


## Article

# Pilot Behavior Recognition Based on Multi-Modality Fusion Technology Using Physiological Characteristics

Yuhan Li , Ke Li <sup>\*</sup>, Shaofan Wang, Xiaodan Chen and Dongsheng Wen <sup>\*</sup>

National key Laboratory of Human Machine and Environment Engineering, School of Aeronautical Science and Engineering, Beihang University, Beijing 100191, China; ffflora@buaa.edu.cn (Y.L.); by2005516@buaa.edu.cn (S.W.); xiaodan\_chen@buaa.edu.cn (X.C.)

<sup>\*</sup> Correspondence: like@buaa.edu.cn (K.L.); d.wen@buaa.edu.cn (D.W.)

**Abstract:** With the development of the autopilot system, the main task of a pilot has changed from controlling the aircraft to supervising the autopilot system and making critical decisions. Therefore, the human–machine interaction system needs to be improved accordingly. A key step to improving the human–machine interaction system is to improve its understanding of the pilots’ status, including fatigue, stress, workload, etc. Monitoring pilots’ status can effectively prevent human error and achieve optimal human–machine collaboration. As such, there is a need to recognize pilots’ status and predict the behaviors responsible for changes of state. For this purpose, in this study, 14 Air Force cadets fly in an F-35 Lightning II Joint Strike Fighter simulator through a series of maneuvers involving takeoff, level flight, turn and hover, roll, somersault, and stall. Electro cardio (ECG), myoelectricity (EMG), galvanic skin response (GSR), respiration (RESP), and skin temperature (SKT) measurements are derived through wearable physiological data collection devices. Physiological indicators influenced by the pilot’s behavioral status are objectively analyzed. Multi-modality fusion technology (MTF) is adopted to fuse these data in the feature layer. Additionally, four classifiers are integrated to identify pilots’ behaviors in the strategy layer. The results indicate that MTF can help to recognize pilot behavior in a more comprehensive and precise way.

**Keywords:** MTF; physiological; behavior recognition; pilot; machine learning; multi-modal



**Citation:** Li, Y.; Li, K.; Wang, S.; Chen, X.; Wen, D. Pilot Behavior Recognition Based on Multi-Modality Fusion Technology Using Physiological Characteristics.

*Biosensors* **2022**, *12*, 404. <https://doi.org/10.3390/bios12060404>

Received: 29 April 2022

Accepted: 8 June 2022

Published: 12 June 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the continuous development of autopilot systems and flight assistance systems, pilot performance and air safety have been improved significantly. For example, the DLR project ALLFlight provides pilot assistance during all phases of flight [1] and includes an automatic aircraft upset-recovery system (AURS) that supports the pilot in recovering from any upset in a manner that is both manually assisted and automatic [2] and an assisting flight control system (FCS) that can simplify piloting, reduce pilot workload, and improve the system’s reliability [3]. Owing to these developments, the main task of the pilot has changed from controlling the aircraft to supervising the autopilot system and making critical decisions [4]. Therefore, it is essential to improve autopilot systems via better human–machine interaction. In addition to optimizing the display interface and improving the automation technology, ensuring that the flight assistance system has a better understanding of the pilots’ physiological status is a key step to improving human–machine interaction. To correctly evaluate the effects of different flight behaviors at different difficulty levels on pilot status, it is necessary to monitor various physiological parameters of pilots under different flight behaviors and predict their workload.

Traditionally, the pilots’ workload under different flight operations is assessed based on expert interviews and professional questionnaires, such as subjective rating scales [5]. However, these indirect analyses have many problems. For example, assessing workload through questionnaire evaluation is subjective, and the result is greatly influenced by

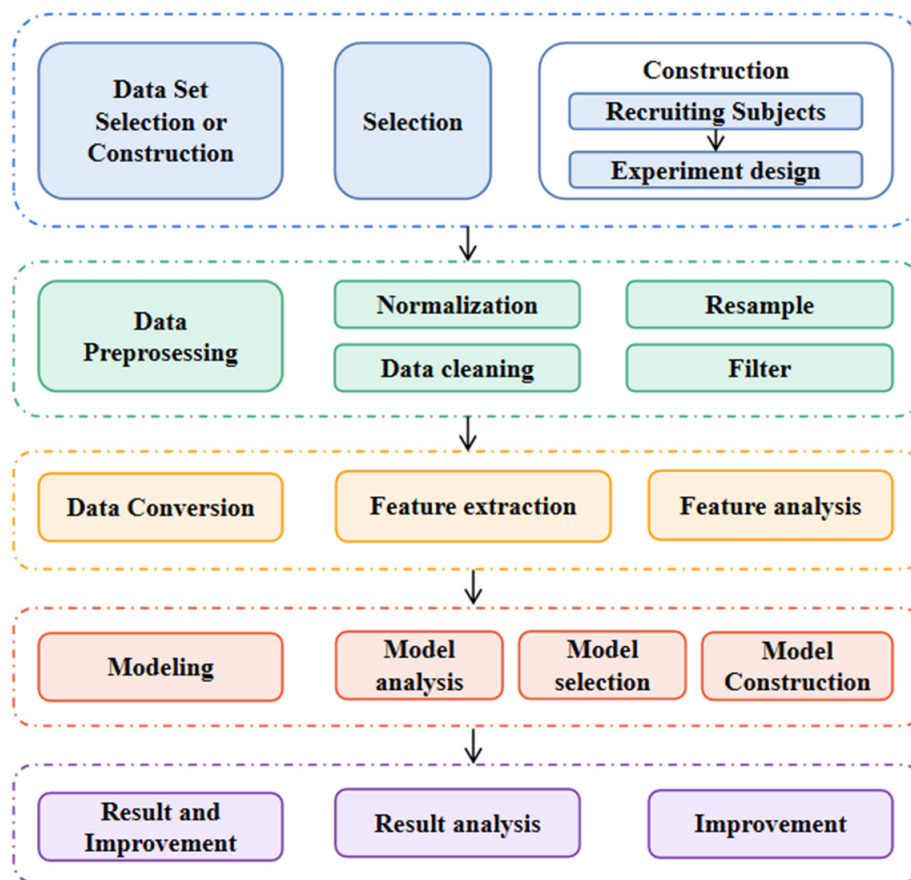
different individuals. Furthermore, the pilot sometimes needs to interrupt the flight in order to answer the questionnaire, which cannot be carried out in real scenarios. Moreover, the questionnaire can only be conducted at discrete time points and cannot provide information on continuous, task-related workload or physiological state changes [6]. To solve these problems, a more objective and time-continuous method, which recognizes the human state according to physiological parameters, has been proposed. For instance, Law [7] assessed pilot workload by analyzing electro cardio (ECG) and electroencephalogram (EEG) data and showed that RR interval (RRI) and the root mean square of successive differences (RMSSD) decreased with increasing flight mission difficulty. Feleke [8] used myoelectricity (EMG) to detect the driver's intention to make an emergency turn and found that the driver's EMG rose sharply before the turn. Some studies adopted a single physiological signal for the identification of human states. For example, Rahul [9] used ECG to estimate pilot fatigue, and Matthew [10] used functional near-infrared spectroscopy to identify the mental load of helicopter pilots. However, recognition by a single physiological signal suffers from poor stability, reduced data information, low reliability, and low discriminatory ability. Multiple physiological signals have therefore been proposed. For example, Pamela et al. [11] used electrodermal and electrocardiographic information to monitor the sympathetic response of drivers. Nevertheless, the research on multi-modal physiological signals for human state recognition still offers great development opportunities. Despite some studies [12–14] involving a combination of physiological signals, there is still no consensus on many elements, for instance, which indicators should be used as input, at which level to fuse, what kind of the classification models should be adopted [15,16], and how to deal with the problem of insufficient available data. Consequently, it is essential to create improved frameworks that provide greater robustness in identifying the status of pilots during different operations.

Besides fusing physiological information, the fusion of classification models to achieve optimal results is also a way to improve the framework. This requires the investigation of several classifiers which are to be integrated. These candidate classifiers have been used frequently in previous studies. For example, Wang [17] analyzed skin conductance, oximetry pulse, and respiration signals using Hilbert transform and a random forest classifier algorithm and evaluated the model by Accurate Rate, MSE, ROC, F1\_score, Precision, and Recall. Hu [18] determined whether a driver is fatigued based on EEG signals using a gradient-boosting decision tree (GBDT), reaching 94% accuracy, and the k-nearest neighbor algorithm, support vector machine, and a neural network were also employed as a comparison. Vargaslopez [19] compared different machine learning algorithms, such as SVM and MLP, and showed that SVM obtained the best result in detecting stress after normalization.

This study has two objectives: (i) to develop a model that recognizes pilots' behavior based on multi-modal fusion technology (MFT) that uses physiological characteristics; and (ii) to analyze the physiological indicators influenced by the pilot's behavioral status and their association with flight difficulty. Firstly, a physiological database of pilots in different behavioral states is formed through experimental protocol design and data collection. Then, the correlation between pilot behavioral states and physiological data is analyzed in terms of raw data, time-domain features, and frequency-domain features, and the difficulty of flight behavior is evaluated. Additionally, the classification model for recognizing pilots' behavior is designed based on MFT. Finally, the performance of each machine learning classification model and the proposed model is analyzed and compared from the perspectives of accuracy, precision, F1 score, and mean squared error (MSE). As a result, this paper makes the following contributions: (i) a new physiological data set from 14 pilots with wearable physiological measurement devices for multi-modal data input; (ii) a classification model based on MFT, the accuracy of which reaches 98.15%; and (iii) an analysis of the effects of flying behaviors on a pilot's physiological measurements from an objective point of view.

## 2. Methods

The study was conducted according to the process of data mining [20], which is shown in Figure 1.



**Figure 1.** The process of data mining.

### 2.1. Data Construction

#### 2.1.1. Participants

In this study, data were collected from 14 Air Force flight cadets (age  $21 \pm 2$  years, weight  $70 \pm 6$  kg, all male). All subjects were professionally trained and proficient in flying the F-35 Lightning II Joint Strike Fighter used for flight testing. All subjects were right-handed, had normal vision, and had no history of cardiac, neurological, or psychiatric disease. All subjects rested well and were healthy during the experiment.

#### 2.1.2. Flight Platform and Task Details

In this study, the pilots performed various flight maneuvers in a flight simulator. The flight simulator consisted of the obutto ergonomic workstation, Saitek Pro Flight X-56 Rhino Stick (for controlling direction, roll, and pitch in the air), Saitek Pro Flight X-56 Rhino Throttle (for controlling throttle), foot pedals (for controlling direction on the ground), and flight simulation software Xplane (as shown in Figure 2). It provided a flight simulation environment with high fidelity and good immersion. The simulation scenario was 10 nautical miles from Beijing Capital Airport.



**Figure 2.** Illustration of the experimental platform.

Before the experiment began, each subject had 5 min to familiarize themselves with the flight simulator. Then, the experiment started officially, and each subject flew as instructed, completing five types of operation in each flight: takeoff, level flight, hover and turn, roll, and somersault, and there was a chance that a stall scenario might occur during the operation. Each subject performed the experiment once. All the maneuvers were carried out for about 4 min to achieve a balanced data set, except for the stall scenes, which occurred randomly and took about 3 min. Operations that met the criteria were counted. After the experiment, all subjects were asked to rate the difficulty of the different flight behaviors (1 to 10, with 10 being the most difficult) to obtain their subjective opinions.

### 2.1.3. Data Acquisition

The ErgoLAB Human Factors Physiological Recorder was used to collect physiological data from the subjects. ECG data were collected from the subject's chest via Ag/AgCl electrodes (sampling rate of 512 Hz) and also from the earlobe via a pulse sensor (PPG) (sampling rate of 64 Hz). GSR data were collected from the palm of the subject's right hand via Ag/AgCl electrodes (sampling rate of 64 Hz); EMG data were collected from the radial carpal extensor muscle of the subject's right lower arm via Ag/AgCl electrodes (sampling rate of 1024 Hz); RESP data were collected from the subject's chest cavity via a chest strap respiratory sensor (sampling rate of 64 Hz); SKT data were collected from the subject's right lower arm via a skin temperature sensor (sampling rate of 32 Hz). All data were transmitted via Bluetooth sensors to the synchronization platform for processing.

To reduce industrial frequency interference and environmental interference, high viscosity electrodes were used to ensure good contact when collecting data. To avoid the effect of temperature and humidity on data acquisition, the laboratory was kept in a dry condition, and the temperature was maintained at 22–24 degrees Celsius. To reduce motion disturbance, the subjects were told to avoid large movements as much as possible.

### 2.2. Data Pre-Processing

Data pre-processing is a time-consuming but essential step and can help improve the classification accuracy and performance [21], as there are obvious problems in raw data, including: missing data, data noise, data redundancy, unbalanced data sets, etc. Data pre-processing methods can be summarized into three categories [22], as shown in Table 1. It is believed that the pre-processing method should be selected and adjusted according to the data set and the specific task [23,24]. Our data set was subject to time continuity,

individual variability, and a high sampling rate, and was affected by motion and noise. As a result, the pre-processing methods described below were adopted.

**Table 1.** Data pre-processing methods.

| Category             | Aim   | Methods   |
|----------------------|---|---|
| Data cleaning        | Handling of anomalies in data values          | Missing value processing (abandon/replacement)<br>Ectopic values processing<br>Outlier and noise handling |
| Data integration     | Increase sample data size                     | Combining multiple data sets into a single data set   |
| Data standardization | Scales the sample values to a specified range | Discretization<br>Dualization<br>Normalization (min–max, z-score)<br>Function transformation              |

### 2.2.1. Normalization

Due to individual differences, each subject's measured physiological signal exhibited a different scale or magnitude. Therefore, it was necessary to normalize the physiological signal for each subject to obtain the same scale or magnitude [25]. Normalization methods included min–max, z-score, baseline, etc. In this study, the changes in physiological indicators varied greatly between individuals during flight and at rest. Adopting the baseline method may have made the data subject to data scale inconsistency. To ensure that the data distribution was not altered and to eliminate the influence of dimensionality between indicators, the ECG signal, EDA signal GSR signal, EMG signal, and SKT signal were normalized separately using Equation (1):

$$\tilde{x} = \frac{x - \mu}{\sigma} \quad (1)$$

### 2.2.2. Ectopic and Missing Value Processing

For ectopic values, ectopic value detection was performed first. The percentage detection method, which defines data points with more than 20% variation from the previous data point as ectopic, was adopted. Time windows (30 s) with more than 20% ectopic or missing values were removed [26,27]. The remaining ectopic and missing values were replaced with the mean value of the 11 adjacent data points centered on the missing or ectopic value using Equation (2):

$$x'(n) = \text{mean}(x(n + m)), \text{ where } |m| \leq \frac{w - 1}{2} \quad (2)$$

### 2.2.3. Downsampling and Filtering

Since the sampling rate of physiological signals is usually much higher than what is needed and the sampling rates vary [28], downsampling is needed. Before downsampling, heart rate variability (HRV) features and frequency domain features were extracted first. Then, to ensure the data consistency in timing and the ability to represent the pilot's behavior [29], all physiological signals were downsampled to 2 Hz based on the minimum time of flight maneuvers (5 s for subject 2's roll maneuver).

In addition, the physiological data were filtered to reduce interference from motion artifacts, human noise, and other factors. Filtering mainly includes noise reduction, high pass, band resistance, and low pass. Wavelet noise reduction was employed to remove the baseline noise and drift signal from the signal. Gaussian filters, on the other hand, transform the data by building a mathematical model, smoothing it, and reducing the effect of noise [30–34]. The detailed data processing methods are shown in Table 2.



**Table 2.** Detailed data processing.

|                 | ECG     | GSR      | EMG     | RESP    | SKT             |
|-----------------|---------|----------|---------|---------|-----------------|
| Noise reduction | Wavelet | Gaussian | Wavelet | Wavelet | Sliding average |
| High pass       | 1 Hz    | /        | 5 Hz    | /       | 5 Hz            |
| Band stop       | 50 Hz   | 50 Hz    | 50 Hz   | 50 Hz   | 50 Hz           |
| Low pass        | 40 Hz   | 5 Hz     | 500 Hz  | 20 Hz   | 200 Hz          |

### 2.3. Data Conversion

Data conversion means transforming the data into a suitable form for data mining through aggregation. To transform the processed data into features that can be input into the classifier and to make the features accurately describe the data, the physiological data were analyzed in the time and frequency domains. To ensure the accuracy and continuity of time- and frequency-domain data analysis, we selected 30 s as the time window and 10 s as the step to obtain the continuous time- and frequency-domain indexes.

#### 2.3.1. Time-Domain Features

Time-domain information is a depiction of the signal waveform with time as the variable. Time-domain features include dimensional characteristic parameters, as well as dimensionless characteristic parameters [35]. In this paper, the main parameters used were the dimensional characteristics, including the indicators associated with heart rate variability, the mean, the standard deviation, and the RMS. The expressions are shown in Table 3.

**Table 3.** Time-domain features mathematical representation.

| Parameters             | Description   |
|------------------------|---|
| Mean                   | $\bar{x} = \frac{1}{N_s} \sum_{i=1}^{N_s} x(i)$                                 |
| Standard Deviation     | $F_s = \left( \frac{1}{N_s} \sqrt{\sum_{i=1}^{N_s} (x(i) - \bar{x})^2} \right)$ |
| Root Mean Square (RMS) | $F_s = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (x(i))^2}$                          |

#### 2.3.2. Frequency-Domain Features

Frequency-domain analysis observes signal characteristics by frequency. The analysis in the time domain is more intuitive, while the representation in the frequency domain is more concise [36]. The frequency-domain feature parameters used in this study included HRV frequency-domain analysis, EMG signal frequency-domain analysis, and RESP signal frequency-domain analysis. The expressions are shown in Table 4.

**Table 4.** Frequency-domain features mathematical representation.

| Parameters           | Description  |
|----------------------|--|
| Power                | Power in the frequency band  |
| Median Frequency     | $\int_0^{MF} P(\omega) d\omega = \int_{MF}^{\infty} P(\omega) d\omega = \frac{1}{2} \int_0^{\infty} P(\omega) d\omega$ |
| Mean Power Frequency | $MPF = \frac{\int_0^{\infty} \omega P(\omega) d\omega}{\int_0^{\infty} P(\omega) d\omega}$                             |

#### 2.3.3. Multi-Modal Features Conversion

The pilot's behaviors are reflected in physiological and biological changes, including changes in heart beat, muscle activity, respiration, reflexes, etc. [37]. Therefore, these features were extracted from ECG, GSR, EMG, RESP, and SKT. Since physiological measurements change as the pilot's behavior changes, statistical features were extracted to

describe the variation in the measurements. As such, the variations of the time-domain features and frequency-domain features commonly used in ECG, GSR, EMG, RESP, and SKT measurements were analyzed to determine the most relevant features. The mathematical representation of these features extracted from the multi-dimensional signal measurements per time window is shown in Tables 3 and 4. Then, all features were fused according to the time series. A total of 28 features, as shown in Table 5, were derived at each time point.

**Table 5.** Multi-modal features.

| ECG       |          | GSR                |  | EMG                  |  | RESP               |  | SKT       |  |
|-----------|----------|--------------------|--|----------------------|--|--------------------|--|-----------|--|
| ECG value | HR value | SC value           |  | EMG value            |  | RESP value         |  | SKT value |  |
| SDSD      | NN       | mean               |  | standard deviation   |  | standard deviation |  |           |  |
| SDNN      | RMSSD    |                    |  | RMS                  |  | power              |  |           |  |
| pNN50     | pNN20    |                    |  | Integral EMG         |  | mean               |  |           |  |
| VLF       | ULF      | standard deviation |  | median frequency     |  |                    |  |           |  |
| LF        | HF       |                    |  | mean power frequency |  |                    |  |           |  |
| LF/HF     |          |                    |  | mean                 |  |                    |  |           |  |

The full details of abbreviations can be found in Abbreviations.

#### 2.3.4. Correlation Analysis

To ensure the independence of the features, the degree of correlation between the features must be analyzed. Pearson correlation coefficient  $r$  reflects the degree of linear correlation between two variables  $x$  and  $y$ . The value of  $r$  is between  $-1$  and  $1$ , and the larger the absolute value, the stronger the correlation [38,39]. The formula for the Pearson correlation coefficient  $r$  is shown in Equation (3):

$$r_{x,y} = \frac{cov(x,y)}{\sigma_x\sigma_y} = \frac{E((x - \mu_x)(y - \mu_y))}{\sigma_x\sigma_y} = \frac{E(xy) - E(x)E(y)}{\sqrt{E(x^2) - E^2(x)}\sqrt{E(y^2) - E^2(y)}} \quad (3)$$

Figure 3 is the correlation heatmap of the features. The different colors in the heatmap correspond to the correlation coefficient; the darker the color, the greater the correlation between the corresponding two features. The model tends to be influenced and to output wrong results if there is high correlation between features [40].

According to the Pearson correlation coefficients among the features, it can be seen that most of the features had low or no correlation, and a moderate correlation existed among SDNN, RMSSD, SDSD, ULF, VLF, and LF. High correlations were observed in HR and NN, SDSD and RMSSD, SC and SC mean, and EMG mean and iEMG. Features with high correlations were then removed in the Classification Improvement part of the procedure.

#### 2.4. Modeling

The features selected in Data Conversion part of the procedure were used as input for the classifier models. The model suitable for this study was selected based on the advantages and disadvantages of machine learning models. Several commonly used classifiers were trained on the data set, and their 10-fold cross-validation results are shown in Table 6. Candidate classifiers were picked by analyzing the validation results and reviewing previous studies [17–19,41–43]. Finally, 4 classifiers were selected due to their high accuracy and low MSE, including the Extra Tree Classifier (ETC), Decision Tree Classifier (DTC), Gradient Boosting Classifier (GBC), and XGBoost (XGBC). Extra Tree is a modification of Random Forest (RF); its principle is the same as RF, but its generalization ability is stronger than RF. DTC is an algorithm that divides the input space into different regions. Compared with other machine learning classification algorithms, DTC is relatively simple and efficient in data processing, which makes it suitable for real-time classification. GBC is an algorithm for regression and classification problems that integrates weak predictive models to produce a strong predictive model. XGBC is its enhanced version. The residuals of prediction are reduced so that the effect can be improved.

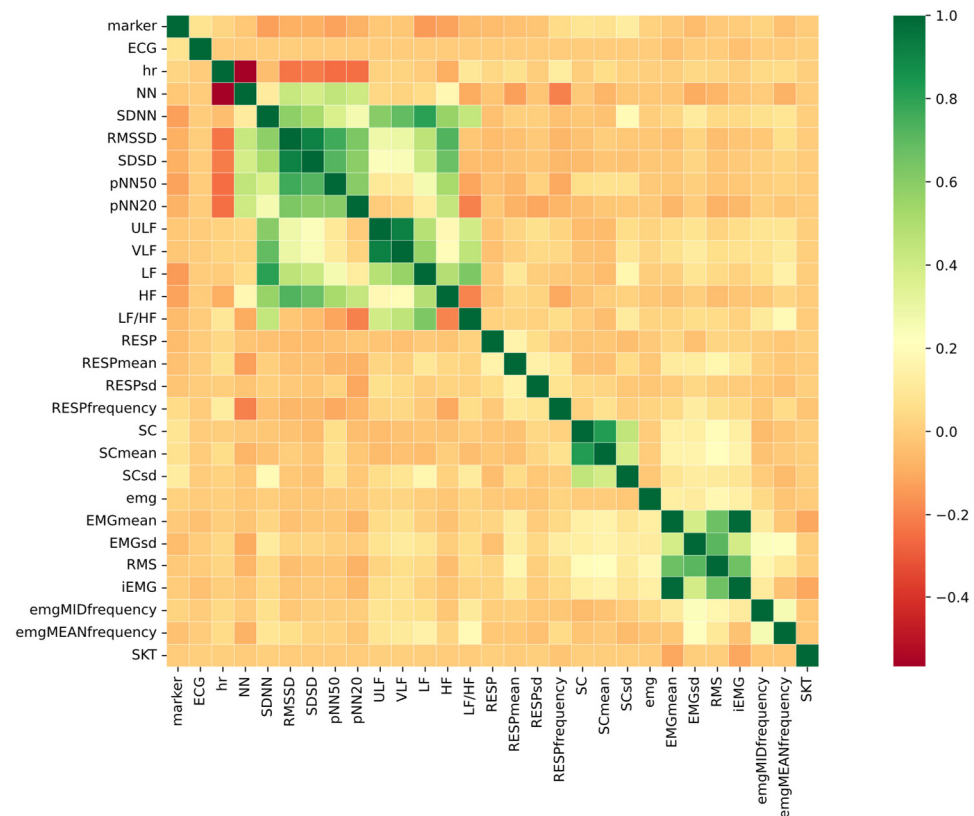


Figure 3. Pearson correlation analysis.

Table 6. The performance of several classification models.

| Model               | Mean Accuracy | Lowest Accuracy | MSE    |
|---------------------|---------------|-----------------|--------|
| Logistic Regression | 0.430         | 0.417           | 4.0109 |
| Naive Byes          | 0.362         | 0.339           | 4.1813 |
| AdaBoost            | 0.373         | 0.355           | 4.0822 |
| SVM                 | 0.441         | 0.432           | 3.9632 |
| K-Nearest Neighbor  | 0.952         | 0.947           | 0.2989 |
| ETC                 | 0.965         | 0.962           | 0.1765 |
| DTC                 | 0.964         | 0.960           | 0.1733 |
| GBC                 | 0.968         | 0.965           | 0.1755 |
| XGBC                | 0.967         | 0.962           | 0.1847 |

The goal of the modeling was to find the best classifier suitable for the study. To integrate the predictions of these classifiers and, thus, improve the prediction performance, a model was proposed that used a decision layer fusion technique. The advantage of decision level fusion is that the errors of the fusion model come from different classifiers, and the errors are often unrelated and independent of each other, which does not cause further accumulation of errors [44]. As such, we created our proposed model by assigning different weights to these classifiers and feeding their predictions to an integrated model, which is also known as ensemble learning. The voting algorithm is one of the simplest, most popular, and effective combiner schemes for ensemble learning [45,46]. It fuses the results from various learning algorithms to achieve knowledge discovery and better predictive performance [47–49]. Generally, there are two types of voting algorithm, majority voting (MV) and weighted voting (WV) [50]. Past studies showed the effectiveness of ensemble learning over the learning of a single learner [51,52]. After careful investigation, we proposed a model based on a weighted version of the simple majority voting, where each classifier contributes to the final output according to a reasonable weight  $\omega$ .  $\omega$  is calculated by the base classifiers along with the related, estimated probability distributions. It is also



known as the confidence level (CL), indicating the degree of support for the prediction. The framework and scheme of the model is shown in Figure 4 and Algorithm 1.

---

**Algorithm 1.** Weighted voting scheme.

---

**Input:**  
 $C_i$ : Classifier  
 $L_j$ : Labels of Data Set  
 $m$ : Ensemble Size  
 $n$ : the Number of Labels

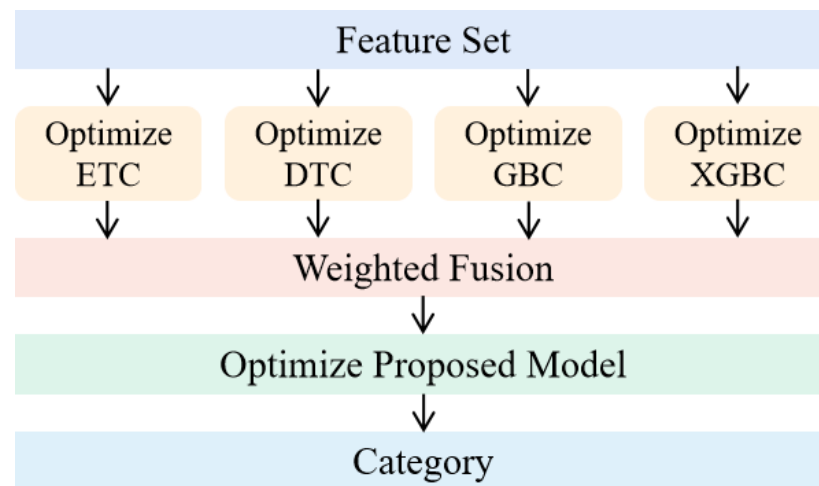
**Output:**  
the predicted class  $y_j$  from a single classifier  $C_i$   
the predicted class  $y^*$

```

for i = 1: m
  for j = 1: n
    compute  $p_{C_i L_j}$ , the probability assigned by  $C_i$  to class  $L_j$ 
     $\mu = \arg \max_{L=1, \dots, r} P_{C_i L_j}$ 
     $y_{C_i} = y_\mu$ 
  for j = 1: n
     $y(L_j) = \{i = 1, \dots, m: y_{C_i} == y_{L_j}\}$ 
    if  $y(L_j) == \emptyset$ 
       $g_{L_j} = 0$ 
    else
      for i in  $y(L_j)$  do
         $\omega_{C_i} = \max_{j=1, \dots, n} P_{C_i L_j}$ 
       $g_{L_j} = \sum_{i=1}^m \omega_{C_i}$ 
     $\mu = \arg \max_{L=1, \dots, r} g_{L_j}$ 
     $y^* = y_\mu$ 
  return  $y^*$ 

```

---



**Figure 4.** The framework of proposed model.

### 3. Results and Discussion

#### 3.1. Data Measures

Table 7 shows the results of the subjective difficulty assessment provided by the participants (assessed while watching a video playback directly following the flight).

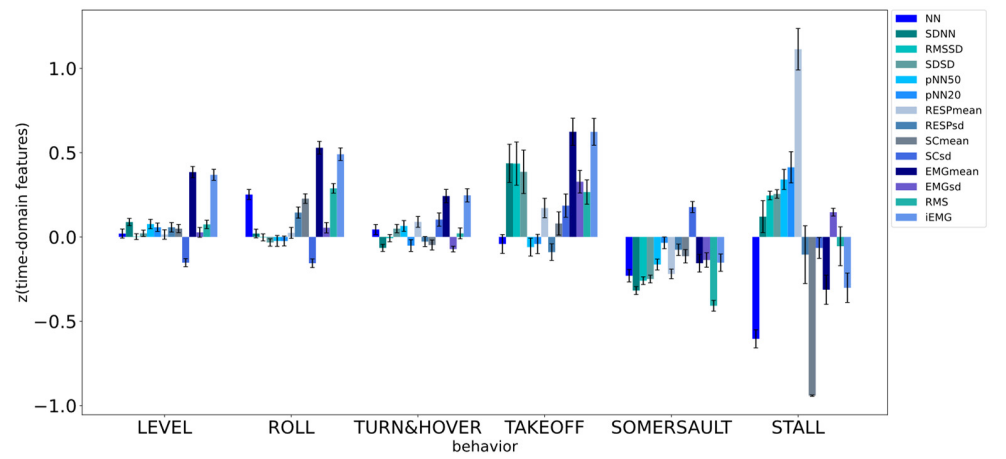
**Table 7.** Difficulty level rating according to subjective ratings.

|            | Stall | Somersault | Takeoff | Turn and Hover | Level Flight | Roll |
|------------|-------|------------|---------|----------------|--------------|------|
| Subject 1  | 10    | 8          | 6       | 6              | 4            | 3    |
| Subject 2  | 7     | 6          | 8       | 5              | 3            | 4    |
| Subject 3  | 8     | 9          | 5       | 6              | 4            | 4    |
| Subject 4  | 9     | 8          | 3       | 5              | 3            | 3    |
| Subject 5  | 8     | 5          | 2       | 3              | 2            | 2    |
| Subject 6  | 8     | 4          | 1       | 4              | 1            | 3    |
| Subject 7  | 6     | 8          | 1       | 5              | 2            | 2    |
| Subject 8  | 3     | 7          | 2       | 3              | 3            | 6    |
| Subject 9  | 8     | 9          | 5       | 6              | 5            | 5    |
| Subject 10 | 7     | 8          | 3       | 6              | 4            | 4    |
| Subject 11 | 8     | 7          | 3       | 5              | 4            | 3    |
| Subject 12 | 5     | 7          | 2       | 6              | 5            | 2    |
| Subject 13 | 7     | 9          | 3       | 7              | 5            | 3    |
| Subject 14 | 8     | 8          | 4       | 5              | 4            | 3    |
| Mean       | 7.29  | 7.36       | 3.43    | 5.14           | 3.50         | 3.36 |

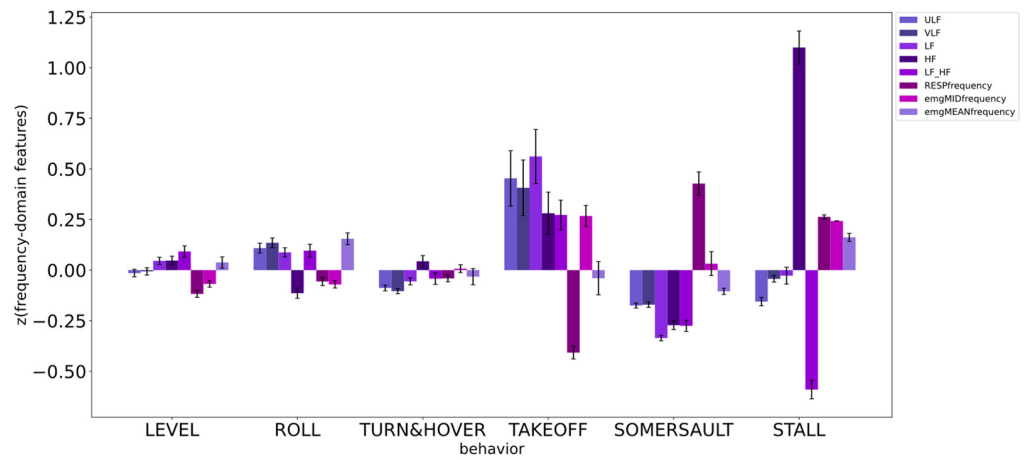
Time-domain measurements, such as NN, SDNN, RMSSD, pNN50, pNN20, the mean of RESP, the standard deviation of RESP, the mean of GSR, the standard deviation of GSR, the mean of EMG, the standard deviation of EMG, the RMS of EMG, and iEMG, and frequency-domain measurements, such as ULF, VLF, LF, HF, LF/HF, the frequency of RESP, the EMG median frequency, and the EMG mean power frequency, as well as raw data, such as ECG, HR, RESP, SC, EMG, and SKT, were recorded according to pilot behavior and plotted as a trend (Figure 5).

To objectively analyze the relationship between these physiological parameters and different behaviors, we analyzed them using the Kendall correlation analysis [53], as shown in Figure 6. It turned out that the ECG value, HR, SDNN, SDD, pNN50, pNN20, VLF, LF, HF, RESP value, REAP mean, RESP standard deviation, SC standard deviation, EMG value, RMS, iEMG, EMG mid frequency, EMG mean frequency, and SKT value had no evident correlations with behavior; LF/HF, SCL, the EMG mean, and EMG standard deviation had moderate correlations with behavior; and NN, ULF, RESP frequency, SCL, and the SCL mean had high correlations with behavior. Parameters with a dark-green background were negatively correlated with the degree of difficulty, and parameters with a light-green background were positively correlated with the degree of difficulty. The trends in physiological parameters were consistent with previous studies [54–58]. Nevertheless, the results showed slight differences in ULF compared to some studies [9,59–62] which believed that ULF cannot reflect a human’s workload level. After further literature research and analysis, our results were found to be justified. The pilot is in a complex state when performing flight maneuvers, inducing complicated physiological responses associated with cognition, fatigue, stress, concentration, or distraction, etc. So, studies which only take workload into consideration are neither comprehensive nor convincing with regard to difficulty rating. As a result, despite ULF being unable to reflect a human’s workload level, it could possibly be used in difficulty level recognition.

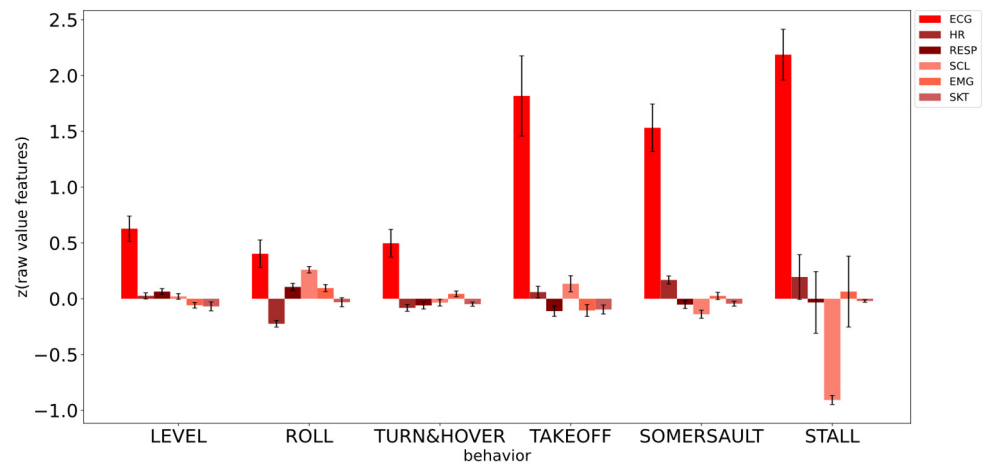
The objective analysis of the correlation between physiological indicators and behaviors helped to select key indicators for research. For example, if the condition was too limited to collect multiple physiological features, the features with higher correlation were selected for analysis. Additionally, this is the visual explanation of the feature importance analysis.



(a). Time-domain features trend according to pilot behavior



(b). Frequency-domain features trend according to pilot behavior



(c). Raw data trend according to pilot behavior

Figure 5. Data trend according to pilot behavior.

|  |                                      |  |                                  |  |   |                                    |
|--|--------------------------------------|--|----------------------------------|--|---|------------------------------------|
| ECG<br>$r = 0.27$<br>$p = 0.44$        | HR<br>$r = 0.27$<br>$p = 0.44$       | NN<br>$r = -0.55$<br>$p = 0.12$            | SDNN<br>$r = 0.14$<br>$p = 0.70$ | RMSSD<br>$r = 0.14$<br>$p = 0.70$              | SDDSD<br>$r = 0.0$<br>$p = 1.0$                 | pNN50<br>$r = -0.27$<br>$p = 0.44$ |
| pNN20<br>$r = -0.27$<br>$p = 0.44$     | ULF<br>$r = -0.55$<br>$p = 0.12$     | VLF<br>$r = -0.14$<br>$p = 0.70$           | LF<br>$r = -0.14$<br>$p = 0.70$  | HF<br>$r = -0.14$<br>$p = 0.70$                | LF/HF<br>$r = -0.41$<br>$p = 0.25$              | RESP<br>$r = -0.14$<br>$p = 0.70$  |
| RESP mean<br>$r = -0.14$<br>$p = 0.70$ | RESP SD<br>$r = -0.14$<br>$p = 0.70$ | RESP frequency<br>$r = 0.83$<br>$p = 0.02$ | SCL<br>$r = 0.41$<br>$p = 0.25$  | SCL mean<br>$r = 0.55$<br>$p = 0.12$           | SC SD<br>$r = 0.14$<br>$p = 0.70$               | EMG<br>$r = 0.0$<br>$p = 1.0$      |
| EMG mean<br>$r = -0.41$<br>$p = 0.25$  | EMG SD<br>$r = -0.41$<br>$p = 0.25$  | RMS<br>$r = 0.0$<br>$p = 1.0$              | iEMG<br>$r = 0.14$<br>$p = 0.70$ | EMG mid frequency<br>$r = -0.14$<br>$p = 0.70$ | EMG mean frequency<br>$r = -0.27$<br>$p = 0.44$ | SKT<br>$r = -0.27$<br>$p = 0.44$   |

Figure 6. Kendall correlation analysis.

### 3.2. Classification Model Performance

Figure 7 and Table 8 depict the 10-fold cross-validation (10-fold CV) results and leave-one-person-out (LOO) cross-validation results of various classifiers. The principles of LOO CV can be found in Appendix A [63]. In this study, the proposed model obtained a higher accuracy and generalization than other machine learning classification models. Although the mean accuracy of the proposed model was not much improved compared to the base model, the results of cross-validation show that it had the best stability. In addition, the robustness and generalization of the model was improved.

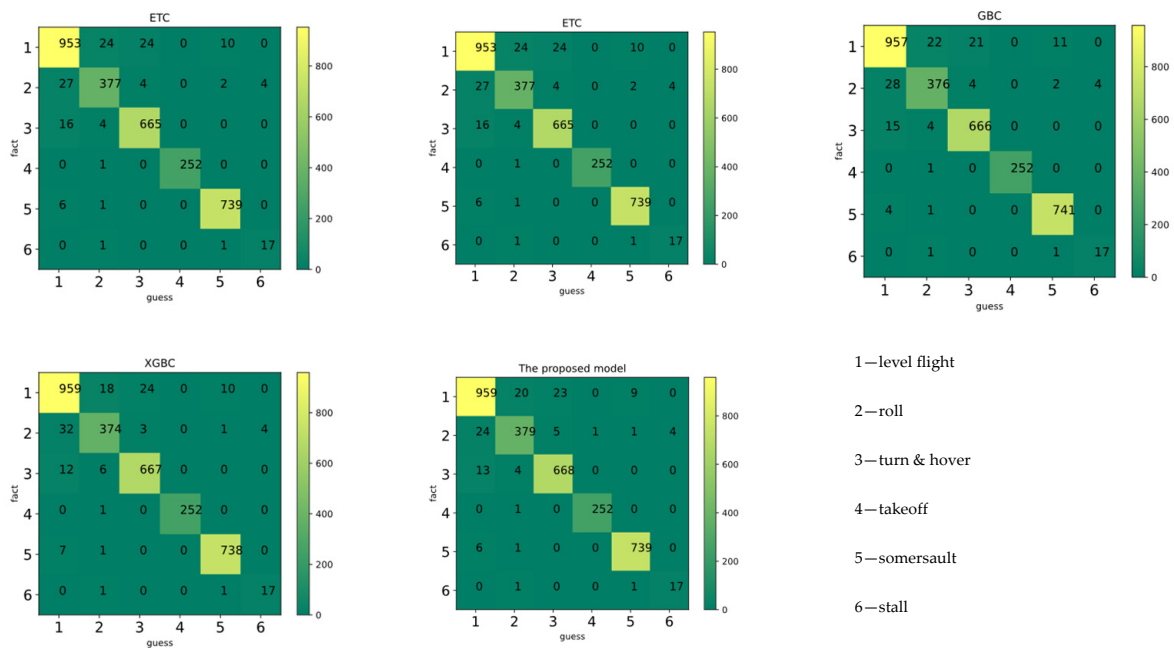


Figure 7. The confusion matrix of classifiers.

Table 8. The classification report of classifiers.

|                                 | ETC   |      |                |         |                    |        |
|---------------------------------|-------|------|----------------|---------|--------------------|--------|
|                                 | level | roll | turn and hover | takeoff | somersault         | stall  |
| precision                       | 0.96  | 0.91 | 0.97           | 1.00    | 0.99               | 0.97   |
| recall                          | 0.95  | 0.92 | 0.98           | 1.00    | 0.99               | 1.00   |
| F1                              | 0.96  | 0.91 | 0.97           | 1.00    | 0.99               | 0.98   |
| average accuracy for 10-fold CV |       |      |                | 0.9652  | MSE for 10-fold CV | 0.1765 |
| average accuracy for LOO CV     |       |      |                | 0.7817  | MSE for LOO CV     | 1.1199 |

Table 8. Cont.

| DTC                                    |       |      |                |                           |            |        |
|--|-------|------|----------------|---------------------------|------------|--------|
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.96  | 0.91 | 0.97           | 1.00                      | 0.99       | 0.97   |
| <b>recall</b>                          | 0.95  | 0.92 | 0.97           | 1.00                      | 0.99       | 1.00   |
| <b>F1</b>                              | 0.96  | 0.91 | 0.97           | 1.00                      | 0.99       | 0.98   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9642         | <b>MSE for 10-fold CV</b> |            | 0.1733 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.7062         | <b>MSE for LOO CV</b>     |            | 1.5760 |

| GBC                                    |       |      |                |                           |            |        |
|--|-------|------|----------------|---------------------------|------------|--------|
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.96  | 0.91 | 0.97           | 1.00                      | 0.99       | 0.97   |
| <b>recall</b>                          | 0.95  | 0.93 | 0.98           | 1.00                      | 0.99       | 0.93   |
| <b>F1</b>                              | 0.96  | 0.92 | 0.98           | 1.00                      | 0.99       | 0.95   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9677         | <b>MSE for 10-fold CV</b> |            | 0.1755 |
| <b>average accuracy for LOOCV</b>      |       |      | 0.7064         | <b>MSE for LOO CV</b>     |            | 1.4856 |

| XGBC                                   |       |      |                |                           |            |        |
|--|-------|------|----------------|---------------------------|------------|--------|
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.96  | 0.91 | 0.97           | 1.00                      | 0.99       | 0.97   |
| <b>recall</b>                          | 0.96  | 0.92 | 0.98           | 1.00                      | 0.99       | 1.00   |
| <b>F1</b>                              | 0.96  | 0.92 | 0.98           | 1.00                      | 0.99       | 0.98   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9674         | <b>MSE for 10-fold CV</b> |            | 0.1847 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.7473         | <b>MSE for LOO CV</b>     |            | 1.4894 |

| Proposed Model                         |       |      |                |                           |            |        |
|--|-------|------|----------------|---------------------------|------------|--------|
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.96  | 0.93 | 0.97           | 1.00                      | 0.99       | 0.97   |
| <b>recall</b>                          | 0.96  | 0.92 | 0.98           | 1.00                      | 0.99       | 1.00   |
| <b>F1</b>                              | 0.96  | 0.93 | 0.98           | 1.00                      | 0.99       | 0.98   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9693         | <b>MSE for 10-fold CV</b> |            | 0.1693 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.8094         | <b>MSE for LOO CV</b>     |            | 1.0606 |

3.3. Classification Improvement

The importance ranking of various classifiers for features was found by feature importance selection, as shown in Figure 8.

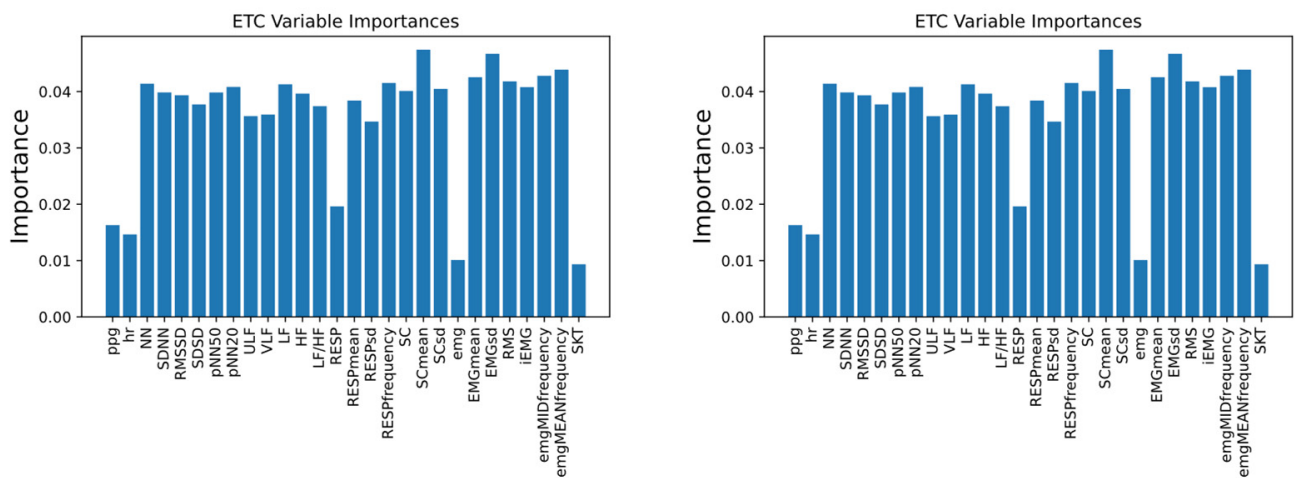


Figure 8. Cont.



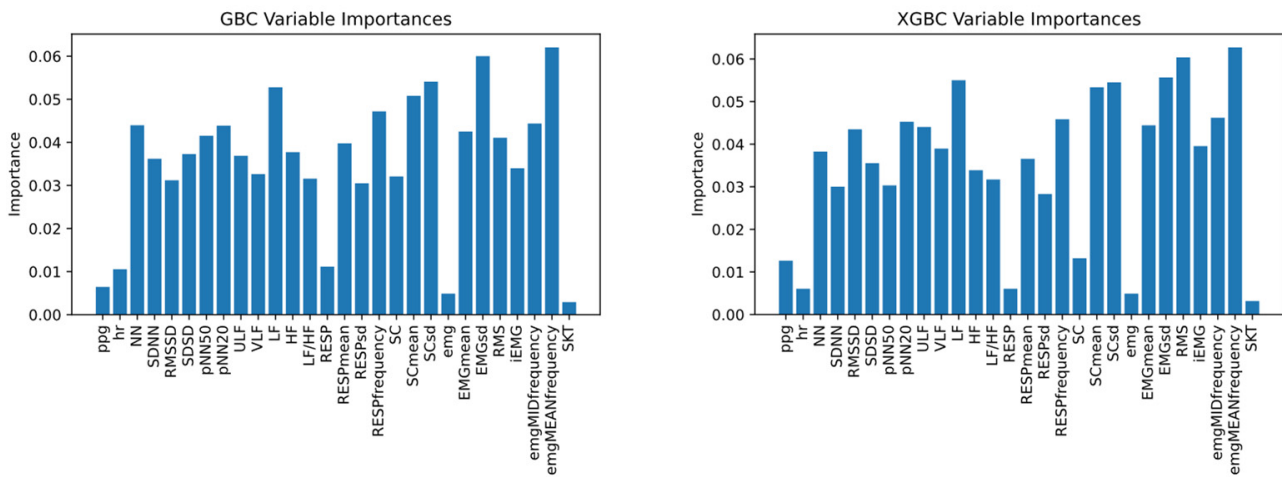


Figure 8. The importance ranking of the four different classifiers.

Removing irrelevant features can help to ease the learning task, make the model simple and reduce the computational complexity. Based on the importance ranking, we chose to keep NN, SDNN, pNN50, pNN20, VLF, LF, HF, RESP, RESP mean, RESP standard deviation, SC, SC mean, SC standard deviation, EMG mean, EMG standard deviation, EMG RMS, iEMG, EMG median frequency, and EMG mean power frequency as features. After the features were selected and re-input into the model, the classification performance was as shown in Figure 9 and Table 9. It was found that, after feature selection, the performance of all models improved.

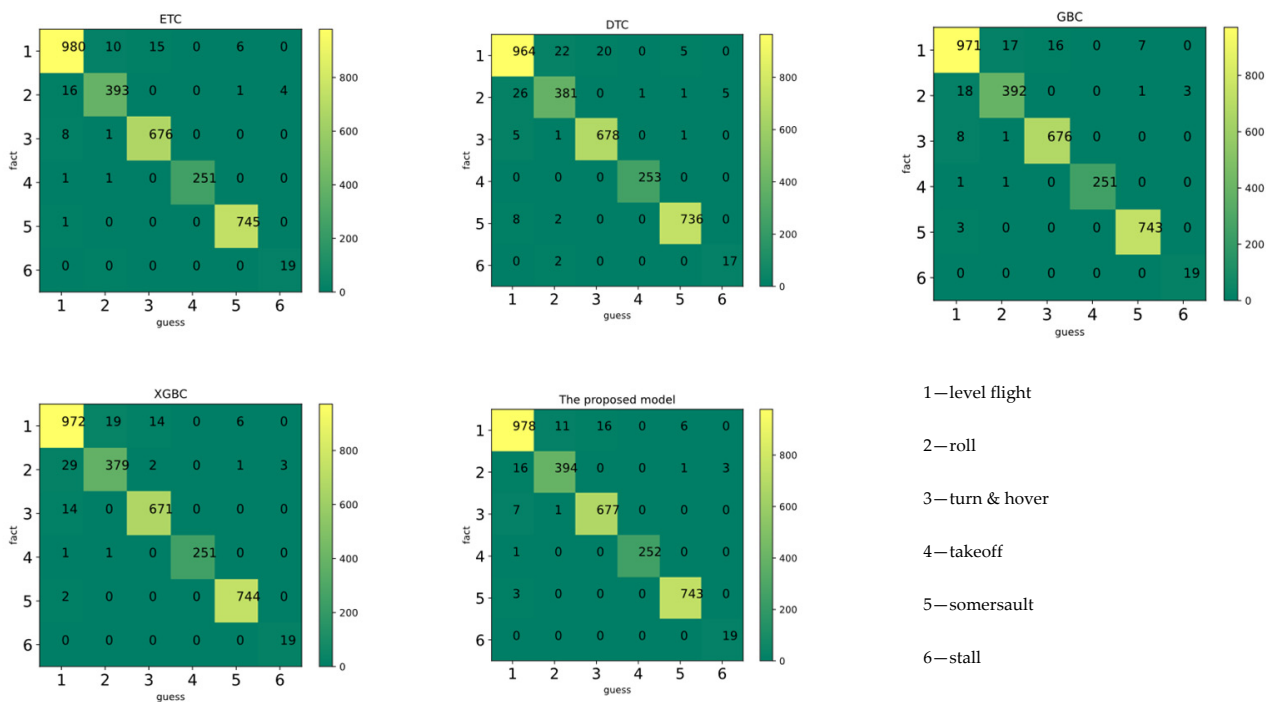


Figure 9. The confusion matrix of classifiers after improvement.

**Table 9.** The classification report of classifiers after improvement.

| ETC                                    |       |      |                |                           |            |        |
|--|-------|------|----------------|---------------------------|------------|--------|
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.97  | 0.95 | 0.99           | 1.00                      | 1.00       | 1.00   |
| <b>recall</b>                          | 0.97  | 0.96 | 0.99           | 1.00                      | 0.99       | 0.91   |
| <b>F1</b>                              | 0.97  | 0.96 | 0.99           | 1.00                      | 0.99       | 0.95   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9792         | <b>MSE for 10-fold CV</b> |            | 0.1093 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.7889         | <b>MSE for LOO CV</b>     |            | 1.1933 |
| DTC                                    |       |      |                |                           |            |        |
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.96  | 0.92 | 0.99           | 1.00                      | 0.99       | 0.97   |
| <b>recall</b>                          | 0.96  | 0.94 | 0.98           | 1.00                      | 0.99       | 0.88   |
| <b>F1</b>                              | 0.96  | 0.93 | 0.98           | 1.00                      | 0.99       | 0.92   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9728         | <b>MSE for 10-fold CV</b> |            | 0.1579 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.7301         | <b>MSE for LOO CV</b>     |            | 1.3759 |
| GBC                                    |       |      |                |                           |            |        |
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.98  | 0.93 | 0.99           | 1.00                      | 0.99       | 0.97   |
| <b>recall</b>                          | 0.97  | 0.96 | 0.98           | 1.00                      | 0.99       | 0.90   |
| <b>F1</b>                              | 0.97  | 0.94 | 0.99           | 1.00                      | 0.99       | 0.93   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9726         | <b>MSE for 10-fold CV</b> |            | 0.1266 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.7306         | <b>MSE for LOO CV</b>     |            | 1.2341 |
| XGBC                                   |       |      |                |                           |            |        |
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.96  | 0.93 | 0.98           | 1.00                      | 1.00       | 1.00   |
| <b>recall</b>                          | 0.97  | 0.94 | 0.98           | 1.00                      | 0.99       | 0.94   |
| <b>F1</b>                              | 0.97  | 0.94 | 0.98           | 1.00                      | 0.99       | 0.97   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9741         | <b>MSE for 10-fold CV</b> |            | 0.1151 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.7697         | <b>MSE for LOO CV</b>     |            | 1.3256 |
| Proposed Model                         |       |      |                |                           |            |        |
|  | level | roll | turn and hover | takeoff                   | somersault | stall  |
| <b>precision</b>                       | 0.98  | 0.95 | 0.99           | 1.00                      | 1.00       | 0.97   |
| <b>recall</b>                          | 0.98  | 0.96 | 0.98           | 1.00                      | 1.00       | 0.93   |
| <b>F1</b>                              | 0.98  | 0.95 | 0.99           | 1.00                      | 1.00       | 0.95   |
| <b>average accuracy for 10-fold CV</b> |       |      | 0.9815         | <b>MSE for 10-fold CV</b> |            | 0.1026 |
| <b>average accuracy for LOO CV</b>     |       |      | 0.8273         | <b>MSE for LOO CV</b>     |            | 0.9601 |

According to the classification report, ‘takeoff’ achieved the highest accuracy, probably because takeoff is at the beginning of the flight when the pilot’s attention is most focused. To obtain more reliable data, in future work, we plan to perform the takeoff operation several times in one trial to avoid the effect of attention on the results. Confusion tended to occur between ‘level flight’, ‘roll’, and ‘turn and hover’, probably due to the similar difficulty of these behaviors and the fact that stalls rarely occur during these maneuvers. To our surprise, the ‘somersault’ was sometimes confused with ‘level flight’. After discussion with the pilots, we thought that it might have been because there was a period of level flight before and after the somersault. So, the definition of the flight behavior still needed further clarification. ‘Stall’ achieved the lowest accuracy. We believe this is due to the minimum number of sample points in the ‘stall’ state, which led to an insufficient amount of data and, thus, affected the results. Therefore, expanding the data set is necessary to train a better model.

Another finding is that the classification model trained on a data set using 10-fold CV does not perform best in leave-one-person-out cross-validation. The reason for the

variation might be attributed to individual differences. Hence, how to select the most suitable classifier based on the results of 10-fold cross-validation and LOO cross-validation and how to build an adaptive model to reduce individual variability will be the focus of subsequent research.

#### 4. Conclusions

This study provided a comparison of the physiological responses of 14 pilots performing flight missions with different flight behaviors, thus, demonstrating the physiological indicators influenced by the pilot's behavioral status and their association with flight difficulty. It was found that ECG value, HR, SDD, pNN50, pNN20, VLF, LF, HF, RESP value, REAP mean, RESP standard deviation, SC standard deviation, EMG value, RMS, iEMG, EMG mid frequency, EMG mean frequency, and SKT had no evident relations with pilots' behaviors. RESP frequency, SCL, and SCL mean rose with an increasing subjective difficulty rating. On the contrary, NN, ULF, LF/HF, EMG mean, and EMG standard deviation decreased with increasing subjective difficulty rating. This provided a clue for selecting key indicators for research, especially as the situation was too limited to collect multiple physiological features. The comparison also visually explained the selection of feature importance, which helped to reduce redundancy, avoid overfitting, and improve the real-time detection capability of the model.

Furthermore, a model that recognizes pilots' behavior based on multi-modal fusion technology (MFT) using physiological characteristics was proposed. Pilot multi-modal physiological parameters and various machine learning classifiers were used to detect pilot behavior. Several machine learning models were employed to recognize the pilot's behaviors. By analyzing and comparing the accuracy, precision, F1 score, and MSE of different models, it was found that the pilot-state-recognition model could be improved in many ways, including feature layer fusion, decision layer fusion, and feature filtering. The experimental results verified the superiority of the proposed model in recognizing flight behavior. The accuracy of the proposed model reached 98.15%, proving that MFT is promising for pilot state recognition.

Due to the limited data set used in this study, the accuracy of the classification network could not be fully verified. In future work, it will be necessary to acquire more data for training and testing to verify the model more precisely. Building an adaptive model to reduce individual variation is also the core of future work. In addition, we plan to apply the model in practice to detect the effects of flight assist systems on the physiological state of pilots when they perform various behaviors. Furthermore, the generalization of the model will be improved to make it applicable to more scenarios.

**Author Contributions:** Conceptualization, Y.L. and S.W.; methodology, Y.L.; algorithm, Y.L.; resources, K.L. and D.W.; data curation, Y.L. and K.L.; writing—original draft preparation, Y.L.; writing—review and editing, D.W. and X.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Chinese National Natural Science Foundation (no. NSFC61773039).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Beihang University for studies involving humans.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

|       |  |
|-------|--|
| MFT   | Multi-modal Fusion Technology                                      |
| ECG   | Electro Cardio   |
| EMG   | Myoelectricity   |
| GSR   | Galvanic Skin Response   |
| RESP  | Respiration  |
| SKT   | Skin Temperature   |
| EEG   | Electroencephalogram   |
| MSE   | Mean Square Error  |
| ROC   | Receiver Operating Characteristic Curve                            |
| GBDT  | Gradient Boosting Decision Tree                                    |
| SVM   | Support Vector Machine   |
| MLP   | Multilayer Perceptron  |
| RRI   | RR Interval  |
| NN    | Normal-to-Normal   |
| SDSD  | Standard Deviation of the Difference between Adjacent NN Intervals |
| SDNN  | Standard Deviation of NN intervals                                 |
| RMSSD | Root Mean Square of Successive Differences                         |
| pNN50 | Percentage of Mean R–R Intervals Greater than 50 MS                |
| pNN20 | Percentage of Mean R–R Intervals Greater than 20 MS                |
| VLF   | Very Low Frequency (0.0033–0.04 Hz)                                |
| ULF   | Ultra Low Frequency (0–0.0033 Hz)                                  |
| LF    | Low Frequency (0.04–0.15 Hz)                                       |
| HF    | High Frequency (0.15–0.4 Hz)                                       |
| LF/HF | Energy Ratio of Low Frequency to High Frequency                    |
| RMS   | Root Mean Square   |
| iEMG  | Integral EMG   |

### Appendix A

The principles of leave-one-person-out cross-validation are as shown.

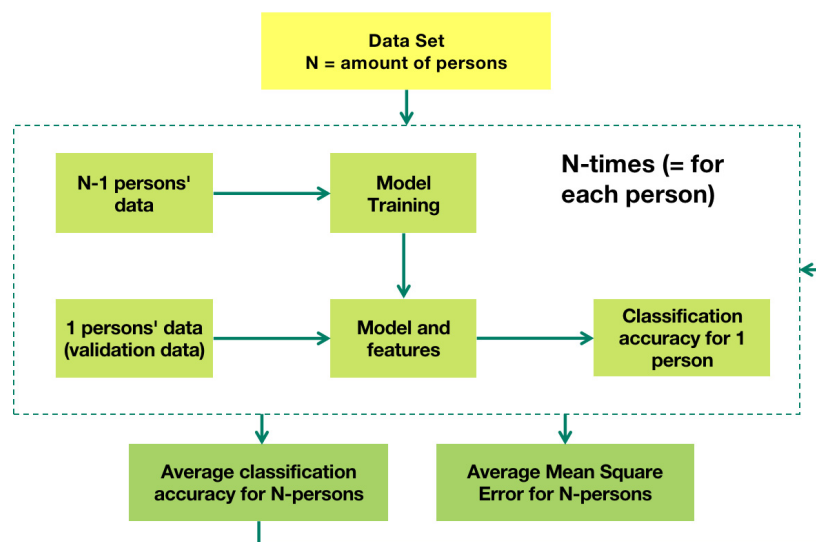


Figure A1. The principles of leave-one-person-out cross-validation.

### References

- Greiser, S.; Lantzsch, R.; Wolfram, J.; Wartmann, J.; Müllhäuser, M.; Lüken, T.; Döhler, H.-U.; Peinecke, N. Results of the pilot assistance system “Assisted Low-Level Flight and Landing on Unprepared Landing Sites” obtained with the ACT/FHS research rotorcraft. *Aerosp. Sci. Technol.* **2015**, *45*, 215–227. [CrossRef]
- Smaili, M.H.; Rouwhorst, W.F.J.A.; Frost, P. Conceptual Design and Evaluation of Upset-Recovery Systems for Civil Transport Aircraft. *J. Aircr.* **2018**, *55*, 947–964. [CrossRef]
- Rydlo, K.; Rzucidlo, P.; Chudy, P. Simulation and prototyping of FCS for sport aircraft. *Aircr. Eng. Aerosp. Technol.* **2013**, *85*, 475–486. [CrossRef]

4. Xi, P.; Law, A.; Goubran, R.; Shu, C. Pilot Workload Prediction from ECG Using Deep Convolutional Neural Networks. In Proceedings of the IEEE International Symposium on Medical Measurements and Applications (IEEE MeMeA), Istanbul, Turkey, 26–28 June 2019.
5. Dai, J.; Luo, M.; Hu, W.; Ma, J.; Wen, Z. Developing a fatigue questionnaire for Chinese civil aviation pilots. *Int. J. Occup. Saf. Ergon.* **2018**, *26*, 37–45. [[CrossRef](#)] [[PubMed](#)]
6. Hajra, S.G.; Xi, P.; Law, A. A comparison of ECG and EEG metrics for in-flight monitoring of helicopter pilot workload. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020.
7. Law, A.; Jennings, S.; Ellis, K. A Comparison of Control Activity and Heart Rate as Measures of Pilot Workload in a Helicopter Tracking Task. In Proceedings of the Vertical Flight Society 75th Annual Forum, Philadelphia, PA, USA, 16 May 2019.
8. Feleke, A.G.; Bi, L.; Fei, W. Detection of Driver Emergency Steering Intention Using EMG Signal. In Proceedings of the 2020 IEEE International Conference on Real-time Computing and Robotics (RCAR), Asahikawa, Japan, 28–29 September 2020; pp. 280–285.
9. Bhardwaj, R.; Natrajan, P.; Balasubramanian, V. Study to Determine the Effectiveness of Deep Learning Classifiers for ECG Based Driver Fatigue Classification. In Proceedings of the 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), Rupnagar, India, 1–2 December 2018.
10. Masters, M.; Schulte, A. Investigating the Utility of fNIRS to Assess Mental Workload in a Simulated Helicopter Environment. In Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS), Rome, Italy, 7–9 September 2020.
11. Pamela, Z.; Antonio, A.; Riccardo, B.; Alessandro, P.; Roberto, R.; Fabio, F.; Diego, M.; Michela, M. Car Driver’s Sympathetic Reaction Detection Through Electrodermal Activity and Electrocardiogram Measurements. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 3413–3424. [[CrossRef](#)]
12. Šalkevicius, J.; Damaševicius, R.; Maskeliunas, R.; Laukienė, I. Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals. *Electronics* **2019**, *8*, 1039. [[CrossRef](#)]
13. Petrescu, L.; Petrescu, C.; Mitruț, O.; Moise, G.; Moldoveanu, A.; Moldoveanu, F.; Leordeanu, M. Integrating Biosignals Measurement in Virtual Reality Environments for Anxiety Detection. *Sensors* **2020**, *20*, 7088. [[CrossRef](#)]
14. Chen, L.-L.; Zhao, Y.; Ye, P.-F.; Zhang, J.; Zou, J.-Z. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Syst. Appl.* **2017**, *85*, 279–291. [[CrossRef](#)]
15. Qi, M.S.; Yang, W.J.; Ping, X.; Liu, Z.-J. Driver fatigue Assessment Based on the Feature Fusion and Transfer Learning of EEG and EMG. In Proceedings of the 2018 Chinese Automation Congress (CAC), Changsha, China, 1 November 2018.
16. He, J.; Zhang, C.Q.; Li, X.Z.; Zhang, D.H. A review of research on multimodal fusion techniques for deep learning. *Comput. Eng.* **2020**, *32*, 186–189.
17. Wang, D.; Shen, P.; Wang, T.; Xiao, Z. Fatigue Detection of Vehicular Driver through Skin Conductance, Pulse Oximetry and Respiration: A Random Forest Classifier. In Proceedings of the 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), Guangzhou, China, 6–8 May 2017.
18. Hu, J.; Min, J. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model. *Cogn. Neurodynamics* **2018**, *12*, 431–440. [[CrossRef](#)]
19. Vargas-Lopez, O.; Perez-Ramirez, C.; Valtierra-Rodriguez, M.; Yanez-Borjas, J.; Amezcua-Sanchez, J. An Explainable Machine Learning Approach Based on Statistical Indexes and SVM for Stress Detection in Automobile Drivers Using Electromyographic Signals. *Sensors* **2021**, *21*, 3155. [[CrossRef](#)] [[PubMed](#)]
20. Jiawei, H.; Micheline, K.; Jian, P. *Data Mining Concepts and Techniques*; Machinery Industry Press: Beijing, China, 2012.
21. Kim, K.H.; Bang, S.W.; Kim, S.R. Emotion recognition system using short-term monitoring of physiological signals. *Med Biol. Eng. Comput.* **2004**, *42*, 419–427. [[CrossRef](#)] [[PubMed](#)]
22. Dan, B. *Data Mining for Design and Manufacturing: Methods and Applications*; Kluwer Academic Publishers: Alphen am Rhine, The Netherlands, 2001.
23. Rahman, N. A Taxonomy of Data Mining Problems. *Int. J. Bus. Anal.* **2018**, *5*, 73–86. [[CrossRef](#)]
24. Bai, J.; Shen, L.; Sun, H.; Shen, B. Physiological Informatics: Collection and Analyses of Data from Wearable Sensors and Smartphone for Healthcare. *Adv. Exp. Med. Biol.* **2017**, *1028*, 17–37.
25. Cheng, K.S.; Chen, Y.S.; Wang, T. Physiological parameters assessment for emotion recognition. In Proceedings of the Biomedical Engineering and Sciences (IECBES), Langkawi, Malaysia, 15 April 2013.
26. Patel, M.; Lal, S.; Kavanagh, D.; Rossiter, P. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Syst. Appl.* **2010**, *38*, 7235–7242. [[CrossRef](#)]
27. Healey, J.; Picard, R.W. Detecting Stress during Real-World Driving Tasks Using Physiological Sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [[CrossRef](#)]
28. Ying, L.; Shan, F. Brief review on physiological and biochemical evaluations of human mental workload. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2012**, *22*, 177–187.
29. Tian, X.T. Evaluation of Driver Mental Load Based on Heart Rate Variability and Its Application. Master’s Thesis, Northern Polytechnic University, Beijing, China, 2016.
30. Chieng, T.M.; Yuan, W.H.; Omar, Z. The Study and Comparison between Various Digital Filters for ECG De-noising. In Proceedings of the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, 3–6 December 2018.



31. Aladag, S.; Guven, A.; Ozbek, H.; Dolu, N. A comparison of denoising methods for Electrodermal Activity signals. In Proceedings of the Medical Technologies National Conference, Sarawak, Malaysia, 15–18 October 2015.
32. Ghalyan, I.; Abouelenin, Z.; Kapila, V. Gaussian Filtering of EMG Signals for Improved Hand Gesture Classification. In Proceedings of the 2018 IEEE Signal Processing in Medicine and Biology Symposium, Philadelphia, PA, USA, 1 December 2018.
33. Akay, M. Wavelets in biomedical engineering. *Ann. Biomed. Eng.* **1995**, *23*, 531–542. [[CrossRef](#)]
34. Arozi, M.; Caesarendra, W.; Ariyanto, M.; Munadi, M.; Setiawan, J.D.; Glowacz, A. Pattern Recognition of Single-Channel sEMG Signal Using PCA and ANN Method to Classify Nine Hand Movements. *Symmetry* **2020**, *12*, 541. [[CrossRef](#)]
35. Behbahani, S.; Ahmadi, H.; Rajan, S. Feature Extraction Methods for Electroretinogram Signal Analysis: A Review. *IEEE Access* **2021**, *9*, 116879–116897. [[CrossRef](#)]
36. Purnamasari, P.D.; Martmis, R.; Wijaya, R.R. Stress Detection Application based on Heart Rate Variability (HRV) and K-Nearest Neighbor (KNN). In Proceedings of the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), Jakarta, Indonesia, 24–25 October 2019.
37. Esener, I.I. Subspace-based feature extraction on multi-physiological measurements of automobile drivers for distress recognition. *Biomed. Signal Process. Control* **2021**, *66*, 102504. [[CrossRef](#)]
38. Edelmann, D.; Móri, T.F.; Székely, G.J. On relationships between the Pearson and the distance correlation coefficients. *Stat. Probab. Lett.* **2021**, *169*, 108960. [[CrossRef](#)]
39. Liu, X.S. A probabilistic explanation of Pearson's correlation. *Teach. Stat.* **2019**, *41*, 115–117. [[CrossRef](#)]
40. Yu, J.; Benesty, J.; Huang, G.; Chen, J. Daily Activity Feature Selection in Smart Homes Based on Pearson Correlation Coefficient. *Int. J. Electr. Power Energy Syst.* **2019**, *5*, 38–40.
41. Mitchell, J. Three machine learning models for the 2019 Solubility Challenge. *ADMET DMPK* **2020**, *8*, 215–250. [[CrossRef](#)] [[PubMed](#)]
42. Molan, M.; Molan, G. *Maintaining Sustainable Level of Human Performance with Regard to Manifested Actual Availability*; Springer: Cham, Switzerland, 2018.
43. Sirikul, W.; Buawangpong, N.; Saphamrer, R.; Siviroj, P. Mortality-Risk Prediction Model from Road-Traffic Injury in Drunk Drivers: Machine Learning Approach. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10540. [[CrossRef](#)] [[PubMed](#)]
44. Veeramachaneni, K.; Yan, W.; Kai, G.; Osadciw, L. Improving Classifier Fusion Using Particle Swarm Optimization. In Proceedings of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, Honolulu, HI, USA, 2007.
45. Aytu, O.A.; Serdar, K.B.; Hasan, B.B. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification—ScienceDirect. *Expert Syst. Appl.* **2016**, *62*, 1–16.
46. Delgado, R. A semi-hard voting combiner scheme to ensemble multi-class probabilistic classifiers. *Appl. Intell.* **2022**, *52*, 3653–3677. [[CrossRef](#)]
47. Dong, X.; Zhiwen, Y.U.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2019**, *14*, 241–258. [[CrossRef](#)]
48. Gomes, H.M.; Barddal, J.P.; Enembreck, F.; Bifet, A. A Survey on Ensemble Learning for Data Stream Classification. *Acm Comput. Surv.* **2017**, *50*, 1–36. [[CrossRef](#)]
49. Parvin, A.; MirnabiBaboli, B.; Alinejad-Rokny, E. Proposing a classifier ensemble framework based on classifier selection and decision tree. *Eng. Appl. Artif. Intell.* **2015**, *37*, 34–42. [[CrossRef](#)]
50. Wang, H.; Suh, J.W.; Das, S.R.; Pluta, J.B.; Craige, C.; Yushkevich, P.A. Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 611–623. [[CrossRef](#)]
51. Pawiak, P.; Acharya, U. Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals. *Neural Comput. Appl.* **2020**, *32*, 11137–11161. [[CrossRef](#)]
52. Shen, H.; Lin, Y.; Tian, Q.; Xu, K.; Jiao, J. A comparison of multiple classifier combinations using different voting-weights for remote sensing image classification. *Int. J. Remote Sens.* **2018**, *39*, 3705–3722. [[CrossRef](#)]
53. Bonett, D.G.; Wright, T.A. Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika* **2000**, *65*, 23–28. [[CrossRef](#)]
54. Olbrich, S.; Sander, C.; Matschinger, H.; Mergl, R.; Hegerl, U. Brain and body: Associations between EEG vigilance and the autonomous nervous system activity during rest. *J. Psychophysiol.* **2011**, *25*, 190. [[CrossRef](#)]
55. Wanyan, X.; Zhuang, D.; Zhang, H. Improving pilot mental workload evaluation with combined measures. *Bio-Medical Mater. Eng.* **2014**, *24*, 2283–2290. [[CrossRef](#)]
56. Machado, A.V.; Pereira, M.G.; Souza, G.G.; Xavier, M.; Aguiar, C.; de Oliveira, L.; Mocaiber, I. Association between distinct coping styles and heart rate variability changes to an acute psychosocial stress task. *Sci. Rep.* **2021**, *11*, 24025. [[CrossRef](#)]
57. Castaldo, R.; Montesinos, L.; Wan, T.S.; Serban, A.; Massaro, S.; Pecchia, L. Heart Rate Variability Analysis and Performance during a Repeated Mental Workload Task. In Proceedings of the EMBEC & NBC 2017, Tampere, Finland, 11–15 June 2017.
58. Satti, A.T.; Kim, J.; Yi, E.; Cho, H.Y.; Cho, S. Microneedle Array Electrode-Based Wearable EMG System for Detection of Driver Drowsiness through Steering Wheel Grip. *Sensors* **2021**, *21*, 5091. [[CrossRef](#)]
59. Wrle, J.; Metz, B.; Thiele, C.; Weller, G. Detecting sleep in drivers during highly automated driving: The potential of physiological parameters. *IET Intell. Transp. Syst.* **2019**, *13*, 1241–1248. [[CrossRef](#)]
60. Skibniewski, F.W.; Dziuda, Ł.; Baran, P.M.; Krej, M.K.; Guzowski, S.; Piotrowski, M.A.; Truszczyński, O.E. Preliminary Results of the LF/HF Ratio as an Indicator for Estimating Difficulty Level of Flight Tasks. *Aerosp. Med. Hum. Perform.* **2015**, *86*, 518–523. [[CrossRef](#)] [[PubMed](#)]

61. Rosa, E.; Lyskov, E.; Grönkvist, M.; Kölegård, R.; Dahlström, N.; Knez, I.; Ljung, R.; Willander, J. Cognitive performance, fatigue, emotional and physiological strains in simulated long-duration flight missions. *Mil. Psychol.* **2022**, *34*, 224–236. [[CrossRef](#)]
62. Borghini, G.; Di Flumeri, G.; Aricò, P.; Sciaraffa, N.; Bonelli, S.; Ragosta, M.; Tomasello, P.; Drogoul, F.; Turhan, U.; Acikel, B.; et al. A multimodal and signals fusion approach for assessing the impact of stressful events on Air Traffic Controllers. *Sci. Rep.* **2020**, *10*, 8600. [[CrossRef](#)] [[PubMed](#)]
63. Koskimaki, H. Avoiding Bias in Classification Accuracy—A Case study for Activity Recognition. In Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 7–10 December 2015; pp. 301–306.