Mental Pressure Recognition Method Based on CNN Model and EEG Signal under Cross Session

Song Zhou 1, Tianhan Gao 1,∗ and Jun Xu 2

1 Software College, Northeastern University, Shenyang 110169, China
2 Science and Technology on Special System Simulation Laboratory, Beijing Simulation Center, Beijing 100854, China
∗ Correspondence: gaoth@mail.neu.edu.cn

Abstract: There is an important application value in assessing an operator’s mental pressure (MP) level in human–computer cooperative tasks through continuous asymmetric electroencephalogram (EEG) signals, which can help predict hidden risks. Due to the different distributions of EEG features in different periods, it is particularly challenging to accurately identify brain states by training and testing asymmetric EEG signals with static pattern classifiers. Due to the limitations of non-stationary neurophysiological data capture technology, cross-session MP recognition schemes can only be used as an auxiliary means in practical applications. Deep learning methods can achieve stable feature extraction at a high level. Based on this advantage, this paper proposes a triplet loss (TL)-based CNN model that can automatically update the weights of shallow hidden neurons in cross-session MP classification tasks. Firstly, the generalization ability of the CNN model under both intra-session and cross-session conditions is evaluated. Moreover, the proposed model is compared with the existing MP classifier under different feature selection and noise destruction modes. According to the results, our TL-based CNN model has high performance in processing cross-session EEG features.

Keywords: mental pressure recognition; electroencephalogram (EEG); triplet loss; deep learning; brain–computer interface

1. Introduction

Generally, MP evaluation methods are designed considering three levels: subjective score, secondary task performance, and neurophysiological signals [1]. Subjective scoring requires users to assess a task in its corresponding stage with a questionnaire. The general Subjective Workload Assessment Technology (SWAT) and Task Load Index (TLX) have been proven to be efficient in evaluating the MP status in many practical applications [2]. A cross entropy (CE) loss function is often used in multiclass classification problems to measure the relationship between the output of multiple classifiers and their predicted value. CE loss can measure the difference between two different probability distributions in the same random variable, which is expressed as the difference between the real probability distribution and the predicted probability distribution in machine learning. In human–machine collaborative work environments with many safety hazards, such as high-altitude hazardous operations and special flights, the psychological state of operators is an important component oflink in ensuring safety during work. Different from machine operation, the limited working memory of the operator does not always guarantee a safe working state. In this case, negligence during work processes due to a high-MP state is the key factor leading to catastrophic accidents. In order to establish a connection between internal psychological stress and external actions in the field of safety production, the MP level has been proven to be related to task performance, vigilance, situation awareness and the capability to handle emergency events induced by automation failures. However, due to the fact that subjective scoring requires the collection of subjective ratings manually, the sampling rate...
of the MP indicators obtained is limited; this is particularly true in specific environments, e.g., during surgery, it is difficult to evaluate the MP of doctors. In order to address the above issues, neurophysiological signal assessment methods are widely used in brain state recognition applications due to their ability to objectively infer inherent cognitive states and highly repetitive independent tasks. Among them, more mature assessment methods include electroencephalograms (EEG), event related potentials (ERP), electrocardiograms (ECG), electroencephalograms (EOG), and functional near-infrared spectroscopy (fNIR) [3]. Of this variety of neurophysiological signal markers, asymmetric electroencephalogram (EEG) is highly utilized as it is easy to obtain via portable devices and possesses high temporal resolution characteristics. Studies have shown that changes in the level of MP are determined on the basis of the EEG power spectral density (PSD) gathered from the multi-cortical region and frequency band. When task demands increase under high-MP states, the EEG signals from the parietal and occipital lobes display changes, such that α (8–13 Hz) power is reduced, while θ (4–7 Hz) power is increased. In addition, the increase in high-frequency outputs from the occipital cortex regarding β and γ power have corresponding influences on different task commands in the operation scenario [4]. The application of multi-band channels for EEG measurements is also attractive for the derivation of comprehensive information related to MP changes [5–7]. There exists an urgent need for methods that can automatically analyze a large amount of neurophysiological data for complex EEG acquisition equipment with high temporal and spatial resolutions. Machine learning pattern recognition methods can effectively solve this problem.

In this paper, we propose a TL-based CNN model to address the cross-session MP classification task to alleviate the above issues. Our proposed model solves the problem of high variability in the EEG signal across sessions, uses the TL algorithm to ensure the time-invariant characteristics of the EEG signal, and achieves the ideal effect of identifying the human’s MP state. We thus design a method that utilizes TL functions to achieve the compression of EEG data such that the distance between subjects is minimized. TL has been applied to different EEG classifications, such as motion image classification and emotion recognition in BCI [8–10]. However, as far as the present situation is concerned, the superiority of TL has not been evaluated in the biometric system of EEG. In this paper, compared with traditional CE loss, BCI spatial filtering and opponent training in EEG biometrics, we achieved advanced results by combining the TL function and a CNN model [10,11]. To identify useful EEG features for EEG feature extraction, we propose a deep neural network-based approach for MP extraction. The benefits of EEGs include their noninvasive nature, as well as the fact that their PSD characteristics can accurately reflect changes in the potential of cortical neurons. However, deeper neuronal activity was associated with altered MP as the signal traveled from its origin to the scalp. Our effective approach thus aims to construct hierarchical deep structures to reconstruct the root of hidden features with the aim of completely exploiting the EEG properties of deep neurons.

The rest of the paper is organized as follows. Section 2 describes the experimental paradigm for the MP task simulation and EEG acquisition. Section 3 presents the proposed framework along with the architecture of the CNN models and the triplet loss function. In addition, the setup for the experimentation and the analysis of the control parameters of the cross-session MP classifier are provided in Section 4. In addition, Section 5 provides the conclusions of the current work.

2. Preliminaries

In general, multi-layer perceptrons (MLPs) have been extensively used in the recognition of MP in machine operators [12]. Linear discriminant analysis (LDA) has been adopted for binary MP evaluation [13]. Support vector machine (SVM) regression and recursive elimination have been conducted to discover stable EEG frequency characteristics in different n-back task settings [14]. A least squares support vector machine (LS-SVM) has been applied to the case where the dimension of the neurophysiological features was superior when compared with the number of training data points [15]. Research shows that when
the training set and test set of a machine learning model are collected from the same subject and the same task, the classification accuracy of MP may be higher than 90%. The use of EEG power characteristics and machine-learning-based methods for MP recognition is expected to show the potential patterns corresponding to the specific MP state hidden in the EEG data distribution [16]. For MP recognition research, the latest related research includes using an improved HCNN (hierarchical convolutional neural networks) network for EEG emotion classification, extracting its differential entropy features at specific time intervals of each channel. In order to maintain the position information of the EEG signal, this method converts one-dimensional EEG time domain information into two-dimensional differential entropy frequency domain features for subsequent HCNN training [17]. On the other hand, using Discrete Wavelet Transform (DWT) and K-Nearest Neighbor (KNN) algorithm is also an impact method, the EEG signal is decomposed into three frequency bands for MP state recognition. Although these traditional methods of manually extracting emotional features combined with machine learning algorithms have achieved good development, most of them require a large amount of prior knowledge to search for the features of EEG signals and construct feature engineering. EEG signals are prone to noise interference, and the differences between different subjects make manual feature selection based on EEG signals time-consuming and labor-intensive [18]. Due to different subjects’ EEG characteristics, the classification accuracy of a static MP classifier may be affected due to different subjects’ EEG characteristics. The major motivation of presenting a deep learning network to recognize useful EEG feature abstractions is that human cortical functions operate in a deep hierarchical structure. Because EEG is recorded in a noninvasive manner, PSD features can accurately represent the voltage fluctuations in neurons in the superficial cortex. However, when the signal propagates from the source to the scalp, the expression of deep neuron activity related to MP changes may be disturbed. In order to make the best use of the EEG characteristics of deep neurons, a feasible method aims to establish a hierarchical deep structure to reconstruct the source of hidden features [19,20]. In this case, the shallow layer of the deep learning model can be regarded to be a filter to capture the best feature combination of the external scalp EEG information.

While many different DL models have been proposed in the past, most leverage standard CE loss and conventional regularization techniques to learn about temporal persistence and topic specificity. Few studies utilize other training methods, such as adversarial learning and comparison losses in generative adversarial networks (GANs) [21,22]. For example, it is very difficult to find a balanced adversarial approach between discriminative and adversarial networks, thus an additional adversarial network is required. Comparing losses, on the other hand, requires a lot of computation to estimate the pairings for the training data. However, the DL method can increase the repeatability of EP/ERP in different sessions to achieve better performance within 0.5 to 5 s. Although the performance of DL algorithms has been greatly improved, in order to obtain a higher recognition rate, a lot of training is still necessary. In addition, traditional methods such as weighted regularization and dropout are not effective for training a DL model on a single EEG data. EEG data augmentation methods are thus important to learn session-invariance and topic unique embeddings that minimize the triplet loss as will be described in the following sections.

3. The Proposed Method
3.1. EEG Signal Pre-Processing

The main purpose of the pre-processing stage is to denoise the continuous asymmetric EEG data and divide them into continuous synchronous frame segments for neural network processing. The denoising operation is mainly to filter out common EEG artifacts, such as common baseline drift, power interference and EEG spikes. Since the Butterworth filter has the advantages of ideal zero ripple passband and stopband characteristics, it is selected for all filtering operations in this paper. The sampling frequency of the original EEG signal is 12 khz, before the down sampling operation, the EEG signal is filtered by the 8th order Butterworth IIR filter with a cut-off frequency of 800 Hz to prevent aliasing,
and then the continuous EEG signal is down-sampled to a sampling frequency of 2000 Hz (the down-sampling factor is 6). When removing EEG baseline drift, compared with the traditional high pass filter, the S-G filter can effectively detect transient or short-term slow drift without discarding the low-frequency EEG frequency band. The 3-order polynomial and 1001 sample (0.5 s) window size S-G filter are designed again for channel correction, and then the estimated baseline is removed from the record. When removing the power supply interference signal, a second-order notch IIR filter is used to remove the 60 Hz power supply interference and its 120 Hz first harmonic. Finally, the spike caused by electrode displacement is clipped from the threshold value estimated from the deviation of continuous EEG signals, which will be greater than $5\sigma$ amplitude frequency cancellation, where $\sigma$ is the standard deviation of the EEG signal. Another important step of EEG signal pre-processing is EEG signal framing and outlier removal. That is, the filtered continuous EEG signal is divided into non-overlapping frames with a length of 1 s to obtain a minimum frequency resolution of 1 Hz and ensure the recognition accuracy of the EEG signal in the cross-session model. Finally, the high variance frame is removed.

3.2. TL-Based CNN Model

3.2.1. Input Layer

The model consists of seven basic layers, in which the input layer is the synchronous frame extracted from continuous EEG data. The data set used for training is defined as $(X^c_n, s)$, where $X^c_n \in \mathbb{R}^{P \times Tc \times 1}$ is the $n$th training EEG frame synchronized to mental component $c$, and $s$ represents the subject. The input of the encoder is represented as a 3D tensor, in which $P$ and $Tc$ are the number of EEG channels ($P = 7$) and the number of time samples synchronized to component $c$, separately. The last dimension of 1 represents the number of input channels of the first convolution layer of the model.

3.2.2. Two Dimensional Convolution

In the convolution layer, a layer of network is added to share the weight of the filter according to the input dimension to train the correlation characteristics of the learning samples. Based on the correlation between adjacent neurons, the convolution filter weight is optimized to extract the potential EEG signal characteristics, which can effectively reduce the number of training parameters and improve the training efficiency. The input–output relationship of the two-dimensional convolution layer is as follows:

$$a^l(n, h, w, m) = \sum_{c=0}^{C} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} a^{l-1}(n, h + i, w + j, c) \times W^m(i, j, c) + b^m$$ (1)

where, $a^l$ is the activation amount of layer $l$, and $W^m$ and $b^m$ are the $m$-th convolution filter (size $H \times W \times C$) and offset. $C$ is the depth of input (i.e., the number of input channels). According to the two-dimensional convolution of EEG signal analysis, a set of spatiotemporal filters is applied to capture the functional connection characteristics of adjacent EEG electrodes.

3.2.3. Activation Layer

Due to the linear mapping characteristics of EEG data, the training model which only depends on the convolution layer will produce a set of simple outputs. In order to increase the complexity of the training model, a nonlinear activation layer is introduced after each convolution layer. The nonlinear activation layer is realized by the activation function, and the common activation functions are the modified linear element (ReLU), tanh function and sigmoid function. Through the experiment, it is found that the use of relu in the proposed model can effectively reduce the gradient disappearance problem, and is superior to other nonlinear excitation functions in convergence. ReLU is a tensor operation as shown in Equation (2).

$$a^l = \text{Max}(a^{l-1}, 0)$$ (2)
3.2.4. Batch Normalization (BN) Layer

Since the CNN model needs to effectively process long EEG sequences in a short calculation time, and in order to prevent the model from over-fitting, it is necessary to introduce a BN layer. The BN layer avoids the activation load by continuously normalizing the data of the output layer to zero mean value and unit standard deviation, so that the activation will not diverge significantly when the mean value and standard deviation remain unchanged. Moreover, the convergence speed of the CNN model is ensured with a large learning rate. In the CNN model, the BN layer is added after each convolution layer to speed up the convergence time. The BN operation is shown as Equation (3).

\[ a_l(n, h, w, c) = \gamma_c a_l^{l-1}(n, h, w, c) - \mu_c \sqrt{\sigma^2_c + k} + \beta_c \]  

(3)

where, \( \mu_c \) and \( \sigma_c \) are the mean and standard deviation of input channel C. \( \gamma_c \) and \( \beta_c \) are rescaling parameters that are learned during training. \( k \) is a small constant to prevent division by zero.

3.2.5. Dropout Layer

During model training, the neurons in the hidden layer tend to associate with other neurons, resulting in redundant features. Dropout, as one of the commonly used techniques to prevent model over-fitting, randomly breaks these dependencies between neurons, and sets the percentage of neurons controlled by the dropout rate to zero to promote relatively independent characteristic performance between neurons. In this model, the dropout rate after each convolution layer is set to 0.1 (that is, 10% of neurons are randomly set to zero). The experiment shows that by increasing dropout, the performance will be improved by about 4–5%.

3.2.6. Depth-Wise 2D Convolution Layer

Similar to 2D convolution, as shown in Figure 1, depth-wise convolution (DC) applies a 2D filter to the input layer while ensuring that the input channel (depth dimension) is processed independently. The depth direction two-dimensional convolution operation first separates the input channels, then applies a set of 2D filters to each channel (the number of filters is controlled by the depth multiplier parameter), and finally ensures that the output of the filter continues along the depth dimension.

Figure 1. TL based CNN model architecture diagram.

Figure 2 illustrates the difference between deep convolution and standard two-dimensional convolution. Using DC is able to significantly reduce the number of trainable parameters, thus improves the generalization ability of invisible examples. Especially when processing long sequence EEG, DC is the ideal choice of high-density layer as the feature aggregation layer. The complete model is shown in Table 1.
### Table 1. Architecture of the proposed CNN model.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Parameters</th>
<th>Output Size</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>input</td>
<td>-</td>
<td>$(P \times T \times C \times 1)$</td>
<td>-</td>
</tr>
<tr>
<td>Conv2D</td>
<td>$(1 \times 5 \times 25 \times 32)$</td>
<td>$(P \times T \times C \times 1)$</td>
<td>activation=RELU, strides=(1,2)</td>
</tr>
<tr>
<td></td>
<td>bias:32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN</td>
<td>$(2 \times 32)$</td>
<td>$(P \times T \times C \times 1)$</td>
<td></td>
</tr>
<tr>
<td>Average pooling</td>
<td>-</td>
<td>$(P \times T \times C \times 1)$</td>
<td>pool size=(1,2), strides=(1,2)</td>
</tr>
<tr>
<td>Dropout</td>
<td>-</td>
<td>$(P \times T \times C \times 1)$</td>
<td>rate=0.1</td>
</tr>
<tr>
<td>Conv2D</td>
<td>$(32 \times 5 \times 15 \times 64)$</td>
<td>$(P \times T \times C \times 1)$</td>
<td>activation=RELU, strides=(1,2)</td>
</tr>
<tr>
<td></td>
<td>bias:64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN</td>
<td>$(2 \times 64)$</td>
<td>$(P \times T \times C \times 1)$</td>
<td></td>
</tr>
<tr>
<td>Average pooling</td>
<td>-</td>
<td>$(P \times T \times C \times 1)$</td>
<td>pool size=(1,2), strides=(1,2)</td>
</tr>
<tr>
<td>Dropout</td>
<td>-</td>
<td>$(P \times T \times C \times 1)$</td>
<td>rate=0.1</td>
</tr>
<tr>
<td>Conv2D</td>
<td>$(64 \times 5 \times 5 \times 128)$</td>
<td>$(P \times T \times C \times 1)$</td>
<td>activation=RELU, strides=(1,2)</td>
</tr>
<tr>
<td></td>
<td>bias:64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BN</td>
<td>$(2 \times 128)$</td>
<td>$(P \times T \times C \times 1)$</td>
<td></td>
</tr>
<tr>
<td>Average pooling</td>
<td>-</td>
<td>$(P \times T \times C \times 1)$</td>
<td>pool size=(1,2), strides=(1,2)</td>
</tr>
<tr>
<td>Dropout</td>
<td>-</td>
<td>$(P \times T \times C \times 1)$</td>
<td>rate=0.1</td>
</tr>
<tr>
<td>DepthwiseConv2D</td>
<td>$P \times T \times C \times 64 \times 2$</td>
<td>$256$</td>
<td>Depth multiplier=2, axis=1</td>
</tr>
<tr>
<td>BN</td>
<td>$(2 \times 256)$</td>
<td>$256$</td>
<td></td>
</tr>
<tr>
<td>L2-norm</td>
<td>-</td>
<td>$256$</td>
<td></td>
</tr>
</tbody>
</table>

The CNN model is composed of four main blocks. The first three blocks represent a set of spatiotemporal filters. The last block is the feature aggregation block to control the size of feature embedding. In the last layer, L2 is to normalize the embedding feature, so that the embedding is constrained on a unit hypersphere in the embedding space. Embedding normalization can speed up the convergence of triple loss and improve the performance of data validation. The total amount of training parameters of the first three convolution blocks (i.e., independent of the input size) is 425730, and the number of trainable parameters of the last layer related to the input size is $128 \times P \times T \times C / 64 \times 2 + 3 \times 256$. For example, for an EEG input with 7 channels and 1250 time samples (5 s at 250 Hz), the total number of trainable parameters of the designed CNN model will be 523,740.

![Figure 2. Illustration of the depth-wise and standard 2D convolution operations.](image-url)

(a) depth-wise 2D convolution: (1) input volume with shape $6 \times 6 \times 3$, (2) channel separation, (3) channel filters with shape $3 \times 3 \times 2$ (depth multiplier = 2), (4) output of the convolution operation for each input channel, (5) channel concatenation to get the final output with shape $6 \times 6 \times 6$. (b) standard 2D convolution: (1) input same as (a), (2) four different convolution filters with shape $3 \times 3 \times 3$, (3) convolution output with shape $6 \times 6 \times 4$. 

![Figure 2. Illustration of the depth-wise and standard 2D convolution operations.](image-url)
3.3. Triple Loss

Compared with the traditional classification task using CE, TL can capture more significant data features from input data to improve the performance of feature extraction and classification. Even when training on a relatively small data set, TL may have a stronger generalization ability for invisible data. For the problem of low convergence of the EEG signal, TL can also ensure its stronger generalization ability. In terms of the above reasons, TL is selected as the objective function of the cross-session state deep learning model. TL calculation needs to redefine the status labels of training samples into three dimensions, namely Anchor (A), Positive (P) and Negative (N). The P dimension is allocated to training samples from the same type (topic). The N dimension is allocated to training samples from different types of topics. TL will calculate the distance embedding function ($d_p$) between A and P and the distance embedding function ($d_N$) between A and N. The triplet loss function is given by Equation 4.

$$L = \sum_{n=1}^{B} \max \left[ d_p^{(n)} - d_N^{(n)} + \alpha \right] = \sum_{n=1}^{B} \max \left[ \left\| e_A^{(n)} - e_P^{(n)} \right\|_2^2 - \left\| e_A^{(n)} - e_N^{(n)} \right\|_2^2 + \alpha \right]$$

where, $d_p$ and $d_N$ are the square of Euclidean distance between triads, $e_A^{(n)}$, and $e_P^{(n)}$ and $e_N^{(n)}$ represent the embedding of the A,P,N triples with index n in batch processing with the size B, respectively. Hyperparameter $\alpha$ represents a positive value representing the boundary between $d_p$ and $d_N$. The minimization of the loss function $L$ to zero means that the average inter-subject distance ($d_N$) of the embedded feature is greater compared with the average within-subject distance of the embedded feature ($d_p$), its minimum value is $\alpha$, and the margins of all tests performed $\alpha$ are set to 0.5.

4. Experimental Analysis

4.1. Model Training and Triplet Acquisition

For different cross-session EEG models, the parameters of the training model are randomly initialized with the Glorot canonical initialization model. The parameters of the model are updated based on the small batch gradient descent method optimized by Adam. The initial learning rate is set to $10^{-3}$, and the total number of training cycles and the size of batch processing are set to 64 and 128, respectively. The triplet loss function provided by formula 5 and 4 is used to train the model. The model is trained to minimize the triplet loss function 4 given in the equation. When dealing with triplet loss, hard triplets are selected for online mining, which can effectively reduce the number of calculations of the model and reduce the amount of memory required to assess the pairwise distance of training examples in batch processing. The proposed CNN model is a $32 \times 128$ EEG matrix. The temporal dimension convolution in each spatiotemporal core block is carried out using a $1 \times 3$ convolution kernel three times, while the spatial dimension convolution is carried out using a $3 \times 1$ convolution kernel one time. The feature map parameters are set to exponential growths of 16, 32, and 64, respectively. The number of neurons in the fully connected layer is set to 50, and the final number of classification units is 2. In addition, the batch size in this article is set to 128, the learning rate is 0.1, the learning decay rate is 0.99, and the Epoch is 100.

During the training process, this article divides the 63 s EEG signal into 63 1-s time periods in the time domain and extends the corresponding data labels. The total number of EEG epochs for each subject in 40 trials is 2520, and ultimately the user data were divided into 128 data points and 32 channels. Based on the range of MP level values for each test from 1 to 9, the median 5 as the threshold divides arousal and valence into two categories. More than 5 indicates a high pressure state; the greater the number, the greater the pressure.

Finally, we obtain $1 \times 2520$ dimensional label data corresponding to the EEG signal. At the same time, $1/4$ of the test data are taken as the validation set to verify the cross session correct recognition rate (CRR) of the DL model in each training cycle and the remaining
3/4 as training data. Here, the k-NN algorithm \((k = 1)\) is adopted for estimating the CRR of the model to save the model with the best validation CRR. The best validation CRR is monitored in a patient window of a 15 epochs cycle. If the validation CRR does not improve in this window, the training process will be terminated and the model parameters with the best validation CRR will be saved. In the end, in order to reduce the impact of random parameter initialization, each test is trained for five times, and then the performance index is averaged.

4.2. Task Setting

The task performance indicators of different MP states are shown in Figure 3. The pre-processing stage was implemented in the MATLAB software. The DL model was implemented in Python using the Keras platform for machine learning. All the computations required for training and testing our model were performed using a Google Colab PRO account. In order to verify the experimental effect, the average time percentage of the four subsystems within the test target range is evaluated, which can be expressed as the system error range (SIE), and the SIE data are analyzed by one-way repeated measurement variance analysis. The SIE of low MP is notably higher than that of high MP, where the SIE of all courses is \(p < 0.001\). Wilcoxon signed rank test is introduced to explore the mean value and medium of SIE of all subjects between stage 1 and stage 2. The results show that the change of the two groups of statistical data is not significant, which can prove that there is no difference in learning effect between different sessions.

![Figure 3. SIE box diagram of MP binary level.](image)

4.3. Significant EEG Characteristics

To verify the relationship between the MP state and EEG characteristics describing different frequency bands and discover the most significant EEG characteristics of MP changes, the linear correlation coefficient \(r\) between EEG power in 55 different frequency bands and the time history of the target MP type, and the correlation coefficient \(r\) in three different situations, Session 1 (Case 1), Session 2 (Case 2) and Double Session (Case 3), are repeatedly calculated, as shown in Figure 4.

Correlation is measured by a Pearson product moment correlation coefficient,

\[
\begin{align*}
    r &= \frac{\sum_{k=1}^{N} (x(k) - \bar{x})(y(k) - \bar{y})}{\sqrt{\sum_{k=1}^{N} (x(k) - \bar{x})^2} \sqrt{\sum_{k=1}^{N} (y(k) - \bar{y})^2}} \\
\end{align*}
\]

(5)

where \(x(k)\) and \(y(k)\) represent EEG eigenvalues and target MP class at time step \(K\). \(y(k) = 0\) and \(y(k) = 1\) represent low MP and high MP States, so the change of \(y(k)\) maps the MP level. In Figure 4, most EEG features are negatively associated \(y(k)\), while the EEG features of theta, beta and gamma are negatively related to \(y(k)\) and the bands show a positive relationship to \(y(k)\). Additionally, the higher the absolute value of \(r\), the stronger the recognition ability of some EEG features. That is, the value of \(R\) can be the basis for feature selection, the maximum value of \(|r|\) is consistent with the most significant EEG feature in Figure 5.
Figure 4. Linear analysis results of EEG characteristics of each frequency band and time process of target MP class.
Intra Session and Cross Session MP Classification

In order to evaluate the classification performance of the MP classifier, the following indicators are introduced. The classification rate of the first class (low MP) is defined as the sensitivity $P_{\text{sen}}$, $P_{\text{sen}} = \frac{N_{lp}}{N_{lp} + N_{hf}}$, $N_{lp}$ represents the number of EEG data points with low MP correctly estimated by the classifier, and $N_{hf}$ represents the number of high MP classes that are misclassified. The classification rate of the second category (high MP) is set as specific $P_{\text{spe}}$, $P_{\text{spe}} = \frac{N_{hn}}{N_{hn} + N_{lp}}$, $N_{hn}$ represents the number of high MP EEG data points correctly estimated by the classifier, and $N_{lp}$ represents the number of low MP data points misdiagnosed. The precision of the low MP class is defined as $P_{\text{pre}} = \frac{N_{lp}}{N_{lp} + N_{hp}}$, and the precision of the high MP class is defined as $P_{\text{npv}} = \frac{N_{hn}}{N_{hn} + N_{hp}}$. The overall classification accuracy is determined as $P_{\text{acc}} = \frac{N_{ln} + N_{lp}}{N_{ln} + N_{hn} + N_{hp}}$. Figure 6 compares the classification results of 10 training and test procedures. In the context of the training intra-session test and cross-session test, the classification algorithm based on CNN is adopted to calculate the classification performance indexes $P_{\text{ppn}}, P_{\text{app}}, P_{\text{pre}}, P_{\text{npv}}$ and $P_{\text{acc}}$. The test result in the session (case 2) is to use 3/5 data of case 1 for training, and the remaining 2/5 data are used for comparative testing. The test results of the cross-session (case 3) are calculated by using the data of session 1 for training and verification, and the data of session 2 for testing. The
test results of cross-session (case 3) are obtained through training and verifying with the
data of case 1 and testing with the data of case 2. According to case 1 in Section 3.2, the
most significant EEG characteristics are derived.

The verification set SVA is only used for determining the number of nodes in the
adaptive classifier in case 3, whereas the number of nodes in case 1 and case 2 is simply
selected to be the same as that in case 3 for comparison. For the average $P_{spn}, P_{spe}, P_{pre}, P_{npe}$
and $P_{acc}$ of each participant, in accordance with Wilcoxon signed rank test, the change
between cases 2 and 3 is not significant. The results show that although the average perfor-
mance indicators in the two cases are not significantly different, the average performance
indicators in case 2 are better than those in case 3 under C, D and G. This means that
intra-session MP classification is an easier task for adaptive classifiers.

![Figure 6](image)

Figure 6. Error bar graph of CNN model for training intra-session testing and cross-session testing.
(a–e) represent the performance of $P_{sen}, P_{spe}, P_{pre}, P_{npe}$ and $P_{acc}$, respectively.

The CNN-based recognition algorithm is run 10 times in the intra-session and cross-
session situations. The optimal classification confusion matrix represented by the classification
rate of each case is summarized in Table 2.

The total classification rate $P_{acc}$ is 0.9496. The values of $P_{sen}, P_{spe}$ and $P_{acc}$ are decreased
to 0.8578, 0.8389, and 0.8146 under cross-session situations. The matrix at the end of the
table is the result of all participators’ correct or incorrect assessments of the EEG samples.
The results show that the accuracy of the intra-session situation is shown to be much higher
than that of the cross-session. After the Wilcoxon signed rank test, the values of $P_{spe}$ and
$P_{acc}$ of all participants are significantly higher in the former case than in the latter case
($(z = -2.67, S = 0.02)$ and $(z = -2.23, p = 0.02)$), and $P_{sen}$ is not significantly different in both
cases ($z = -1.79, p = 0.07$).
Table 2. Optimal classification confusion matrix in P_{spe} in intra session and cross session situations.

<table>
<thead>
<tr>
<th>Test Object</th>
<th>Estimated Grade</th>
<th>Intra Session</th>
<th>Cross Session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target Category</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Object1</td>
<td>Low</td>
<td>0.98</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.02</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.9683</td>
<td></td>
</tr>
<tr>
<td>Object2</td>
<td>Low</td>
<td>1</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.9950</td>
<td></td>
</tr>
<tr>
<td>Object3</td>
<td>Low</td>
<td>0.9167</td>
<td>0.0789</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.0833</td>
<td>0.967</td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.9217</td>
<td></td>
</tr>
<tr>
<td>Object4</td>
<td>Low</td>
<td>0.91</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.08</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Object5</td>
<td>Low</td>
<td>0.96</td>
<td>0.0167</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.04</td>
<td>0.9866</td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Object6</td>
<td>Low</td>
<td>10.8</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Object7</td>
<td>Low</td>
<td>0.95</td>
<td>0.0036</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.05</td>
<td>0.9989</td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>Average value</td>
<td>Low</td>
<td>0.96</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.03</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>(P_{acc})</td>
<td>0.9496</td>
<td></td>
</tr>
</tbody>
</table>

4.5. Classification Results

This paper verifies the performance of the CNN model designed for cross-session MP classification. As shown in Figure 7, the average classification performance index of the 10 times adaptive model is compared with seven classification algorithms commonly used in EEG stress recognition for all classification algorithms, the training set and verification set are from the non-overlapping three-fifths and two-fifths data of session 1, and the test sample is all the data of session 2.

ANN represents a three-layer artificial neural network based on MLP. Each participant repeats the training and testing process of the neural network for 10 times, calculates its average value for comparison and statistical analysis, and sets the number of hidden neurons and input feature dimensions to 55. NB represents a naive Bayesian classifier without any pre-set parameters. KNN means K-nearest neighbor classifier with K = 30. SVM_{lin} and SVM_{rbf} represent standard support vector machines using linear and nonlinear kernels with radial basis functions (RBFs), respectively. For linear support vector machines, the model selection of regularization parameters can be obtained with a training set, a validation set and 15 candidate parameter sets composed of \(\{2^{-7}, 2^{-6}, \ldots, 2^{7}\}\). For RBF-SVM, grid search regularization parameters and kernel width are used in the set \(\{2^{-7}, 2^{-6}, \ldots, 2^{7}\}\). BSVM indicates that the bounded support vector machine is a variant of the standard SVM. Similar to the implementation of the standard support vector machine, the linear kernel and RBF kernel use the same model selection criteria, respectively, which are represented as BSVM_{lin} and BSVM_{rbf}. It can be found from Figure 7 that among all classifiers of \(P_{sen}, P_{npv}\) and \(P_{acc}\), the median of CNN model is the highest, while the median of the neural network classifier is the highest. Table 3 lists the results of comparing the five performance
indicators between the CNN model and the other seven MP classifiers with Wilcoxon signed rank test. It shows that $P_{\text{sen}}, P_{\text{spe}}$ and $P_{\text{acc}}$ have been significantly improved, and the values of $P_{\text{pre}}$ and $P_{\text{npv}}$ are equivalent to the performance of other classifiers.

The classification performance index in the case of single channel is shown in Figure 8. For all subjects, the five power characteristics of the P4, O1, F3, and O2 channels are conducted for MP classification, respectively. The parameter settings corresponding to each classification algorithm are the same as those defined in Section 4.5. In terms of different classification algorithms, the median of the CNN algorithm indexes $P_{\text{sen}}, P_{\text{npv}}$ and $P_{\text{acc}}$ based on TL is the highest. The detailed statistical results of the TL-based CNN model and other classifiers by Wilcoxon signed rank test are shown in Table 3. Compared with BSVM$_{\text{rbf}}, P_{\text{sen}}$ and $P_{\text{npv}}$ have significant anti-preventive effects. Nevertheless, regarding
other classification performance indicators, the improvement of adaptive CNN is not significant. The classification performance in the single frequency band is shown in Figure 8. The specific bands of participants A-G are $\sigma$, $\gamma$, and $\alpha$, and 11 EEG features are conducted, respectively. Compared with BSVM$_{lin}$, KNN, BSVM$_{rff}$, SVM$_{rff}$, SVM$_{lin}$, and ANN, $P_{sen}$ and $P_{npo}$ are significantly improved. Compared with Nb, $P_{pre}$ is also significantly improved. For the overall classification accuracy $P_{acc}$, adaptive CNN is better than BSVM$_{lin}$, Nb, SVM$_{lin}$ and ANN, and is equivalent to BSVM$_{rff}$, SVM$_{rff}$ and KNN. Various DL models have been previously proposed; however, most of these works used the standard CE loss to learn subject-unique features. Meanwhile, a few papers adopted other training approaches such as adversarial training using Generative Adversarial Network (GAN) and contrastive loss to learn invariant representations of EEG. The generator/encoder in the GAN model was trained to learn session-invariant representations by hiding the session information from a discriminator, but this approach achieved poor CRR values ranging between 66.6% and 71.6% over 10 subjects only due to the short EEG epochs used for testing (0.5 s) [23,24]. Compared with the existing research, we put forward the scheme on the inheritance of the EEG deep learning model signal processing advantage, focus on solving the problem of the EEG signal high transition challenges, using the TL algorithm to ensure the invariance characteristics of EEG signal, to identify that the psychological stress state has obtained the ideal effect. The limitation of the model is that the automatic separation of EMG signals ensures the integrity of EEG signals, which is the focus of our next research.

**Figure 8.** Box diagram of cross session MP classification performance index related to the selected channel.

4.6. Noise Robustness Analysis

The noise robustness of eight different MP classifiers in cross-session is evaluated by artificially adding different numbers of Gaussian distribution time processes (i.e., white Gauss) to EEG features. The purpose of the examination is to test whether the classifier is effective when the constraints controlled by the experimental paradigm are not available. Classifier input and parameter settings are the same as defined in Section 4.5. The detailed
results of classification accuracy under different noise conditions are shown in Figure 9. The number of noise signals increased from 1 to 15. The standard deviation of the Gaussian distribution of noise also increases in 15 noise signals, where the mean value is zero. From the 1st to the 15th additional noise signal, the standard deviation values are 0.1, 0.2, . . . , 1.4, 1.5, respectively. Therefore, in the worst case, 15 noise signals with increasing pollution levels are superimposed on the EEG characteristics. The average classification performance is then calculated and compared across all participants.

Figure 9. Relationship between the average classification performance index and the number of noise signals of different MP classifiers.

The results show that the classification performance will decrease with the increase in the noise number of all classifiers. From the Figure 9e, it can be observed that the CNN classification algorithm based on TL achieves higher overall classification accuracy ($P_{acc}$) than the other seven classification methods under all noise conditions. Thanks to the specially designed hierarchical structure, the CNN classification algorithm based on TL owns better noise robustness. For the performance index of single class classification, the CNN classification algorithm based on TL is shown to have superiority to other classifiers of $P_{sen}$, $P_{pre}$ and $P_{npv}$ under all noise conditions. As the number of noise signals increases, the classification rate of high MP class (i.e., $P_{spe}$) decreases significantly, see Figure 9b. When EEG features are seriously polluted, this degradation will weaken the corresponding overall classification accuracy.

5. Conclusions

This paper designs a CNN deep learning model for brain stress state recognition in cross session situations with EEG biometrics, and classifies cross session MP through EEG signals. The model encodes EEG data into a feature space, takes the triplet loss (TL) function to maximize the distance between subjects and minimize the distance between embedded intra-subjects. The weights of the first hidden layer connected to the input layer in the CNN model are iteratively updated to track the continuous changes of EEG power characteristics. Through different feature selection and noise destruction paradigms, further performance comparison between CNN classification model and classical MP classifier shows that when taking comprehensive cortical information as network input, the proposed method is superior to shallow and static classifiers. The analysis and test of EEG characteristics of hidden neurons with different depths show that the hierarchical structure of the CNN model based on TL can describe the clear distribution of EEG data at a higher level. Moreover, in future studies, we will explore the depth classifiers to simulate more EEG internal noise information to better capture the non-stationary distribution of neurophysiological data, so as to more accurately utilize EEG for human state perception prediction.
Author Contributions: Conceptualization, S.Z. and T.G.; methodology, S.Z.; software, S.Z.; validation, S.Z. and T.G.; formal analysis, S.Z. and J.X.; data curation, S.Z.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z., T.G. and J.X. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the National Natural Science Foundation of China under Grant Number 52130403.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to this paper only uses machine learning technology to analyze and study EEG data, and does not involve human ethics, so the ethical approval is not necessary in this paper.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

23. Upadhyaya, V.; Salim, M. Effect of sensing matrices on quality index parameters for block sparse bayesian learning-based EEG compressive sensing. *Int. J. Wavelets* 2023, 21, 31–53. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.