

## Article

# Real-Time Psychological Stress Detection According to ECG Using Deep Learning

Pengfei Zhang<sup>1,2</sup>, Fenghua Li<sup>3</sup>, Rongjian Zhao<sup>1,2,\*</sup>, Ruishi Zhou<sup>1,2</sup>, Lidong Du<sup>1</sup> , Zhan Zhao<sup>1</sup>, Xianxiang Chen<sup>1</sup> and Zhen Fang<sup>1,2,\*</sup>

- <sup>1</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100000, China; zhangpengfei17@mails.ucas.ac.cn (P.Z.); zhouruishi17@mails.ucas.ac.cn (R.Z.); lddu@mail.ie.ac.cn (L.D.); zhaozhan@mail.ie.ac.cn (Z.Z.); chenxx@aircas.ac.cn (X.C.)
- <sup>2</sup> School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100000, China
- <sup>3</sup> Institute of Psychology, Chinese Academy of Sciences, Beijing 100000, China; lifh@psych.ac.cn
- \* Correspondence: zhaorij@aircas.ac.cn (R.Z.); zfang@mail.ie.ac.cn (Z.F.)

**Abstract:** Today, excessive psychological stress has become a universal threat to humans. That stress can heavily affect work and study when a person repeatedly is exposed to high stress. If that exposure is long enough, it can even cause cardiovascular disease and cancer. Therefore, both monitoring and managing of stress is imperative to reduce the bad outcomes from excessive psychological stress. Conventional monitoring methods firstly extract the characteristics of the RR interval of an electrocardiogram (ECG) from a time domain and a frequency domain, then use machine learning models, like SVM, random forest, and decision tree, to distinguish the level of that stress. The biggest limitation of using these methods is that at least one minute of ECG data and other signals are indispensable to ensure the high accuracy of the results. This will greatly affect the real-time application of the models. To satisfy real-time detection of stress with high accuracy, we proposed a framework based on deep learning technology. The proposed monitoring framework is based on convolutional neural networks (CNN) and bidirectional long short-term memory (BiLSTM). To evaluate the performance of this network, we conducted the experiments applying conventional methods. The data for the 34 subjects were collected on the server platform created by the group at the Institute of Psychology of the Chinese Academy of Sciences and our group. The accuracy of the proposed framework was up to 0.865 on three levels of stress using a 10 s ECG signal, a 0.228 improvement compared with conventional methods. Therefore, our proposed framework is more suitable for real-time applications



**Citation:** Zhang, P.; Li, F.; Zhao, R.; Zhou, R.; Du, L.; Zhao, Z.; Chen, X.; Fang, Z. Real-Time Psychological Stress Detection According to ECG Using Deep Learning. *Appl. Sci.* **2021**, *11*, 3838. <https://doi.org/10.3390/app11093838>

Academic Editor: Pasi Fränti

Received: 10 March 2021

Accepted: 19 April 2021

Published: 23 April 2021

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



**Copyright:** © 2021 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/>).

**Keywords:** psychological stress; deep learning; CNN; BiLSTM; real-time

## 1. Introduction

Economic development has led to fierce competition among people. People are now more prone to be affected by high psychological stress. Excessive psychological stress reduces work efficiency, affects relationships and transportation safety [1]. Long-term stress can even induce depression, addiction, and cardiovascular and cerebrovascular diseases [2,3]. Excessive emotional stress has become a major problem affecting human physical and mental health. MAE (ecological momentary assessment) [4] and JITAI (just-in-time adaptive interventions) [5] are two effective methods that can be used to deal with the negative consequences of excessive psychological stress. However, the two methods both require real-time monitoring of psychological stress, and the lack of a method that can monitor emotional stress in real time has become the main problem.

It is fortunate that there is an important relationship between psychological stress and the autonomic nervous system, i.e., SNS (sympathetic nervous system) and PNS (parasympathetic nervous system). The exposure to stress will enhance SNS's function,

including shortening the RR interval, increasing the low-frequency energy of HRV (heart rate variability), increasing the respiration rate, decreasing HRV's high-frequency energy, and more [6]. After stress, the PNS will be activated to generate the opposite effect. Therefore, the features of the RR interval and other physiological parameters could be used to detect stress. As an electrocardiogram (ECG) signal contains RR interval information and is easy to obtain, a lot of research has been undertaken on judging psychological stress via the ECG signal. Traditionally, five-minute ECG data have been used to detect stress [7]. Although the accuracy of these studies is reliable, the 5 min time phase is too long for real-time monitoring. To promote real-time application, some researchers have successfully detected stress on one-minute ECG and respiration [8] with high accuracy. However, the acquisition of breathing signals requires wearing a bandage, which will greatly reduce the comfort. Furthermore, the interval of one minute is not short enough to satisfy the requirements of real-time monitoring.

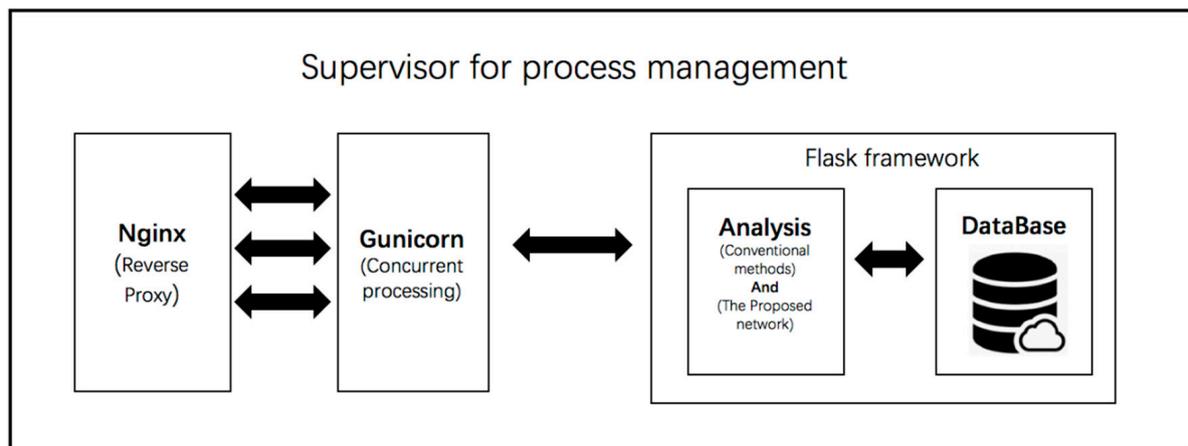
Deep learning technology, proposed in recent years, has greatly promoted the development of artificial intelligence, especially for the CNN (convolutional neural network) and RNN (recurrent neural network). There have been some studies that have used deep learning techniques to monitor psychological stress. Winata used the LSTM (long short-term memory) model with an attention mechanism to classify stress for spoken language, and the accuracy achieved was 0.741 [9]. Li Jin built models using DNN (deep neural network) and a conditional random field to classify stress regarding information on adolescent social behavior [10]. Lin used a deep sparse neural network to monitor the stress level from cross-media microblog data [11]. Bosun Hwang [12] categorized stress using 10 s ECG data with CNN and LSTM, but there are only two categories. Se-Hui Song [13] categorized stress into four classes, using SBP (systolic blood pressure), DBP (diastolic blood pressure), sleep time, heart rate, and age with DBN (deep belief network) and the accuracy only reached 0.66, and the data sleep time, SBP, DBP used did not satisfy a real-time application well. Although these studies have proven their models to be effective, they did not consider the need for real-time application.

To address this issue, considering the effectiveness of deep learning technology and the easy accessibility of ECG signals, we proposed a network according to ECG using deep learning technology. For this proposed network, we also built a psychological stress monitoring platform on which data collection and analysis function were integrated. Using it, the ECG data of 34 people were collected and used for training and testing the models. To show the proposed network's performance, we also conducted the conventional feature extracting method applying nine machine learning methods (SVM, decision tree, random forest and so on). In the current study, we divided mental stress into three levels, i.e., low, medium and high. To the best of our knowledge, it is the first time that deep learning technology has been applied to detect three levels of psychological stress by only using a 10 s ECG signal from a data set that is collected based on the Montreal model. The proposed framework with deep learning is more suitable for real-time applications.

## 2. Materials and Methods

### 2.1. Experiments and Data Acquisition

To acquire ECG signal acquisition, storage, management, and real-time analysis, we built an emotional stress-monitoring platform that contained server and front-ends. After receiving the ECG data through the 4G net, the server could analyze and then send the results back to the front-end in a timely manner. We designed the server based on Flask and used MongoDB [14] for the database, Nginx [15] for a reverse proxy, and Gunicorn [16] for concurrent processing. The software supervisor was also used to monitor the programs of the main program and the database so the system could automatically restart for continuous running when stopped for any bugs. The structure of the server is shown in Figure 1.



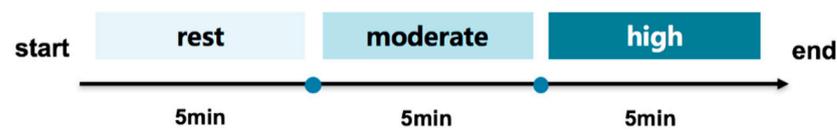
**Figure 1.** The server based on Flask for data collecting and processing with Nginx, Gunicorn MongoDB and Supervisor technology.

The front end consists of sticky devices and mobile apps. The sticky devices we used were created by our lab using the chip ADS1292 for ECG signal collection and CC2640 for data transmission as shown in Figure 2. Compared to other devices, the two-electrode structure makes the device more comfortable to wear due to small volume and a light weight. When using the sticky device was placed horizontally on the center of the chest in the fourth rib or fifth rib area. The ECG sampled at 250 Hz would be transmitted to the mobile app via Bluetooth 4.0, and finally. The server would receive the data through the 4G net.



**Figure 2.** Front and back surface of the sticky device.

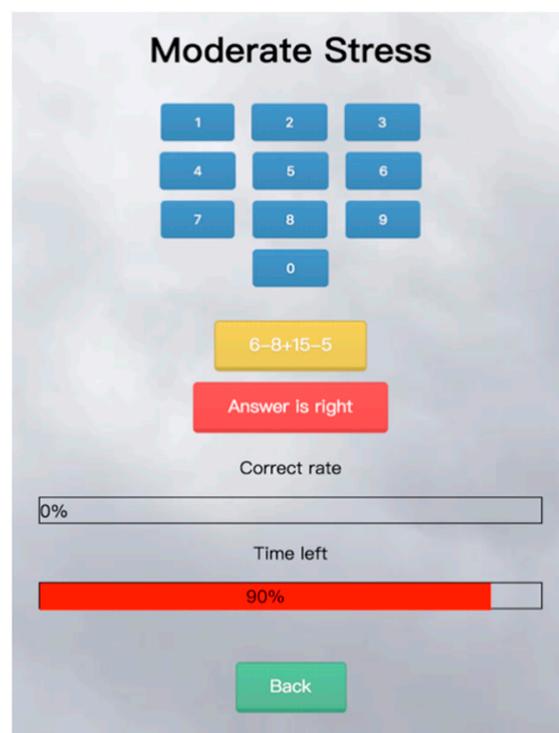
Common methods for stress inducement include color word experiments, ice water simulation, public speaking mathematical calculations, and watching horror videos. In view of the experimental feasibility and controllability of the induced degree, we chose the Montreal stress model with calculations as the main method [17]. The Montreal Imaging Stress Model was originally designed by psychologists to assess psychological stress, and it contains three processes—rest, moderate stress, and high stress inducing. The experimental procedures were developed by the group from the Institute of Psychology, Chinese Academy of Sciences. It contained three periods which last as long as 15 min, the first five minutes are a rest phase, and then a light stress phase, and heavy stress phase alternately arranged. The detailed process is shown in Figure 3. The VAS (visual analogue scale) report [18] of psychological stress was gathered from the subjects every 30 seconds as the ground truth for labeling the data.



**Figure 3.** The process of emotional stress induced experiment.

None of the computing tasks were scheduled during the rest period. At that time, we played melodious music and humorous moving pictures to the subjects to make them as relaxed as possible. This design could simulate the none-stress situations that the subjects encountered in daily life.

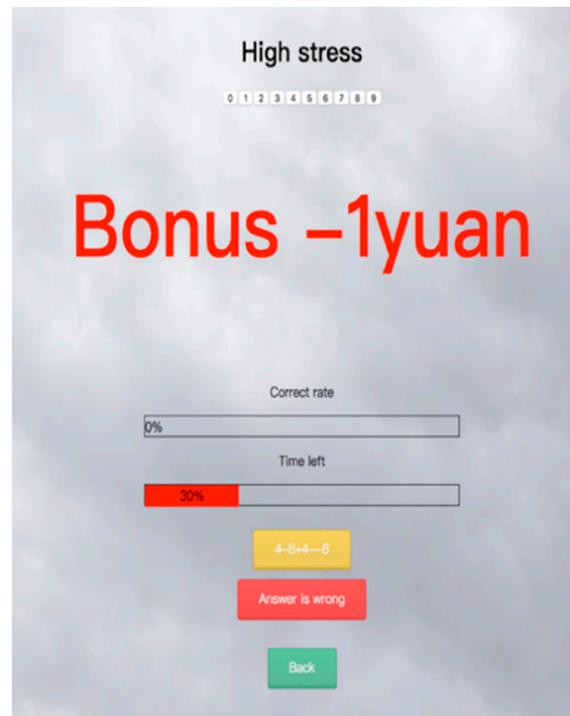
During the moderate stress phase, we improved the Montreal stress model by incorporating more evoking factors. Among these, we made the background color gray to give the participants a psychological hint of negativity. The subjects were asked to participate in simple double-digit addition and subtraction. The formula contained four variables, each variable ranged from a negative 20 to 20, the operators ( $-$ ,  $+$ ,  $-$ ) are fixed. To cause moderate stress in the participant, below the formula was a result prompt, a correct rate and a 5 s time bar that implies sufficient time. If the time ran out and the subject still has not answered, the system would judge the answer wrong and let the timing bar recount. This design simulated the moderate-stress situations encountered by the subjects when working or studying. Therefore, it is easy to finish the tasks with moderate difficulty and the time required is sufficient. The actual effect is shown as shown in Figure 4.



**Figure 4.** The period of moderate stress, simple operation, sufficient answer time, answer hint, and gray background to induce moderate stress.

During the stage of high stress stage, to induce the subject's psychological stress as much as possible, we added bounty missions, where the participants would get a reward of 1 yuan for each correct answer and, a 1 yuan punishment for each wrong answer. The system also displays the amount of money in large red font in the middle of the interface attracting enough attention to form a sense of oppression in the subjects. The answer buttons were also zoomed out and randomly arranged. When the answer has been figured out, the subject also has to find the correct button in the remaining time, and during that

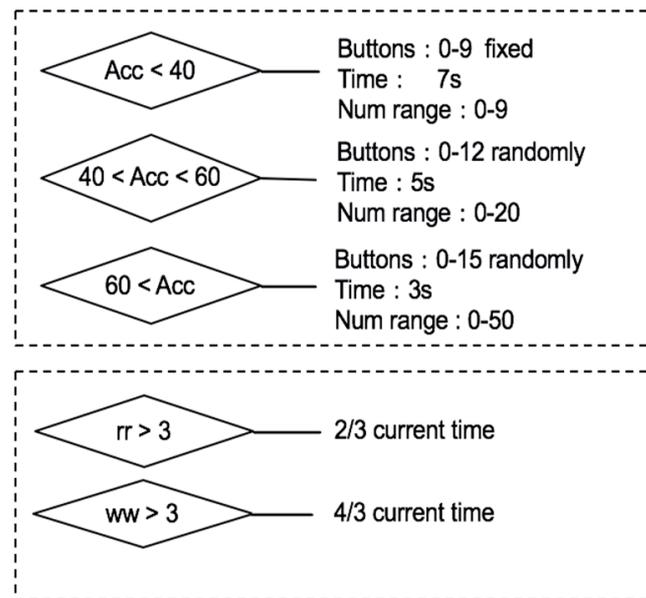
finding time, the stress would increase significantly. The actual display effect is shown in Figure 5.



**Figure 5.** The period of high stress, simple operation, bounty task, zoom out button.

Different subjects have different computing capabilities. In order to avoid problems of fixed difficulty lacking enough of a challenge for the subjects with strong ability, there was a lack of participation for those subjects with weak computing ability. The system integrated the adaptive function of problem difficulty, adaptive function of the button layout, and the adaptive function of the time left. When the answer accuracy was less than 40% which implied the subject's computing ability is weak, the system would generate the correct answer between 0 and 9, made the button's sequence fixed and set the answer time to 7 s. When the answer accuracy was between 40% and 60%, which implied the subject's ability is normal, the system would generate the correct answer between 0 and 12, made the button's sequence random, and set the answer time to 5 s. If the answer accuracy was larger than 60%, which implied the subject's ability is strong, the system would generate the correct answer between 0 and 15, made the button's sequence random and set the answer time to 3 s.

In addition, the program could judge the participant's state from another dimension. When 3 consecutive questions were answered correctly, the answer time would become  $2/3$  of the current time, conversely, when 3 consecutive questions were answered wrong, the answer time would become  $4/3$  of the current time. This design could simulate the high-stress situations encountered in the subjects' work or study. Namely, it was hard to reach the passing line that even if the subject worked as well as possible which was equivalent to 60% accuracy in this experiment. The adaptive algorithm is shown in Figure 6 below.



**Figure 6.** The adaptive function. RR is the number of consecutive correct answers, WW is the number of consecutive wrong answers.

A total of 34 people without cardiovascular and cerebrovascular diseases were recruited by poster from University of Chinese Academy of Sciences to participate in the experiment. Among them were 20 males and 14 females. The subjects' ages ranged from 20 to 35, the average being 23.4. It was confirmed that they had not participated in similar experiments before and had no history of smoking or drinking in the prior two days. We informed the subjects of the experimental process and gave a demonstration before they submitted their written informed consent. The electrocardiogram (ECG) was recorded using the sticky device with the sampling rate set at 250 Hz, then a 10-point moving average filter was used to filter out burrs, and every 10 seconds the data were uploaded to the server. The actual test scene and the collected ECG wave form are shown in Figure 7. The research ensured that the rights and interests of the subjects were fully protected, and the research content would not cause harm or risk to the subjects. It was reviewed and approved by the Institutional Review Board of Beijing Tiantan Hospital, Capital Medical University.



(A)



(B)

**Figure 7.** (A) is the actual measurement, (B) is the recorded electrocardiogram (ECG) signal as displayed on the website. The horizontal axis represents the time in seconds, the vertical axis represents the amplitude in millivolts.

### 2.2. Conventional Methods

To compare the conventional methods to the deep learning model proposed in this article, conventional methods were conducted, the characters of conventional algorithms were manually extracted from the time domain and frequency domain of the RR interval or HRV. HRV was defined as the change in the difference of successive RR intervals. To get the RR interval, the R-peaks' position was detected from the ECG signal using the Pan Tompkins algorithm [19]. To reduce the impact of individual differences, this article used the z-score algorithm as Formula (1) below to normalize the RR interval, so that its average would become 0 and the variance becomes 1.

$$z = ((x - \mu)) / \delta \tag{1}$$

The time domain features influenced by momentary ANS (autonomic nervous system) activities were extracted. Such statistical features included mean-RR (the mean value of RR interval), SD (standard deviation of HRV), MED (the median value of HRV), QD (quartile deviation of RR-interval), percent 20th (the 20th percentile's value of RR-interval), percent 80th (the 80th percentile's value of RR-interval) and average heart rate. As the research [20] has shown that the high-frequency energy of HRV is related to the activity parasympathetic nerve and the low-frequency energy is related to the activity of the sympathetic nerve, then HRV was computed from the RR intervals time-series and the features of HRV were extracted, including HF (high-frequency energy value of HRV signal), LF (low-frequency energy value of HRV signal) and LF\_HF (the ratio of HF power and LF power) as shown in Figure 8. Finally, the 10 features would be trained and tested using nine machine learning classifiers (XG-boost [21], logistic regression [22], RBF Radial Basis Function-SVM Support Vector Machine [23], random forest, decision tree, Linear SVM, K-nearest neighbors, Ada Boost, Naïve byes) as Table 1 shows. The performance would be further illustrated in the results section.

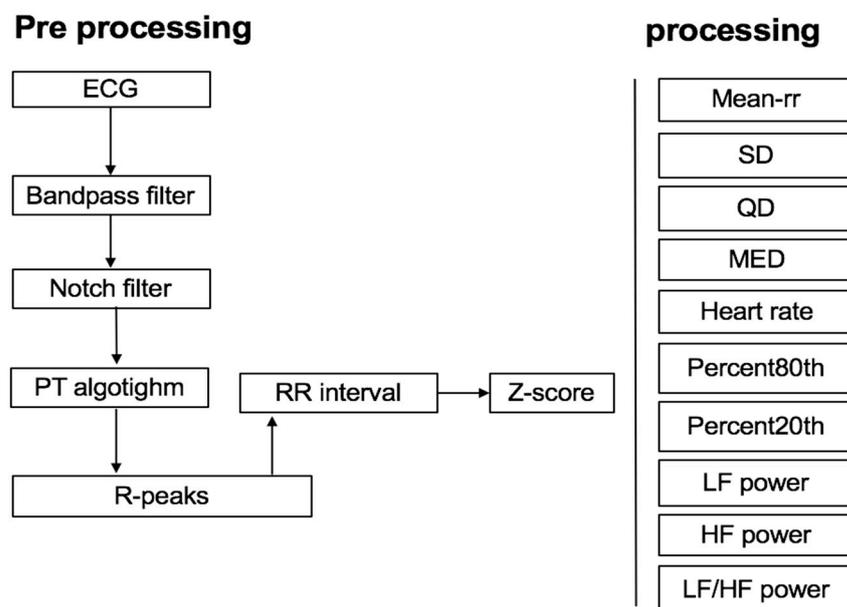


Figure 8. Signal preprocessing and extraction of statistical features from time domain and frequency domain.

Table 1. The features and the methods used in classification.

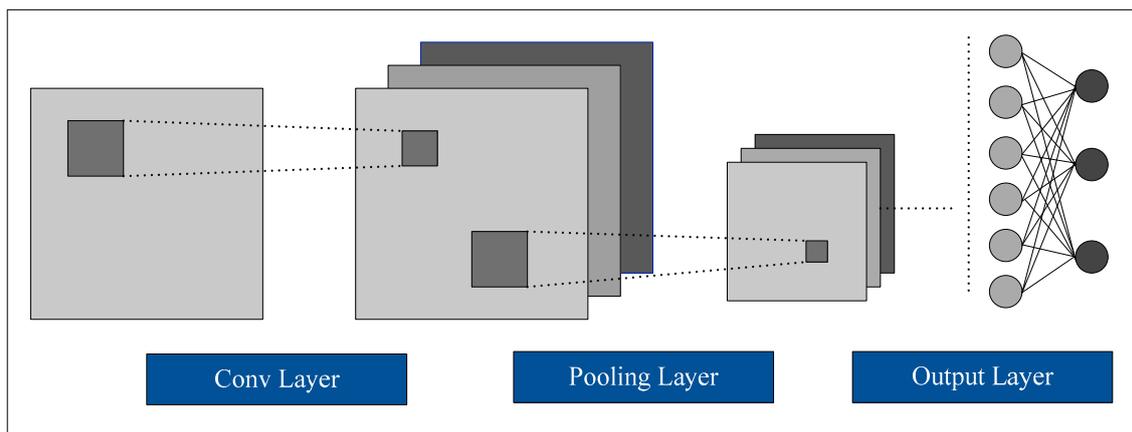
Features	Machine Learning Methods	
	Probabilistic Model	Non-Probabilistic Model
RR-interval's statistic features	Decision Tree, Logistic regression, Naïve Bayes byes, Random forest, XG-boost	K-Nearest Neighbors, Ada Boost, Linear-SVM, RBF-SVM

### 2.3. The Proposed Network

Most conventional machine learning methods used prior knowledge to manually extract features. Due to the limitations of prior knowledge, conventional feature extraction methods would inevitably ignore some non-linear relationship between ECG and stress. The purpose of this article was to avoid the limitations of prior knowledge by using deep learning technology.

With deep learning, we could extract features from the original ECG signal automatically and establish mapping from the original ECG for the psychological pressure to meet the real-time application (using less than 1 min ECG to speculate stress with high accuracy).

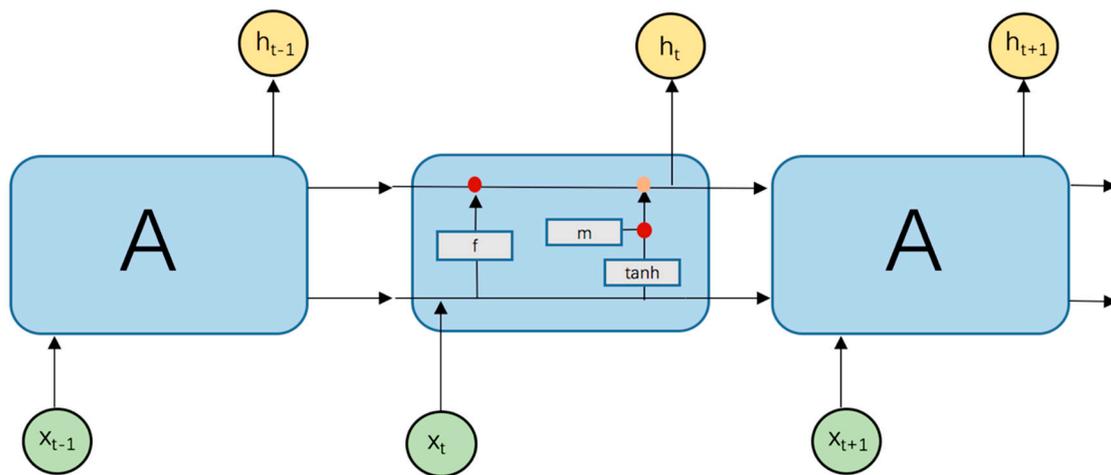
CNN is one of the most popular neural networks, and it has greatly promoted the development of image processing. Unlike RNN, its features are local connecting and weight sharing, using local connection and weight sharing CNN could automatically extract the structural features of the image in space, which makes it good at recognizing the data's displacement, scaling and rotation [24]. On the other hand, the model's complexity is greatly reduced, so it is easy to optimize compared with a fully connected network. A typical CNN net usually contains a convolution layer, a pooling layer, and a fully connected layer as shown in Figure 9. The convolutional layer performs simplified convolution operations between the input data that is intercepted by the window function and the convolution kernel. Therefore, the convolution layer can extract the features from the input data that relate to the convolution kernel.



**Figure 9.** The typical convolutional neural network (CNN) with one conv layer which has one channel and three filters, one pooling layer and one fully connected layer.

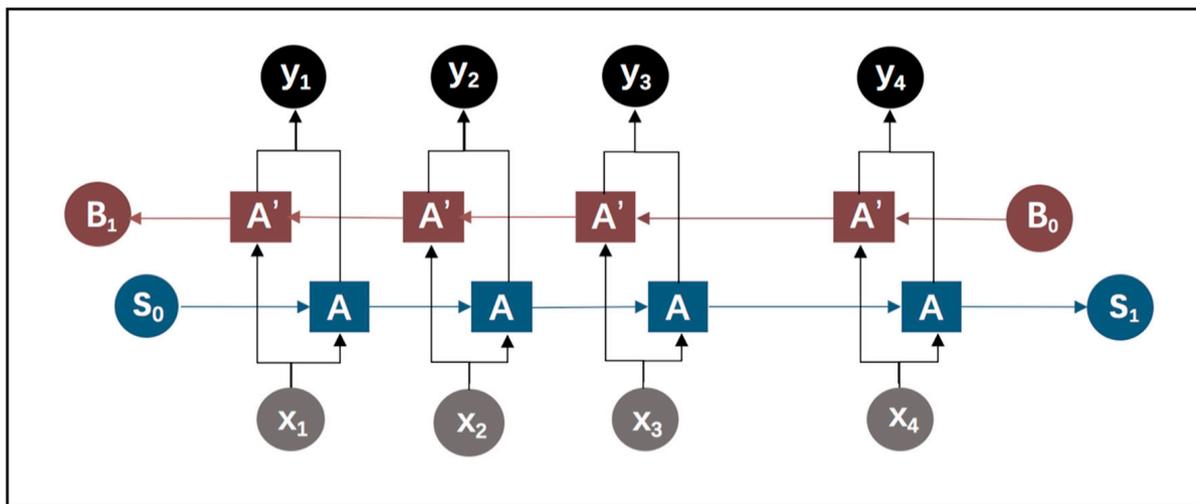
As different convolution kernels can extract different features, increasing the number of feature kernels within a certain range can increase the computing time. The pooling layer after convolution mainly plays the role of down sampling, which can highlight the effective ingredient of the features and reduce the amount of calculations. Taking into account the CNN's ability at extracting local features, we designed the process to use CNN instead of manually extracting features via conventional methods and also by increasing the number of convolution kernels to increase the number of features. The choice of CNN was based on the fact that it does not require feature engineering.

Traditional neural networks assumed that all the data entered before and after are independent. Therefore, it was not possible to take advantage of the progressive relationship within the time series signals, such as language. RNN [25] is one of the most popular networks, as it advances through time by a fixed step size and continuously generates historical status, in the process of advancement, the historical information of the previous moment is also taken as a part of the input, so it can use the back-and-forth connection of time-series signals. It showed excellent performance in the field of natural language processing. LSTM (as shown in Figure 10) is a kind of RNN which uses the forget gate and memory gate to address the disadvantages of short-term memory.



**Figure 10.** The structure of long short-term memory (LSTM) unit,  $f$  is the forget gate,  $m$  is the memory gate. ● represents multiply by element. ● represents add by element.

Throughout the entire time series of ECG, the state of the current moment was also affected by the future moment. The characteristics of the next moment could also be used to predict the current moment’s status. BiLSTM (shown in Figure 11) is a widely used variant of LSTM [26]. It consists of two LSTM units that can extract the features from both positive and negative sequences separately and then jointly determine the output. There are generally three ways to combine the outputs of two LSTM units, including addition, multiply and series connection (as shown in Figure 12). BiLSTM were used to further extract features from the whole time, considering that ECG belongs to time series signal that is suitable for BiLSTM. The three methods have been tried and the results are given in the result section.



**Figure 11.** The unit of bidirectional long short-term memory (BiLSTM), an extracts features from positive sequence, an extracts features from negative sequence.

The proposed network we designed is shown in Figure 13. First, one-dimensional convolution was used to extract the local temporal features. In CNN, LRN (local response normalization) was used to refine the classification boundary of the model and highlight the contrast of features. The mechanism of LRN [23] mimics the lateral inhibition phenomenon of neuroscience in that the activated neurons will suppress nearby neurons, which could

generate competition in the network where the effective features are strengthened and the ineffective features are diminished during training.

$$b_{x,y}^i = a_{x,y}^i / (k + \alpha \sum_{j=\max(0,(i-n)/2)}^{\min(N-1,(i+2)/2)} (a_{x,y}^j)^2) \tag{2}$$

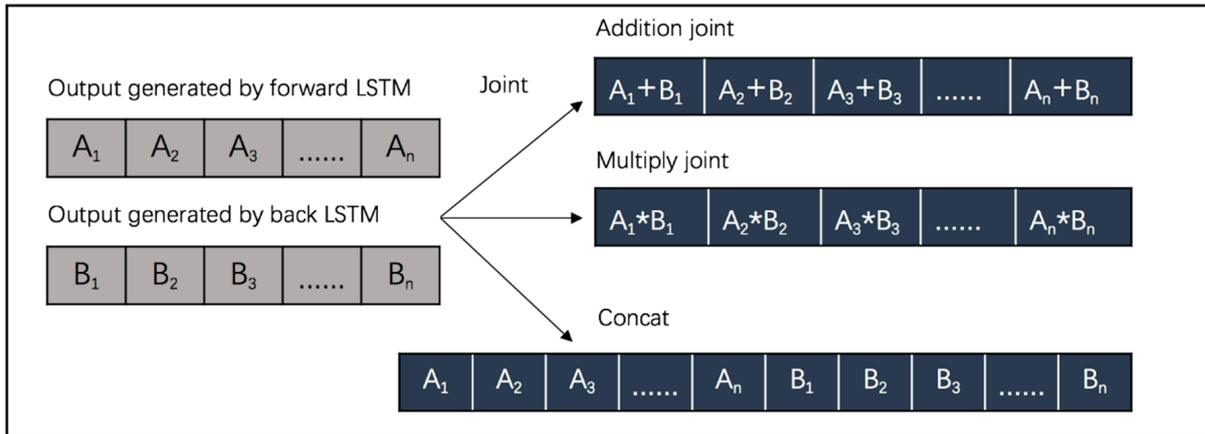


Figure 12. The three kinds of output of BiLSTM generated by three kinds of joint from LSTM’s output.

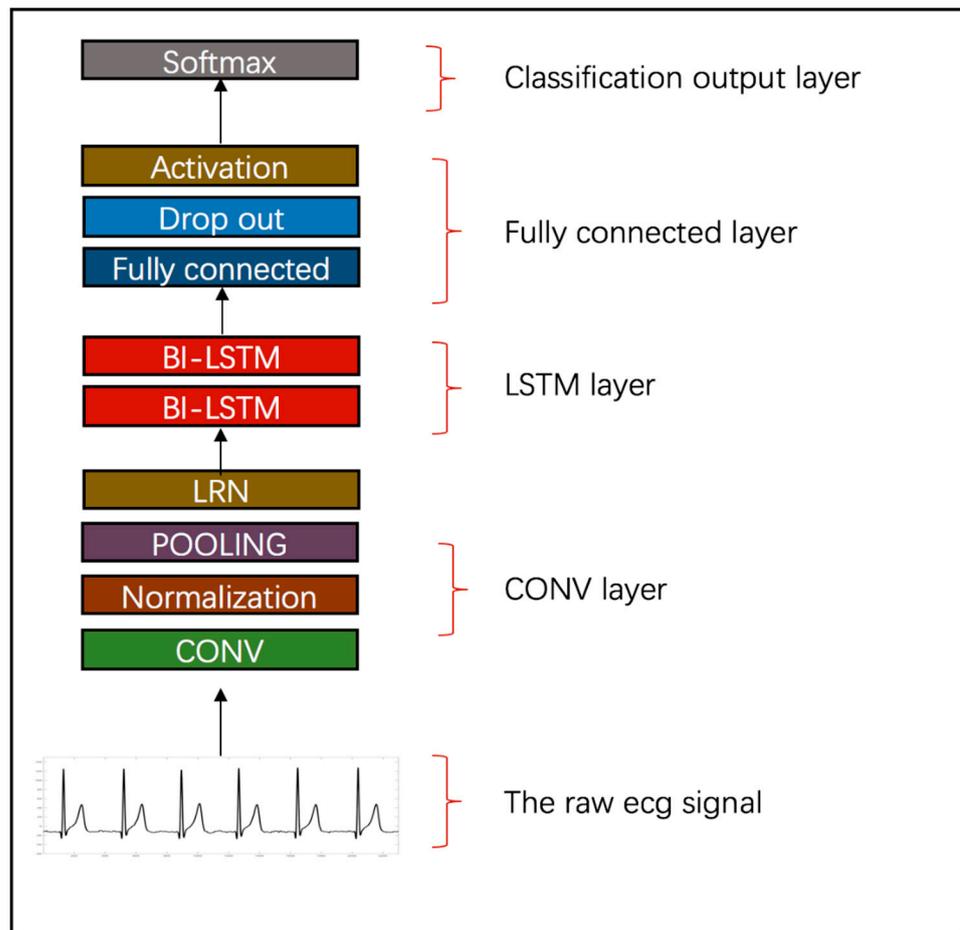


Figure 13. The proposed network using BiLSTM and CNN

In the activation layer, RELU (rectified linear unit) was used for non-linear transformation of the model. As the generalization ability and the computational complexity were determined by the parameters of BiLSTM and CNN, this article through repeated attempts to determine the optimal parameters. After the BiLSTM layers, a fully connected layer was used to generate maps between the extracted features and the three levels of psychological stress. At the end, the softmax layer calculated the probability of the individual to whom the current sample belongs. For training, cross information entropy was selected as the loss function. To maximize the generalization ability of the model, the L1 regular term [27] of the fully connected layer's weights was also added to the total loss, and that could increase the scarcity of the model.

Before the training, the initial learning rate was set to 0.6, the batch size was 30 and the epochs, 3000. This setting implied that the model randomly selects 30 samples for training each time, 3000 times in total. During the training, the network updated parameter values through adaptive stochastic gradient descent algorithm. With the Adadelta algorithm, the learning rate could be adaptively adjusted for better training results. The dropout layer method was used to randomly select a part of the nodes to participate in training each time, so that the model could get more diverse training. The proposed methods would be described in detail in the 'Results' section.

We evaluated our deep neural network on 3098 10 s ECG signals divided randomly into training set (80% of the data set) and testing set (20%). Each deep learning model was evaluated by five-fold cross-validation and the results were averaged. The experiments were performed on a computer with an Intel Core i7 processor, 16 GB RAM and Tensorflow 1.14.0

### 3. Results

To achieve real-time stress level detection, we designed a deep neural network and more conventional methods which rely on hand-crafted features and compared their performance.

To conduct conventional methods, we preprocessed the ECG signal and extracted 10 features from the time and frequency domains as shown in Figure 8. Then 9 machine learning methods were used to detect the stress level. The results of these models are evaluated based on different window sizes of 1 min and 10 s, respectively. The result based on 1 min is shown in Table 2 where these nine algorithms finally obtained an average accuracy of 0.647, an average recall of 0.557, and an average specificity of 0.810. When using the window of 10 s, the performance of these models decreased. As shown in Table 3, the average accuracy rate dropped to 0.563, the average recall rate dropped to 0.457 and the average specificity dropped to 0.731. The reason is that extracting features of frequency for a shorter window is less effective.

**Table 2.** The accuracy and recall of conventional method on 1 min ECG.

Algorithms	Accuracy	Recall	Specificity
XGboost	0.620	0.517	0.769
Logistic	0.667	0.564	0.792
LinearSVM	0.676	0.604	0.784
Randomforest	0.649	0.559	0.837
Decision Tree	0.636	0.523	0.816
Bayes	0.666	0.681	0.829
Gauss SVM	0.676	0.459	0.799
KNN	0.636	0.581	0.834
Adaboost	0.600	0.527	0.830
Average	0.647	0.557	0.810

**Table 3.** The accuracy and recall of conventional method on 10 s ECG.

Algorithms	Accuracy	Recall	Specificity
XGboost	0.637	0.492	0.765
Logistic	0.614	0.462	0.760
LinearSVM	0.624	0.468	0.748
Randomforest	0.502	0.424	0.708
Decision Tree	0.482	0.407	0.694
Bayes	0.459	0.497	0.741
Gauss SVM	0.632	0.492	0.747
KNN	0.575	0.483	0.755
Adaboost	0.547	0.393	0.667
Average	0.563	0.457	0.731

Since the goal of this article is to build a model for real-time application, the performances of the proposed framework and the conventional methods for the 10 s window will be compared and analyzed. Among the conventional models, XGboost obtained the highest accuracy of 0.637, the highest recall of 0.492, and the highest specificity of 0.765. The confusion matrices of XGboost is shown in Figure 14. It could be seen from the confusion matrix that XGboost achieved a recall for moderate stress at 0.916, but produced a poor recognition rate on low stress. It also misclassified most high stress states into moderate stress. It was especially difficult to correctly detect both stress and high stress for the conventional models in this particular experiment. The ROC curve of XGBoost (as Figure 15 shows) also indicated that its ability of recognizing high pressure is lower than that of the medium pressure and the low pressure.

<b>Low</b>	<b>98</b>	<b>75</b>	<b>0</b>
<b>Medium</b>	<b>23</b>	<b>295</b>	<b>4</b>
<b>High</b>	<b>7</b>	<b>112</b>	<b>0</b>
	<b>Low</b>	<b>Medium</b>	<b>High</b>

**Figure 14.** The confusion matrix of XGboost.

To determine the optimal parameters of the proposed network, we conducted various experiments. Initially, we preset the number of  $n_{\text{inputs}}$  to 5, the stride of the pooling layer to 8 and use stitching to deal with the output of BiLSTM. Because the ECG signal is quasi-periodic, we set the length of convolutional filter to 200 which corresponds to the average RR interval (0.8 s) at the sampling rate of 250 Hz.

The performance of our initial network was as shown in Figure 16 left, and the accuracy with 4 conv filters was only 0.819. We then tried to increase the number of features by augmenting the number of convolution kernels. In Figure 16, it can be seen that as the number of filters increases, the accuracy rate continues to increase, and when the number of filters reaches 32, the network achieved the maximum accuracy of 0.863. Then, the number of convolutions further increased, the accuracy decreased instead, which implied that when the number of filters exceeded 32, no more valid features could be extracted, and the exceeded filters even caused over fitting, which reduced the accuracy of the model in the test set. The 32 was used as the optimal parameter for the conv filter.

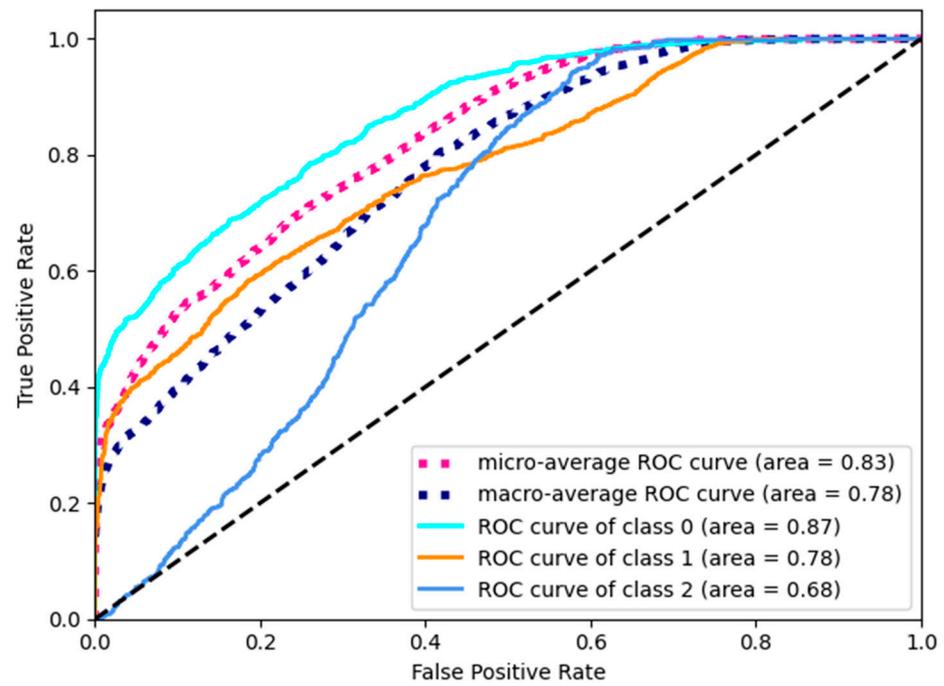


Figure 15. The Receiver Operating Characteristic (ROC) Curve of XGBoost

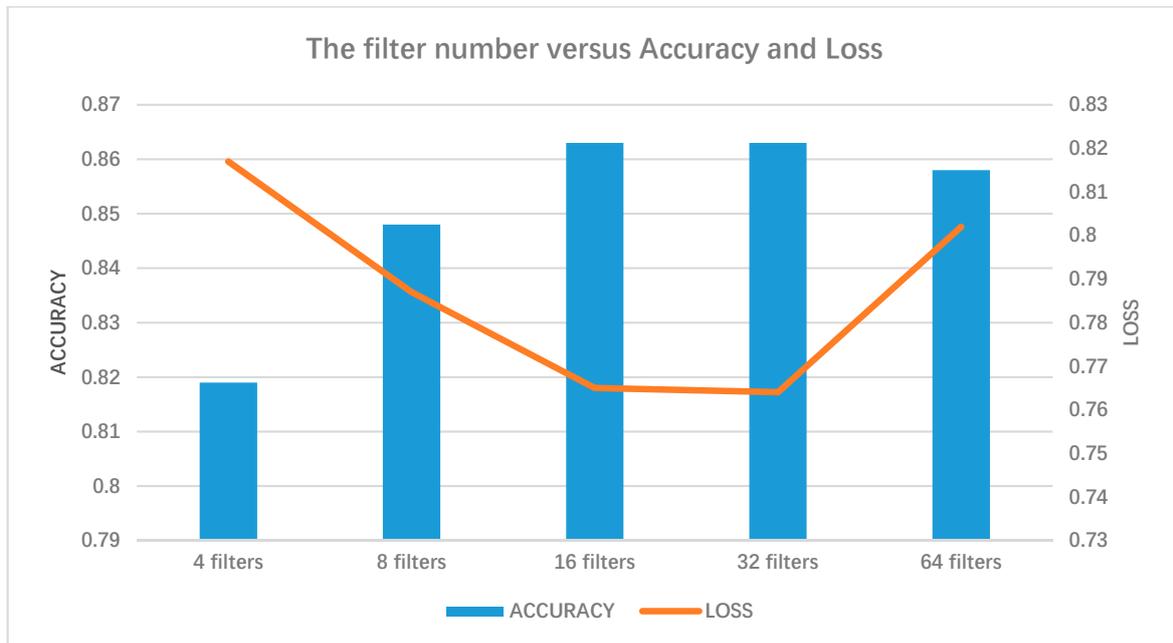


Figure 16. The relation of filter number and learning accuracy and learning loss.

Based on the fixed optimal conv filter number, other experiments were conducted, and a similar phenomenon could also be observed in the experiment of BiLSTM size. In the experiment of BiLSTM size, as shown in Figure 17, the accuracy is not increased at the size of 64 compared with the size of 32. However, the time for training did greatly increase. Considering the accuracy and training time, 32 BiLSTM units were used for the proposed network.

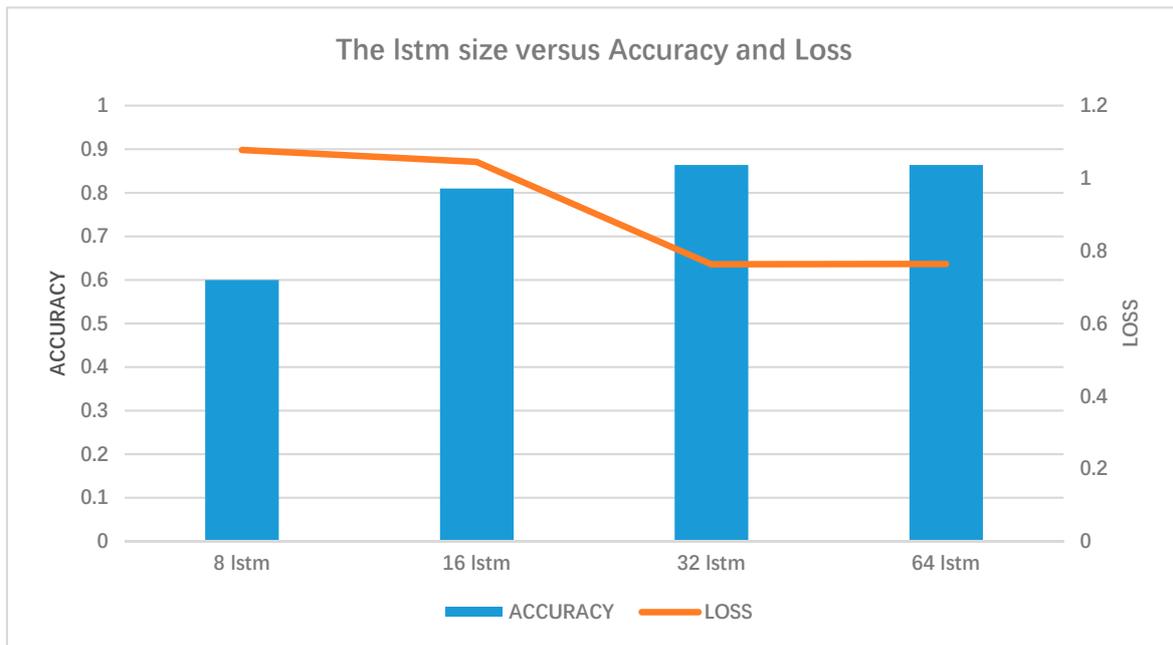


Figure 17. The relation of LSTM size and learning accuracy and learning loss.

The size of pooling window and  $n$ -inputs of BiLSTM was determined based on the maximum accuracy. As shown in Figure 18, the proposed network achieved the highest accuracy when the pool length was set to 8. The set implies that for the features generated by the convolutional layer, the pooling layer will downsample using a window of size 8. As shown in Figure 19, the proposed network achieved the highest accuracy when the  $n_{input}$  was set to 50. The set implies that BiLSTM could further extract more abstractive features from each 50 initial features that CNN was able to extract.

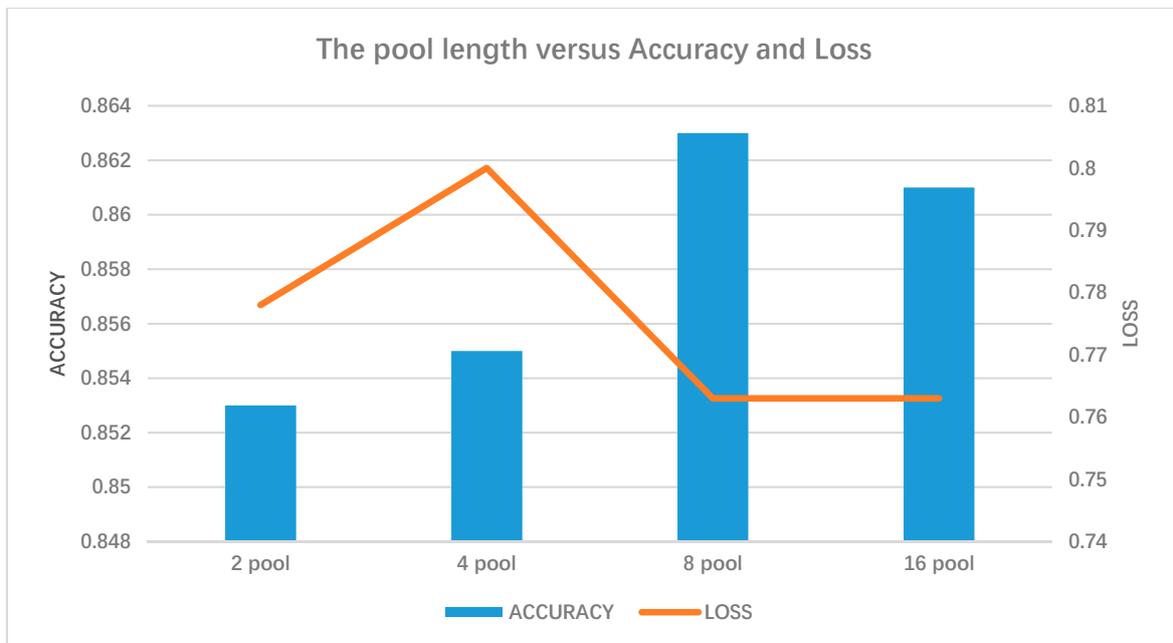


Figure 18. The relation of pool length and learning accuracy and learning loss.

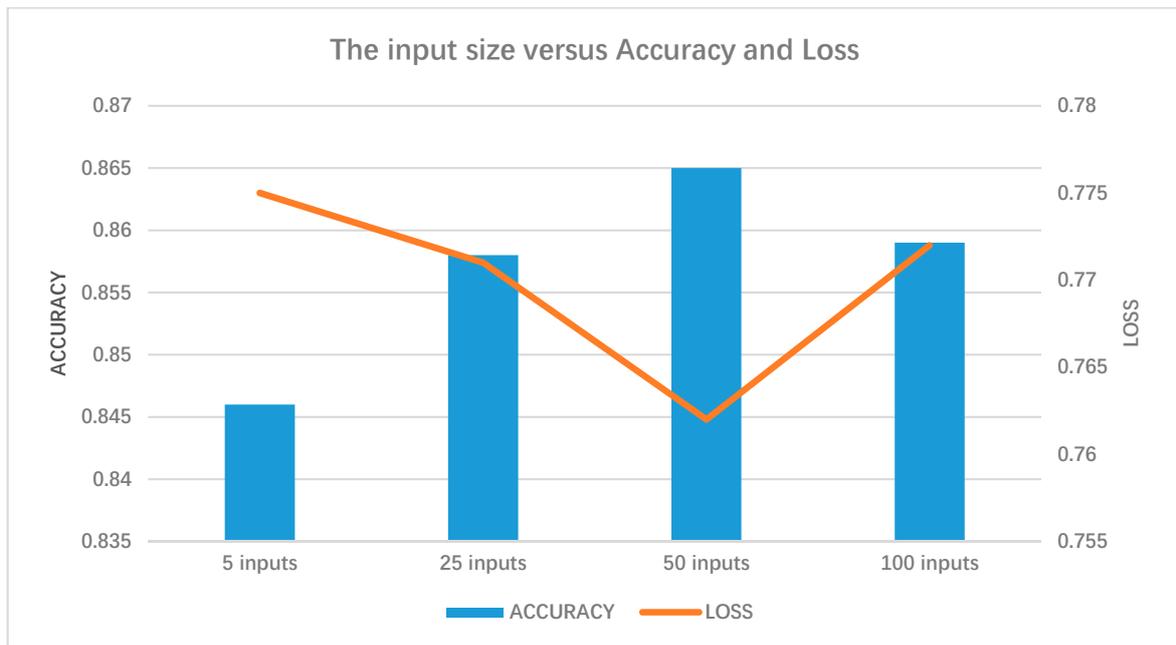


Figure 19. The relation of input size and learning accuracy and learning loss.

To reduce training time without affecting the performance of the proposed network, we also attempted to reduce redundant features and save training time by adjusting the size of the stride in the pooling layer. At the beginning, we set the size of the stride to 2. This implies that there will be 0.5 times features that the Conv layer extracted when inputting to the BiLSTM layer. As seen in Figure 20, until the stride size increased to 8, the accuracy of the model began to decline, which implies that the size of 8 could filter out redundant features to the maximum. Finally, for the outputs of BiLSTM, we tried three stitching methods based on the previously obtained optimization parameters. As shown in Figure 21, the method of addition which was adopted by the proposed network obtained the highest accuracy.

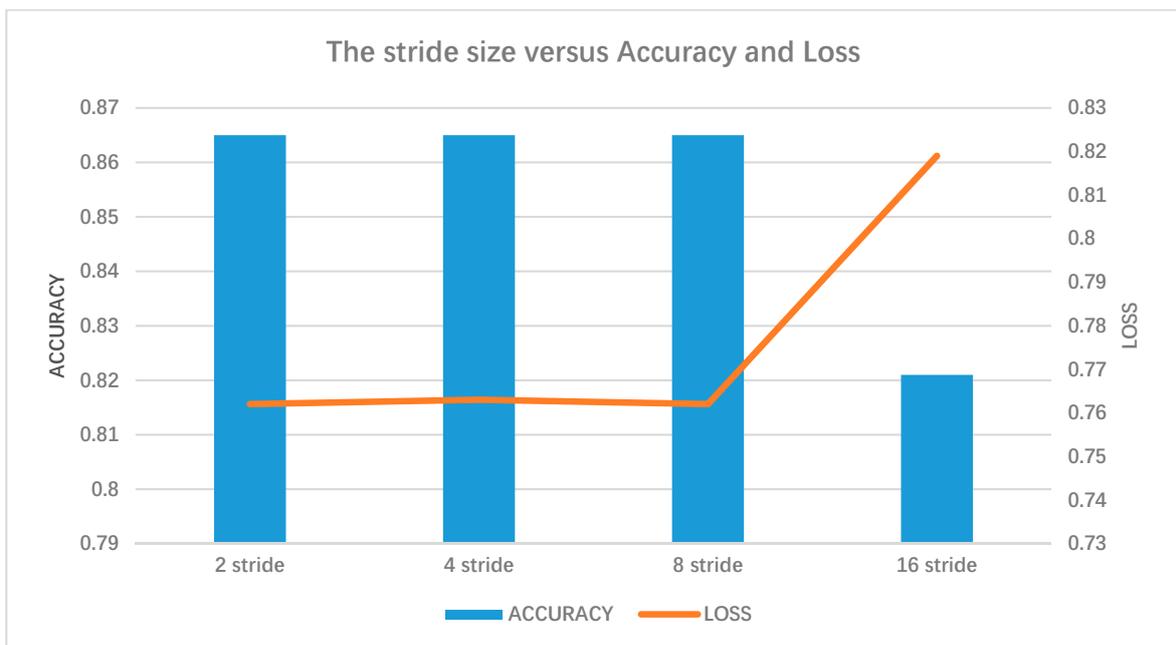
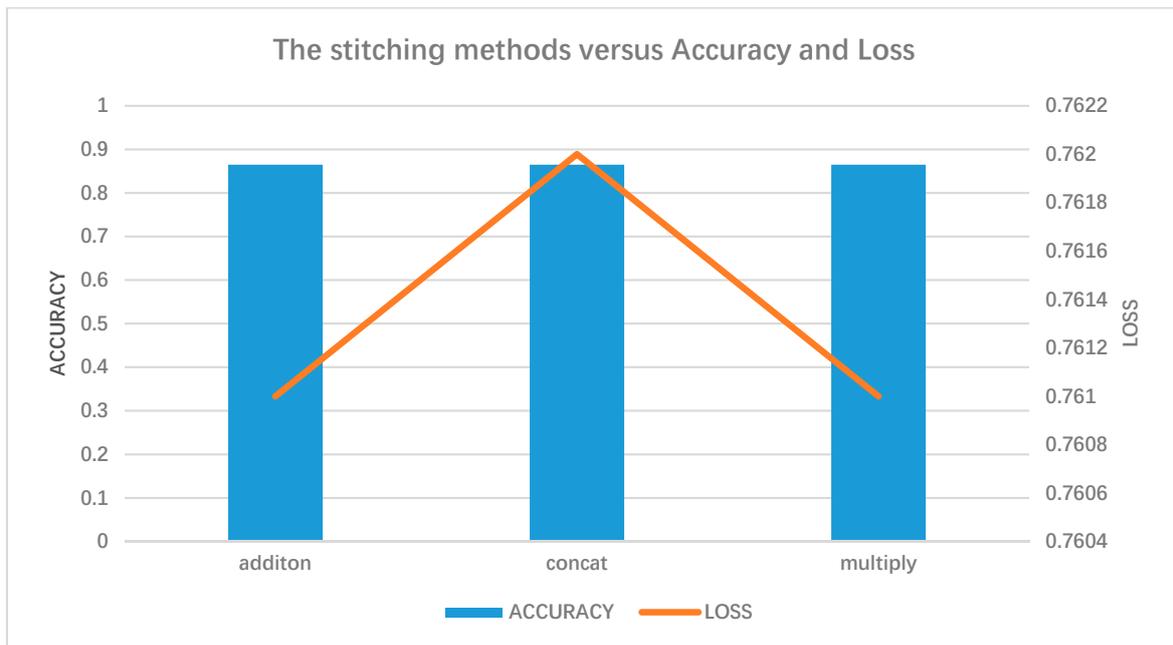
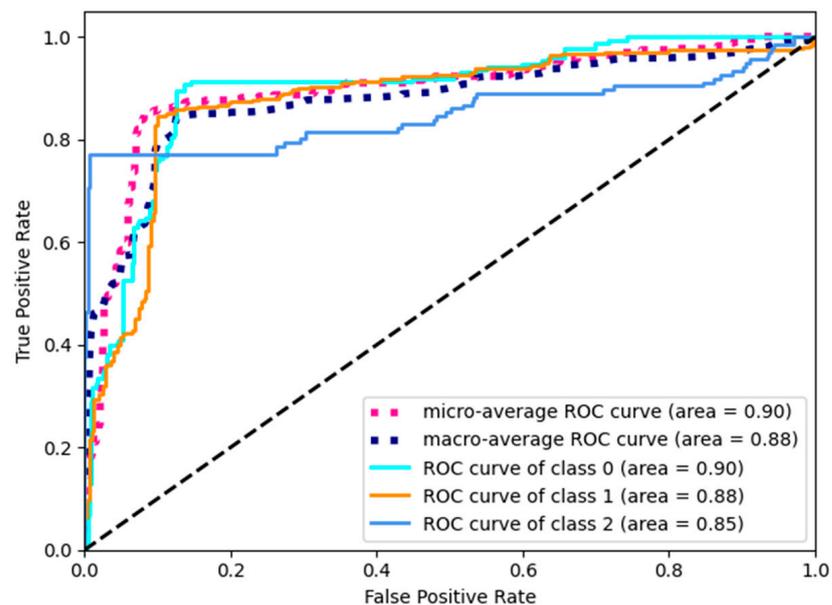


Figure 20. The relation of stride size and learning accuracy and learning loss.



**Figure 21.** The relation of stitching methods and learning accuracy and learning loss.

The proposed network obtained the accuracy of 0.865, and a specificity of 0.928. Compared to the conventional methods (XGboost), our proposed network improves accuracy by 0.228. The ROC curves of the proposed network and the XGBoost are as shown in Figures 15 and 22, the area under the curve (AUC) in moderate stress was 0.88 improved by 0.1 compared with XGboost’s, the AUC in high stress was 0.85 improved by 0.17. The confusion matrix of the proposed network was as shown in Figure 23, the recall of low stress was 0.913, the recall of moderate stress was 0.894, and the recall of high stress was 0.798. It could thus be concluded that the proposed network produces a significant increase in the classification of the stress on three levels compared to the conventional methods.



**Figure 22.** The curve of the proposed network.

Low	159	15	0
Medium	28	280	5
High	12	12	95
	Low	Medium	High

Figure 23. The confusion matrix of the proposed network.

#### 4. Discussion and Conclusions

This paper proposes a model for psychological stress detection by using deep learning technology and obtaining the optimal parameters of that model by various experiments. For a comparison, we also implemented the conventional methods, and the results showed our proposed network obtained significant improvement in both medium stress and high stress detection compared to the conventional methods using 10 s ECG. This finding implies our model is more satisfactory for real-time application. That conclusion is mainly due to the use of CNN and BiLSTM. CNN's excellent ability when dealing with local features and BiLSTM's excellent ability when dealing with long-term sequence signals enable our network extracting features automatically to overcome the limitations of conventional methods by using prior knowledge for extracting key features.

Individual differences, especially personality differences are important factors affecting emotional stress. In the process of building the proposed network, we did not take into account the personality differences of the experimenter, exploring how to improve the performance of the proposed network using personality differences will become our future work.

**Author Contributions:** Conceptualization, P.Z. and F.L.; methodology, P.Z.; software, R.Z. (Ruishi Zhou); validation, Z.Z., Z.F., X.C. and L.D.; formal analysis, F.L., Z.Z. and Z.F.; investigation, R.Z. (Rongjian Zhao) and L.D.; writing—original draft preparation, P.Z.; writing—review and editing, Z.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Key Research and Development Project 2018YFC2001101, 2018YFC2001802, 2020YFC2003703, 2020YFC1512304, National Natural Science Foundation of China (Grant 62071451), and CAMS Innovation Fund for Medical Sciences (2019-I2M-5-019).

**Institutional Review Board Statement:** This research was reviewed and approved by the Institutional Review Board of Beijing Tiantan Hospital, Capital Medical University (no. KYSQ 2019-013-01).

**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. As the data were generated during this study, we did not find an appropriate platform to share the data.

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

#### References

1. Beanland, V.; Fitzharris, M.; Young, K.L. Driver inattention and driver distraction in serious casualty crashes: Data from the Australian National Crash In-depth Study. *Accid. Anal. Prev.* **2013**, *54*, 99–107. [[CrossRef](#)] [[PubMed](#)]
2. Sauter, S.L.; Murphy, L.R.; Hurrell, J.J. Prevention of work-related psychological disorders: A national strategy proposed by the National Institute for Occupational Safety and Health (NIOSH). *Am. Psychol.* **1990**, *45*, 1146. [[CrossRef](#)] [[PubMed](#)]
3. Hillebrandt, J. *Work-Related Stress and Organizational Level Interventions Addressing the Problem at Source*; GRIN: München, Germany, 2008.

4. Ebner-Priemer, U.W.; Trull, T.J. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychol. Assess.* **2009**, *21*, 463–475. [CrossRef] [PubMed]
5. Spruijt-Metz, D.; Nilsen, W. Dynamic Models of Behavior for Just-in-Time Adaptive Interventions. *IEEE Pervasive Comput.* **2014**, *13*, 13–17. [CrossRef]
6. Tsigos, C.; Chrousos, G.P. Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. *J. Psychosomat. Res.* **2002**, *53*, 865–871. [CrossRef]
7. Electrophysiology. Task Force of the European Society of Cardiology the North American Society of Pacing. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation* **1996**, *93*, 1043–1065. [CrossRef]
8. Hovsepian, K.; Al'Absi, M.; Ertin, E.; Kamarck, T.; Nakajima, M.; Kumar, S. CStress: Towards a gold standard for continuous stress assessment in the mobile environment. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15). Association for Computing Machinery, New York, NY, USA, 10 September 2015; pp. 493–504. [CrossRef]
9. Winata, G.I.; Kampman, O.P.; Fung, P. Attention-based lstm for psychological stress detection from spoken language using distant supervision. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; IEEE: New York, NY, USA, 2018; pp. 6204–6208.
10. Jin, L.; Xue, Y.; Li, Q. Integrating human mobility and social media for adolescent psychological stress detection. In Proceedings of the International Conference on Database Systems for Advanced Applications; Springer: Cham, Switzerland, 2016; pp. 367–382.
11. Lin, H.; Jia, J.; Guo, Q.; Xue, Y.; Huang, J.; Cai, L.; Feng, L. Psychological stress detection from cross-media microblog data using deep sparse neural network. In Proceedings of the 2014 IEEE International Conference on Multimedia and Expo (ICME), Chengdu, China, 14–18 July 2014; IEEE: New York, NY, USA, 2014; pp. 1–6.
12. Hwang, B.; You, J.; Vaessen, T.; Myin-Germeys, I.; Park, C.; Zhang, B.-T. Deep ECGNet: An Optimal Deep Learning Framework for Monitoring Mental Stress Using Ultra Short-Term ECG Signals. *Telemed. e-Health* **2018**, *24*, 753–772. [CrossRef] [PubMed]
13. Song, S.H.; Kim, D.K. Development of a stress classification model using deep belief networks for stress monitoring. *Healthc. Inform. Res.* **2017**, *23*, 285. [CrossRef] [PubMed]
14. Zhu, W.; Li, M.; Chen, H. Using MongoDB to implement textbook management system instead of MySQL. In Proceedings of the IEEE 3rd International Conference on Communication Software and Networks, Xi'an, China, 27–29 May 2011; IEEE: New York, NY, USA, 2011; pp. 303–305.
15. Chi, X.; Liu, B.; Niu, Q.; Wu, Q. Web load balance and cache optimization design based nginx under high-concurrency environment. In Proceedings of the Third International Conference on Digital Manufacturing & Automation, Guilin, China, 31 July–2 August 2012; IEEE: New York, NY, USA, 2012; pp. 1029–1032.
16. Pan, L.; Lee, S.; Zhang, J.; Tang, B.; Zhai, C.; Jiang, J.H.; Wang, W.; Bao, Q.; Qi, M.; Kubar, T.L.; et al. Software architecture and design of the web services facilitating climate model diagnostic analysis. In Proceedings of the AGU Fall Meeting Abstracts, San Francisco, CA, USA, 14 December 2015. IN31A-1753.
17. Dedovic, K.; Renwick, R.; Mahani, N.K.; Engert, V.; Lupien, S.J.; Lupien, J.C. The Montreal Imaging Stress Task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *J. Psychiatr. Neurosci.* **2005**, *30*, 319.
18. Cella, D.F.; Perry, S.W. Reliability and concurrent validity of three visual-analogue mood scales. *Psychol. Rep.* **1986**, *59*, 827–833. [CrossRef] [PubMed]
19. Pan, J.; Tompkins, W.J. A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **1985**, *3*, 230–236. [CrossRef] [PubMed]
20. Piccirillo, G.; Vetta, F.; Fimognari, F.L.; Ronzoni, S.; Lama, J.; Cacciafesta, M.; Marigliano, V. Power spectral analysis of heart rate variability in obese subjects: Evidence of decreased cardiac sympathetic responsiveness. *Int. J. Obes. Relat. Metab. Disord.* **1996**, *20*, 825–829. [PubMed]
21. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H. Xgboost: Extreme Gradient Boosting, R Package Version. 0.4-2; 2015. Available online: <https://mran.microsoft.com/web/packages/xgboost/vignettes/xgboost.pdf> (accessed on 23 April 2021).
22. Lemeshow, S. A review of goodness of fit statistics for use in the development of logistic regression models. *Am. J. Epidemiol.* **1982**, *115*, 92–106. [CrossRef] [PubMed]
23. Lin, S.; Liu, Z. Parameter selection in SVM with RBF kernel function. *J. Zhejiang Univ.-Sci. B* **2007**, *85*, 1–4.
24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1 (NIPS'12), Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.
25. Williams, G.; Baxter, R.; He, H.; Hawkins, S.; Gu, L. A comparative study of RNN for outlier detection in data mining. In Proceedings of the IEEE International Conference on Data Mining, Maebashi City, Japan, 9–12 December 2002; pp. 709–712.
26. Nguyen, N.K.; Le, A.C.; Pham, H.T. Deep bi-directional long short-term memory neural networks for sentiment analysis of social data. In Proceedings of the International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making; Springer: Cham, Switzerland, 2016; pp. 255–268.
27. Huan, J.; Ma, S.; Zhang, C.H. Adaptive LASSO for sparse high-dimensional regression. *Stat. Sin.* **2006**, *18*, 1603–1618.