

Article

Fatigue Detection of Air Traffic Controllers Based on Radiotelephony Communications and Self-Adaption Quantum Genetic Algorithm Optimization Ensemble Learning

Nan Wu and Jingjuan Sun *

Air Traffic Control and Navigation School, Air Force Engineering University, Xi'an 710051, China

* Correspondence: jjuansun@163.com



Citation: Wu, N.; Sun, J. Fatigue Detection of Air Traffic Controllers Based on Radiotelephony Communications and Self-Adaption Quantum Genetic Algorithm Optimization Ensemble Learning. *Appl. Sci.* **2022**, *12*, 10252. <https://doi.org/10.3390/app122010252>

Academic Editors: Filipe Moura and Manuel Marques

Received: 28 August 2022

Accepted: 28 September 2022

Published: 12 October 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/>).

Abstract: Air traffic controller (ATC) fatigue has become a major cause of air traffic accidents. Speech-based fatigue-state detection is proposed in this paper. The speech signal is preprocessed to further extract the Mel frequency cepstrum coefficient (MFCC) from speech discourse. The machine learning method is used in fatigue detection. However, single machine learning fatigue detection methods often have low detection accuracy. To solve this problem, an ensemble learning method based on self-adaption quantum genetic algorithm (SQGA) heterogeneous learning methods is proposed. Pattern-level and feature-level resampling are used to increase the differences in the base learner's training dataset. To enlarge the diversity of single learners, k-nearest neighbor (KNN), Bayesian network (BN), back propagation neural network (BPNN) and support vector machine (SVM) are adopted for the heterogeneous ensemble. On this basis, finally, the detection result is obtained by weighted summation. The weight of each base learner was determined by SQGA. The SQGA method combines the quantum genetic algorithm with the adaptive strategy. The adaptive strategy includes adaptive adjustment of the quantum rotation gate, adaptive generation of crossover probability and adaptive generation of mutation probability. The experiments on real civil aviation radio land-air communication show that the proposed method can obtain 98.5% detection accuracy, with a 1.2% false and 3.0% missing report rate, whereas the SVM only obtains 94.0% detection accuracy, with a 5.4% false and 9.0% missing report rate.

Keywords: fatigue detection; MFCC; heterogeneous ensemble; quantum genetic algorithm; self-adaption

1. Introduction

As air traffic volume increases, air traffic controller (ATC) fatigue has become a major cause of air traffic accidents. This state makes the operating ability of the controller decline, which makes it easy for them to cause work errors and even lead to accident symptoms or accidents. According to data from the American Aviation Safety Voluntary Reporting System, there are 52,000 unsafe incidents caused by fatigue, accounting for 21% of operational errors. According to data from the British Aviation Safety Voluntary Reporting System, 13% of the operational errors were directly related to controller fatigue, and according to data from China's ATC Safety Reporting System, 18% of work errors are related to controller fatigue [1,2]. It can be seen that nearly one-fifth of the unsafe events in the operation of civil aviation are related to the fatigue of controllers. For this reason, it has become a hot topic in recent years to detect the fatigue state of ATC quickly and accurately.

Fatigue detection usually judges the fatigue state according to some features related to the tested person [3]. At present, fatigue state detection methods of air traffic controllers can be divided into subjective detection methods and objective detection methods, according to different implementation paths. The subjective test method is mainly in the form of filling in a fatigue scale by the tested personnel [4–10]. According to the score on the fatigue scale, the fatigue state can be judged. This method is easy to implement but has poor applicability. The objective detection method refers to the use of auxiliary tools to record the changes

in some indicators related to the human body, or human factor engineering, in order to judge the fatigue state. Wang Lili et al. [4] evaluated the fatigue state of controllers under different shift scheduling modes through the analysis and research of EEG signals. The study found that when the EEG frequency of the controller decreased to a certain extent, it appeared to be θ wave, the ATC experienced decreased alertness and weakened operational ability. Mandal et al. proposed a fatigue classification method based on eyelid closure using simulation tower control software and an eye tracker. Jin Huibin [8] and others extracted six indicators, such as the rate of a saccade, the time span of fixation and the percentage of eye closure time (PERCLOS) as the discrimination index of fatigue. The disadvantage of this method is that an image acquisition device needs to be set in front of the subject's face to collect facial images. On the one hand, it will make the subject more prone to fatigue under psychological pressure. On the other hand, it needs to process the image, and the detection results will be delayed to a certain extent.

In recent years, the fatigue detection method based on speech has become a research hotspot of fatigue detection methods because of its high detection accuracy and strong applicability [11–13]. Speech signal data from the controller can be directly obtained by embedding a sensor in the controller's radio communication device. In the process of speech signal processing, compared with the linear analysis method, the application of the speech nonlinear dynamics analysis method can more completely obtain the impact of fatigue on speech signaling. Specifically, the fatigue state detection method based on speech features will not have any impact on the normal work of controllers, and it can greatly save the detection cost. The most important point is that the fatigue state detection method based on speech features can effectively improve the accuracy of the detection results. At present, the research on the fatigue state detection method based on speech features mainly focuses on time domain and frequency domain features. The commonly used time domain features include pitch, short-time energy, short-time zero crossing rate, average amplitude, harmonic noise ratio, autocorrelation coefficient, etc. Frequency domain features include formant frequency, Mel frequency cepstrum coefficient (MFCC), linear prediction cepstrum coefficient (LPCC) and line spectrum pair (LSP) [14,15]. In this paper, we chose the MFCC as the detection feature.

Fatigue detection, in essence, can be attributed to a two-classification problem: normal/fatigue. At present, the main research direction in the field of fatigue detection is the machine learning method. Some typical machine learning methods, such as k-nearest neighbor (KNN), back propagation neural network (BPNN) and deep learning, have been successfully applied in this field [16–24]. However, for machine learning, in order to obtain an ideal detection model, an unavoidable problem is the parameter setting. Although many scholars have performed a lot of work on parameter selection, it is difficult to find the optimal parameters. A poor parameter may affect the detection effect and has become a bottleneck in the application of machine learning methods. In recent years, an ensemble learning algorithm has attracted more and more attention [25,26]. The ensemble learning algorithm improves the generalization performance of the whole algorithm by integrating multiple base classifiers. In ensemble learning, in order to obtain a good generalization ability, the decision accuracy of each base classifier is required to be higher than that of a random guess. In this condition, the SVM, KNN, Bayesian and other methods with good generalization ability almost do not need to consider the parameter problem, thus avoiding parameter selection. In this way, the current bottleneck problem can be broken through. So, this paper chooses ensemble learning to research fatigue detection.

The process of ensemble learning mainly includes two parts: the construction of a base classifier and the fusion method of multi-classifier results [27]. The greater the difference in each base classifier, the stronger the overall performance that can be obtained. In order to improve the differences between the base classifiers, this paper adopts the method of heterogeneous integration and multi-sampling. In addition, the fusion method of the base classifier is the other key problem affecting the detection model. At present, the most commonly used and effective methods are the voting method and the weighting method.

In contrast, the weighted method is an improvement on the voting method. It weighs the decision-making output according to the classification performance of different base classifiers. However, how much weight is assigned and whether the result is optimal is another problem. In order to solve this problem, this paper uses a quantum genetic algorithm to optimize the weights of each base classifier and improve the accuracy of the detection model.

To sum up, this paper takes the ensemble learning method as the core detection algorithm and improves the detection accuracy by optimizing the construction of the base classifier and result fusion methods. The specific arrangement of the paper is as follows:

In the second section, the speech feature extraction method MFCC is introduced. The third section introduces the whole algorithm, heterogeneous integration and sampling are introduced to improve the difference in base classifiers, and then optimizes the weighted method of decision-making with SQGA; the fourth section tests the proposed algorithm on some basic datasets to verify the effectiveness of the algorithm; in the fifth section, this method is applied to ATC fatigue detection to test the performance of the algorithm; and finally, the full text is summarized.

2. ATC Speech Extract by MFCC

Speech-based fatigue-state detection is a new hotspot in the research on fatigue state management. The implementation process of this method is similar to that of speech emotion recognition. The research on speech-based fatigue-detection technology first needs to establish a speech fatigue database.

When the human body is tired, attention and reaction ability decline. The direct manifestation of controller fatigue is that when issuing control instructions, the controller often gets stuck, hesitates, or even issues the wrong instructions, or cannot find the error in the pilot's repeated instructions in time. Therefore, when the controller's voice is stuck, fuzzy and wrong, the controller's state corresponding to the voice is labeled as "fatigue". Figure 1 shows the waveform of the same command sent by controllers in different statuses.

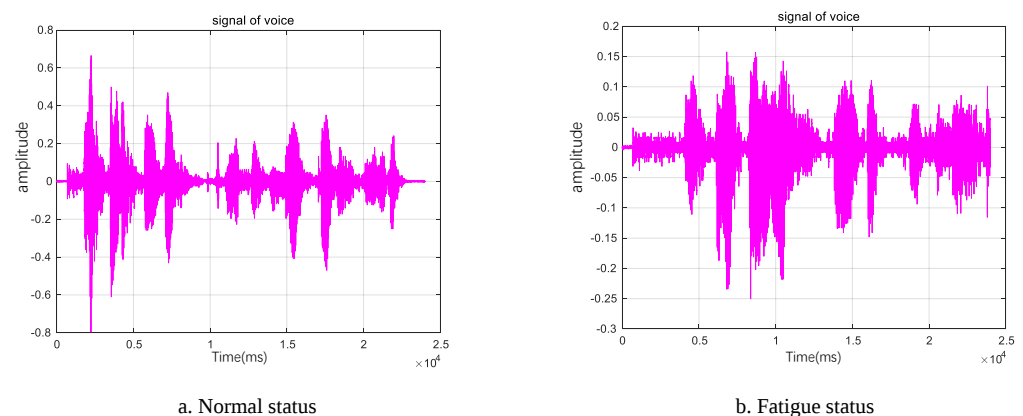


Figure 1. The signal of voice at different statuses.

In recent years, the most widely used feature parameters in speech recognition are spectral correlation features. The commonly used spectral correlation feature is the Mel cepstrum coefficient (MFCC). MFCC was proposed by Stevens, Volkman and Newman in 1937. It is a perceptual scale judged by the audience at an equal distance from each other. Mel is a measure of perceived tone or tone frequency. Above about 500 Hz, the listener judges larger and larger intervals to produce equal pitch increments. The four octaves on the hertz scale above 500 Hz are judged to include about two octaves on the scale. The Mel scale is approximately linear below 1 kHz and logarithmic above. The following formula is used to calculate the hertz (in Hz) of a given frequency f . MFCC has a strong anti-noise

ability, and the relationship between its Mel frequency (f_{mel}) and the actual frequency f of voice is:

$$f_{mel}(f) = 2595 \lg(1 + f/700). \quad (1)$$

The MFCC feature extraction process is as follows:

The Figure 2 shows the MFCC feature extraction process. The specific process is as follows:

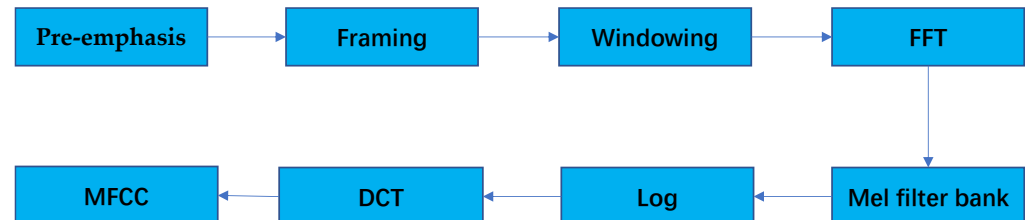


Figure 2. MFCC feature extraction process.

(1) Sample continuous signal (voice) at sampling frequency: In the first stage of MFCC, feature extraction is performed to increase the energy at high frequency. This pre-emphasis is accomplished by a digital filter, and the equation is:

$$H(Z) = 1 - \alpha Z^{-1}. \quad (2)$$

where α represents the pre-weighting factor.

(2) Framing: The speech signal is divided into frames and the short-time characteristics of the speech signal are extracted. It is a process of dividing speech samples obtained from analog-to-digital conversion (ADC) into small frames with a time length in the range of 20~40 ms. Framing enables the non-static speech signal to be segmented into quasi-static frames, and the speech signal can be Fourier transformed because the speech signals exhibit quasi-static behavior in a short time of 20~40 ms. The smoothening of the voice signal is performed with the Hamming window:

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N-1. \quad (3)$$

(3) Add windows: The windowing step is used to windowize each individual frame, to minimize signal discontinuities at the beginning and end of each frame.

(4) FFT: The Fast Fourier transform (FFT) algorithm is ideally used to evaluate the spectrum of speech. FFT converts each frame of n samples from time domain to frequency domain. The linear spectrum of speech signal is obtained by discrete Fourier transform:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N} \quad 0 \leq n, k \leq N-1. \quad (4)$$

(5) Extract the spectrum, Mel filter bank and frequency envelope through Mel frequency filter: Mel filter bank is composed of overlapping triangular filters, and the cut-off frequency is determined by the center frequency of two adjacent filters. The filter has a linearly spaced center frequency and a fixed bandwidth on the Mel scale. Mel frequency filter is designed as follows:

$$H_p(k) = \begin{cases} 0 & k < f(p-1) \\ \frac{k-f(p-1)}{f(p)-f(p-1)} & f(p-1) \leq k \leq f(p) \\ \frac{f(p+1)-k}{f(p+1)-f(p)} & f(p) \leq k \leq f(p+1) \\ 0 & k > f(p+1) \end{cases}; \quad (5)$$

where p is the index of filters. Each filter has triangular filtering characteristics, and its center frequency is $f(p)$:

$$f(p) = \left(\frac{N}{f_s}\right) f_{mel}^{-1}(f_{mel}(f_l)) + p \frac{f_{mel}(f_h) - f_{mel}(f_l)}{P + 1}. \quad (6)$$

where f_h and f_l are the highest frequency and the lowest frequency of the filter banks, respectively; f_s is the sampling frequency, $f_s = 44,100$; P is the number of filter banks, here $P = 26$; and N is the number of points of FFT transformation, and $N = 4096$.

(6) Take the logarithm of Mel spectrum: This has the effect of changing multiplication into addition. Therefore, this step simply converts the multiplication of amplitude in Fourier transform into addition to obtain logarithmic spectrum $S(p)$. Its transfer function is:

$$S(p) = \left(\ln \sum_{k=0}^{N-1} |X(k)|^2 H_p(k) \right) \quad 0 \leq p \leq P. \quad (7)$$

(7) Discrete cosine transforms (DCT): Orthogonalize the energy vector of the filter. Due to this orthogonalization step, the information of the filter energy vector is compressed. The MFCC parameter $c(n)$ is obtained by discrete cosine transform of $S(p)$:

$$c(n) = \sum_{p=1}^{P-1} S(p) \cos\left(\frac{\pi n(p + 1/2)}{P}\right) \quad 0 \leq p \leq P. \quad (8)$$

From Figure 3, we can see that the speech signal in the time domain with 24,000-dimensional features is extracted by MFCC to only 150-dimensional features. The dataset of fatigue detection can be obtained by processing the voice of controllers in different statuses.

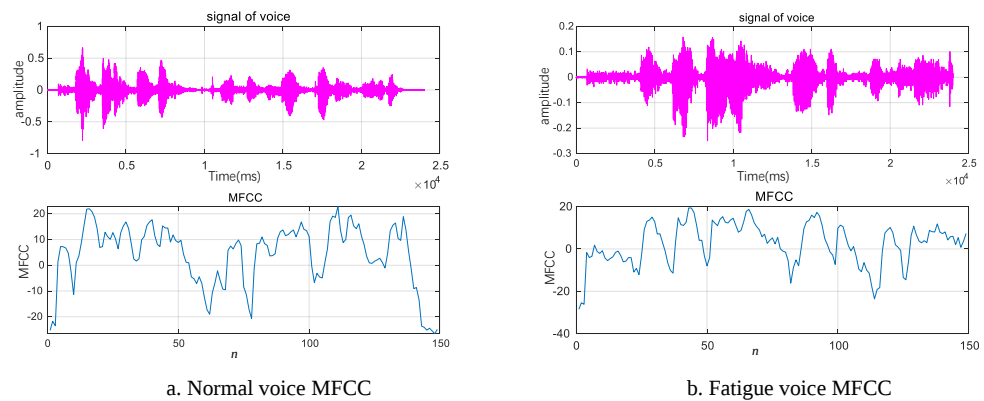


Figure 3. Normal/fatigue MFCC.

3. Heterogeneous Ensemble Learning by SQGA

After obtaining the characteristic, the fatigue is detected by the ensemble learning method. In order to improve the effect of ensemble learning, this paper uses the heterogeneous ensemble learning method to integrate a variety of learning machines. In addition, different sample methods are adopted to improve the differences between training samples. These methods can improve the generalization of ensemble learning and ensure detection accuracy.

3.1. Heterogeneous Base Classifiers

Ensemble learning is combining multiple learned models to solve a problem. The final output of input is determined by the output of each learning machine in this input [27]. In reference [27], the necessary and sufficient condition for the ensemble classifier to be more accurate than the individual classifier is given: the individual classifier is accurate and diverse. As we mentioned earlier, the current machine learning methods with good

generalization can meet the accuracy requirements without considering the parameter selection. Here we need to focus on how to improve the differences between base classifiers.

According to the similarities and differences of their own classification algorithms, ensemble learning can be divided into isomorphic and heterogeneous ensembles. In an isomorphic ensemble, all base classifiers adopt the same learning algorithm, but the parameters and samples selected by the algorithm are different. A heterogeneous ensemble uses different classifiers, and different learning algorithms can effectively ensure the differences between base classifiers. Therefore, in this paper, a heterogeneous ensemble is chosen to study fatigue detection.

In addition to the classification method, the selection of training datasets for base classifiers influences the learning machine significantly. According to the method of dataset selection, ensemble learning can be divided into pattern-level integration and feature-level ensemble. Pattern-level ensemble uses simple repeated sampling or changes the distribution of samples to resample the original dataset so that each base classifier can obtain different training sets to improve the differences between them. At present, the common bagging and boosting methods are pattern-level ensembles. Feature-level ensemble mainly aims at the scene with more features. It selects feature subsets reflecting different properties of the problem to form the training set for the base classifier. Generally, the sample dimension of fatigue detection is high. It is feasible to use a feature-level ensemble here. In order to improve the accuracy of the base classifier, we combine the pattern level and feature level to construct the heterogeneous ensemble.

3.2. Weight Optimization by SQGA

On the basis of obtaining the output of multiple base classifiers, it is necessary to fuse these outputs to get the final detection result. Because fatigue detection is a binary classification problem, we set the output results of each base classifier to +1 and −1, respectively, indicating normal and fatigue detection. Suppose there are n base classifiers, the detection result of the i -th base classifier is y_i and the weight of fusion is x_i , then the final detection result is:

$$y = \text{sgn}\left(\sum_{i=1}^n x_i y_i\right). \quad (9)$$

We need to find the optimal $\mathbf{x}$ to ensure the highest detection accuracy of the model, the optimization objective function should be:

$$f(\mathbf{x}) = \max(\text{acc}), 0 < \mathbf{x} < 1. \quad (10)$$

Among them, acc represents the test accuracy of the whole ensemble model under different weights, vector $\mathbf{x}$. In order to solve this problem, we choose quantum genetic algorithm (QGA) to solve it. QGA is an optimization method combining quantum computing with a genetic algorithm [28], which has the advantages of fast convergence speed and strong global optimization ability.

The difference between quantum genetic algorithm and traditional genetic algorithm is mainly reflected in two aspects:

(1) chromosome is represented by qubit code, and gene is expressed in the form of probability amplitude.

The m -bit qubit code can be expressed as:

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m \\ \beta_1 & \beta_2 & \cdots & \beta_m \end{bmatrix}; \quad (11)$$

where (α_i, β_i) is the i -th qubit in the chromosome, which is a pair of complex numbers, representing the probability amplitude of the corresponding bit, and satisfying $|\alpha|^2 + |\beta|^2 = 1$, $|0\rangle$ and $|1\rangle$ represent two different bit states. When the qubit is not observed, the qubit is in the basic state. After observing the qubit, the qubit will collapse to 0 or 1 with the probability of $|\alpha|^2$ or $|\beta|^2$.

(2) The quantum rotation gate is used to replace the selection, crossover and mutation operations, and the population is updated by changing the quantum rotation angle. The qubit is updated to

$$\begin{vmatrix} \alpha_i' \\ \beta_i' \end{vmatrix} = \begin{vmatrix} \cos \Delta\theta_i & -\sin \Delta\theta_i \\ \sin \Delta\theta_i & \cos \Delta\theta_i \end{vmatrix} \begin{vmatrix} \alpha_i \\ \beta_i \end{vmatrix}. \quad (12)$$

where: (α_i', β_i') denotes the updated i -qubit; $\Delta\theta_i$ is the quantum rotation angle.

A Quantum genetic algorithm has the characteristics of fast convergence speed and strong global optimization ability; however, the quantum angle rotation strategy can only rotate at a fixed size, it is easy to fall into local optimization. In this paper, we use adaptive adjustment of quantum rotation angle to improve the quantum genetic algorithm. When the individual fitness approaches the optimal individual, the quantum rotation angle decreases, which makes the individual evolve in a good direction; when the individual fitness is far away from the optimal individual, the quantum rotation angle increases, expands the search space and jumps out of the local optimum.

In the process of quantum gate rotation and update, it cannot reach the global optimal solution with probability "1", which can easily fall into premature convergence. Therefore, this paper introduces crossover and mutation operations in traditional GA to increase the search space and avoid falling into the local optimal solution. The selection of crossover probability and mutation probability in the genetic algorithm is the key to affecting the behavior and performance of the genetic algorithm. If the crossover probability is too large, the probability of a genetic mode being destroyed is greater, which makes the individual structure with high fitness be destroyed quickly; if the crossover probability is too small, the search process is slow, so as to stagnate. For the mutation probability, if the value is too small, it is not easy to produce a new structure; if the value is too large, it will become a pure random search algorithm. The adaptive algorithm can solve this problem well, and the crossover probability and mutation probability can be adjusted automatically with the fitness. For the individuals whose fitness is higher than the average fitness of the population, the lower crossing probability and mutation probability should be adopted to protect the individual into the next generation, while the individual with fitness lower than the average value should be selected with a higher cross probability and mutation probability to make the individual eliminated.

In the adaptive strategy, the quantum rotation angles $\Delta\theta$, the crossover probability P_c and the mutation probability P_m are calculated as follows:

$$\Delta\theta = \theta_{\min} + \frac{f_{\max} - f_x}{f_{\max} - f_{\min}}(\theta_{\max} - \theta_{\min}); \quad (13)$$

$$P_c = \begin{cases} \frac{k_1(f_{\max} - f)}{f_{\max} - f_{avg}} & f \geq f_{avg}; \\ k_2 & f < f_{avg} \end{cases}; \quad (14)$$

$$P_m = \begin{cases} \frac{k_3(f_{\max} - f')}{f_{\max} - f_{avg}} & f' \geq f_{avg} \\ k_4 & f' < f_{avg} \end{cases}. \quad (15)$$

where: $f_{\max}$ is the maximum fitness of the population; f_{avg} is the maximum fitness of the population; f is the average fitness of the population; f is the larger fitness of the two individuals to be crossed; f' is the fitness of the mutant individual; f_x is the current individual fitness; $\theta_{\max}$ is the maximum value of $\Delta\theta$ interval; $\theta_{\min}$ is the minimum value of $\Delta\theta$ interval; and k_1, k_2, k_3 and k_4 are constants. Compared with QGA, the SQGA's crossover and mutation probabilities are generated adaptively by Formulas (14) and (15) to perform quantum crossover mutation operation. The algorithm can dynamically adjust the key parameters of the algorithm, improve the search efficiency and find the optimal solution conveniently, which is suitable for determining the weight of decision-making level.

3.3. Algorithm Flow

According to the previous analysis and the actual situation of fatigue detection, this paper adopts heterogeneous integration. The current mature machine learning methods with better generalization performance: SVM, KNN, Bayesian network (BN) and BPNN are selected as the learning methods of the base classifier. In order to further improve the differences between base classifiers, pattern-level and feature-level selection training set methods are adopted. An ensemble learning detection algorithm based on SQGA optimized weights is proposed. The specific algorithm process is shown in Figure 4.

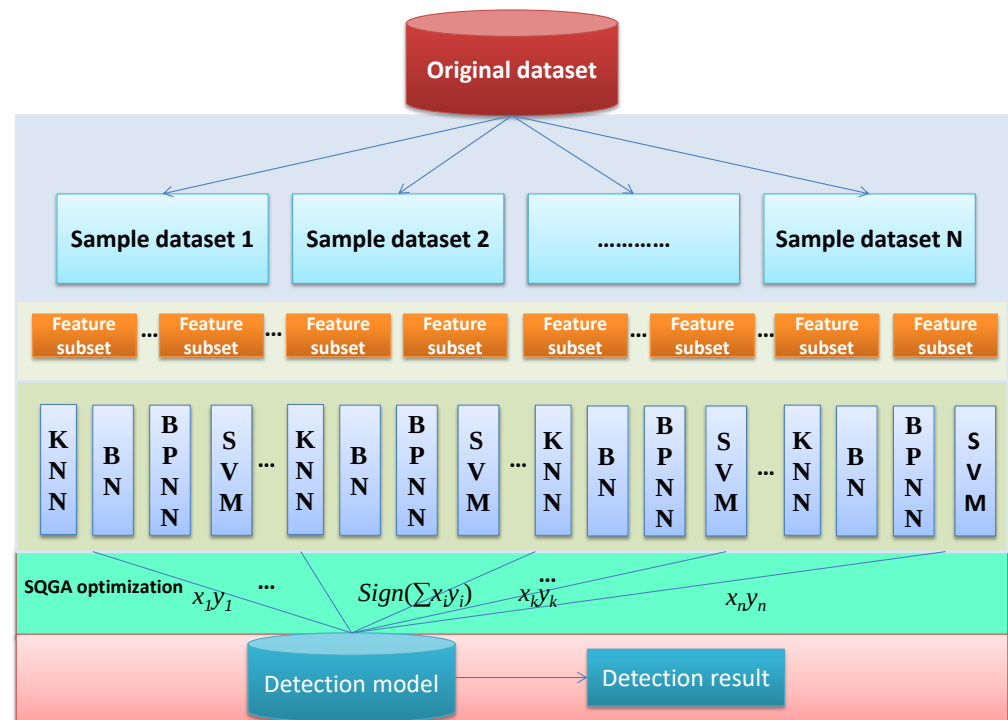


Figure 4. Flow chart of SQGA optimized ensemble learning.

The specific steps of the algorithm are as follows:

Step 1: According to the bagging method, samples are resampled to determine multiple sample subsets;

Step 2: Feature-level method is used to randomly select features of the previously obtained sample subsets;

Step 3: Use different learning methods to learn the subsets obtained in step 2. The parameters of each base classifier are arbitrarily set;

Step 4: Weight the results of each base classifier. The weights are obtained by SQGA in Section 3.2, and the final detection results are determined by symbolic function.

4. Algorithm Verification

In order to verify the effectiveness of the proposed method, the general dataset [29] is applied to investigate the classification performance. Here, we select Thyoid, Ionosphere, Image and German datasets. The performance comparison between the adopted method and the traditional classification methods, as well as the bagging and boosting ensemble algorithms, are investigated.

In reference [27], the paper selects 40 base classifiers and constructs the base classifiers based on the proposed method, bagging and boosting, respectively. Samples are resampled in the first layer five times. On this basis, the feature selection is carried out twice for each subset at the proportion of 80%. We obtained 10 training sets, SVM, KNN, BN and BPNN

were adopted, and 40 base classifiers were obtained. Finally, the majority voting method was used to get the final classification results.

Firstly, the convergence performance of SQGA was investigated. The parameters of each base classifier were arbitrarily set, as the learning ability of the selected learning machine is strong. In the SQGA optimization algorithm, the number of populations was set to 50 and iterations set to 100. The experimental results are shown in Figure 5.

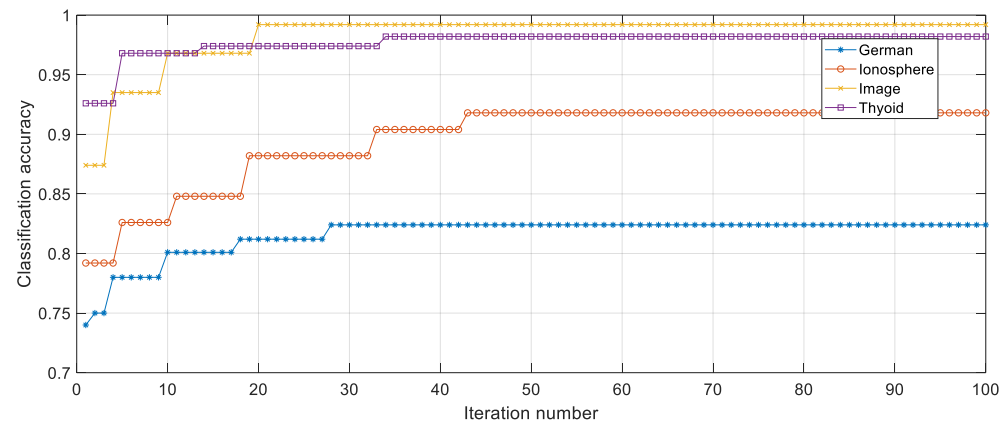


Figure 5. SQGA optimization results of different datasets.

From Figure 5, we can see that, at the beginning, the classification accuracy of each dataset is relatively high. This shows that the ensemble learning algorithm can achieve better learning results, and the whole algorithm is stable. On different datasets, the algorithm converges rapidly, which shows that the convergence of the SQGA algorithm is better, and each algorithm is finally stable at a relatively high classification accuracy.

In order to verify the SQGA's advantages of convergence, the QGA, genetic algorithm (GA) [30] and self-adaptive genetic algorithm (SAGA) [31,32] were selected for comparison. The Image dataset was picked for testing, and the convergence curves are shown in Figure 6.

From Figure 6, we can find that: SQGA converges in 20 generations, accuracy is 99.2, GA converges in the 52th generation, and accuracy is 98.8. Compared with the other three methods, SQGA has the characteristics of fast convergence and good search results.

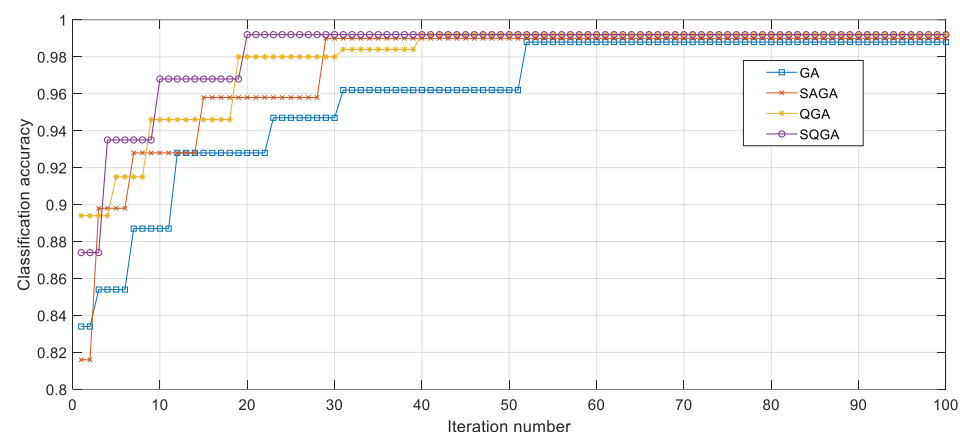


Figure 6. Optimization results from different methods.

Next, we compared the results of the proposed method with other single classification methods. The single classifier adopts the base classifier method in this paper: SVM, KNN, BN and BPNN. Both SVM and BPNN select Gaussian function as kernel function and activation function, and the parameters of each learner were determined by the 5-fold cross-validation and grid search methods. The classification results are given in Table 1.

Table 1. Experimental accuracy on general datasets.

Dataset	Classification Accuracy (%)				
	KNN	BN	BPNN	SVM	Proposed Method
Thyroid	94.2	95.8	95.3	96.0	98.2
Ionosphere	82.4	83.7	85.5	87.2	91.8
Image	93.6	95.4	96.3	96.9	99.2
German	75.6	77.8	76.9	78.6	82.4

From the experimental results, we can see that the classification accuracy of KNN is basically the lowest. Compared with other learning methods, the KNN algorithm is simple and efficient but it extracts the least classification information. SVM, BN and BPNN achieve similar classification results. These three methods use different learning mechanisms to learn samples from different angles, and they can better deal with nonlinear classification problems. The proposed method can achieve optimal classification accuracy, which is because the errors and omissions of some individual classifiers can be compensated by other individual classifiers through the ensemble of each classifier. This ensemble learning method can further improve the generalization ability of the overall classification.

Then, we analyzed the performance between the proposed method and the traditional ensemble learning algorithm. The traditional ensemble algorithm selects the classical bagging and boosting algorithms, and the number of base classifiers was 40. The differences between base classifiers can reflect the effect of ensemble learning. Reference [33] gives the definition of difference degree: the calculation formula of the difference degree of T base classifiers on N samples is as follows:

$$Div = \frac{1}{TN} \sum_{t=1}^T \sum_{i=1}^N d_t(\mathbf{x}_i) \quad (16)$$

$$d_t(\mathbf{x}_i) = \begin{cases} 0 & \text{if } h_t(\mathbf{x}_i) = f(\mathbf{x}_i) \\ 1 & \text{if } h_t(\mathbf{x}_i) \neq f(\mathbf{x}_i) \end{cases} \quad (17)$$

where $h_t(\mathbf{x}_i)$ is the prediction label of the t -th individual classifier on sample $\mathbf{x}_i$, and $f(\mathbf{x}_i)$ is the prediction result after integrating all individual classifiers. The difference threshold is not less than 80% of the difference degree of the previous individual SVM classifiers. The larger the difference degree is, the smaller the correlation between the base classifiers is, and the final integration effect will be better. We calculated the average difference of the three ensemble algorithms on different datasets, and the results are shown in Table 2.

Table 2. Average difference degree between base classifiers.

Dataset	Average Difference		
	Bagging	Boosting	Proposed Method
Thyroid	0.075	0.084	0.104
Ionosphere	0.181	0.192	0.223
Image	0.014	0.013	0.021
German	0.214	0.227	0.283

As can be seen from Table 2, the difference degree of the bagging and boosting integration methods is not particularly large. In comparison, the average difference degree of the proposed method is significantly higher than that of the previous two methods. This is because the method in this paper uses heterogeneous integration. Next, we compared the classification accuracy of several ensemble algorithms to investigate their generalization performance. Here, we focused on the SQGA optimization and decision fusion results only by voting. Table 3 shows the classification accuracy of the four methods:

Table 3. Classification accuracy results of ensemble algorithm.

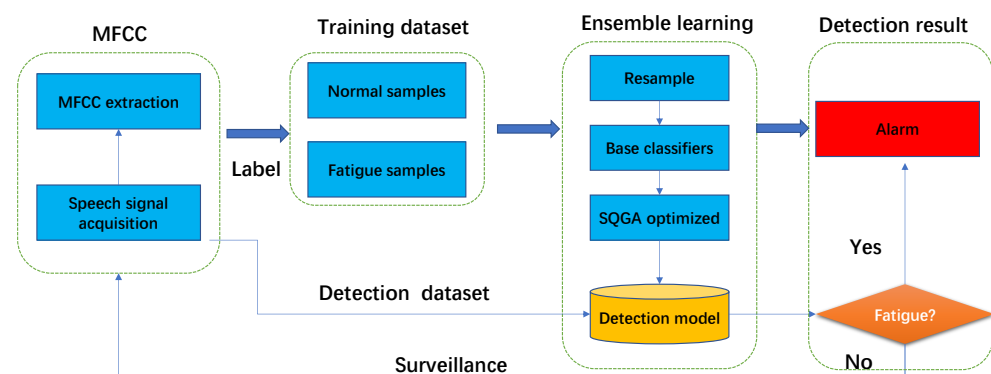
Dataset	Classification Accuracy (%)			
	Bagging	Boosting	Vote	SQGA
German	79.5	79.5	80.4	82.4
Ionosphere	88.1	89.4	90.4	91.8
Image	98.4	98.2	98.8	99.2
Thyoid	96.4	96.9	97.5	98.2

From the experimental results, we can see that the classification accuracy of ensemble algorithms is improved. The classification accuracy of voting fusion is higher than traditional bagging and boosting, due to the increased differences between base classifiers. Compared with the simple voting method, the SQGA optimization method can obtain the highest classification accuracy, which shows that the method in this paper can further improve the performance of ensemble learning. This also lays a foundation for its application in fatigue detection.

5. The Application in Fatigue Detection

The above experiments verify that the SQGA optimized ensemble learning algorithm proposed in this paper can obtain better accuracy in dealing with the binary classification problem. In this section, the algorithm is applied to real fatigue detection.

The Figure 7 gives the flow chart of fatigue detection. The whole process is mainly divided into four stages: MFCC extraction, training dataset acquisition, ensemble learning to establish the detection model and detection stage.

**Figure 7.** The flow chart of fatigue detection.

MFCC extraction stage: collect the voice of controllers and extract the speech by MFCC.

Training dataset acquisition: label the MFCC processed information according to the controller's own feelings and obtain the original training dataset.

Ensemble learning stage: use the ensemble learning method proposed in Section 3 for establishing the detection model.

Detection stage: Collect real-time voice data from the controller and extract it by MFCC. Input the processed data into the detection model, which then judges the controller's fatigue state.

The voice data from land–air communication used in this paper comes from a front-line control unit in China, including the voice data from the unit's tower, approach and controllers of regional control positions. The date of voice data recordings is from 3–7 April 2022 and from 16–21 April 2022, from 2:00 to 24:00 every day. According to the structural characteristics of land–air communication, the voice data from the controller in the land–air communication data was intercepted. Then, the corresponding MFCC was obtained by speech preprocessing technology. The fatigue state of voices and the relevant information corresponding to the controllers are marked according to the controllers' own feelings.

The original dataset includes 158,620 samples. The ratio of normal samples and fatigue samples is about 10:1. In this experiment, the training samples were selected according to the ratio of 10:1. A total of 5000 normal samples and 500 fatigue samples were randomly selected from the training set. The test samples consisted of 1000 normal samples and 200 fatigue samples. The detection experiments were carried out based on the single machine learning method (SVM is selected here), ensemble learning algorithms, bagging and boosting and the method in this paper. The base classifiers ensemble by bagging and boosting algorithms are SVM. The false positive rate, false negative rate and detection accuracy of the whole body were compared. Table 4 shows the test results:

Table 4. Experimental results of fatigue detection.

Methods	Normal Detection Number	False Positive Rate (%)	False Detection Number	False Negative Rate (%)	Accuracy (%)	Train Time (s)	Test Time (s)
SVM	946	5.40	182	9.00	94.00	3.2	0.13
Bagging	962	3.80	187	6.50	95.75	1.4	0.33
Boosting	964	3.60	186	7.00	95.83	2.1	0.35
Proposed method	988	1.20	194	3.00	98.50	3.5	0.36

From the detection results obtained in the above table, it can be seen that the false positive rate and false negative rate of fatigue detection with the SVM algorithm are higher than those of the ensemble algorithm. We can also see the advantages of ensemble learning. Among the three ensemble algorithms, the proposed algorithm can achieve the best detection effect, which is consistent with the experimental results obtained on the UCI dataset. By increasing the differences between base classifiers, the generalization performance of the algorithm is improved. The training time of SVM is 3.2 s, and the grid search to find the optimal parameters costs a lot of time. Ensemble learning methods are not sensitive to parameters and can greatly reduce computational complexity. For the testing time, as the single SVM obtains the simplest model, it only cost 0.13 s, while the proposed method cost 0.36 s. This is also good proof of the “no free lunch” theory. Nevertheless, we still feel that the cost is acceptable.

Next, we analyzed the performance of the detection algorithm from the receiver operation characteristic curve (ROC) [34]. The ROC curve is used to judge whether the detection model is good or bad, mainly from the curve shape and AUC (area under the curve). If the curve is closer to the upper left corner, the better the detection effect is. On the contrary, it means that the detection effect of the model is worse. AUC refers to the area under the ROC curve, which can reflect the diagnostic value of the detection model. The closer the AUC value is to 1, the better the performance of the model is. Figure 6 shows the ROC curves of several algorithms.

From Figure 8, we can see that, on the whole, the ROC curves of the previous algorithms are more inclined to the upper left corner, indicating that they are better detectors. Among them, the algorithm proposed in this paper has the best bias, which is obviously better than other detection methods. In terms of AUC, the offline area of this method is larger, reaching 0.992, which is higher than the other three methods. In conclusion, the advantages of the proposed method can be seen from the ROC curve.

In the detection method proposed in this paper, there is a certain randomness in the construction of base classifiers and the optimization of weights. In order to verify the stability of the algorithm, we used SQGA to do 50 Monte Carlo simulation experiments and analyzed the experimental results with a box diagram. In the same way, we made a box diagram for bagging, boosting and vote, respectively. We counted the Monte Carlo simulation experiments’ false positive rates, false negative rates and detection accuracy of these methods, respectively. The experimental results are shown in Figures 9–11.

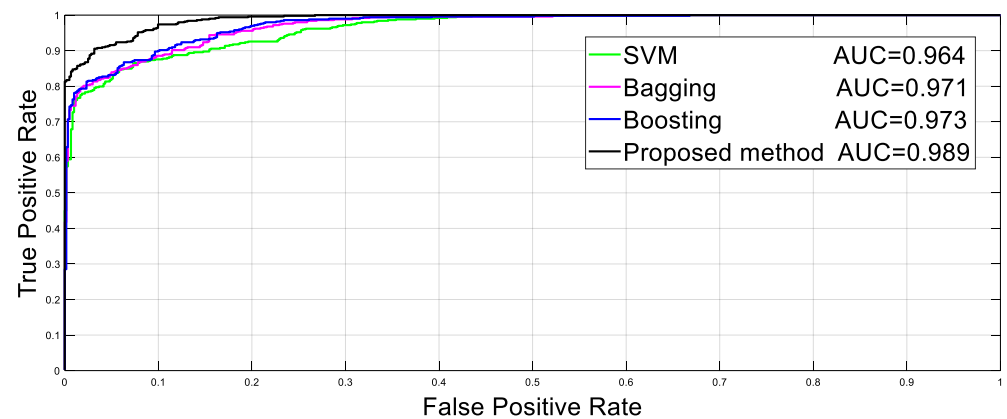


Figure 8. ROC curves of different detection methods.

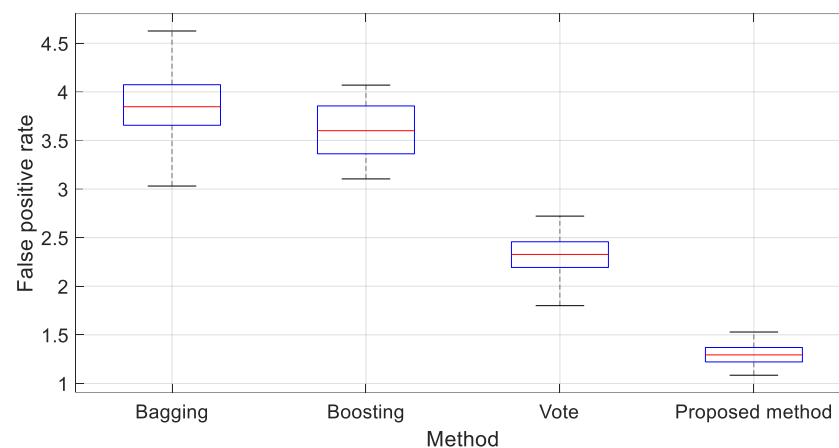


Figure 9. Box diagram of false positive rate.

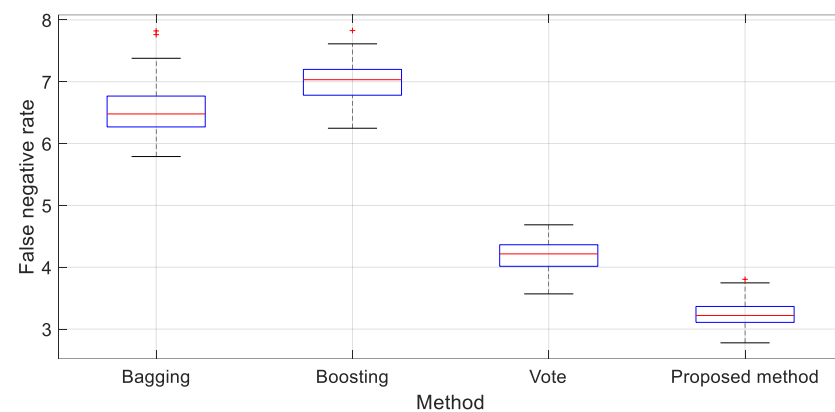


Figure 10. Box diagram of false negative rate.

The horizontal line in the box diagram represents the statistical median, the height of the box represents the distribution interval, and '+' stands for abnormal value in different methods. From the experimental results, we can clearly see that the height of the proposed method box is the lowest among the three statistics, which shows that the method proposed in this paper is very stable. Compared with bagging and boosting, the vote method can also achieve better stability, which also reflects that the heterogeneous integration method in this paper can increase the diversity of base classifiers. The experimental results show that the SQGA can not only obtain good search results (minimum false positive rate, false negative rate and maximum detection accuracy) but also has good convergence characteristics and minimal outliers.

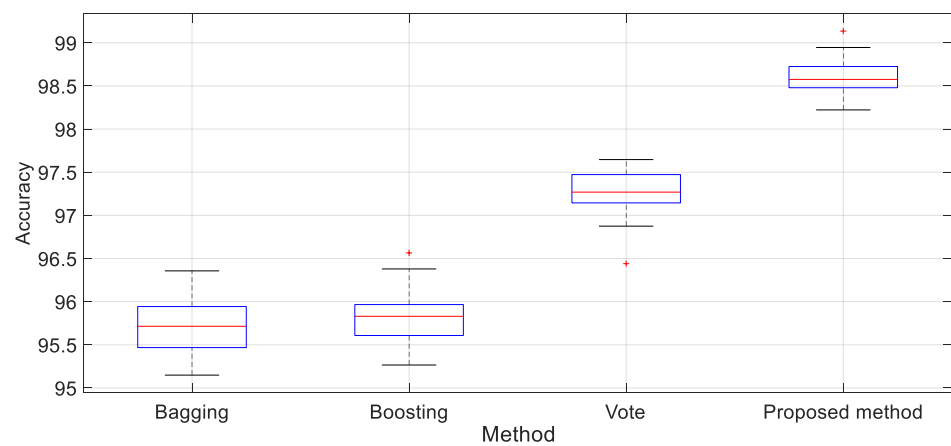


Figure 11. Box diagram of detection accuracy.

In reality, there may be more fatigue happening during some periods and less during other periods. Corresponding to the dataset, it shows the proportion of normal samples and fatigue samples. This also poses some challenges to the training data collection process. In theory, the larger the dataset, the better the detection effect of the model. However, it is also possible that in the early detection stage, the number of fatigue events collected is relatively small. In this case, can this detection model perform well? Here, we investigate the detection effects of various detection models under different fatigue intensities (the ratio of fatigue samples to normal samples). The values of 5%, 10%, 15% and 20% are used to simulate the samples under different fatigue intensities, and the detection model is trained with these four fatigue samples. The detection accuracy under different learning modes is shown in Figure 12.

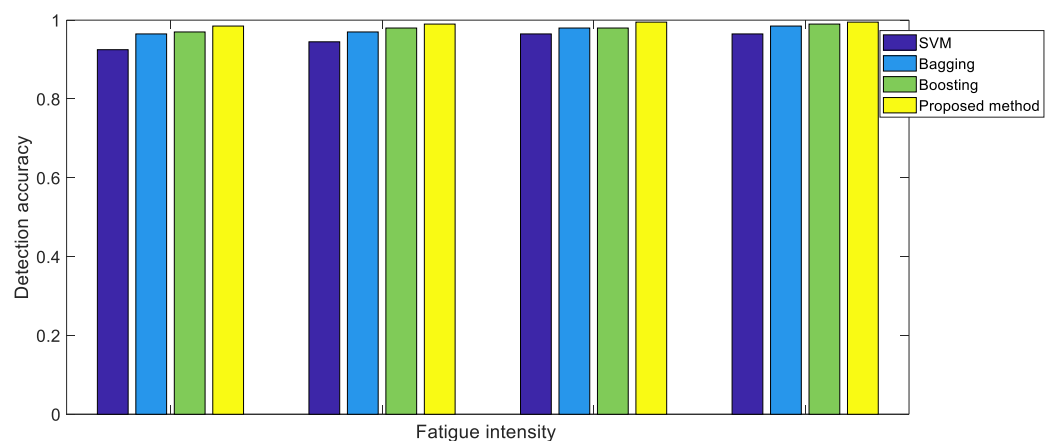


Figure 12. Detection accuracy under different fatigue intensities.

It can be seen from Figure 10 that under different fatigue intensities, the training detection model can detect fatigue events better. The detection accuracy of SVM is 92.5% when there are fewer fatigue samples (fatigue intensity is 5%) and the miss detection rate is relatively high. This is mainly due to the imbalance of samples, which makes the classification plane of SVM shift. The imbalance of samples affects the effect of machine learning. Compared with other algorithms, the detection accuracy of this method can be maintained at a relatively high level (the lowest is 98.6%), which shows that the proposed algorithm is stable and can also obtain better detection results in the case of unbalanced samples.

6. Discussion

In this paper, speech-based fatigue-state detection is proposed. Compared to other objective detection methods using auxiliary tools, the speech-based method is more applicable due to its convenient collection method.

The speech signal is extracted by MFCC. On this basis, fatigue is detected by the ensemble learning method. The experiment result shows that the method proposed in this paper can detect fatigue of ATC effectively. Because the idea of ensemble learning is adopted, the disadvantage of a single learning machine is overcome. The following stability analysis and asymmetric training set experiments prove the superiority of the method. However, due to the multiple base classifiers, establishing the complexity of computation in the training process is increased. Overall, the experiment results verify that the proposed method can accurately monitor the fatigue state of controllers and realizes the original intention of the method design.

The data used in the experiment is from only one control unit. To improve the generalization of the detection model, more data needs to be collected. Anyway, for the future, we are ready to further explore the methods of speech signal extraction, hoping to find better features and improve detection accuracy.

7. Conclusions

In order to improve the practicability of the controller's fatigue state detection, the paper detects the controller's fatigue state through the controller's radiotelephony communications voice. ATC fatigue detection is a typical binary classification problem, and the detection accuracy of a single learning machine is not high, so we adopted the heterogeneous ensemble learning method. By increasing the differences between the training set and the training method in base classifiers, the overall integration effect was improved. At the same time, in order to further improve the detection accuracy, we improved the result of the fusion part of ensemble learning. This was achieved through SQGA algorithm optimization, to find the best base classifier weights, and through the weighting to get the final demerit detection. The experimental results show that the proposed method is stable and accurate. Therefore, this method has a certain practical value in the practical application of ATC fatigue detection.

Author Contributions: Conceptualization, N.W. and J.S.; methodology, N.W.; figure, N.W.; data analysis, N.W. and J.S.; writing—original draft preparation, N.W.; writing—review and editing, J.S.; supervision, N.W. and J.S.; project administration, J.S.; and funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Science Foundation of China (Grant number 71801221).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, J.S., upon reasonable request.

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

References

1. Bendak, S.; Rashid, H.S.J. Fatigue in aviation: A systematic review of the literature. *Int. J. Ind. Ergon.* **2020**, *76*, 102928. [\[CrossRef\]](#)
2. Yen, J.-R.; Hsu, C.-C.; Ho, H.; Lin, F.-F.; Yu, S.-H. Identifying flight fatigue factors: An econometric modeling approach. *J. Air Transp. Manag.* **2005**, *11*, 408–416. [\[CrossRef\]](#)
3. Chang, Y.-H.; Yang, H.-H.; Hsu, W.-J. Effects of work shifts on fatigue levels of air traffic controllers. *J. Air Transp. Manag.* **2019**, *76*, 1–9. [\[CrossRef\]](#)
4. Wang, L.L.; Chen, F.L. Study on relationship between controllers' cognitive behavior and fatigue based on EEG. *China Saf. Sci. J.* **2018**, *28*, 1–6.
5. Xinyun, H.; Gabriel, L. Exploration of the effects of task-related fatigue on eye-motion features and its value in improving driver fatigue-related technology. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *80*, 150–171.
6. Chai, R.; Naik, G.R.; Nguyen, T.N.; Ling, S.H.; Tran, Y.; Craig, A.; Nguyen, H.T. Driver Fatigue Classification with Independent Component by Entropy Rate Bound Minimization Analysis in an EEG-Based System. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 715–724. [\[CrossRef\]](#)

7. Mandal, B.; Li, L.; Wang, G.S.; Lin, J. Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 545–557. [CrossRef]
8. Jin, H.B.; Zhu, G.L. Effectiveness of eye Movement Indicators on Air Traffic Controllers' Fatigue Detection. *Sci. Technol. Eng.* **2018**, *18*, 136–140.
9. Shen, Z.; Pan, G.; Yan, Y. A High-Precision Fatigue Detecting Method for Air Traffic Controllers Based on Revised Fractal Dimension Feature. *Math. Probl. Eng.* **2020**, *2020*, 4563962. [CrossRef]
10. Bu, J.; Liu, Y.X.; Wang, Y.J. Relationship between air traffic controllers' eye movement and fatigue. *Acta Aeronaut. Astronaut. Sin.* **2017**, *38*, 721525.
11. Sun, R.; Ma, G.F.; Yuan, L.P. Analysis of risk of controller fatigue based on characteristics of speech reaction time. *China Saf. Sci. J.* **2016**, *26*, 7–12.
12. Krajewski, J.; Sommer, D.; Schnupp, T.; Laufenberg, T.; Heinze, C.; Golz, M. Applying nonlinear dynamics features for speech-based fatigue detection. In Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research, Eindhoven, The Netherlands, 24–27 August 2010; pp. 1–5.
13. Greeley, H.; Friets, E.; Wilson, J.; Raghavan, S.; Picone, J.; Berg, J. Detecting fatigue from voice using speech recognition. In Proceedings of the 2006 IEEE International Symposium on Signal Processing and Information Technology, Vancouver, BC, Canada, 27–30 August 2006; pp. 567–571.
14. Pawar, M.D.; Kokate, R.D. Convolution neural network based automatic speech emotion recognition using Mel-frequency Cestrum coefficients. *Multimed. Tools Appl.* **2021**, *80*, 15563–15587. [CrossRef]
15. Fahad, M.S.; Deepak, A.; Pradhan, G.; Yadav, J. DNN-HMM-based speaker-adaptive emotion recognition using MFCC and epoch-based features. *Circuits Syst. Signal Process.* **2021**, *40*, 466–489. [CrossRef]
16. Shahin, I.; Nassif, A.B.; Hamsa, S. Emotion recognition using hybrid Gaussian mixture model and deep neural network. *IEEE Access* **2019**, *7*, 26777–26787. [CrossRef]
17. Mohanty, M.N.; Palo, H.K. Child emotion recognition using probabilistic neural network with effective features. *Measurement* **2020**, *152*, 107369. [CrossRef]
18. Inoue, K.; Hara, S.; Abe, M. Model architectures to extrapolate emotional expressions in DNN-based text-to-speech. *Speech Commun.* **2021**, *126*, 35–43. [CrossRef]
19. Xiao, Z.; Hu, Z.; Geng, L.; Zhang, F.; Wu, J.; Li, Y. Fatigue Driving Recognition Network: Fatigue Driving Recognition Via Convolutional Neural Network and Long Short-term Memory Units. *IET Intell. Transp. Syst.* **2019**, *13*, 1410–1416. [CrossRef]
20. Ye, C.H.; Yin, Z.; Zhao, M.Y.; Tian, Y.; Sun, Z. Identification of mental fatigue levels in a language understanding task based on multi-domain EEG features and an ensemble convolutional neural network. *Biomed. Signal Process. Control.* **2022**, *72*, 103320. [CrossRef]
21. Rohit, H.; Vedant, J.; Manan, S. A comprehensive review of approaches to detect fatigue using machine learning techniques. *Chronic Dis. Transl. Med.* **2021**, *8*, 26–35.
22. Liu, Y.S.; Lan, Z.R.; Cui, J.; Sourina, O.; Müller-Wittig, W. Inter-subject transfer learning for EEG-based mental fatigue recognition. *Adv. Eng. Inform.* **2020**, *46*, 101157. [CrossRef]
23. Shen, Z.; Wei, Y. A high-precision feature extraction network of fatigue speech from air traffic controller radiotelephony based on improved deep learning. *ICT Express* **2021**, *7*, 403–413. [CrossRef]
24. Hasni, H.; Alavi, A.H.; Jiao, P.; Lajnef, N. Detection of fatigue cracking in steel bridge girders: A support vector machine approach. *Arch. Civ. Mech. Eng.* **2017**, *17*, 609–622. [CrossRef]
25. Zhou, G.; Guo, F.L. Research on ensemble learning. *Comput. Technol. Autom.* **2018**, *37*, 148–153.
26. Ghaemi, A.; Safari, A.; Afsharirad, H.; Shayeghi, H. Situational awareness and deficiency warning system in a smart distribution network based on stacking ensemble learning. *Appl. Soft Comput.* **2022**, *128*, 109427. [CrossRef]
27. Xu, J.W.; Yang, Y. A survey of ensemble learning approaches. *J. Yunnan Univ.* **2018**, *40*, 1082–1092.
28. Zhou, C.; Pan, P.; Huang, L. Authenticity identification of speaker digital recording data based on quantum genetic algorithm. *Multimed. Tools Appl.* **2018**, *77*, 19399–19413. [CrossRef]
29. Available online: <https://archive.ics.uci.edu/ml/datasets.php> (accessed on 2 August 2022).
30. Renato, T. NK Hybrid Genetic Algorithm for Clustering. *IEEE Trans. Evol. Comput.* **2018**, *22*, 748–761.
31. Kivijarvi, J.; Franti, P.; Nevalainen, O. Self-Adaptive Genetic Algorithm for Clustering. *J. Heuristics* **2003**, *9*, 113–129. [CrossRef]
32. Lu, H.-L.; Wen, X.-S.; Lan, L.; An, Y.-Z.; Li, X.-P. A self-adaptive genetic algorithm to estimate JA model parameters considering minor loops. *J. Magn. Magn. Mater.* **2015**, *374*, 502–507. [CrossRef]
33. Dietterieh, T.D. Ensemble Methods in Machine Learning. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
34. Tadeusz, P. On the Use of Roc Analysis for the Optimization of Abstaining Classifiers. *Mach. Learn.* **2007**, *68*, 137–169.