

Article

Research on Emotion Recognition for Online Learning in a Novel Computing Model

Mengnan Chen , Lun Xie ^{*}, Chiqin Li and Zhiliang Wang

School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China; g20198772@xs.ustb.edu.cn (M.C.); lcqin111@163.com (C.L.); wzl@ustb.edu.cn (Z.W.)

^{*} Correspondence: xielun@ustb.edu.cn; Tel.: +86-136-8156-0734

Abstract: The recognition of human emotions is expected to completely change the mode of human-computer interaction. In emotion recognition research, we need to focus on accuracy and real-time performance in order to apply emotional recognition based on physiological signals to solve practical problems. Considering the timeliness dimension of emotion recognition, we propose a terminal-edge-cloud system architecture. Compared to traditional sentiment computing architectures, the proposed architecture in this paper reduces the average time consumption by 15% when running the same affective computing process. Proposed Joint Mutual Information (JMI) based feature extraction affective computing model, and we conducted extensive experiments on the AMIGOS dataset. Through experimental comparison, this feature extraction network has obvious advantages over the commonly used methods. The model performs sentiment classification, and the average accuracy of valence and arousal is 71% and 81.8%, compared with recent similar sentiment classifier research, the average accuracy is improved by 0.85%. In addition, we set up an experiment with 30 people in an online learning scenario to validate the computing system and algorithm model. The result proved that the accuracy and real-time recognition were satisfactory, and improved the online learning real-time emotional interaction experience.

Keywords: physiological signals; affective computing; online learning; galvanic skin response; electrocardiogram; terminal-edge-cloud; feature selection



Citation: Chen, M.; Xie, L.; Li, C.; Wang, Z. Research on Emotion Recognition for Online Learning in a Novel Computing Model. *Appl. Sci.* **2022**, *12*, 4236. <https://doi.org/10.3390/app12094236>

Academic Editors: Hüseyin Kusetogullari and Cih-Hsiung Tu

Received: 24 March 2022

Accepted: 19 April 2022

Published: 22 April 2022

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



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

1. Introduction

Emotions are considered to be human emotional states, it is a response to certain stimuli in the external environment or interpersonal interactions. Understanding and quantifying human emotional states, which have major implications for intelligent human-machine systems. In 1997 Picard and Healey proposed to equip sensors to record physiological signals, identify the emotional state of the wearer through signals, improve human-computer interaction experience through affective computing [1], the paper indicates that in the future, sensors will become small enough, thus, a wearable device for real-time emotion recognition is designed.

The accessibility, non-fakeability, and continuous detectability of physiological signals [2] is a hot topic of current research. Physiological signals can be divided into two categories: signals originating from the peripheral nervous system and signals from the central nervous system. Compared to Electroencephalogram (EEG) signals, the combination of Electrocardiogram (ECG) and Galvanic Skin Response (GSR) is less explored in the literature. Since ECG and GSR are rich in emotional information and can be obtained by low-cost, non-invasive devices, making them highly significant in terms of affective computing. Among them, ECG has been shown to be a reliable source of information for emotion recognition systems [3–5], ECG analysis can identify the emotional state of the users, such as happiness, sadness, and stress. The GSR signal is a non-smooth signal, usually measured at the palm of the hand, and is a collection of two different components:

Tonic and phase components. Tonic indicates the general level of skin conductivity, its horizontal value varies slowly with time. Phase components shows a sharper peak on the tidal drift of the tonic GSR, phase components usually caused by the presence of instantaneous sympathetic activation of the stimulus causing, and that can reflect changes in cognitive and emotional processes [6–8]. Several studies [9–12] have shown that an adequate combination of information extracted from multiple models may improve robustness (for noisy inputs). Therefore, this paper mainly focuses on GSR and ECG to analyze. In many real-life scenarios, a key factor in decision-making (e.g., healthcare) is the classification model. For applications in these areas, affective computing systems must be able to describe the uncertainty of their emotional state outputs, and arousal and valence dimensions are the best options [13,14]. Therefore, the binary high/low classification problem is considered in this study [15]. The affective computing task can be accomplished using two types of models, one is a deep learning model and the other is a traditional machine learning model. Deep learning related methods have had great success in the field of pattern recognition. More and more researchers are using it in affective computing tasks [16]. For example, new deep learning models [17], and many innovative models have been generated in machine learning models. Affective Computing has an important role in healthcare [18], education [19], and entertainment [20], and its deeper value deserves to be explored.

Currently, deep learning, machine learning methods for affective computing has their own advantages, the deeper reasons for the advantages and disadvantages between the two models need to be further summarized. The effectiveness of feature selection in affective computing directly affects the level of accuracy, the joint mutual information (JMI) dimension of multidimensional features is used as a direct factor for feature validity, can effectively improve the rationality and effectiveness of the selection of features. Recent research in affective computing has focused on improving accuracy while ignoring the importance of the time dimension of affective computing, the time required for affective computing is an important factor in enhancing the human-computer interaction experience. In response to the above proposed deficiencies, the following work has been carried out in this paper. We used deep learning model and a machine learning model to process the ECG and GSR modalities in the AMIGOS dataset, respectively. Focus on exploring the advantages and disadvantages of both models, derive a model architecture with high recognition accuracy. The JMI-based greedy feature selection algorithm is proposed to feature-level fusion, to analyze which features extracted from ECG and GSR are more compatible with the affective computing task. In addition, focus on the time dimension of affective computing, propose a new type of terminal-edge-cloud computing architecture. Organize realistic scenario experiments based on the proposed computing architecture, using online education as an experimental scenario, the method proposed above was used to analyze the experimentally collected physiological database, the more promising results were obtained.

The paper is organized as follows: Section 2 reviews the literature related to physiological signal-based emotion computing. Section 3 describes the analysis of the feature selection algorithms proposed in this paper, based on machine learning and deep learning algorithm for sentiment classification methods, using the AMIGOS dataset to validate their effectiveness. Section 4, describes the novel computing architecture, verifies its advantages in the affective computing times dimension, and designs an online learning scenario experiment to build an emotional database to verify the advantages of the proposed method and computing architecture. Finally, Sections 5 and 6 present the results and conclusions of the experiments generated during this study.

2. Related Work

Changes in physiological signals can be influenced by human emotions, and since the proposal of non-invasive devices that can collect human physiological signals in real-time, many efforts have been made to analyze physiological signals. First, public datasets, DEAP [21], SEED (2015) [22], AMIGOS [23], etc., are proposed, and then a series of senti-

ment computation models are proposed to analyze them. Zheng [24] studied the mental arousal space in four quadrants and solved a four-category task using the graph regularization extreme learning machine (GELM) method, these two obtained about 70% accuracy in the polynomial classification task study. When data are incomplete, semi-supervised learning methods can be used to integrate Stack Auto-Encoder (SAE) with deep belief networks (DBN) using a decision fusion method and based on Bayesian inference classification [25], yielding 73.1% accuracy in arousal and 78.8% accuracy in valence. Another recent GSR-based framework ref. [26] proposed temporal and spectral features of SVM (RBF kernel) under the AMIGOS dataset, reporting 83.9% and 65% arousal and valence recognition accuracy, respectively. New trends in emotion-evoking computing use deep neural networks (DNNs) to process physiological signals and improve recognition rates. One of the earliest attempts was [27], which proposed a multimodal residual LSTM for emotion recognition (MMResLSTM) yielded encouraging results, with their classification accuracy of 92.87% for arousal and 92.30% for valence on the DEAP dataset. Ref. [28] processed ECG and GSR data from the AMIGOS dataset and proposed to use machine learning methods and DCNN to process the data, obtained better results of 0.76 for valence and 0.75 for arousal. A recent study Yang [29] presented the fusion of statistical features extracted from EEG, ECG, and GSR of the AMIGOS dataset. They reported recognition rates of 67% and 68.8% for valence and arousal, respectively, which using the SVM classifier. LSTM-RNN was recently proposed [30] using an attention-based mechanism for the AMIGOS dataset and reported recognition rates of 79.4% and 83.3% for binary classification of valence and arousal. Four-category emotion results also became progressively more common, however, the reported recognition rates decreased more in the case of four categories of emotions [31]. Granados [32] proposed a one-dimensional convolutional neural network model to analyze ECG and GSR signals in the AMIGOS dataset with an accuracy of 65.25% for the A-V four-category emotion recognition task.

The features extracted from physiological signals are the most important aspect in emotion recognition. The processing is carried out in the time domain, frequency domain, or nonlinear domain. Time domain methods include the use of various mathematical/statistical features such as mean [33], median, etc., or the use of methods such as sample differences, zero-crossing, etc. In the frequency domain, the Fourier transform (FT) [34] and the wavelet transform [35] are widely used. The FT allows one to use time-based features on the signal (e.g., its mean or DC component and dominant frequency component) represented in the spectrum. The nonlinear domain approaches require the conversion to the sensor signals to discrete symbolic strings, and the key to performing this conversion is the discretization process. Once these signals are mapped to strings, exact or approximate matches and edit distances [36]. Compared three feature selection algorithms Joint Mutual Information (JMI), Conditional Mutual Information Maximization (CMIM), and Dual Input Symmetric Correlation (DISR) on the AMIGOS dataset and concluded that the three feature selection algorithms are similar, and it is important to have the same number of features to obtain the best accuracy for arousal recognition and valence recognition. Therefore, which features are better, need to be further explored.

In order to accelerate the response speed of model systems, and use the resources at the edge efficiently, it has triggered a boom in edge computing among researchers, and the use of new computing architectures for sentiment analysis are gradually attracting attention, Chen [37] designed a medical artificial intelligence framework based on data-width evolution and self-learning, which aim to provide medical services for skin diseases that meet the requirements of real-time, scalability and personalization. This computational framework allows physicians to quickly obtain patient skin conditions. Edge AI technology to analyze thermal imaging image data of buildings, for rapid analysis of building house occupancy information [38]. Ref. [39] the authors proposed Smart Edgent, a collaborative on-demand DNN co-inference framework with device edge synergy, that can split the network to run the network faster and efficiently use other node resources. Few studies have proposed methods related to the use of affective computing for education, and in [40],

a dynamic difficulty adjustment mechanism for computer games is proposed to provide a tailored gaming experience in individual users by analyzing ECG and GSR.

Our Contribution

We propose to extract features from the ECG and GSR, and use the proposed JMI-Score algorithm to compute the best set of features that match the current sentiment classification task. Machine learning models parameters were optimized to obtain the optimal model, and the CNN model automatically extracted features are compared with extracted manual features, and the accuracy of sentiment classification results was improved compared with the state of art. Propose a new computing architecture that leverages both edge-side and terminal-side computing resources, to speed up the recognition of emotions, and reduce network bandwidth, recognition latency. We also organize field experiments to verify the effectiveness of the novel computational architecture and the proposed affective computing model in the context of online learning.

3. Method

3.1. Experimental Data Description

The following paragraph describes the AMIGOS dataset in a condensed form. In this paper, the newly released dataset AMIGOS are used to validate the model, not only because it is widely used in the recent literature on physiological signal-based emotion elicitation, but also because they use low-cost physiological signal acquisition devices for data collection and the non-invasive nature of the whole process. AMIGOS applied a 14-channel Emotiv Epoc wireless headset to acquire EEG signals, peripheral physiological signals (ECG Right, ECG Left and, GSR physiological data pre-processed at a sampling frequency of 128 Hz) recorded with non-invasive devices such as the ECG Shimmer 2R5, as well as frontal video (RGB), stimulus material from the MAHNOB-HCI [41] dataset as emotionally stimulating material. The dataset used both individually and GrOups scenarios, the first with 40 participants watching 16 short videos (<250 s in length); the second, with 17 people in an individual setting and 5 groups of 4 people each, where participants watched long videos (>14 min in length). Each trial first contained a 5 s baseline signal, with the signal depending on the duration of the video. After viewing the video, participants rated self-assessments of arousal, potency, liking, and dominance on a scale of 1 to 9 in the self-assessment of potency (SAM) [42]. A total of 12,580 video clips were annotated in this way (340 clips from 37 participants in both short and long-video experiments). The arousal and valence scales used for these annotations are continuous, ranging from 1 (low arousal or potency) to 9 (high arousal or potency), and there is a high degree of agreement between annotators. There were 800 records in the dataset and 7 subjects (ID numbers 33, 24, 23, 22, 21, 12, 9) had missed data and were considered invalid.

The dataset can be divided into four classes: low arousal low valence (LALV), high arousal low valence (HALV), low arousal high valence (LAHV), and high arousal high valence (HAHV), and the threshold values of the valence and arousal dichotomy is 5. The kmeans algorithm are applied to cluster the distribution of the data, and Figure 1 shows the distribution of the sentiment classes in AMIGOS, purple represents LALV, blue represents HALV, green represents LAHV, yellow represents HAHV in the Figure 1.

3.2. Preprocessing

In this paper, we use deep learning and machine learning to process physiological signal data separately, and achieve effective emotion detection. GSR is a non-stationary signal, and in this study, the signal is first decomposed by smoothing through empirical modal decomposition (EMD) to obtain the effective frequency, and then the low-pass butterworth filter is used to pre-process the GSR signal since the skin electrical signal changes slowly, and the effective frequency is between 0–0.3 Hz, the cutoff frequency of the low-pass filter is set to 0.5 Hz and the sampling frequency is 128 Hz, and then the SCR and SCL are decomposed. The ECG signal frequency is usually 0.05–100 Hz, firstly, the

baseline drift is eliminated by discrete wavelet transform, which is to eliminate unnecessary low-frequency noise in the frequency range of 0.05 and 1 Hz, and then the end frequency is set to 1 Hz using butterworth high-pass filter which the sampling frequency is 128 Hz. Then to acquire the denoised ECG signal. The noise-reduced signal is first normalized by Z-Score (Equation (1)) using a sliding window of 2 s and an offset of 1 s to capture the subtle changes in emotional motion and derive the feature vectors. Then, the data enters the display or implicit feature extraction phase. The first one extract implicit features by convolutional networks through deep learning, and the second one, by machine learning methods, which extracts time and frequency domain manual features, in three steps of preprocessing, classification and multimodal fusion, respectively.

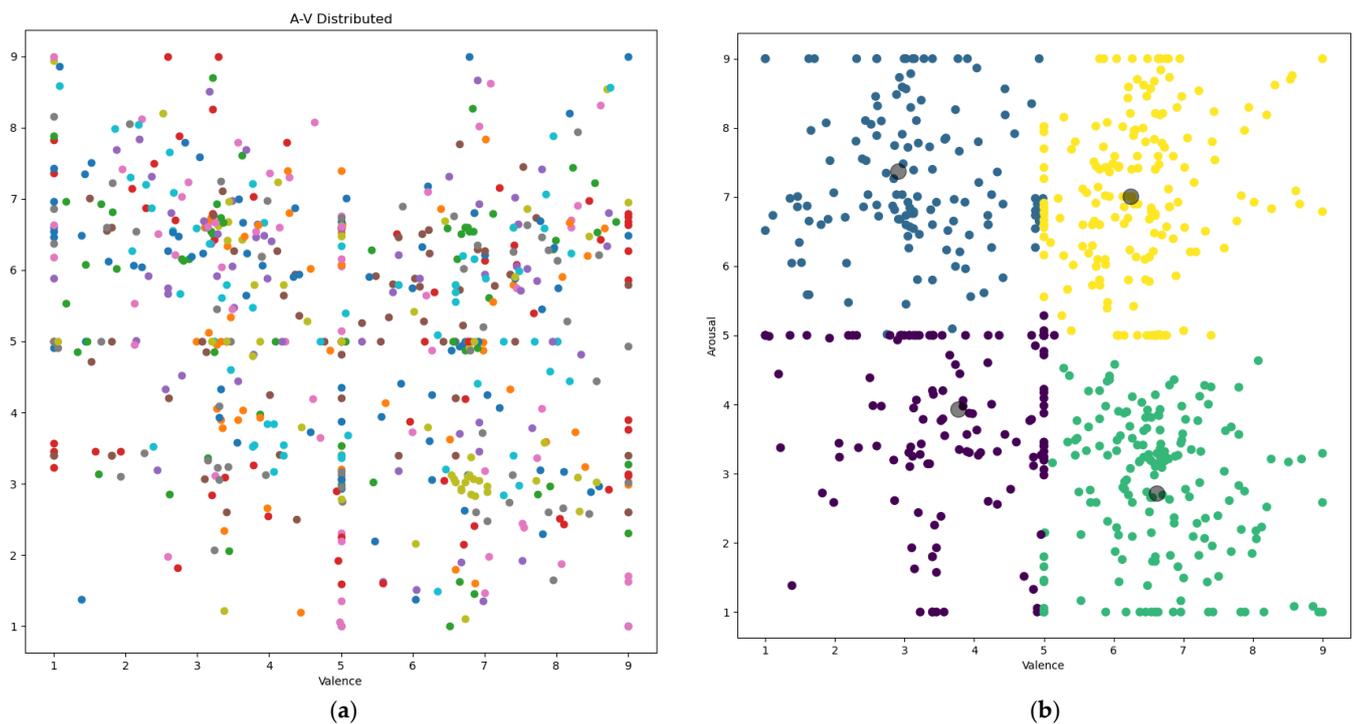


Figure 1. (a) Distribution of Arousal/Valence tags in AMIGOS dataset (b) Distribution of sentiment labels after processing with kmeans.

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

where \tilde{X} is the standardized data, μ and σ are the mean and standard deviation of the data, respectively.

3.3. Detailed Analysis

3.3.1. Deep Learning Methods

Deep learning is an algorithm-based, difficult-to-interpret machine learning field, which used to model high-dimensional features in datasets. In recent research on emotion recognition based on physiological signals, more and more studies use deep learning models to process them, and achieved good results [43]. The deep network structure we used is shown in Figure 2 in this study, CNN is considered as a blur filter, which can automatically discover SCR peaks or SCLs in GSR signals, specific morphological patterns of the QRS complex in the ECG. The signal dimension after CNN processing is 2304 * 528. We believe that the obtained features have noise or invalid features, and SVD is often used in dimension reduction algorithms in deep learning [44]. SVD processing is performed on the extracted features [45], and the signal dimension becomes 268 * 528, which is fed into the fully connected layer.

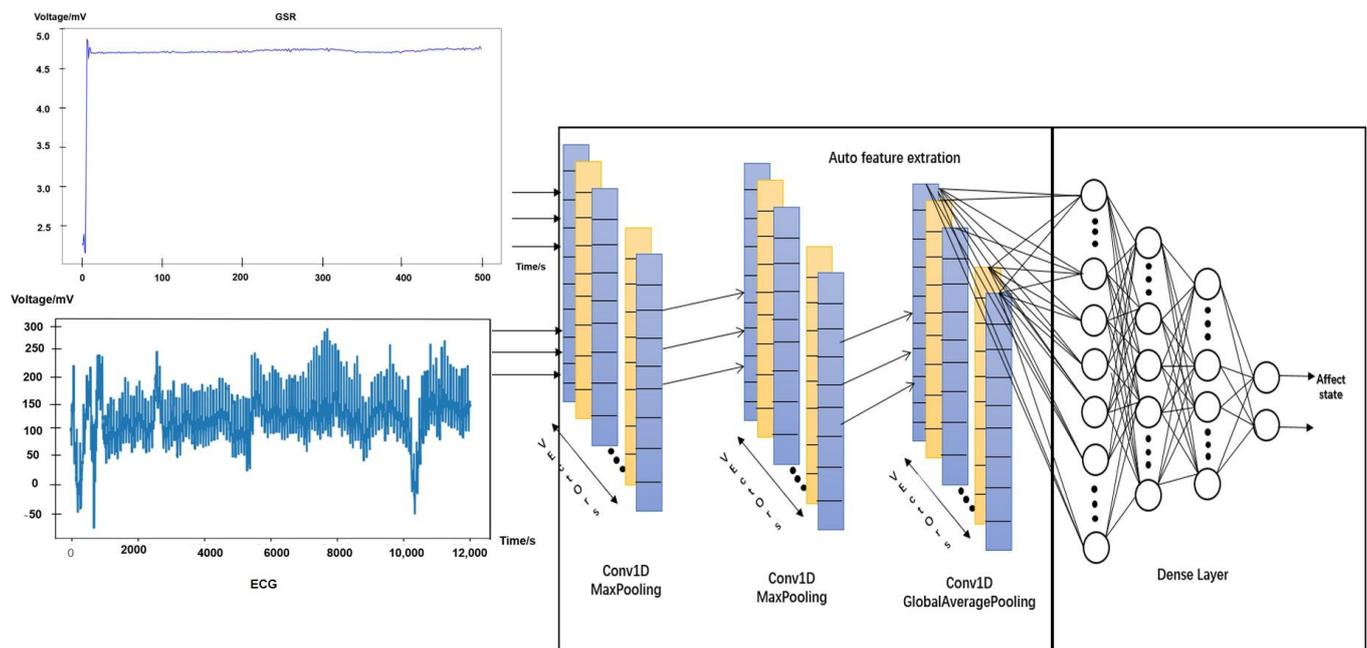


Figure 2. Schematic diagram of deep convolutional neural network.

The maxpooling layer alternates between CNNs as a regularization technique to reduce transition fitting in neural networks, and finally, to evaluate sentiment recognition. A cross-entropy loss function is set in the fully connected layer, which determines how well the target output vector y_i correspond to the predicted output vector c_j , as shown in Equation (2).

$$E = \frac{1}{2} \sum_{j=1}^N (y_i - c_j)^2, \tag{2}$$

Our multi-task signals conversion recognition network consists of 3 convolutional blocks and 3 pooling layers. The convolutional layers are shared among different tasks, while the dense layers are task-specific, as shown in Figure 3. Each convolutional block consists of 2×1 D convolutional layers with ReLu activation function, and followed by a maximum pooling layer of size 8. In the convolutional layers, we gradually increase the number of filters from 32 to 64 and 128. After each convolution blocks, the kernel size decreases from 32 to 16 and 8, respectively. Finally, at the end of the convolutional layer, global maximum pooling is performed. The dense layer immediately following consists of 2 fully connected layers and 128 hidden nodes, followed by a sigmoid layer.

3.3.2. Machine Learning Methods

In order to design reliable emotion recognition systems, it is particularly important to select appropriate and effective signal features. When designing affective computing systems, one of the most important considerations for application functionality are their simplicity and acceptable computational speed, thus making them suitable for real-time applications. Therefore, we use simple time-domain and frequency-domain features that do not require complex transformations and heavy computations.

Most of the characteristics of the ECG signal are based on the analysis of the P, Q, R, S and T waves of the recorded signal, including several statistical features calculated from the amplitude and width of the P, Q, R, S and T wavelets. Subsequently, heart rate variability (HRV) is calculated based on the detected R peak, and further features are extracted from the resulting signal, including the mean and root-mean-square deviation from HRV. In addition, the slope of the linear regression fitted to the appearance of the R-peak was calculated IBI. Based on [46], wavelet transformed decomposition coefficients were also extracted, using 8th order Daubechies wavelets applied to detect and align the R-peaks.

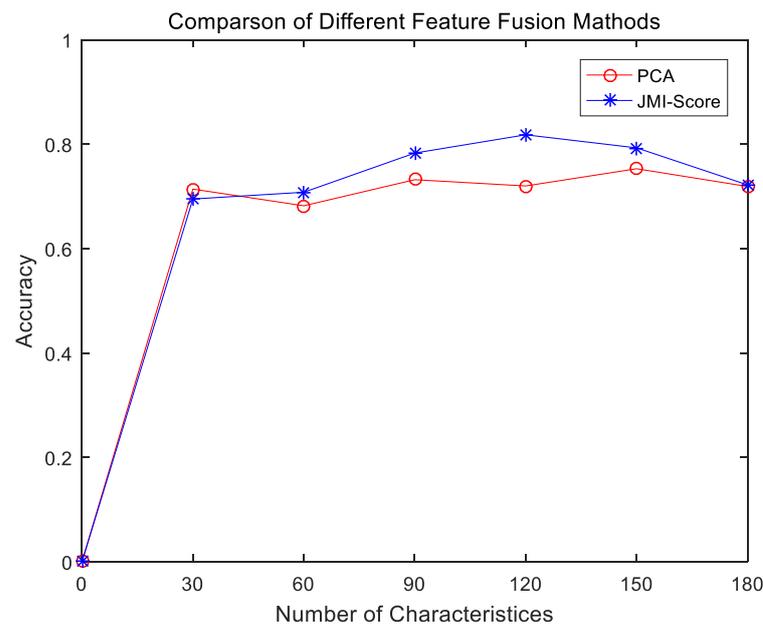


Figure 3. Feature layer fusion recognition accuracy of different feature dimensions.

For GSR, include features such as signal mean, standard deviation, kurtosis, or skewness (e.g., [47,48]). In other cases, researchers focused on event-related features of GSR. Event-related features refer to the properties of short-term responses, such as the presence or absence of an SCR, when seconds after the presence of a stimulus (such as an image or sound). In this sense, SCR can be automatically detected and features extracted from longer time windows. Phases skin conductance response (SCR) and the sum of SCR amplitude, SCR peak count, mean SCR rise time [49,50]. Furthermore, tonic skin conductance level (SCL). Power Spectral Density (PSD) estimation in the frequency domain using Welch's method, which is the most commonly used algorithms to obtain a frequency domain representation of the signal. Previous studies have considered the statistical aspects (variance, range, signal amplitude region, skewness, kurtosis, harmonic summation) and spectral power of the five frequency bands, as well as their minimum, maximum, and variance [51].

The physiological signal changes without a specific pattern and is highly random. Much of the information cannot be judged on the time domain, so it is also analyzed in the frequency domain. The signal frequency band is generally divided into very low frequency band (VLF = [0.0022–0.04] Hz), low frequency band (LF = [0.04–0.15] Hz) and high frequency band (HF = [0.15–0.40] Hz). The PSD method extracts the spectral power of each frequency band as the spectral characteristics of the original signal with the following equation.

$$\text{Power} = \int_{f_1}^{f_2} \text{PSD}(f)df, \quad (3)$$

Power calculated in the VLF, LF and HF band, total power in the entire frequency range (TP), power calculated in the power range LF band as a proportion of that calculated in the HF band (LF/HF). The proportion of power LF band calculated in the power range to that calculated in the whole band (LF/TF), LF power normalized to the sum of LF and HF power (nLF), and HF normalized to the sum of LF and HF power (nHF).

The nonlinear entropy domain feature, which can reflect the complexity and uncertainty of physiological signals, and has a wide range for applications in computational studies of emotions based on physiological signals. The extracted entropy values help to quantify the regularity of the signal, which can be applied to emotion recognition. This section applies three types of entropy domain features, including information entropy, multi scale entropy, and refined composite multi scale dispersion entropy (RCMDE) [52]. The extracted features are shown in the Table 1.

Table 1. Summary of extracted features.

Signal	Feature Group	Description of the Extracted Features
ECG	Time Domain	R, P, Q, S, T (mean, max, min, std), IBI (mean, max, min, Peaks, std, kurtosis, skewness), HRV (mean, std, kurtosis, skewness), RR interval mean, RR interval std
	Frequency Domain	PSD-VLF, PSD-LF, PSD-HF, TP, pLF, pHF, LFHF, LFHF, nHF, nLF for all values above(mean, min, max, range, median, std)
	NonLinear	Shannon Entropy, RCMDE, MSE
GSR	Time Domain	Mean, Max, Min, Std, 1Diff (abs, mean, std), 2Diff (abs, Mean, std), kurtosis, skewness, SCR RiseTime, SCR Width (mean, max, min, std), Skin Conductance (Mean, Max, Min, Std, 1Diff, 2Diff), SCSR, SCVSR (mean, std, 1Diff, 2Diff)
	Frequency Domain	PSD-VLF, PSD-LF, PSD-HF, TP, pLF, pHF, LFHF, LFHF, nHF, nLF for all values above (mean, min, max, range, median, std)
	NonLinear	Shannon Entropy, RCMDE, MSE

To sum up, there are 33 time domains, 60 frequency domains, and 3 nonlinear ECG signal features, 32 time domains, 60 frequency domains, and 3 nonlinear GSR signal features. The total number of physiological signal features per window are 191.

3.4. JMI-Based Greedy Feature Selection Algorithm (JMI-Score)

In the task of emotional feature classification and recognition, it is necessary to perform feature dimensional reduction processing on the obtained high-dimensional features, and to avoid overfitting caused by too high dimensional. Therefore, a greedy feature selection algorithm based on JMI is proposed here to select features, as shown in the Algorithm 1, the specific steps of the Joint Mutual Information (JMI)-based greedy feature selection algorithm proposed in this paper are shown below.

JMI Introduction

The mutual information is a measure X and Y between two (possibly multidimensional) random variables, which quantifies the amount of information about one random variable obtained through another random variable. The mutual information is given by the following equation:

$$I(x, y) = \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}, \tag{4}$$

x_i, y_j are the X, Y components, and with N and M values, respectively.

JMI provides the best trade-offs in terms of accuracy, stability, and flexibility based on two assumptions:

- (1) After removing a feature given the removed feature itself, any unselected feature is conditionally independent of the union of the selected features:

$$p(x_k; x_{s/i} | x_i) = p(x_k | x_i)p(x_{s/i} | x_j), \tag{5}$$

- (2) Any unselected feature is conditionally independent of the union of selected features after removing any feature of a given class label and the removed feature itself:

$$p(x_k; x_{s/i} | x_i, y) = p(x_k | x_i, y)p(x_{s/i} | x_i, y), \tag{6}$$

Assuming the above two, the JMI score of the feature x_k is obtained according to the mutual information formula,

$$J_{JMI}(X_k) = \operatorname{argmax}_{X_j \in S} \left(\sum I(X_k X_j; Y) \right), \tag{7}$$

This is the information between the target Y and the joint random variables $X_k X_j$, defined by pairing the candidate X_k with each of the previously selected features X_j . The candidate feature X_k that maximizes this mutual information is selected and added to the feature subset S .

The maximum joint mutual information is defined as: Let $F = \{f_1, f_2, \dots, f_N\}$ be the full feature set, let S is a subset of the selected features. Let $f_i \in F - S$, $f_s \in S$. MaxJMI is the maximum value of the joint mutual information shared with the class label C by the candidate feature f_i when each feature in the subset S is individually connected, therefore

$$\max_{s=1,2,\dots,k} I(f_i, f_s; C).$$

Algorithm 1 JMI-Score

Input:

All feature sets $F = \{f_1, f_2, \dots, f_N\}$, Classification Tags C , Number of features D , Simple classification model model

Selected feature set subscript: S .

JMI-Score (F, C, model, S, D):

```

1.  Score = []
2.  maxJMI = 0
3.  for i from 1 to D
4.    S[i] = [f1]
5.    Temp = S[i]
6.    for j from 1 to D
7.      S[i]. add (fj)
8.      Score[i] = model.fit(tmp, C)
9.      If Score[i] > Score[i - 1]
10.         S[i] = S[i]. add(fj)
11.      else
12.         S[i] = Temp
13.    End for
14.  End for
15.  Sort the Score, select the top ten largest, and record all subscript IDX
16.  for i from 1 to 10
17.    If MI < JMI(S[IDX[i]], C) :
18.      maxJMI = JMI(S[IDX[i]], C)
19.      Ans = IDX[i]
20.  S = S[Ans]

```

Output: S .

The algorithm first iterates through each feature, using a single feature as the starting set, and iterates through the features, other than the original features. Feeding the selected set of features into the pre-trained model, if the model score improves on this feature set, the newly traversed features are added to the feature set, otherwise they are not added. Therefore, all features are traversed, the set of features with the highest model score is selected, assuming that these feature combinations are most relevant to the labels and the features in each feature combination are complementary information. The ten features that make the highest model score are combined with the labels, and then to calculate the joint mutual information. Since JMI indicates the selection of the candidate selection features that maximize the cumulative sum of the joint mutual information, and the selected subset of features add them to the subset. The method performs well in terms of classification accuracy and stability. Therefore, the final optimal subset with the largest joint mutual information is selected and the algorithm ends.

The feature types proposed in the previous section are selected to help reduce the features used for feature processing. The final size of manual features is 123×528 .

3.5. Results and Verification

3.5.1. Feature Selection Algorithm Verification

All the extracted multimodal physiological features are subjected to Principal Component Analysis (PCA) feature dimension reduction, and then input to XGBoost for feature classification. A total of 10 folds of independent experiments are carried out, and the samples are randomly scrambled in each fold. Taking Valence as the classification label, the recognition accuracy of the two feature dimension reduction methods is

It can be seen from the Figure 3 that at the beginning of dimension reduction, the recognition effects of the two algorithms are not much different. With the decline in later features, the recognition rates of both have increased. For the PCA algorithm, the best recognition effect is when the feature dimension is 150, the recognition rate is 75.3%; for the JMI-Score feature selection algorithm. The best recognition effect is when the feature dimension is only 120, the recognition rate is 81.8%.

3.5.2. Model Validation

Table 2 illustrates the computational results of the AMIGOS dataset. In the second method, after comparison, it is concluded that the accuracy obtained by using the XGBoost algorithm is 81.8%, which is higher than other algorithms because XGBoost uses multi-classifier stacking, which can achieve a better classification effect.

Table 2. Classification accuracy of each model.

Model	CNN		SVC		XGBoost		Naive Bays	
Modality Accuracy	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
	81.3%	73.2%	70.2%	78.7%	71.0%	81.8%	54.7%	64.1%

From the Table 2, it can be concluded that, use of deep neural networks takes longer than machine learning methods, but due to its model characteristics, the accuracy is better than machine learning. Using the computing framework proposed in this paper has obvious advantages in reducing the running time of the model and determining the response delay rate, decentralization, and rational use of edge resources.

3.6. Accuracy Description

Table 3 shows the comparative results of studies similar to this study. The types of features, feature selection algorithms, and optimal model parameters proposed in this paper are extracted from physiological data, and their results are compared with other studies:

Table 3. Accuracy comparison with other similar studies.

Research	Model	Signal Type	Emotion Classification	Accuracy (AMIGOS)
[26]	SVM-RBF	GSR	2	65%(A) 83.9%(V)
[29]	SVM	EEG, ECG	2	68.8%(A) 67%(V)
[28]	DCNN	ECG, GSR	2	75%(A), 76%(V)
[52]	XGBoost	EEG, ECG, GSR	2	68%(A), 80%(V)
	Proposed Model	ECG + GSR	2	70.9%(A), 81.8%(V)

4. New Computing Architectures

New computing architecture to accelerate computing: Often when processing data, we rely too much on cloud servers, which wastes network bandwidth and consumes time. Thanks to the development of Tensor Processing Unit (TPU), which have become conveniently portable computing devices, we propose novel computing architectures to accelerate emotion recognition and shorten recognition time, in contrast with inputting features directly into the model, we used TPU.

The computing framework of this study includes three layers: terminal-side, edge-side, and cloud-side, which effectively integrate the computing resources of the terminal-side

and edge device, to make them work together to complete the computational process of deep learning. Achieve accelerated processing of data, while ensuring data security, user experience, and system availability. Reduce the latency of human-computer interaction, and decentralization. At the same time, effective and reasonable use of terminal-side idle computing resources, edge-side proximity computing resources.

Terminal-side: When the raw physiological data are obtained, run the pre-processing decision algorithm, including three values Computing Resource Utilization (CRU) as equation 8 (local-side computing resources, cloud-side current computing resources, and cloud-side predicted resource usage), when the terminal-side (CRU) is more than 0.7 then the raw physiological data will be directly uploaded to the cloud server, the pre-processing and algorithm decision will be run by the cloud, conversely, when the terminal-side computing resource is sufficient, the feature extraction in data pre-processing will be performed by the terminal-side.

$$CRU = \frac{(CPU \text{ Occupancy rate} + \text{Memory Occupancy})}{2 \times 100\%} \tag{8}$$

Terminal-side: On the edge-side, we deploy several feature selection algorithms to process features from deep learning or machine learning and pass the streamlined features to the cloud for model decisions

Cloud side: On the cloud side, we need to collect cloud server computing resources in three seconds, and use machine learning models to predict the resource occupation in the next time period, then calculate the average CRU, in addition to deploying corresponding decision models such as CNN, XGBoost.

The data flow is shown in the Figure 4a. When the original data is on the terminal-side, the data pre-processing process will have two cases: when (computing resources) are sufficient, the data pre-processing is performed on the terminal-side, and then passed into the edge-side for feature selection, and finally into the cloud model to produce results; When (capacity value) is insufficient, the data pre-processing is performed directly on the cloud side, after feature selection on the edge-side, and finally the decision is made in the cloud without the participation of the end-side.

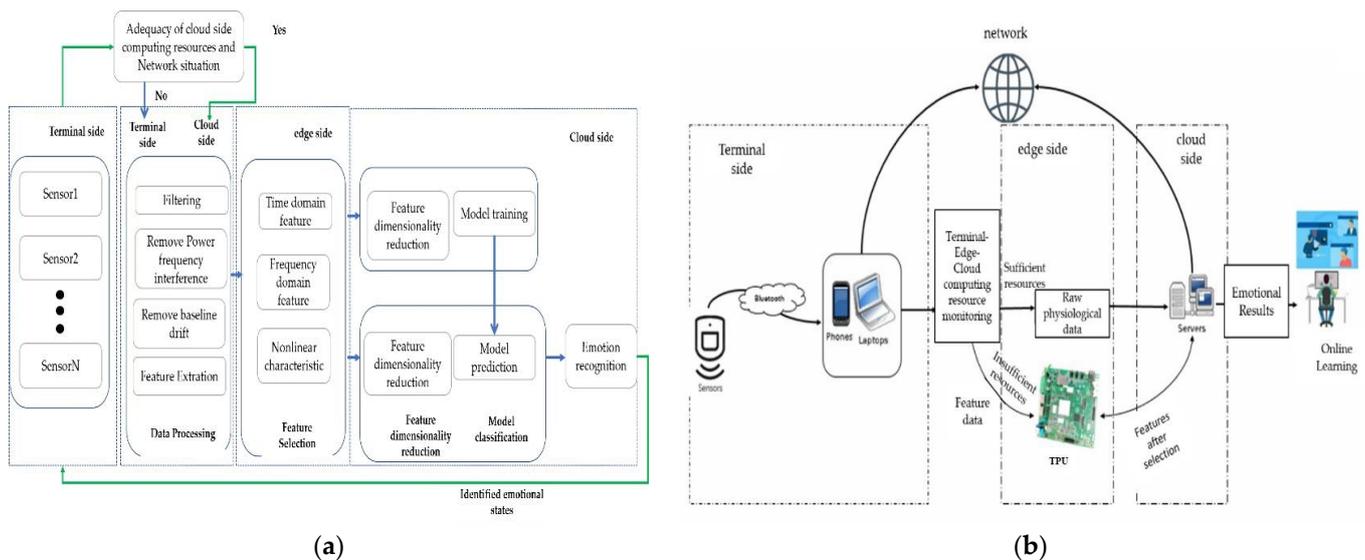


Figure 4. (a) Novel Affective computing Collaboration Architecture (b) Novel Computing Architecture Data Flow.

As Figure 4b shows that the edge side is based on the network situation and cloud-side computing resources to decide whether to participate in affective computing, if the cloud-side computing capacity is sufficient, it is Faster processing directly on the cloud side

as opposed to going back and forth between the cloud and the edge, but the cloud side is often heavily loaded, the edge side is taken into account, so that the cloud side and the edge side can compute together. The edge side will run the feature selection algorithm and input the selection results to the cloud side.

The Table 4 shows the time elapsed between data collection and input into the pre-trained model, when analyze the sentiment results in the same network environment, and shown based on the proposed computing architecture and in the same hardware environment (The hardware configuration is shown in Table 5), the time required to perform the same emotional computing task in. In terms of time consumption, we use two parameters to measure, (1) Running time: The time it takes to obtain analysis results from raw data under the same network environment; (2) Determine Response Latency Rate (DRLR): In the case of the same emotion calculation time and network transmission time. Sentiment recognition takes up the percentage of time it takes to send from the sensor, send the sentiment data to the user within the fine edge of the network, and correctly identify it.

Table 4. The response times of our two calculation modes are compared, and the comparison results are as follows.

Exponential/Time(s)	CNN (200 epoch)		SVC		XGBoost		Naïve Bays	
Running on the cloud-edge-device system	Yes	No	Yes	No	Yes	No	Yes	No
Runtime(s)	18.74	21.58	11.31	14.36	17.65	19.36	10.53	13.24
Determine Response Latency Rate	80.4%	82.6%	60.7%	66.5%	75.6%	79.3%	53.1%	56.2%

Table 5. The experimental equipment is as follows.

Terminal side	Inter(R)Core(TM) i7-4790CPU, 3.60 GHz, 4 core, 8 GB RAM, NVIDIA GeForce GTX 1050Ti (4 GB)
Edge side	ARM Coretex A73 1.6 GHz, 2 core, ARM Mail G71@900 MHz (256 K)
Cloud side	Inter(R)Core(TM) i7-4790CPU, 3.60 GHz, 8 core, 16 GB RAM, NVIDIA GeForce GTX 1050Ti (4 GB)

From the data in the Table 4, we can see that the new computing mode can give feedback on emotional results faster. Compared with the traditional cloud-centric computing mode, the advantages of such computing are: 1. Speed up the operation without affecting the accuracy of the model; 2. It can not only ensure the security of data, but also realize a decentralized computing model, and make rational use of edge resources; 3. Reduce the use of network bandwidth, and innovatively integrate and use cloud and edge resources.

Online Learning Experiment

In order to verify the effectiveness of the computing architecture and algorithms, taking online learning as a scenario, collecting physiological data of students during online learning, and running with the new computing architecture proposed in this paper, considerable results were obtained, which is of great significance to the future online education and medical fields. We invited 30 subjects as shown in Figure 5b (age range 22–26 years), 17 males and 13 females, all of whom had received more than six years of formal EFL education. The experimental equipment is placed as in Figure 5a (Shimmer3 ECG device, E4 wristband, windows core i5, ASHU 603, Hi3559A TPU, ubuntu 32 G/4 T Server), before participating in the experiment, sign the required process description and give informed consent, and the acquisition process complies with the ethical requirements of the Human Biobanking Educational Exam. Establish the context as offline.

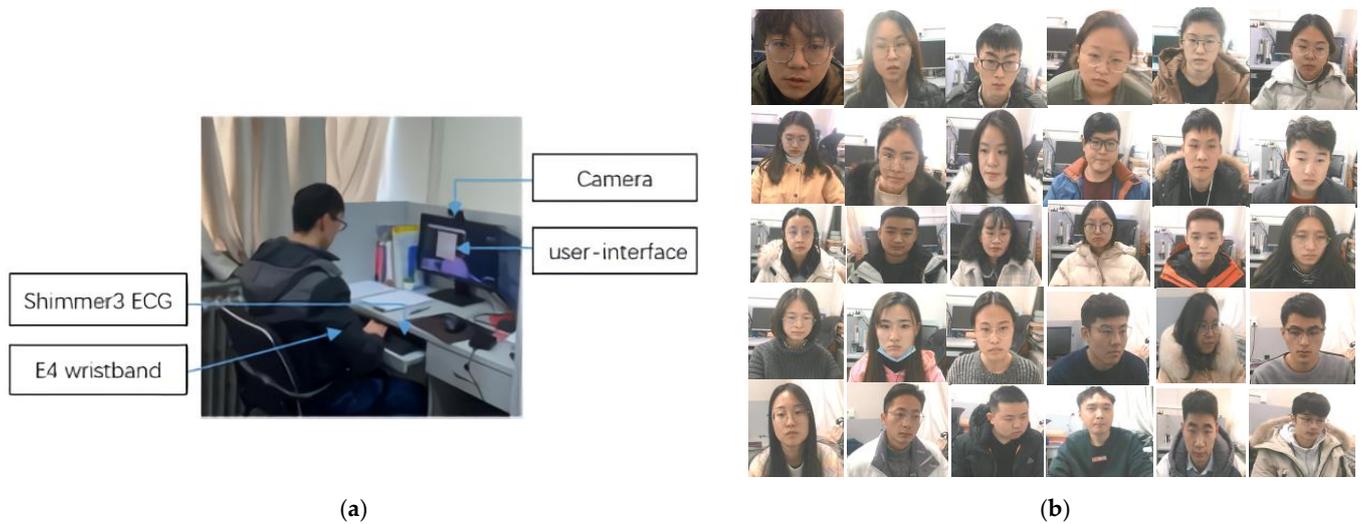


Figure 5. (a)Equipment distribution; (b) 30 subjects.

The experimental flow is shown in Figure 6, and the detailed procedure is as follows:

1. Make sure the subjects remain calm, take a five-minute baseline test, and fill in the familiarity of the test questions before the experiment, and evaluate the difficulty level of the test according to the familiarity;
2. Show multiple-choice questions to the subjects. After each answer, the participants self-assess their arousal level and valence, and the background selects the difficulty level of the next question according to the subject’s emotional score;
3. The test paper contains 30 questions, and 30 min of ECG (Shimmer3 ECG equipment) and GSR (E4 wristband) data are collected;
4. After the experiment, annotators performed annotations based on video clips, first for valence and then for arousal.

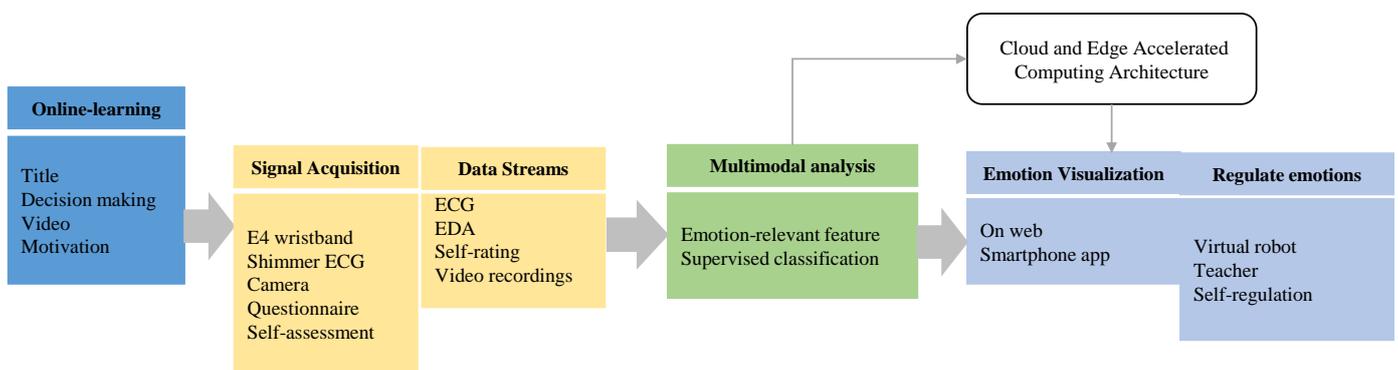


Figure 6. Experimental flow chart.

The collection frequency of the Shimmer3 ECG device is 256 Hz, the amount of ECG data collected is (13,824,000) per subject, the E4 wristband is used to collect GSR frequency of 4 Hz, and the quantity is (216,000) per subject. The emotion annotation includes user self-assessment Valence and Arousal external annotation. We performed variable statistical analysis on the collected data. The degree of influence between these variables was measured using the Pearson correlation coefficient, defined as:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \tag{9}$$

X carries the ECG or GSR physiological data vector and Y represents emotional decision making. The correlation between ECG and affective state is usually lower than the correlation between GSR and affective state, which proves that different subjects stimulates different control factors for affective state. The system can adjust the difficulty of the questions according to the emotions fed by different subjects. The scientific validity and rationality of analyzing affective states from ECG, GSR is illustrated according to Pearson coefficients in Figure 7.

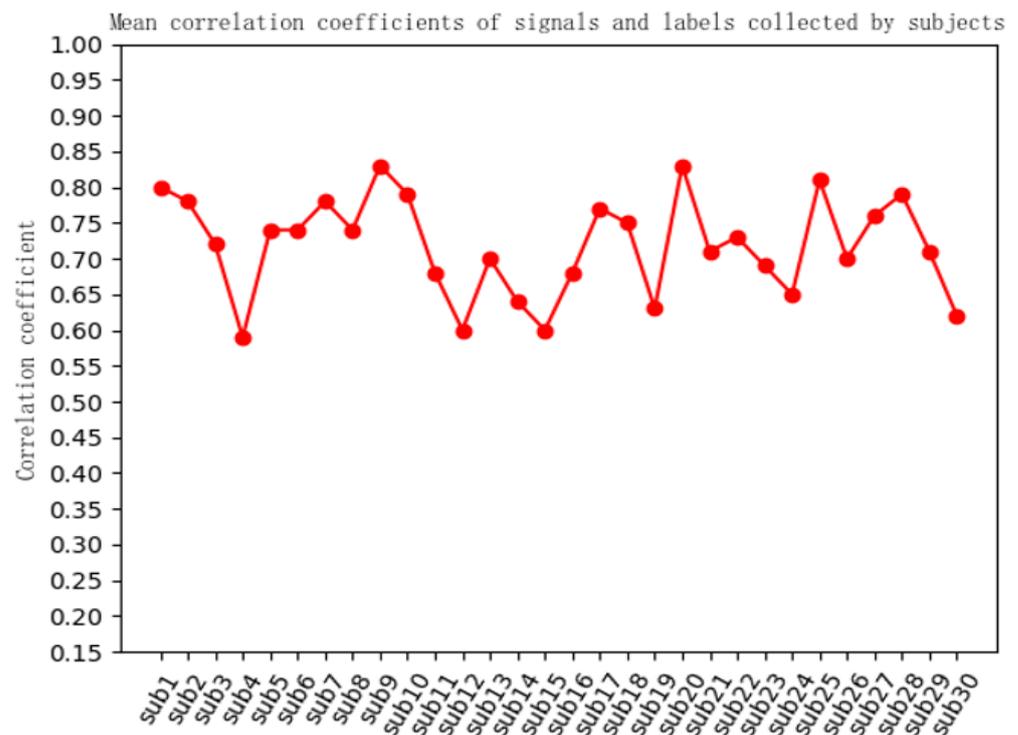


Figure 7. Overall mean correlation coefficient of different subjects.

We use the optimal model method proposed above to analyze the data. The XGBoost model has the highest accuracy rate of 80.6% for the second classification of arousal. Under the new computing architecture's operating model, Affective computing takes an average of 5 s less time than under the usual cloud-centric architecture.

5. Discussion

After analyzing the results, it was observed that the method using XGBoost performed better compared to the other method, for one reason: EEG, ECG and, GSR are continuous time signals with large memory content, manual feature engineering and, the better features can be obtained by using JMI-Score algorithm. The second reason are: machine learning can remove irrelevant features from feature sets, which deep learning cannot do. In this study, machine learning has great advantages. In addition to basic interpretability, the combined use of user device and edge device resources can accelerate computing, while affecting accuracy, reduce processing time, and achieve decentralization processing methods. The time reduction is not very significant. The reason for the analysis is that, the main purpose of the new computing architecture are to reduce the load on the cloud center, and effectively use the computing resources on the edge and the terminal, while the overall computing resources have not increased significantly. Of course, the advantages of deep learning are also obvious, which can avoid complex feature extraction, extract high-dimensional features, and obtain better results. In order to seek the choice of better features, this paper extracts many features, including time domain, frequency domain, and nonlinear features. According to the Spearman correlation coefficient, it is more stable in GSR. The ECG signal

has higher inter-class variability. According to the correlation coefficient, some features are low and the jump is serious, so it is very necessary for the feature selection of ECG.

Compare with other studies, in this experiment, the amount of data collected has increased, and the uniqueness of the decision labels needs to be further verified. The proposed method and framework are used to obtain promising results, which are expected to solve the problem in the epidemic era. The majority of teachers and students encounter the problem of interaction channels in distance education.

It is an experiment to move Affective computing based on physiological signals towards life. Of course, in this study, the research method was applied to the actual scene, and subjective factors such as subjects' different educational backgrounds, and different answering backgrounds were not considered. It is an important factor, but because it is difficult to express mathematically, it is not considered in training data and needs to be studied in the future.

The model selection of machine learning is also very critical. This article selects several representative models. JMI-Score is an iterative version of JMI, which is relatively new. XGBoost is a widely used stacking ensemble algorithm, which can solve the limitations of a single model. Naïve Bayes is a traditional basic algorithm model and the origin of machine learning, which is very representative. In this study, XGBoost performed better, indicating that it is more appropriate to use traditional optimized machine learning when the amount of data is not large.

6. Conclusions

This work shows that emotion recognition can be performed with high accuracy from ECG and GSR signals. In addition, using a new MSE-based feature RCMDE, we found that the derived features of GSR along with the energy, and zero-crossing rate of its EMD patterns, allows for the correct classification of target emotional states. For the GSR signal, its stability characteristics can be used to predict the stress value, while the ECG has a strong mutation, and its frequency characteristics are more important to emotion recognition. Several classification models are trained in the machine learning method to select the model that maximizes the accuracy. In practical applications, the emotion recognition model should not only focus on accuracy, but also on timeliness. Only faster feedback can improve the human interaction experience.

In this paper, the public multi-physiological signal database AMIGOS is used as the experimental data to perform preprocessing, feature extraction, and feature selection to verify the effectiveness of the method proposed in this paper. The stimulation materials and acquisition processed of the physiological signal dataset proposed in this paper are briefly introduced. The acquired dataset is used to verify the effectiveness of the proposed method in real scenarios. When analyzing physiological data, it is first proposed to use discrete wavelet analysis, butterworth filter and empirical mode decomposition method to denoise the data. The feature engineering is divided into two categories: manual feature extraction and deep network automatic extraction. Machine learning uses time domain, frequency domain and nonlinear feature analysis to perform traditional feature extraction for two physiological signals: ECG signal and GSR signal. With a 3 s sliding window. Deep network methods are automatically extracted with convolutional neural networks, but their interpretability is not high. The shallow emotional features extracted by machine learning and the deep emotional features obtained by deep learning are, respectively.

We present a novel computational framework for affective computing, and the proposed system helps make affective computing applicable to solve problems in our lives, and helps bridge the gap between the representation of low-level physiological signal sensors and high-level contextually relevant interpretations of human emotions. The experimental results obtained from the two optimal algorithms using public datasets show that our feature selection processes use the JMI-Scores algorithm proposed in this paper for feature selection, and the dimension reduction effects is obvious. Feature sets, model parameter are set better than outperform state-of-the-art recognition rates. In fact, we observed an average

0.85% improvement in accuracy. There is also an extensive analysis of feature selection, model selection, time dimensions. Physiology is processed separately with deep learning and machine learning. It turns out that after feature selection and parameter tuning, the two architectures based on new computing systems are effective in emotion recognition, that is, better than previous methods, and in time dimension, the computational space dimension has been optimized.

Future work includes optimizing protocols in cloud-side computing systems, taking more into account security, and coordination. Applying more intelligent algorithms to new computing architectures, and developing real-time sentiments detection of wearable systems.

Author Contributions: Conceptualization, M.C., Z.W. and L.X.; methodology, M.C.; software, M.C.; validation, M.C. and C.L.; formal analysis, L.X. and Z.W.; investigation, M.C.; resources, M.C.; data curation, M.C.; writing—original draft preparation, M.C.; writing—review and editing, M.C., L.X. and Z.W.; visualization, M.C.; supervision, M.C. and L.X.; project administration, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key R&D Program of China (Grant No. 2018YFC2001700) and Beijing Natural Science Foundation (No. L192005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly and only available upon reasonable request.

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

References

- Picard, R.W.; Vyzas, E.; Healey, J. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1175–1191. [\[CrossRef\]](#)
- Fairclough, S.H.J.I.W.C. Fundamentals of physiological computing. *Interact Comput.* **2009**, *21*, 133–145. [\[CrossRef\]](#)
- Nardelli, M.; Valenza, G.; Greco, A.; Lanata, A.; Scilingo, E.P. Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Trans. Affect. Comput.* **2015**, *6*, 385–394. [\[CrossRef\]](#)
- Sarkar, P.; Ross, K.; Ruberto, A.J.; Rodenbura, D.; Hungler, P.; Etemad, A. Classification of cognitive load and expertise for adaptive simulation using deep multitask learning. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction, Cambridge, UK, 3–6 September 2019.
- Nussinovitch, U.; Elishkevitz, K.P.; Katz, K.; Nussinovitch, M.; Segev, S.; Volovitz, B.; Nussinovitch, N. Reliability of Ultra-Short ECG Indices for Heart Rate Variability. *Ann. Noninvasive Electrocardiol.* **2011**, *16*, 117–122. [\[CrossRef\]](#) [\[PubMed\]](#)
- Greco, A.; Valenza, G.; Lanata, A.; Scilingo, E.P.; Citi, L. cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 797–804. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dawson, M.E.; Schell, A.M.; Fillion, D.L. *The Electrodermal System*; Berntson, G., Ed.; Cambridge University Press: Cambridge, UK, 2017; pp. 217–243.
- Boucsein, W. *Electrodermal Activity*, 2nd ed.; University of Wuppertal: Wuppertal, Germany, 2012; pp. 186–192.
- Käthele, M.; Schels, M.; Thiam, P.; Schwenker, F. Fusion mappings for multimodal affect recognition. In Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence, Mother City, South Africa, 8–10 December 2015.
- Käthele, M.; Thiam, P.; Palm, G.; Schwenker, F.; Schels, M. Ensemble methods for continuous affect recognition: Multi-modality, temporality, and challenges. In Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge, New York, NY, USA, 26 October 2015.
- Kessler, V.; Thiam, P.; Amirian, M.; Schwenker, F. Pain recognition with camera photoplethysmography. In Proceedings of the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 28 November 2017.
- Thiam, P.; Kessler, V.; Walter, S.; Palm, G.; Schwenker, F. Audio-visual recognition of pain intensity. In Proceedings of the IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction, Cancun, Mexico, 5 December 2016.
- Thayer, R.E. *The Biopsychology of Mood and Arousal*, 1st ed.; Oxford University Press: Toronto, UK, 1992; pp. 65–67.
- Yik, M.; Russell, J.A.; Steiger, J.H. A 12-point circumplex structure of core affect. *Emotion* **2011**, *11*, 705–731. [\[CrossRef\]](#) [\[PubMed\]](#)
- Harper, R.; Southern, J. A Bayesian Deep Learning Framework for End-To-End Prediction of Emotion from Heartbeat. *IEEE Trans. Affect. Comput.* **2020**, *1*, 1–8. [\[CrossRef\]](#)
- Rouast, P.V.; Adam, M.T.P.; Chiong, R. Deep Learning for Human Affect Recognition: Insights and New Developments. *IEEE Trans. Affect. Comput.* **2019**, *12*, 524–543. [\[CrossRef\]](#)

17. Kim, B.H.; Jo, S. Deep Physiological Affect Network for the Recognition of Human Emotions. *IEEE Trans. Affect. Comput.* **2018**, *11*, 230–243. [[CrossRef](#)]
18. Anandhi, B.; Jerritta, S.; Anusuya, I.; Das, H. Time Domain Analysis of Heart Rate Variability Signals in Valence Recognition for Children with Autism Spectrum Disorder (ASD). *IRBM* **2021**, *681*, 118–129. [[CrossRef](#)]
19. Wen, W.-H.; Liu, G.-Y.; Mao, Z.-H.; Huang, W.-J.; Zhang, X.; Hu, H.; Yang, J.; Jia, W. Toward Constructing a Real-time Social Anxiety Evaluation System: Exploring Effective Heart Rate Features. *IEEE Trans. Affect. Comput.* **2018**, *11*, 100–110. [[CrossRef](#)]
20. Gedik, E.; Cabrera-Quiros, L.; Martella, C.; Englebienne, G.; Hung, H. Towards Analyzing and Predicting the Experience of Live Performances with Wearable Sensing. *IEEE Trans. Affect. Comput.* **2018**, *12*, 269–276. [[CrossRef](#)]
21. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [[CrossRef](#)]
22. Zheng, W.-L.; Lu, B.-L. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [[CrossRef](#)]
23. Miranda-Correa, J.A.; Abadi, M.K.; Sebe, N.; Patras, I. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Trans. Affect. Comput.* **2018**, *12*, 479–493. [[CrossRef](#)]
24. Zheng, W.L.; Zhu, J.Y.; Peng, Y.; Lu, B.L. EEG-based emotion classification using deep belief networks. In Proceedings of the 2014 IEEE international conference on multimedia and expo (ICME), Chengdu, China, 14 July 2014.
25. Kawde, P.; Verma, G.K. Multimodal affect recognition in V-A-D space using deep learning. In Proceedings of the 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Chicago, IL, USA, 17–19 August 2017.
26. Shukla, J.; Barreda-Angeles, M.; Oliver, J.; Nandi, G.C.; Puig, D. Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. *IEEE Trans. Affect. Comput.* **2019**, *12*, 857–869. [[CrossRef](#)]
27. Cheng, B.; Liu, G. Emotion recognition from surface EMG signal using wavelet transform and neural network. In Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE), Tianjin, China, 16–18 May 2008.
28. Santamaria-Granados, L.; Munoz-Organero, M.; Ramirez-Gonzalez, G.; Abdulhay, E.; Arunkumar, N. Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS). *IEEE Access* **2018**, *7*, 57–67. [[CrossRef](#)]
29. Yang, H.C.; Lee, C.C. An Attribute-invariant Variational Learning for Emotion Recognition Using Physiology. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12 May 2019.
30. Li, C.; Bao, Z.; Li, L.; Zhao, Z. Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. *Inf. Process. Manag.* **2020**, *57*, 102185. [[CrossRef](#)]
31. Lan, Z.; Sourina, O.; Wang, L.; Liu, Y. Real-time EEG-based emotion monitoring using stable features. *Vis. Comput.* **2015**, *32*, 347–358. [[CrossRef](#)]
32. Siddharth, S.; Jung, T.-P.; Sejnowski, T.J. Utilizing Deep Learning towards Multi-Modal Bio-Sensing and Vision-Based Affective Computing. *IEEE Trans. Affect. Comput.* **2019**, *13*, 96–107. [[CrossRef](#)]
33. Bao, L.; Intille, S.S. Activity Recognition from User-Annotated Acceleration Data. In *Pervasive Computing*; Springer: Berlin, Germany, 2004; pp. 1–17. [[CrossRef](#)]
34. Ho, J.J.C. Interruptions: Using activity transitions to trigger proactive messages. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2004.
35. Sekine, M.; Tamura, T.; Fujimoto, T.; Fukui, Y. Classification of walking pattern using acceleration waveform in elderly people. In Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 23 July 2000.
36. Lin, J.; Keogh, E.; Lonardi, S.; Chiu, B. A symbolic representation of time series, with implications for streaming algorithms. In Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, New York, NY, USA, 13 June 2003.
37. Chen, M.; Zhou, P.; Wu, D.; Hu, L.; Hassan, M.M.; Alamri, A. AI-Skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework. *Inf. Fusion* **2020**, *54*, 1–9. [[CrossRef](#)]
38. Metwaly, A.; Queralta, J.P.; Sarker, V.K.; Gia, T.N.; Nasir, O.; Westerlund, T. Edge computing with embedded ai: Thermal image analysis for occupancy estimation in intelligent buildings. In Proceedings of the INTelligent Embedded Systems Architectures and Applications Workshop 2019, New York, NY, USA, 13–18 October 2019.
39. Li, E.; Zhou, Z.; Chen, X. Edge intelligence: On-demand deep learning model co-inference with device-edge synergy. In Proceedings of the 2018 Workshop on Mobile Edge Communications, New York, NY, USA, 20 August 2018.
40. Chanel, G.; Rebetz, C.; Bétrancourt, M.; Pun, T. Emotion Assessment from Physiological Signals for Adaptation of Game Difficulty. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2011**, *41*, 1052–1063. [[CrossRef](#)]
41. Wiem, M.B.H.; Lachiri, Z. Emotion classification in arousal valence model using MAHNOB-HCI database. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 318–323. [[CrossRef](#)]
42. Morris, J.D. Observations: SAM: The Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. *J. Advert. Res.* **1995**, *35*, 63–68.
43. Dar, M.N.; Akram, M.U.; Khawaja, S.G.; Pujari, A.N. CNN and LSTM-Based Emotion Charting Using Physiological Signals. *Sensors* **2020**, *20*, 4551. [[CrossRef](#)] [[PubMed](#)]

44. Demmel, J.; Kahan, W. *Computing Small Singular Values of Bidiagonal Matrices with Guaranteed High Relative Accuracy: LAPACK Working Note Number 3*; Argonne National Lab.: DuPage County, IL, USA, 1988.
45. Basheera, S.; Ram, M.S.S. A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI. *Comput. Med. Imaging Graph.* **2020**, *81*, 101713. [[CrossRef](#)]
46. Zhao, Q.; Zhang, L. ECG feature extraction and classification using wavelet transform and support vector machines. In Proceedings of the 2005 International Conference on Neural Networks and Brain, Beijing, China, 13 October 2005.
47. Kim, J.; Andre, E. Emotion recognition based on physiological changes in music listening. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 2067–2083. [[CrossRef](#)]
48. Wang, J.; Gong, Y. Recognition of multiple drivers' emotional state. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8 December 2008.
49. Zhai, J.; Barreto, A.; Chin, C.; Li, C. Realization of stress detection using psychophysiological signals for improvement of human-computer interactions. In Proceedings of the IEEE SoutheastCon, Ft, Lauderdale, FL, USA, 8–10 April 2005; pp. 415–420. [[CrossRef](#)]
50. O'Connell, R.G.; Bellgrove, M.; Dockree, P.M.; Lau, A.; Fitzgerald, M.; Robertson, I.H. Self-Alert Training: Volitional modulation of autonomic arousal improves sustained attention. *Neuropsychologia* **2007**, *46*, 1379–1390. [[CrossRef](#)]
51. Azami, H.; Rostaghi, M.; Abasolo, D.; Escudero, J. Refined Composite Multiscale Dispersion Entropy and its Application to Biomedical Signals. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2872–2879. [[CrossRef](#)]
52. Wang, S.H.; Li, H.T.; Chang, E.J.; Wu, A.Y.A. Entropy-assisted emotion recognition of valence and arousal using XGBoost classifier. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Rhodes, Greece, 25 May 2018.