculated to obtain the final performance evaluation of the model.

Tenfold cross-validation was used to assess model performance. To prevent data from the same participant being used in both the training and testing sets, each iteration used data from 81 participants for training and data from the remaining nine participants for testing. Average performance across the 10 iterations was reported as the final result. Additionally, to ensure that the test samples were representative, each testing set was required to include at least nine samples for each of the four flight duration categories (less than 30 min, 30–60 min, 60–90 min, and 90–120 min). In the above-mentioned method, all four models employed the same proportion of training and test sets. This was done to eliminate performance evaluation differences caused by different data division ratios, evaluate the performance stability of the models more accurately, and reduce the errors caused by the randomness of data division. As a unified data division ratio can simulate the same data environment, this enabled a better comparison of the model’s performance during practical use.

Using SPSS® Statistics version 26.0 (IBM Corp, Armonk, NY, United States), we performed statistical analyses to examine the differences in ECG and respiratory features across different states of fatigue. First, normality tests were performed to determine whether the data conformed to normal distribution. For features that met the assumption of normality, one-way repeated-measures analysis of variance (ANOVA) was used to assess differences among flight fatigue states. For features that did not meet the assumption of normality, the Kruskal–Wallis H-test was used to examine differences across multiple groups. A *p*-value of < 0.05 was considered statistically significant.

3 Results

3.1 Feature selection for model input

As shown in Tables 3, 4, homogeneity of variance tests was conducted for all ECG indicators. Mean HR, mean RR, SDNN, RMSSD, and SD2 met the assumption of homogeneity and were analyzed using ANOVA. In contrast, SDSD, pNN50, pNN20, LF, HF, LF/HF, LFn, HFn, SD1, SD1/SD2, and S did not meet the assumption and were analyzed using the Kruskal–Wallis *H*-test. Among these features, mean HR, mean RR, SDNN, RMSSD, pNN50, pNN20, LF, HF, LF/HF, LFn, HFn, and SD1/SD2 showed statistically significant differences across different states of fatigue. For the respiratory indicator, mean Rsp met the homogeneity of variance assumption and was tested using ANOVA, which indicated significant differences across fatigue states.

Based on the results presented above, 12 ECG features and one respiratory indicator exhibited statistically significant differences across the various flight fatigue states.

3.2 Training results of the flight fatigue classification model

A total of 12 ECG features and one respiratory indicator that demonstrated statistically significant differences across different

fatigue states were used as input variables for the models. The ECG features included mean HR, mean RR, SDNN, RMSSD, pNN50, pNN20, LF, HF, LF/HF, LFn, HFn, and SD1/SD2. The respiratory feature was mean Rsp. The DT, SVM, KNN, and LightGBM algorithms were used to construct the classification models.

During model development, the grid search algorithm was used to optimize the hyperparameters of the classifier to systematically search for the best hyperparameter combination within the specified range to improve the performance of the model. For each parameter combination, the average accuracy for that combination was calculated using tenfold cross-validation, and the parameter combination with the highest average accuracy of the classifier was selected as the final result. For the DT model, the parameter space was defined in terms of “max_depth”: [None, 3, 6, 9] and “criterion”: [“Gini,” “entropy”]; the optimal parameter combination was a maximum tree depth of 3 and a splitting criterion of Gini index. For the SVM model, the parameter space was defined in terms of “C”: [0.1, 1, 10] and “kernel”: [“linear,” “rbf,” “poly”], with the optimal regularization parameter set to C = 1 and the kernel parameter set to “rbf.” For the KNN model, the parameter space was defined as follows: “n_neighbors”: [3,5,7] and weights”: [“uniform,” “distance”]. The optimal configuration was found with K = 5 and “distance” as the weighting method. For the LightGBM model, the parameter space included “num_leaves”: [10, 20, 30, 40, 50, 60] and “learning_rate”: [0.01, 0.05, 0.1, 0.2, 0.3]. The optimal parameters were a maximum of 60 leaves and a learning rate of 0.01.

Table 5 lists the average accuracy, precision, recall, and F1 score for each of the four machine learning models using the same feature set to evaluate flight fatigue. The results indicated that the LightGBM model outperformed all other models across all performance metrics, achieving an accuracy of 0.886 ± 0.057, precision of 0.837 ± 0.064, recall of 0.861 ± 0.089, and F1 score of 0.849 ± 0.067. These results demonstrate the feasibility of using HRV features and the respiratory indicator to perform a three-level classification of flight fatigue.

Figure 2 shows the overall confusion matrix of the LightGBM model in the 10-fold cross-validation process. The row labels represent the true class of each sample, while the column labels represent the predicted class. The color shading of each cell represents the proportion of samples in that row assigned to the corresponding prediction category.

TABLE 3 Results for differences in ECG and respiratory features (ANOVA test).

Feature	Unit	F-value	P-value
Mean HR	Times/min	6.262	< 0.001**
Mean RR	Ms	5.731	< 0.001**
SDNN	Ms	4.152	0.028*
RMSSD	ms	3.621	0.031*
SD2	ms	1.061	0.513
Mean Rsp	Times/min	5.268	< 0.001**

**p < 0.01, *p < 0.05.

TABLE 4 Results of differences in ECG features (Kruskal–Wallis *H*-test).

Feature	Unit	H-value	P-value
SDSD	ms	5.182	0.075
pNN50	%	8.752	0.007**
pNN20	%	9.281	0.002**
LF	ms ²	6.723	0.047*
HF	ms ²	6.983	0.042*
LF/HF	/	20.678	< 0.001**
LFn	/	19.525	< 0.001**
HFn	/	15.627	< 0.001**
SD1	ms	5.293	0.073
SD1/SD2	/	16.739	< 0.001**
S	/	4.134	0.161

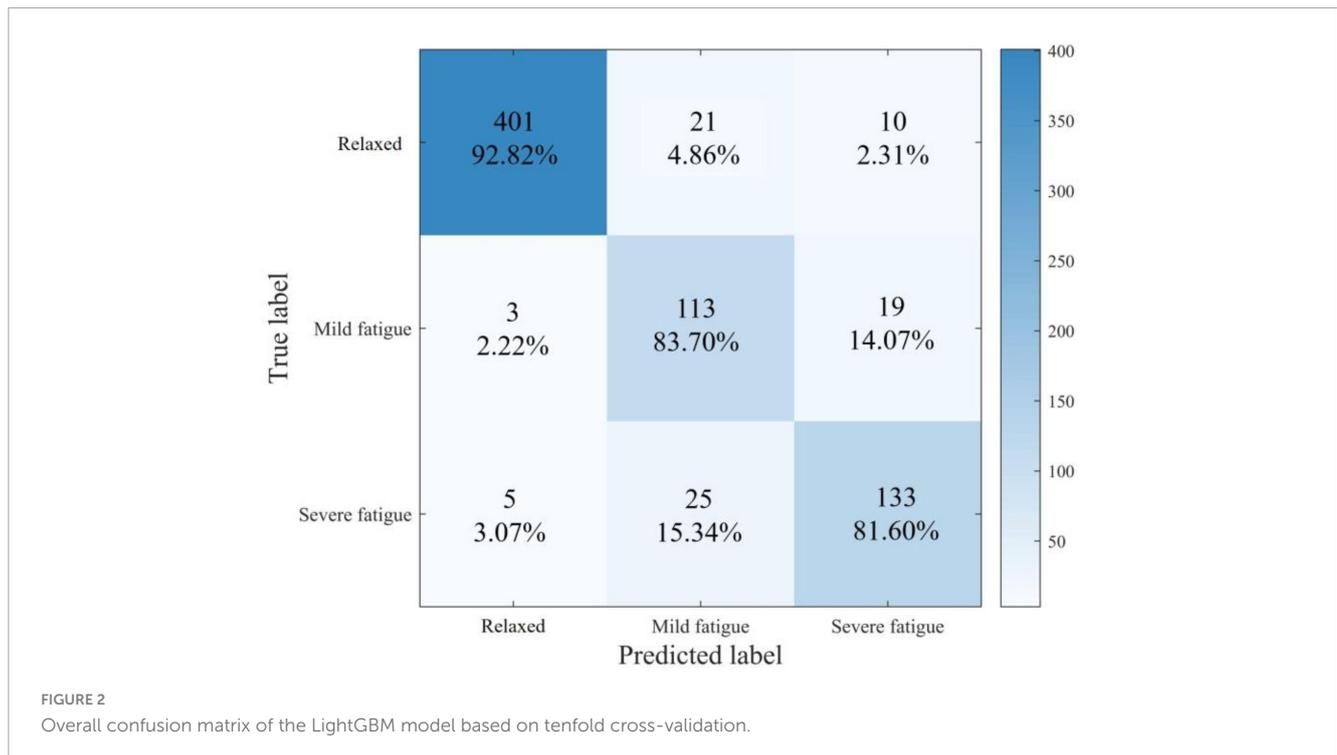
**p < 0.01, *p < 0.05.

4 Discussion

This study employed several HRV features—mean HR, mean RR, SDNN, RMSSD, pNN50, pNN20, LF, HF, LF/HF, LFn, HFn, and SD1/SD2—and a respiratory indicator, MeanRsp to assess

TABLE 5 Performance of four machine learning models using different feature sets.

Model	Accuracy	Precision	Recall	F1
DT	0.851 ± 0.073	0.821 ± 0.086	0.759 ± 0.087	0.783 ± 0.091
KNN	0.863 ± 0.062	0.831 ± 0.084	0.776 ± 0.069	0.789 ± 0.078
SVM	0.851 ± 0.068	0.796 ± 0.072	0.823 ± 0.079	0.824 ± 0.063
LightGBM	0.886 ± 0.057	0.837 ± 0.064	0.861 ± 0.086	0.849 ± 0.067



flight fatigue. While these features have been widely used in prior studies to assess physiological fatigue induced by exercise or mental fatigue resulting from prolonged cognitive tasks (Rudroff, 2024), we demonstrate their applicability in evaluating flight-related fatigue. This is plausible given the nature of flight fatigue, which is a complex state resulting from high workloads, demanding flight tasks, and circadian rhythm disruptions – factors that contribute to both the physiological and psychological components of fatigue (Sun et al., 2024).

As recorded in Table 5, LightGBM outperformed other models across all performance metrics. This superiority is likely attributable to its gradient-boosting DT framework, which captures nonlinear relationships and allows for more precise data partitioning and fitting, thereby enhancing model accuracy. In many real-world classification tasks, prior studies have also demonstrated that LightGBM achieves higher classification accuracy than traditional models, such as logistic regression and SVM, in various tasks (Wei et al., 2023). Moreover, previous studies have shown that LightGBM is effective for classifying muscle and mental fatigue caused by prolonged physical activity or driving (Zeng et al., 2019).

A previous study (Ni et al., 2022) combined LightGBM with a 12-lead ECG device for fatigue classification, achieving a maximum accuracy of 0.855 and an F1 score of 0.801. These results are comparable to those of the current study despite using a single-lead ECG device. This can primarily be attributed to comprehensive utilization of ECG and respiratory indicators, as well as the more refined optimization of these indicators. This supports the feasibility of using wearable ECG devices for accurate flight fatigue classification.

Based on the results exhibited in Figure 2, the model performed best when predicting the “non-fatigue” state, achieving 92.82% accuracy. It performed the worst in predicting “severe fatigue”

states, with only 81.60% accuracy. Examination of the confusion matrix reveals that this is mainly attributable to the model’s reduced sensitivity in distinguishing between “mild fatigue” and “severe fatigue” samples, misclassifying 15.34% of “severe fatigue” samples as “mild fatigue” and 14.07% of “mild fatigue” samples as “severe fatigue.” A key factor contributing to this limitation is class imbalance within the dataset. The dataset contained a higher number of “non-fatigue” samples than “mild fatigue” and “severe fatigue” samples. The limited number of fatigue samples restricted the classifier’s ability to learn robust features for those states, reducing its generalization performance when predicting fatigue states.

In terms of detecting flight fatigue, the LightGBM model in this study demonstrated an accuracy of 0.886 ± 0.057 , precision of 0.837 ± 0.064 , recall rate of 0.861 ± 0.086 , and F1 score of 0.849 ± 0.067 . These results indicate that a model trained using HRV and respiratory rate exhibits high overall performance in the three-class classification task of flight fatigue and is capable of accurately distinguishing samples across different fatigue levels. Previous studies focused on flight fatigue detection based on EEG have achieved similar high accuracy rates, with some reporting accuracies ranging from 0.8 to 0.9 (Lee et al., 2020; Wu et al., 2021; Alreshidi et al., 2023; Lee et al., 2023; Taheri Gorji et al., 2023). However, the acquisition of EEG signals typically necessitates complex equipment and specialized electrode placement procedures, and is susceptible to external electromagnetic interference and other factors, leading to relatively inferior signal stability and repeatability. By contrast, the chest strap method employed in this study for detecting HRV and respiratory rate enables more stable data collection, is less vulnerable to external interference, and achieves comparable model performance to the EEG-based approach, thereby demonstrating certain advantages.

Several other studies have utilized eye movement indicators, such as blink frequency and eyelid closure time, for flight fatigue detection, achieving accuracy rates that range from 0.7 to 0.8 (Borghini et al., 2014; Qin et al., 2021; Xue et al., 2022). However, eye-tracking devices face significant challenges related to installation and operational usability in real-world flight environments. Additionally, pilots may wear goggles or helmets during flights, which could interfere with the acquisition of eye movement indicators. The chest strap detection method employed in this study was unaffected by visual assistive devices, thereby enhancing the reliability of data collection and surpassing some studies based on eye movement indicators in terms of detection accuracy.

Finally, another category of research focused on detecting fatigue by analyzing pilots' operational actions, reaction times, and other behavioral characteristics, achieving accuracy rates that typically range from 0.6 to 0.75 (Zhou et al., 2022; Wang et al., 2023; Ghaderi and Saghafi, 2024). However, behavioral and operational indicators are susceptible to various factors, such as flight mission types and flight environments, and thus exhibit relatively low specificity. By contrast, the physiological indicators (HRV and MeanRsp) examined in this study can reflect the status of the pilot's autonomic nervous system and fatigue level more directly, thereby demonstrating an advantage in terms of detection accuracy.

This study employed the chest strap method for measuring HRV and MeanRsp, which offers high measurement convenience. The chest strap is a lightweight and easy-to-wear device that does not interfere with the pilot's operation during flight nor impose any additional burden. Its installation and use are straightforward, requiring only that it be wrapped around the chest and secured, without the need for complex settings or calibration processes. In addition, chest strap devices typically possess wireless data transmission capabilities, enabling real-time data transfer to the monitoring system and thereby facilitating continuous flight monitoring. Conversely, EEG detection necessitates the placement of multiple electrodes on the scalp, along with the use of auxiliary materials such as conductive paste to ensure optimal signal quality. This procedure is not only cumbersome and time-consuming, but may also cause discomfort for the pilot (Lee et al., 2020; Lee et al., 2023). In actual flight scenarios, particularly in emergency flight missions or during long-duration flights, pilots may be unwilling to allocate a significant amount of time for the installation and debugging of EEG equipment (Alreshidi et al., 2023; Taheri Gorji et al., 2023).

Eye-tracking devices generally require precise installation, either on the pilot's head or within the cockpit, along with a rigorous calibration process to ensure the accurate collection of eye movement data. During flight operations, the pilot's head movements and changes in the cockpit environment, such as fluctuations in lighting conditions, may interfere with the performance of the eye-tracking device, potentially resulting in data loss or an increase in measurement errors (Qin et al., 2021; Xue et al., 2022). However, the chest strap detection method remains unaffected by factors such as head movement and lighting conditions, thus ensuring a more stable and convenient measurement process. Additionally, there have been studies utilizing blood oxygen sensors placed on the fingers or earlobes to detect indicators such as pulse oxygen saturation for assessing flight fatigue (Sammito et al., 2022; Ibrahim et al., 2024). Although these

sensors provide a certain level of convenience, they are typically limited to detecting a single physiological indicator, whereas the chest strap employed in this study is capable of simultaneously measuring two critical physiological indicators. This provides more comprehensive fatigue-related information, thereby facilitating a more accurate assessment of pilots' fatigue states.

Furthermore, the model developed in this study was both trained and validated under actual in-flight scenarios, yielding results with high reliability. Given that flight fatigue is influenced by a multitude of factors, including flight altitude, pilot posture, and task load, the actual flight environment can effectively capture the true impact of these variables on pilots' fatigue states (Jun-Ya and Rui-Shan, 2023). Many prior studies on flight fatigue have been conducted using ground-based simulators that primarily collected physiological data under controlled laboratory conditions for fatigue analysis (Alzehairi et al., 2021; Chen et al., 2024). While they may enable in-depth exploration of the relationship between physiological indicators and fatigue, the absence of actual flight-related background and environmental factors (such as turbulence and air current variations) may limit the reliability and applicability of their findings in real-world flight fatigue monitoring. This study performed measurements and model training during actual in-flight operations, thereby avoiding the discrepancies that may arise between simulator-based environments and real flight conditions. Consequently, the findings gained greater reliability and practical applicability.

The automatic three-classification method for flight fatigue developed in this study holds broad application potential. Its high detection accuracy and convenient measurement process enable seamless integration into existing aviation flight monitoring systems, providing real-time fatigue state monitoring and early warning capabilities for pilots. By promptly identifying the fatigue state of pilots, particularly during the early stages of mild fatigue, appropriate measures (such as adjusting flight task schedules or reminding pilots to rest) can be implemented to prevent fatigue from escalating, thereby significantly reducing the risk of fatigue-induced flight accidents. Some prior studies only achieved binary classification of fatigue versus non-fatigue states, which, while capable of identifying fatigue to a certain extent, may have limitations in effectively managing flight fatigue (Ramos et al., 2020). As pilot fatigue is a gradual process, the binary classification method fails to precisely quantify the severity of fatigue and struggles to detect fatigue in its early stages, thereby hindering timely implementation of targeted measures.

Moreover, some studies rely solely on a single fatigue indicator for detection (Lee et al., 2023; Taheri Gorji et al., 2023). Due to the lack of integration of multiple fatigue-related factors, these methods may yield incomplete or inaccurate results. In real-world flight scenarios, fatigue arises from the synergistic effects of various physiological and psychological factors. Consequently, detection methods that incorporate multiple indicators can provide a more accurate assessment of the fatigue state. This study integrated two critical physiological indicators and performed a comprehensive analysis using the LightGBM model. This approach enables more effective detection of flight fatigue and better aligns with the requirements of real-world flight fatigue monitoring, demonstrating broader application potential.

This study utilized a chest strap to measure pilots' HRV and MeanRsp during actual flight operations, and applied the

LightGBM machine learning model to achieve automated three-class classification of flight fatigue. This approach demonstrates certain advantages in terms of detection accuracy, measurement convenience, result reliability, and application potential. Compared with prior studies that relied on EEG, eye movement indicators, behavioral and operational metrics, as well as ground-based simulators, the method employed in this study is better suited for real flight environments and can provide more precise, reliable, and timely results for flight fatigue monitoring. Furthermore, while this study only utilized MeanRsp among the respiratory indicators, the chest strap employed can also measure inspiratory time, expiratory time, the ratio of inspiratory to expiratory time, and the high-frequency and low-frequency components of the respiratory wave. In future research, these additional indicators could be incorporated comprehensively to train machine learning models, potentially leading to even more accurate models.

This study has several limitations. Owing to the unique nature of the profession, the number of participants was relatively small, which might have affected the model's stability and generalizability. In the future, it is advisable to monitor the HRV and respiratory indicators of a larger number of pilots, particularly those involved in long-duration flights and high-complexity missions. Additionally, we did not separately investigate physical and mental fatigue but instead adopted a general classification of flight fatigue. As a result, the HRV features used in the model might not represent the optimal input set for model training. Future research should focus on identifying the most effective HRV features specific to the characteristics of flight fatigue. In future research, it is essential to differentiate between these two distinct fatigue states and to monitor various forms of flight-related fatigue independently. Finally, this study focused exclusively on the physiological parameters of male pilots. Given that women possess distinct physiological structures and are potentially more susceptible to fatigue, future research should also consider conducting separate classification analyses of flight fatigue among female pilots.

5 Conclusion

This study explored the use of HRV features and respiratory indicators to automatically assess flight fatigue involving a sample of 90 pilots based on machine learning. Statistical analysis was first used to select relevant features, which included 12 HRV features—mean HR, mean RR, SDNN, RMSSD, pNN50, pNN20, LF, HF, LF/HF, LFn, HF_n, and SD1/SD2—as well as one respiratory indicator, mean Rsp. These were used as inputs for four machine learning algorithms: DT, SVM, KNN, and LightGBM. Among these, LightGBM demonstrated the best performance, confirming the feasibility of using HRV and respiratory features for flight fatigue assessment.

Data availability statement

The original contributions presented in this study are included in this article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Air Force Medical Center. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

DG: Writing – original draft, Writing – review and editing, Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Supervision. CW: Conceptualization, Data curation, Writing – original draft. YQ: Data curation, Formal Analysis, Investigation, Writing – original draft. LS: Conceptualization, Data curation, Writing – original draft. AG: Data curation, Formal Analysis, Investigation, Writing – original draft. BT: Data curation, Writing – original draft. YZ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Writing – original draft, Writing – review and editing. GW: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Writing – original draft, Writing – review and editing.

Acknowledgments

We thank Editage (www.editage.com) for English language editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abe, E., Fujiwara, K., Hiraoka, T., Yamakawa, T., and Kano, M. (2014). "Development of drowsy driving accident prediction by heart rate variability analysis," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, (Piscataway, NJ: IEEE), 1–4.
- Alreshidi, I., Moulitsas, I., and Jenkins, K. W. (2023). Multimodal approach for pilot mental state detection based on EEG. *Sensors (Basel)* 23:7350. doi: 10.3390/s23177350
- Alzehir, A., Alhejaili, F., Wali, S., AlQassas, I., Balkhyour, M., and Pandi-Perumal, S. R. (2021). Sleep disorders among commercial airline pilots. *Aerosp. Med. Hum. Perform* 92, 937–944. doi: 10.2257/amhp.5809.2021
- Boateng, E. Y., Otoo, J., and Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review. *J. Data Anal. Information Process.* 8, 341–357. doi: 10.4236/jdaip.2020.84020
- Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cogn. Brain Res.* 25, 107–116. doi: 10.1016/j.cogbrainres.2005.04.011
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., and Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* 408, 189–215. doi: 10.1016/j.neucom.2019.10.118
- Chen, L., Dubrawski, A., Wang, D., Fiterau, M., and Hrvanek, M. (2016). Using supervised machine learning to classify real alerts and artifact in online multisignal vital sign monitoring data. *Crit. Care Med.* 44, e456–e463. doi: 10.1097/CCM.0000000000001660
- Chen, Y., Ge, H., Su, X., and Ma, X. (2024). Classification of exercise fatigue levels by multi-class SVM from ECG and HRV. *Med. Biol. Eng. Comput.* 62, 2853–2865. doi: 10.1007/s11517-024-03116-w
- Cosoli, G., Iadarola, G., Poli, A., and Spinsante, S. (2021). "Learning classifiers for analysis of blood volume pulse signals in IoT-enabled systems," in *Proceedings of the 2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT*, (Piscataway, NJ: IEEE), 307–312.
- Dauner, D. G., Leal, E., Adam, T. J., Zhang, R., and Farley, J. F. (2023). Evaluation of four machine learning models for signal detection. *Ther. Adv. Drug Saf.* 14:20420986231219472. doi: 10.1177/20420986231219472
- Faber, L. G., Maurits, N. M., Lorist, M. M., and De Lange, F. P. (2012). Mental fatigue affects visual selective attention. *PLoS One* 7:e48073. doi: 10.1371/journal.pone.0048073
- Gaines, A. R., Morris, M. B., and Gunzelmann, G. (2020). Fatigue-related aviation mishaps. *Aerosp. Med. Hum. Perform.* 91, 440–447. doi: 10.3357/amhp.5515.2020
- Ghaderi, A., and Saghafi, F. (2024). Enhancing pilot vigilance assessment: The role of flight data and continuous performance test in detecting random attention loss in short IFR flights. *J. Air Transport Manag.* 120:102673. doi: 10.1016/j.jairtraman.2024.102673
- Hu, X., and Lodewijks, G. (2020). Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *J. Safety Res.* 72, 173–187. doi: 10.1016/j.jsr.2019.12.015
- Ibrahim, M. S., Kamat, S. R., and Shamsuddin, S. (2024). An investigation of heart rate and oxygen saturation level (SpO2) in indicating driving fatigue. *Malaysian J. Med. Health Sci.* 20, 97–103. doi: 10.47836/mjms.20.3.14
- Jiang, J., Chen, G., Song, X., Lu, J., Wang, J., Ding, F., et al. (2024). Effects of chronotype on sleep, mood and cardiovascular circadian rhythms in rotating night shift medical workers. *Int. Arch. Occup. Environ. Health* 97, 461–471. doi: 10.1007/s00420-024-02060-4
- Jun-Ya, S., and Rui-Shan, S. (2023). Pilot fatigue survey: A study of the mutual influence among fatigue factors in the "work" dimension. *Front. Public Health* 11:1014503. doi: 10.3389/fpubh.2023.1014503
- Kirodiwal, A., Jahnavi, D., Dash, A., Ghosh, N., and Patra, A. (2022). Normal and abnormal classification of electrocardiogram: A primary screening tool kit. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2022, 2001–2004. doi: 10.1109/embc48229.2022.9871835
- Lee, D.-H., Jeong, J.-H., Kim, K., Yu, B.-W., and Lee, S.-W. (2020). Continuous EEG decoding of pilots' mental states using multiple feature block-based convolutional neural network. *IEEE Access* 8, 121929–121941. doi: 10.1109/ACCESS.2020.3006907
- Lee, D.-H., Jeong, J.-H., Yu, B.-W., Kam, T.-E., and Lee, S.-W. (2023). Autonomous system for EEG-based multiple abnormal mental states classification using hybrid deep neural networks under flight environment. *IEEE Trans. Syst. Man Cybernetics Syst.* 53, 6426–6437. doi: 10.1109/TSMC.2023.3282635
- Li, Y., Li, K., Wang, S., Chen, X., and Wen, D. (2022). Pilot behavior recognition based on multi-modality fusion technology using physiological characteristics. *Biosensors (Basel)* 12:404. doi: 10.3390/bios12060404
- Lin, C.-T., Chang, C.-J., Lin, B.-S., Hung, S.-H., Chao, C.-F., and Wang, I.-J. (2010). A real-time wireless brain-computer interface system for drowsiness detection. *IEEE Trans. Biomed. Circuits Syst.* 4, 214–222. doi: 10.1109/TBCAS.2010.2046415
- Ni, Z., Sun, F., and Li, Y. (2022). Heart rate variability-based subjective physical fatigue assessment. *Sensors* 22:3199. doi: 10.3390/s22093199
- Pan, T., Wang, H., Si, H., Li, Y., and Shang, L. (2021). Identification of pilots' fatigue status based on electrocardiogram signals. *Sensors* 21:3003. doi: 10.3390/s21093003
- Patel, M., Lal, S. K. L., Kavanagh, D., and Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Syst. Appl. Int. J.* 38, 7235–7242. doi: 10.1016/j.eswa.2010.12.028
- Qin, H., Zhou, X., Ou, X., Liu, Y., and Xue, C. (2021). Detection of mental fatigue state using heart rate variability and eye metrics during simulated flight. *Hum. Factors Ergon. Manufacturing Serv. Industries* 31, 637–651. doi: 10.1002/hfm.20927
- Ramkumar, M., Alagarsamy, M., Balakumar, A., and Pradeep, S. (2023). Ensemble classifier fostered detection of arrhythmia using ECG data. *Med. Biol. Eng. Comput.* 61, 2453–2466. doi: 10.1007/s11517-023-02839-6
- Ramos, G., Vaz, J. R., Mendonça, G. V., Pezarat-Correia, P., and Gamboa, H. (2020). Fatigue evaluation through machine learning and a global fatigue descriptor. *J. Healthc Eng.* 2020:6484129. doi: 10.1155/2020/6484129
- Rudroff, T. (2024). Revealing the complexity of fatigue: A review of the persistent challenges and promises of artificial intelligence. *Brain Sci.* 14:186. doi: 10.3390/brainsci14020186
- Sammito, S., Cyrol, D., and Post, J. (2022). Fatigue and ability to concentrate in flight attendants during ultra-long-range flights. *High. Alt. Med. Biol.* 23, 159–164. doi: 10.1089/ham.2021.0173
- Sun, J. Y., Liao, Y., Guo, H., and Jia, H. (2024). Fatigue risk assessment for flight crews flying across time zones in different directions to the east or west during the COVID-19 pandemic in China. *BMC Public Health* 24:3561. doi: 10.1186/s12889-024-20856-4
- Taheri Gorji, H., Wilson, N., VanBree, J., Hoffmann, B., Petros, T., and Tavakolian, K. (2023). Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight. *Sci. Rep.* 13:2507. doi: 10.1038/s41598-023-29647-0
- Vest, A. N., Poian, G. D., Li, Q., Liu, C., Nemati, S., Shah, A. J., et al. (2018). An open source benchmarked toolbox for cardiovascular waveform and interval analysis. *Physiol. Meas.* 39:105004. doi: 10.1088/1361-6579/aae021
- Wang, H., Han, M., Avouka, T., Chen, R., Wang, J., and Wei, R. (2023). Research on fatigue identification methods based on low-load wearable ECG monitoring devices. *Rev. Sci. Instrum.* 94:045103. doi: 10.1063/5.0138073
- Wei, W., Li, Y., and Huang, T. (2023). Using machine learning methods to study colorectal cancer tumor micro-environment and its biomarkers. *Int. J. Mol. Sci.* 24:11133. doi: 10.3390/ijms241311133
- Wilson, N., Guragain, B., Verma, A. K., Archer, L., and Tavakolian, K. (2020). *Blending Human and Machine: Feasibility of Measuring Fatigue Through the Aviation Headset*. Los Angeles, CA: SAGE.
- Wingelaar-Jagt, Y. Q., Wingelaar, T. T., Riedel, W. J., and Ramaekers, J. G. (2021). Fatigue in aviation: Safety risks, preventive strategies and pharmacological interventions. *Front. Physiol.* 12:712628. doi: 10.3389/fphys.2021.712628
- Wu, E. Q., Zhu, L.-M., Li, G.-J., Li, H.-J., Tang, Z., Hu, R., et al. (2021). Nonparametric hierarchical hidden semi-Markov model for brain fatigue behavior detection of pilots during flight. *IEEE Trans. Intell. Transportation Syst.* 23, 5245–5256. doi: 10.1155/2022/3575130
- Xiang, H., Zhang, Q., Wang, J., Wang, S., Liao, Z., Sun, L., et al. (2022). Driving fatigue recognition model based on heart rate variability and respiratory rate. *J. Army Med. Univers.* 44, 1299–1306. doi: 10.16016/j.2097-0927.202203057
- Xue, H., Chen, X., Zhang, X., Ma, Z., and Qiu, K. (2022). "Pilot fatigue evaluation based on eye-movement index and performance index," in *Proceedings of the International Conference on Human-Computer Interaction*, (Berlin: Springer), 446–460.
- Zahid, M. U., Kiranyaz, S., and Gabbouj, M. (2023). Global ECG classification by self-operational neural networks with feature injection. *IEEE Trans. Biomed. Eng.* 70, 205–215. doi: 10.1109/tbme.2022.3187874
- Zeng, H., Yang, C., Zhang, H., Wu, Z., Zhang, J., Dai, G., et al. (2019). A LightGBM-based EEG analysis method for driver mental states classification. *Comput. Intell. Neurosci.* 2019:3761203. doi: 10.1155/2019/3761203
- Zhang, Y., Yang, Z., Lan, K., Liu, X., Zhang, Z., Li, P., et al. (2019). "Sleep stage classification using bidirectional lstm in wearable multi-sensor systems," in *Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, (Piscataway, NJ: IEEE), 443–448.
- Zhou, B., Chen, B., Shi, H., Xue, L., Ao, Y., and Ding, L. (2022). SEMG-based fighter pilot muscle fatigue analysis and operation performance research. *Med. Novel Technol. Devices* 16:100189. doi: 10.1016/j.medntd.2022.100189