IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 16, 2019, accepted February 17, 2019, date of publication February 27, 2019, date of current version March 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901950

Multimodal Mild Depression Recognition Based on EEG-EM Synchronization Acquisition Network

JING ZHU, YING WANG, RONG LA, JIAWEI ZHAN, JUNHONG NIU, SHUAI ZENG, AND XIPING HU

Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China

Corresponding author: Xiping Hu (huxp@lzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61632014, Grant 61210010, and Grant 61402211, in part by the National Basic Research Program of China (973 Program) under Grant 2014CB744600, in part by the Fundamental Research Funds for the Central Universities under Grant lzujbky-2017-it74 and Grant lzujbky-2017-it75, in part by the International Cooperation Project of Ministry of Science and Technology under Grant 2013DFA11140, and in part by the Program of Beijing Municipal Science and Technology Commission under Grant Z171100000117005.

ABSTRACT In this paper, we used electroencephalography (EEG)-eye movement (EM) synchronization acquisition network to simultaneously record both EEG and EM physiological signals of the mild depression and normal controls during free viewing. Then, we consider a multimodal feature fusion method that can best discriminate between mild depression and normal control subjects as a step toward achieving our long-term aim of developing an objective and effective multimodal system that assists doctors during diagnosis and monitoring of mild depression. Based on the multimodal denoising autoencoder, we use two feature fusion strategies (feature fusion and hidden layer fusion) for fusion of the EEG and EM signals to improve the recognition performance of classifiers for mild depression. Our experimental results indicate that the EEG-EM synchronization acquisition network ensures that the recorded EM and EEG data require that both the data streams are synchronized with millisecond precision, and both fusion methods can improve the mild depression recognition accuracy, thus demonstrating the complementary nature of the modalities. Compared with the unimodal classification approach that uses only EEG or EM, the feature fusion method slightly improved the recognition accuracy by 1.88%, while the hidden layer fusion method significantly improved the classification rate by up to 7.36%. In particular, the highest classification accuracy achieved in this paper was 83.42%. These results indicate that the multimodal deep learning approaches with input data using a combination of EEG and EM signals are promising in achieving real-time monitoring and identification of mild depression.

INDEX TERMS EEG, eye movement, mild depression, network, classification, multimodal deep learning.

I. INTRODUCTION

Depression is one of the most common mental illnesses, affecting more than 350 million people worldwide [1]. The World Health Organization (WHO) lists depression as the fourth most significant cause of disability in the world, in particular, depression increases the risk of suicide by about 20 times, resulting in up to 850,000 deaths per year [2]. In addition, new data indicate that the prevalence of depression may be on the rise, especially among college students [3]. Ibrahim et al. suggested that the depression rate in college students was between 10% and 85% [4]. However, although depression is a common mental illness, its diagnosis is difficult owing to subjective biases associated with self-reports and clinical opinions [5]. There is no objective method to diagnose depression because of the absence of dedicated laboratory tests to do so. It has been observed that, in the daily life of individuals, mild depression is more common than depression, and increase in severity over time [6]. Nevertheless, compared with depression, researchers have paid less attention to studies on mild depression [2]. To our knowledge, only few studies have provided effective detection methods for mild depression [7]. Thus, developing a method that can objectively and effectively detect mild depression in individuals to help them manage it by taking precautions and to avoid it from evolving into major depression is an urgent requirement.

The associate editor coordinating the review of this manuscript and approving it for publication was Yin Zhang.

A. RELATED WORK

At present, computer scientists globally are increasingly interested in using physiological signals for depression recognition [8]. Electroencephalography (EEG), an objective and reliable method for the evaluation of brain function, is often used in depression [9]. The advantages of EEG include high sensitivity, relatively low-cost, and convenience of recording [7]. Considering this, recently, several researchers have explored the use of EEG for depression recognition. Bachmann and Lass studied depression detection based on the analysis of single channel short-term EEG signals; in this study, the researchers achieved a classification accuracy of 76.5% and 70.6% using the spectral asymmetry index (SASI) and detrended fluctuation analysis (DFA), respectively [10]. Hosseinifard et al. used the power spectrum of three frequency bands (alpha, beta, and theta) as well as whole bands of the EEG signals as features; they studied the performance of different classification techniques to identify depression patients from normal subjects. Their results indicated that classification accuracies of 71.7% and 88.6% can be achieved using the Support Vector Machines (SVM) approach without and with feature selection, respectively [11].

Furthermore, aside from EEG signals, eye movements (EM) data can be used to identify the focus of users' attention, in order to determine their subconscious behaviors [12]. Moreover, in recent times, EM data are more readily available and accessible than in the past; not only are EM data being used in several areas of medicine, but also their popularity is growing among researchers from different disciplines [13]. Emslie *et al.* [14] studied the EM signals of depressed children and normal control; they observed that the rate of EM is slower in children with depression than in normal control. Duque and Vázquez [15] found a dual attention bias in clinical depression patients when watching positive and negative emotional faces. Alghowinem *et al.* [16] used simple machine learning classification algorithms to analyze the EM signals extracted from face videos; their results showed that using the low-level features of EM led to an accuracy of 70% when a hybrid classifier of Gaussian Mixture Models and SVMs were used, whereas an accuracy of 75% was achieved when using statistical measures with SVM classifiers.

Although most researchers have studied single modality, there is increasing interest in using different modalities to handle information. Gupta *et al.* [17] suggested that the signal from single modality provided only partial information, while a combination of different modality signals can be used to form a more realistic model for recognizing depression than the former. Said *et al.* [18] exploited the intra- and inter-correlation among multiple modalities to achieve efficient classification. Furthermore, many researchers have analyzed data of different modalities using deep learning approaches to exploit the correlation of data from the multiple modalities [18]. Said *et al.* [18] and Ngiam *et al.* [19] proposed a multimodal deep learning approach for cross modality feature learning from video and speech data. Said *et al.* [18] and Srivastava *et al.* [20] developed a multimodal deep belief network to learn multimodal representation from image and text data for image annotation and retrieval tasks. In addition, in another study, researchers designed a deep Boltzmann machine based architecture to extract a meaningful representation from multimodal data for classification and information retrieval tasks [18], [21]. Said *et al.* [18], Hinton and Salakhutdinov [22], and Liu *et al.* [23] proposed a multimodal autoencoder approach for video classification based on audio, image, and text data, and Said *et al.* [18] proposed a multimodal autoencoder approach for joint EEG-EMG data compression and classification.

Further, multimodal depression detection has also attracted significant attention from researchers [24], [25]. Several studies have been conducted to identify depression based on voice [26]–[28], event-related potential [29], facial expressions [30], [31], EEG [7], [32] and EM [3], [33]. Though these data might seem quite different, these can be used to describe the same phenomena [18]. For example, in case of a depressed person, when a stimulus is presented, voice data showed a longer response time and lower pronunciation rate, while EM data showed increased blink rate and longer average blink duration [34]; thus, it is considerably likely that both modalities are correlated. Nevertheless, each modality has its own advantages. It seems obvious that multimodal fusion of different modalities can improve classification performance, because it provides more useful information compared with using only a single modality [34]. Williamson *et al.* [35] fused speech features with facial action unit features using the score fusion method, which yielded good results for predicting depression severity in patients. Scherer *et al.* [36] proposed a depression recognition method based on fusion of audio and visual features; their results indicated that the fused modalities approach significantly outperformed the use of individual modalities, resulting in a 90% accuracy (compared with the individual accuracies of 51% and 64% for acoustic and visual modalities, respectively). Meng *et al.* [37] investigated the fusion of facial and vocal expressions, and used a weighted sum decision fusion; their result showed a slight improvement compared with individual channels.

B. OUR WORK

To our knowledge, there is no prior work reported in the literature related to depression recognition based on multiple physiological signals using multimodal deep learning. In addition, although there is a fast growing interest in the use of co-registration of EEG and EM during free viewing, few researchers focus on the issues that stem from the temporal alignment of EEG and EM data recorded with different devices. In fact, accurate time synchronization is a basic requirement for simultaneous analysis of EEG and EM data. In this study, we used an EEG-EM synchronization acquisition network that allowed us to simultaneously record both the EM and the EEG physiological signals of mild depression and normal controls during free viewing.

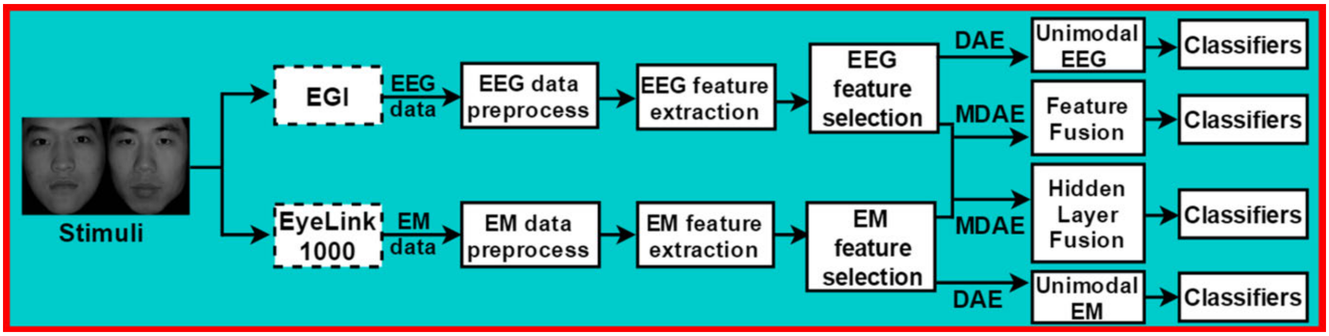


FIGURE 1. **The framework of our experiment processing.**

The long-term objective of our proposed approach based on the EEG-EM synchronization acquisition network is to offer a promising non-invasive method for automatic mild depression assessment and develop an objective and effective multimodal system in a classification framework to assist medical personnel during the diagnosis and monitoring of mild depression. In particular, we propose a new multimodal data based mild depression recognition method using deep learning techniques. There are four primary contributions of our study. First, we used the EEG-EM synchronous acquisition network to ensure that the EEG and EM data recorded simultaneously during the experiment were synchronized with millisecond precision, which is the basis for meaningful analysis of EEG and EM data. Second, we innovatively used signal processing methods to process the EM feature of pupil size, which was divided into five frequency bands, namely delta (0 to 0.2 Hz), theta (0.2 to 0.4 Hz), alpha (0.4 To 0.6 Hz), beta (0.6 to 0.8 Hz) and gamma (0.8 to 1 Hz), and for each band, we extracted 14 features (12 non-linear and 2 linear features). Third, in regard to the multimodal autoencoder, we studied two feature fusion methods (Feature Fusion and Hidden Layer Fusion) to achieve the fusion of EEG and EM, and compared the differences in the improvement of the classification performance of the two fusion methods. Finally, we used the five bands (delta, theta, alpha, beta, and gamma) as well as the whole band of EEG signals and discussed the improvement of the classification results on using each band. In addition, we used different feature selection algorithms and different classification algorithms for data processing and compared the classification accuracies in the case of different classification algorithms.

The remainder of this paper is organized as follows. In Section 2, we describe the autoencoder, denoising autoencoder, and multimodal autoencoder. Section 3 presents our EEG-EM synchronization acquisition network, experimental design, data pre-processing, feature extraction, feature selection, and classification algorithms. The experimental results and our analyses are discussed in Section 4. Section 5 includes an overall discussion of our study. Finally, our conclusions and information regarding future work are provided in Sections 6 and 7, respectively.

II. METHODS

The framework of our experiment processing is shown in Figure 1. For EEG and EM data, we performed preprocessing, feature extraction, feature selection operations. In addition, we considered the classification of unimodal EEG and unimodal EM based on autoencoder, and applied the two fusion strategies (Feature Fusion and Hidden Layer Fusion) combining EEG signals and EM data based on multimodal denoising autoencoder.

A. DATA PREPROCESSING

In our study, the EEG data were recorded using Geodesic 128 electrodes; however, owing to time performance and computational efficiency considerations, we only used 16 electrodes (Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2) according to International Standard 10/20 systems in reference to Cz. In addition, the selection of these electrodes was based on previous studies related to depression in which these electrodes were used extensively [11], [38], [39]. The recorded EEG data were exported to a MATLAB file format using Net Station software. Next, we removed the bad channels and performed baseline corrections. Given that the signal between two trials was not valid, each subject's continuous EEG signals were divided into 30 segments based on the TTL marks in time series, with each segment being 6 s long. For noise reduction, all EEG signals were filtered using a high-pass filter with a cut-off frequency of 1 Hz and a low-pass filter with a cut-off frequency of 40 Hz. Net Station Waveform Tools were used to discard artifacts due to EM and muscle activity. Ocular artifacts (OAs) occur in the frequency band of 0–16 Hz, leading to their overlap with the alpha rhythm frequency band of 8–13 Hz. FastICA [40] was used to eliminate these OAs, as it has been shown to be effective in delineating overlapping frequency bands [41]. MATLAB R2010a was the data processing tool used in our study.

EM data were collected by the EyeLink 1000 Desktop Eye Tracker with a remote camera (SR Research, Ontario, Canada, 250 Hz). The EM data preprocessing methods included performing two classic data mining methods, filling missing data, and data standardization. For example, if the participant did not focus on the picture or blinked too frequently, the eye tracking devices might lose capture and the recorded data will have tuples with empty values; however, there are different strategies to handle tuples with empty values. If a tuple contains many attributes (more than 50% of the attributes) with missing values, the tuple will be deserted. Nevertheless, if a tuple contains only a few attributes (less than 50% of the attributes) with missing values, the empty value is filled by the mean or median. In practice, the average value can be used in case of symmetric data distribution, whereas the median value can be used in the case of skewed data distribution [42]. In our study, records with more than 50% of the values missing were abandoned; in addition, the mean value was considered as the missing value for data completion. Furthermore, the measurement unit used can affect data analysis. In general, expressing an attribute in smaller units will lead to a larger range for that attribute, leading to a higher ''weight'' or ''effect'' of such an attribute. To avoid dependence on the selection of measurement units, the data should be normalized. Here, we used the z-score normalization method to do so before the analysis.

B. FEATURE EXTRACTION

After the raw EEG data were segmented, a Hanning filter was used to filter out five frequency bands, namely delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–31Hz), and gamma (31–40 Hz) for further feature extraction. In particular, we calculated 10 linear features such as maximum power, variance, and sumpower. In addition, 12 nonlinear features were extracted based on previous studies: Approximate Entropy (ApEn), Lempel–Ziv. Complexity (LZC), Kolmogorov Entropy (Kol), Permutation Entropy (Per_en), Correlation Dimension (CD), Lyapunov Exponent (LLE), C0-complexity (C0), Singular-value Deposition Entropy (SVDen), Shannon Entropy, Min-entropy, Hartley Entropy, and Spectral Entropy [38], [39], [43]–[46]. Therefore, a total of 1760 features was considered (22 EEG features ×5 frequency bands ×16 electrodes).

For the preprocessed EM data, first, we used signal processing methods to handle the pupil size feature. In particular, we used the EDF2ASC software to convert EDF files (source file recorded by EyeLink 1000 Desktop Eye Tracker) to ASC files; to extract pupil size signal data in the form of ASC files. Thus, the preprocessing method used for pupil size signal data was the same as that used for EEG signals. In particular, we filtered out five frequency bands, namely delta (0–0.2 Hz), theta (0.2–0.4 Hz), alpha (0.4–0.6 Hz), beta (0.6–0.8 Hz) and gamma (0.8–1 Hz), and extracted 12 nonlinear features (consistent with EEG) and two linear features (power spectral density mean and power spectral density variance) for each band. Second, we exported 16 traditional features available in the EyeLink Data Viewer software, including blink_count, ave_blink_duration, and fixation duration average, among others. In particular, the EyeLink Data Viewer is a tool that allows users to display, filter, and create output reports from EyeLink 1000 EDF data files. Thence, we extracted a total of 86 EM features (5 frequency bands ∗14 features + 16 traditional features).

C. FEATURE SELECTION

Classifiers tend to yield unsuitable results when the number of training samples is less than the number of feature vectors [47]; thus, feature selection is used to overcome this issue. In our study, we have 1760-dimensional EEG features, therefore, it is necessary to perform feature selection. Based on some previous studies, we used five common search algorithms implemented in WEKA (version 3.8.1): BestFirst (BF) [48], GeneticSearch (GS) [49], RankSearch (RS) [50], LinearForwordSelection (LFS) [51], and GreedyStepwise (GSW) [49], based on correlation features selection (CFS) [52].

D. MULTIMODAL DENOISING AUTOENCODER

1) AUTOENCODER

An autoencoder (AE) is a special type of neural network, which generally comprises two parts, an encoder $h = f(Wx + b)$ and a decoder that produces a reconstruction $r = g(W'h + b')$. The encoder converts the data vector set x to a hidden representation h by activating the function f, and the decoder rebuilds the data r using the hidden function h. The autoencoder shown in Figure 3 consists of three layers. First, the data are fed to the input layer; then, the encoder converts the data vector to a hidden layer h by activating the function f. Finally, the decoder rebuilds the data r using the hidden function h in reconstruction layer [53].

2) DENOISING AUTOENCODER

The denoising autoencoder (DAE) was proposed by Vincent *et al.* [54] in 2008. It is based on the autoencoder; however, noise is added to the input data to prevent problems of overfitting. It is trained to predict the original undamaged data as the output. This approach leads to increased robustness and generalization in the learning model.

First, the initial input x is corrupted into x′ by mapping (1).

$$x' \sim qD(x'|x) \tag{1}$$

The corrupted input x′ is then mapped to the hidden layer, as in the case of the basic autoencoder. In particular, the corrupted input x′ is mapped to the hidden representation (2).

$$y = f_\theta\left(x'\right) = s(Wx' + b) \tag{2}$$

From this, we then reconstruct (3).

$$z = f_{\theta'} = s(W'x' + b) \tag{3}$$

The parameters θ and θ' are trained such that the average reconstruction error is minimized. The complete process in the case of a denoising autoencoder is shown in Figure 2 [53].

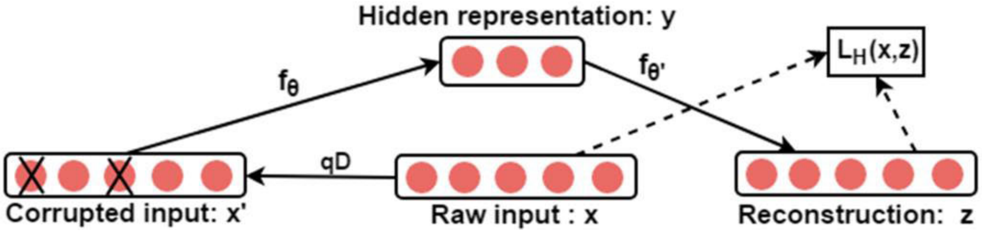


FIGURE 2. Schematic structure of a denoising autoencoder.

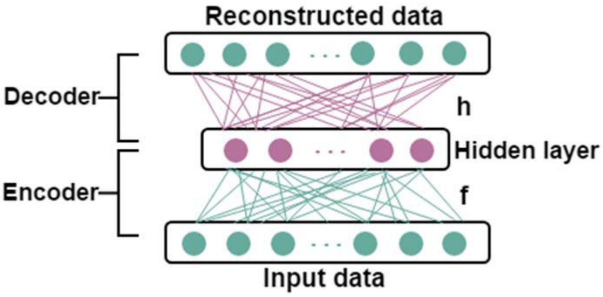


FIGURE 3. Schematic structure of an autoencoder.

3) MULTIMODAL DENOISING AUTOENCODER

In order to improve the recognition accuracy of mild depression by fusing EEG and EM data, we used a multimodal denoising autoencoder (MDAE) to learn a shared representation (high- level features) of EEG and EM data [55].

In particular, in the training step, we used two structures as shown in Figure 4. In the first structure, the features selected from both EEG signals and EM were directly linked in a record, and then input into the autoencoder to generate a shared representation (hidden layer). Finally, an unsupervised back-propagation algorithm was used to fine-tune the weights and biases of the autoencoder. In the second structure, first, we input the EEG and EM features individually into the autoencoder to generate two hidden layers. Then, we directly linked the EEG hidden layer with the EM hidden layer to synthesize a new shared representation. Finally, we used unsupervised back-propagation algorithms to fine-tune weights and biases of the autoencoder.

E. CLASSIFIERS

It is well known that there is no universal classification method that yields the best performance for all applications; therefore, it is often useful to consider different methods. In particular, we need to take into account the computation time, flexibility, and complexity of different classification methods, as well as the applications they were used for in other studies [7], [11], [39]. Therefore, we selected different classifiers to classify the data, namely the Linear SVM [56], Radial Basis Function SVM (RBF SVM) [56], Gradient Boosting Decision Tree (GBD tree) [57], Random Forest (RF) [58], Self Normalizing Neural Networks (SNN) [59], and Batch Normalized Multilayer Perceptron (BNMLP) [60].

In research involving the application of classification algorithms to the recognition of human mental states (such as emotions, mental disorders and motor imagery), two basic schemes for classification exist: subject-dependent and

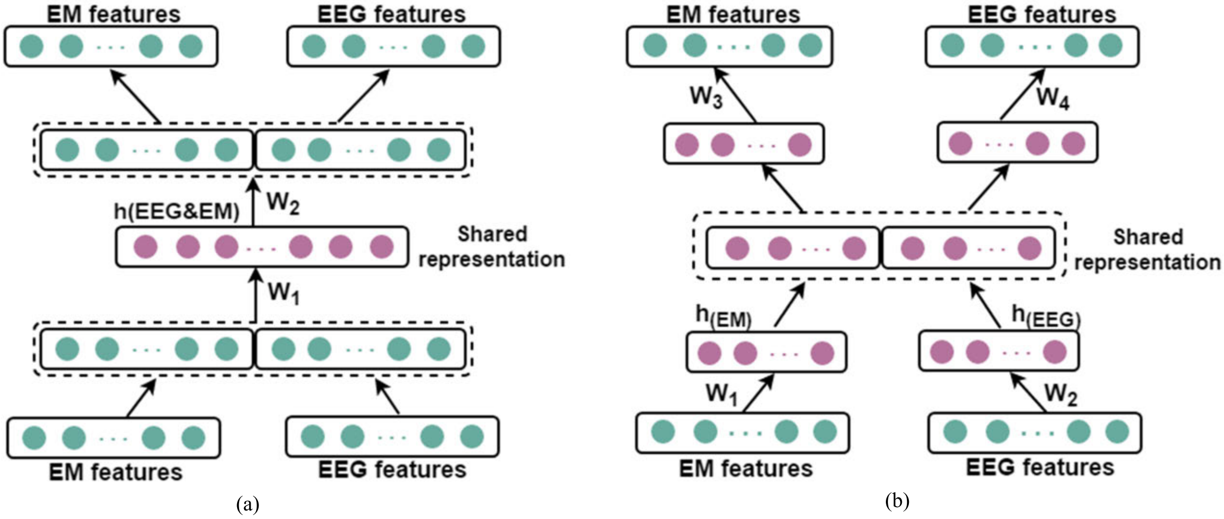


FIGURE 4. Schematic structure of a multimodal denoising autoencoder. (a) Structure of feature fusion. (b) Structure of hidden layer fusion.

subject-independent strategies. The subject-dependent algorithms require a classifier to be trained for each subject, whereas subject-independent algorithms train the classifier using data from several subjects [61]. It should be noted that depression recognition is considered a subject-independent classification case.

In regard to subject-independent classification, a crucial rule must be followed for the selection of training and test data. To eliminate the effect of individual difference on classification results, the training and test data need to be divided based on subjects, i.e., samples from the same subjects should not be used as both training data and test data, as this will lead to a falsely high classification accuracy.

Thus, in order to avoid a falsely high classification accuracy, we used a subject-independent scheme as well as the leave-one subject-out cross-validation method for classification.

III. EXPERIMENT

A. SUBJECTS

Fifty-one college students (36 males, 15 females) from Lanzhou University (Lanzhou, Gansu, China), aged between 18 and 24, participated in this study. All the participants were right-handed with normal or corrected-to-normal vision and had no prior history of psychopathology. Before the experiment, the participants were asked to complete the Beck Depression Inventory Test II (BDI-II) [62]. BDI-II is a widely used instrument that provides information on the presence and severity of depressive symptoms and can be used for diagnostic purposes, clinical decision making or evaluation of treatment effects. All subjects signed informed consent before the experiment and received rewards for participating in it. Our experiment was approved by the Ethics Committee of Lanzhou University Second Hospital (No. 2015 A -037).

In particular, 24 subjects (6 females, 18 males) with BDI scores of 14-28 were considered to have mild depression, while 27 subjects (9 females, 18 males) with BDI scores below 14 were considered normal. After removing some bad data due to EM calibration failure or head movement, we selected 19 subjects (14 males, 5 females) from the normal group and 20 subjects (14 males, 6 females) from the mild depression group to balance the sample size of the two groups. Basic data of the mild depressive and normal control groups are shown in Table 1.

TABLE 1. Basic information of the mild depression group and normal control group.

	Mild depression	Normal control
Cases(n)	20	19
age	21.1±1.95	20.11±2.07
BDI-II(means ± S.D)	18±3.56	4.74±3.04
Sex		
Male	14	14
Female	6	5

B. STIMULI AND DEVICES

In our work, the stimuli used in the experiment were derived from the Chinese Affection Image System (CFAPS) [63]. We selected 45 neutral faces and 15 negative pictures including 3 angry faces, 3 sad faces, 3 surprised faces, 3 disgusted faces, and 3 frightened faces, from the CFAPS.

EEG data were collected with a 128 channel HydroCel Geodesic Sensor Net (HCGSN) and used a signal amplifier provided by EGI. The data collection software was Net Station, with the sampling frequency set as 250 Hz, and electrode impedance maintained below 60 kΩ [64]. EM data were collected by the EyeLink 1000 Desktop Eye Tracker using a remote camera (SR Research, Ontario, Canada, 250 Hz). It should be noted that we only recorded the EM from the left eye of the subjects, because both eyes have the same movement pattern in the case of individuals without eye diseases.

In order to realize synchronous acquisition of EEG and EM data, we used the TTL signal pulse to send a corresponding TTL signal to Net Station when the EyeLink program executed certain steps.

C. EEG-EM SYNCHRONIZATION ACQUISITION NETWORK

In this study, we used an EEG-EM synchronization acquisition network that allowed us to simultaneously record both the EM and the EEG physiological signals of mild depression and normal controls during free viewing. However, meaningful analyses of simultaneously recorded EM and EEG data requires that both data streams are synchronized with millisecond precision, so the use of this method involves challenges such as precise synchronization between EM and EEG data. There are at least three ways to synchronize both systems, as shown in figure 5 [65].

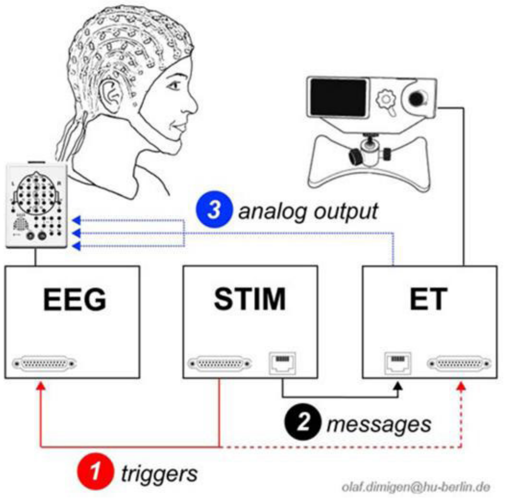


FIGURE 5. Three methods to synchronize both systems: 1. Shared triggers: trigger pulses are sent frequently from the stimulation computer to both eye tracking computer and EEG recording computer. 2. Messages + triggers: triggers are still sent to the EEG, and messages are used as the corresponding events for the eye tracking. 3. Analogue output: a copy of the eye track is fed directly into the EEG. A digital-to-analogue converter card in the Eye tracking outputs (some of) the data as an analogue signal. With SMI, this signal can be fed directly into the EEG headbox.

TABLE 2. Unimodal eeg classification results on neu_block.

	Accuracy % (mean ± std. dev.)					
Classifiers	Delta	Theta	Alpha	Beta	Gamma	All
Linear SVM	62.39±0.82	74.87±0.54	75.04±1.18	67.86±9.50	63.59±3.82	80±0.31
RBF SVM	47.69±0.83	73.33±14.97	76.41±0.81	74.19±2.55	59.32±0.38	**81.03**±1.02
GBD Tree	49.40±1.35	67.52±1.19	75.73±1.05	71.62±2.85	72.48±3.88	75.04±1.67
RF	51.28±1.25	66.67±0.70	75.21±0.67	70.09±3.48	67.52±3.35	72.48±0.86
SNN	55.04±1.36	68.72±10.10	68.21±1.61	68.72±3.99	55.73±1.08	75.38±1.70
BNMLP	53.33±2.84	58.46±6.72	68.72±2.93	68.72±3.33	60.17±3.39	79.32±1.57

Bold indicates the highest classification accuracy obtained among all the algorithms in the Neu_block

TABLE 3. Unimodal eeg classification results on emo_block.

	Accuracy % (mean ± std. dev.)					
Classifiers	Delta	Theta	Alpha	Beta	Gamma	All
Linear SVM	52.14±2.01	76.58±0.25	60.85±1.80	70.43±0.82	64.44±1.85	**76.92**±1.53
RBF SVM	41.71±0.97	75.56±0.11	57.09±1.87	70.26±0.77	66.15±0.74	65.81±0.71
GBD Tree	51.79±2.33	67.01±1.89	63.08±1.42	68.21±2.68	68.03±2.22	60.51±1.12
RF	52.14±0.62	65.47±1.03	63.93±0.95	66.67±1.75	66.50±1.54	59.83±1.32
SNN	50.94±1.91	73.33±0.52	59.49±2.62	69.40±2.61	68.21±1.17	56.41±1.66
BNMLP	50.09±2.19	72.82±1.21	52.82±2.23	65.81±2.24	67.01±2.63	59.15±1.69

Bold indicates the highest classification accuracy obtained among all the algorithms in the Emo_block

The first method is called "Shared triggers". Trigger pulses are sent frequently from the stimulation computer to both eye tracking computer and EEG recording computer. This is achieved via a Y-shaped cable that is attached to the parallel port of the stimulation computer and splits up the pulse so it is looped through to EEG and ET. The disadvantage of this method is the need for an extra cable. And the second method is called "Messages + triggers". Messages are short text strings that can be inserted into the eye tracking data. Triggers are still sent to the EEG, and messages are used as the corresponding events for the eye tracking. The eye tracking computer is given a command to insert an ASCII text message into the eye tracking data. The third method is called "Analogue output". A copy of the eye track is fed directly into the EEG. A digital-to-analogue converter card in the Eye tracking outputs (some of) the data as an analogue signal. With SMI, this signal can be fed directly into the EEG headbox. This requires a custom cable and resistors to scale the output voltage of the D/A converter to the EEG amplifier's recording range. However, this method has some disadvantages. For example, the quality of the eye tracking signal suffers considerably from the D/A and subsequent A/D conversion, and the eye tracking signal may exceed the amplifier's recording range. Therefore, we use the second synchronization method, the flow of which is shown in Figure 6 [65].

D. PROCEDURE

Our experiment was conducted in a light-dimmed, sound-attenuated and comfortable environment. EyeLink 1000 was paired with a 17-inch display with a resolution of 1024 × 768. The participants' eyes were kept at a distance of approximately 60 cm from the monitor and 60 cm from the eye tracker. A joystick with a fixed chin rest was used to keep the participants' heads steady. Before the experiment, calibration was performed to ensure that the eye tracker could capture the pupil and record the EM data accurately. We achieved an error of below 0.5° in our experiment. In addition, to guarantee the effectiveness of the experimental process, four practice trials were conducted, which followed the same procedure as the actual trials, to ensure that participants understood the experiment procedure before beginning the experiment.

The entire experiment consisted of a total of two blocks, Neu_block and Emo_block, and each block contained 15 trials adding up to a total of 30 trials. The participants were asked to view them freely and were allowed to close their eyes to relieve visual fatigue after finishing each consecutive block. In Neu_block, each trial contained two pictures of neutral facial expressions, while in the Emo_block, each trial contained one neutral and one negative facial expression picture. Each picture appeared randomly on the left or right side of the screen. The two facial expressions in each trial were combined into one image and presented on the screen

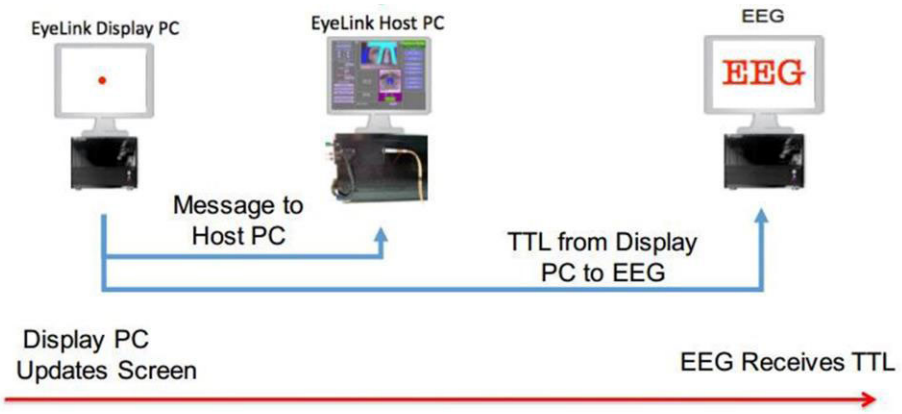


FIGURE 6. EyeLink Display PC displays stimuli via e-prime software, and EM and EEG signals are recorded by EyeLink Host PC and EEG computer respectively. The two systems were coupled by sending a synchronization signal (TTL trigger) as soon as the stimulus was presented on the monitor. The synchronization signals enabled the EM and EEG data to be recorded simultaneously and produced an accurate timestamp matching the offline data.

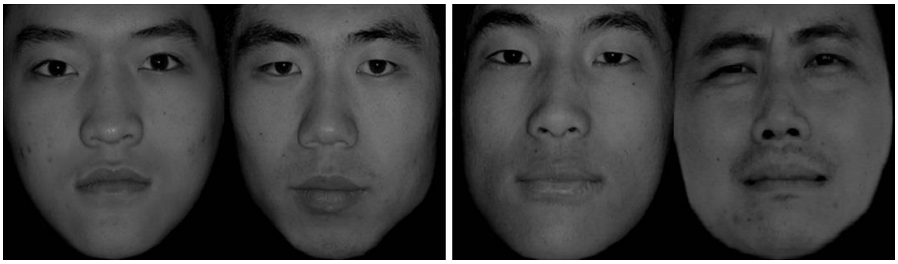


FIGURE 7. Example picture of the Neu_block (left) and Emo_block (right).

for 6 s with a black background, followed by a black background for 2 s. All 30 images were processed using Photoshop software and the size, gradation and resolution were changed accordingly for uniformity(i.e. image size 1280 × 738 pixels; 10.84 × 6.25 cm). Examples of the Emo_block and Neu_block are shown in Figure 7.

IV. RESULTS

For each subject, 30 samples were recorded, including 15 samples from the Neu_block and 15 samples from the Emo_block. Therefore, a total of 1,170 samples were recorded from 39 subjects (19 normal control subjects +20 mild-depression subjects). We performed the five feature selection algorithms described above (BestFirst, Genetic Search, Greedy Stepwise, Linear Forward Selection, and Rank Search) on the EEG and EM features. Our results indicated that the five algorithms could perform effective feature selection and significantly improved the data processing results. Relatively speaking, the BestFirst algorithm led to the best performance; the selected features are listed in Table 4. Therefore, we only show the processing conditions based on the features selected using the BestFirst approach.

TABLE 4. Features selected using the bestfirst algorithm.

	Block1	Block2
EM	9: saccade amplitude maximum, saccade amplitude standard deviation, saccade latency average, saccade latency maximum, p0_c, apen_a, psd_std_a, p0_e, p1_a	7: saccade amplitude maximum, saccade latency average, saccade latency maximum, p1_c, p1_b, apen_a, psd_std_e
EEG	35: c0complex_22, kolmgolov_52, PPmean_83, PPmean_124, mobility_9, complexity_33, complexity_108, f0_83, lyapunov_24, hartley_33, singular_22, permutation_96, min-entropy_58 …	49: ApEn_33, c0complex_36, PPmean_83, meanSquare_104, mobility_9, complexity_92, f0_92, order_24, lyapunov_70, hartley_45, permutation_58, min-entropy_83, singular_96, spectral_70 …

EEG features are denoted as <feature name>_<channel number>

A. DEPRESSION RECOGNITION BASED ON UNIMODAL AUTOENCODER

Single modal data were used as the input of the unimodal autoencoder to generate a shared representation of the EEG or EM data (hidden layer), and each classifier was trained using the shared representation generated from the denoising autoencoder network.

1) EEG-BASED UNIMODAL DEPRESSION RECOGNITION

In the two blocks (Neu_block and Emo_block), delta, theta, alpha, beta, gamma, and the whole band were respectively input into an autoencoder to generate a shared representation (hidden layer) which was then used as the input of the

TABLE 5. Classification results of feature fusion on neu_block.

	Accuracy % (mean ± std. dev.)					
Classifiers	Delta+EM	Theta+EM	Alpha+EM	Beta+EM	Gamma+EM	All+EM
Linear SVM	77.44±2.22	**81.88**±4.09	76.44±6.73	**82.05**±0.21	79.32±4.58	**81.88**±0.19
RBF SVM	**63.08**±1.22	56.92±3.67	62.12±9.97	57.44±12.01	57.78±0.37	67.25±12.02
GBD Tree	70.26±2.24	76.75±1.52	74.43±1.50	74.87±1.46	75.73±1.76	73.16±1.63
RF	69.57±1.59	71.45±0.81	71.25±0.52	71.45±1.25	71.28±0.65	71.97±0.37
SNN	**70.26**±1.34	66.84±2.16	**72.00**±2.67	65.81±6.80	**69.06**±1.64	73.91±6.16
BNMLP	**70.94**±1.79	68.03±3.19	67.93±2.87	66.32±1.81	64.10±0.95	68.99±4.68

Bold indicates that the fusion results are higher than the result of the unimodal EEG and unimodal EM classification.

TABLE 6. Classification results of feature fusion on emo_block.

	Accuracy % (mean ± std. dev.)					
Classifiers	Delta+EM	Theta+EM	Alpha+EM	Beta+EM	Gamma+EM	All+EM
Linear SVM	**73.16**±0.58	72.14±4.07	**72.31**±2.00	**72.31**±3.47	70.77±2.67	70.09±5.21
RBF SVM	74.02±0.26	74.02±0.43	74.02±0.08	74.02±0.20	74.02±0.21	74.02±3.17
GBD Tree	71.79±1.79	75.73±1.43	73.33±3.22	72.48±1.97	**78.46**±1.60	76.75±1.93
RF	72.82±0.86	72.48±1.15	72.31±2.00	70.94±2.38	73.85±0.54	**76.24**±0.92
SNN	**73.68**±1.40	**75.38**±0.67	**74.53**±0.61	**73.16**±2.69	**74.70**±0.80	**74.19**±2.03
BNMLP	69.91±2.40	72.31±3.23	**72.99**±2.21	70.09±1.16	**72.48**±1.71	68.55±1.01

Bold indicates that the fusion results are higher than the result of the unimodal EEG and unimodal EM classification.

six classifiers. The mean accuracy and standard deviation of the six classification results are listed in Tables 2 and 3.

It can be observed that EEG data from both blocks could be used to effectively classify the mild depression from the normal control, and the EEG data of the whole band performed the best, with the highest classification accuracy and smallest standard deviation among all cases. In particular, it achieved the highest classification accuracy of 81.03% and 76.92% in the case of the Neu_block and Emo_block, respectively. In the case of sub-band EEG data, theta and alpha in the Neu_block showed good classification results reaching classification accuracy values of more than 70% with the highest value recorded at 76.41%. The performance of beta and gamma bands were the second best, whereas the delta band was the worst, with almost no effective result. The theta band in the case of the Emo_block had a better classification result, with the highest classification accuracy of 76.58%, followed by alpha, beta, and gamma bands, whereas the delta band had the worst result, in that it could hardly classify the data effectively.

2) EM-BASED UNIMODAL DEPRESSION RECOGNITION

Following the same procedure as in the case of EEG data processing, the mean accuracy and standard deviation of the five classification results for EM data were processed, which are listed in Table 7.

As can be seen from the table, in the case of EM data, the six classification algorithms could effectively identify mild depression for both blocks. In particular, the linear SVM achieved the highest accuracy of 80.17% in the case of the Neu_block with a standard deviation of 4.40%, whereas the highest accuracy in the case of the GBD Tree was 77.44% for the Emo_block with standard deviation of 2.53%.

B. BIMODAL FUSION DEPRESSION RECOGNITION

In our study, two feature fusion methods based on the bimodal autoencoder were used for data fusion; the results for this are discussed in detail in the following subsubsections.

1) FEATURE FUSION

For the Feature Fusion method, we directly linked features from EEG signals of each band and EM features selected using the BestFirst approach together, which were then used as the input for the autoencoder. The structure of the autoencoder used is shown in Figure 4(a). The mean accuracy and standard deviation of the six classification results with Feature Fusion are listed in Tables 5 and 6.

We observed that in the case of both the blocks, after the fusion of EEG (delta, theta, alpha, beta, gamma, and all bands) and EM features, the results obtained using some of the classification algorithms and some EEG bands led to better performance than the unimodal results;

TABLE 7. Unimodal em classification results.

Classifiers	Accuracy % (mean ± std. dev.) Neu_block	 Emo_block
Linear SVM	**80.17** ± 4.40	71.79 ± 0.40
RBF SVM	60.51 ± 1.38	74.36 ± 0.14
GBD Tree	77.26 ± 0.85	**77.44** ± 2.53
RF	73.16 ± 0.40	74.36 ± 1.36
SNN	68.89 ± 0.83	72.82 ± 0.40
BNMLP	69.91 ± 0.93	71.45 ± 2.71

Bold indicates the highest classification accuracy obtained in all algorithms in the case of both the Neu_block and Emo_block

however, no obvious rules were determined for specific classification algorithms and bands. In particular, in the Neu_block, after the fusion of the beta band and EM, the highest classification accuracy of 82.05% was achieved using the Linear SVM, with a standard deviation of 0.21%, whereas in the case of the Emo_block, after the fusion of the gamma band and EM, the GBD Tree approach led to the highest accuracy of 78.46%, with a standard deviation of 1.60%. It is noteworthy that in the Emo_block, after the fusion of the EEG (delta, theta, alpha, beta, gamma, all) and EM features, the classification performance in the case of the SNN algorithm was better than the performance of the unimodal EEG or unimodal EM; however, it must also be noted that there was a case where the bimodal classification was worse than the unimodal classification. However, not every band in the EEG data fused with EM would be better than single mode classification in case of each classification algorithm, in general, the feature fusion strategy performed better than the unimodal classification, leading to satisfactory results.

2) HIDDEN LAYER FUSION

In this feature fusion method, we input the EEG and EM features selected using the BestFirst algorithm separately into an autoencoder to generate a shared representation of the EEG and EM, and then linked the shared representation of the two modes together, as the input of the classifiers. The structure of the autoencoder used in this case is shown in Figure 4(b). The mean accuracy and standard deviation values of the five classification results with Hidden Layer Fusion are listed in Tables 8 and 9.

As can be seen from Tables 8 and 9, in both the blocks, after the fusion of the EEG (delta, theta, alpha, beta, gamma, all) hidden layer and EM hidden layer, the classification performance significantly improved. In particular, in the Neu_block, the alpha band fused with EM data could effectively improve classification performance in the case of all classification algorithms, achieving the highest classification

TABLE 8. Classification results of hidden layer fusion on neu_block.

Classifiers	Accuracy % (mean ± std. dev.) Delta+EM	 Theta+EM	 Alpha+EM	 Beta+EM	 Gamma+EM	 All+EM
Linear SVM	73.68±0.79	77.44±1.94	**83.42**±2.09	74.70±4.45	76.07±3.53	80.17±0.99
RBF SVM	**68.21**±0.26	65.81±1.98	**78.80**±0.32	73.33±1.28	**69.40**±1.05	79.83±0.95
GBD Tree	76.41±1.27	72.65±1.81	**79.66**±2.39	75.38±1.61	73.50±0.95	**79.15**±1.76
RF	72.48±1.48	**73.33**±1.49	**76.58**±2.48	**75.73**±0.25	**73.85**±1.21	**74.53**±0.95
SNN	**72.14**±1.30	**73.33**±3.22	**78.29**±2.44	**74.19**±1.68	**74.19**±1.65	**77.95**±2.04
BNMLP	66.50±1.65	**71.28**±2.27	**76.92**±1.30	**73.16**±1.75	68.89±2.61	77.09±2.20

Bold indicates that the fusion results are higher than the result of the unimodal EEG and unimodal EM classification.

TABLE 9. Classification results of hidden layer fusion on emo_block.

Classifiers	Accuracy % (mean ± std. dev.) Delta+EM	 Theta+EM	 Alpha+EM	 Beta+EM	 Gamma+EM	 All+EM
Linear SVM	**73.85**±0.83	71.11±0.68	**73.85**±0.92	**79.49**±0.79	**79.15**±0.74	**77.26**±0.96
RBF SVM	69.06±0.28	66.84±0.53	70.60±0.49	**75.90**±0.98	**80.85**±1.43	70.94±1.04
GBD Tree	73.33±1.25	75.04±0.56	**77.61**±2.38	**77.95**±1.44	**79.83**±0.57	72.14±1.34
RF	68.55±1.26	**75.73**±0.50	73.68±0.67	**75.73**±1.01	**77.26**±1.23	**75.73**±0.48
SNN	**73.16**±1.02	65.81±1.25	71.45±2.48	**75.66**±3.11	**76.24**±1.28	**73.33**±2.32
BNMLP	70.22±0.72	72.14±1.74	**74.02**±1.39	**72.14**±2.28	**75.38**±0.75	70.09±1.95

Bold indicates that the fusion results are higher than the result of the unimodal EEG and unimodal EM classification.

accuracy of 83.42% using the Linear SVM, with a standard deviation of 2.09%. For the SNN classification algorithm, the performance of each band fused with EM was better than that of unimodal EEG or EM. Furthermore, in the case of the Emo_block, classification performance of the beta and gamma bands improved considerably for all classification algorithms. After the fusion of the beta band and EM, Linear SVM led to a classification accuracy of 79.49%; in contrast, after the fusion of the gamma band and EM, the RBF SVM obtained the highest classification accuracy of 80.85%. Compared with the Feature Fusion method, we observed that the Hidden Layer Fusion method achieved a more significant improvement in terms of classification.

V. DISCUSSION

A. WHICH CLASSIFICATION ALGORITHM LED TO THE BEST PERFORMANCE?

In the case of Feature Fusion, the highest classification accuracy of 82.05% was achieved by the Linear SVM for the Neu_block, while the GBD Tree approach led to the highest classification accuracy of 78.46% for the Emo_block. In a similar manner, in the Hidden Layer Fusion, the highest classification accuracy of 83.42% was achieved by the Linear SVM for the Neu_block, and the RBF SVM led to the highest classification accuracy of 80.85% for the Emo_block. Considering this, the Linear SVM approach achieves the highest classification accuracy in both fusion strategies; thus, the Linear SVM was the best performing classification algorithm. Furthermore, there is evidence that indicates that a Linear SVM is often selected for classification if the data size is insufficiently large, because a Linear SVM might be beneficial in avoiding overfitting as well as realizing good classification performance and robustness [66], [67].

B. WHICH EEG BAND FUSED WITH THE EM LED TO THE BEST PERFORMANCE?

In the case of Feature Fusion, we did not find a specific band that led to significantly better performance than any other bands in terms of classification. However, in comparison, the beta band led to a better classification accuracy of 82.05% for the Neu_block, while the gamma band led to a better accuracy of 78.46% for the Emo_block compared with the other bands in both cases. In contrast, in the case of Hidden Layer Fusion, the alpha band led to the best performance in each classification algorithm in the case of the Neu_block; in addition, the classification accuracy with fusion showed a remarkable improvement compared with the unimodal classification results, achieving the highest classification accuracy of 83.42%. The beta and gamma bands both performed better in each classification algorithm in the case of the Emo_block, and the classification accuracy with fusion was significantly higher than the unimodal results; in particular, the gamma band achieved the highest classification accuracy of 80.85%. Therefore, the EEG bands that led to the best classification performance were the alpha and gamma band in the Neu_block and Emo_block, respectively, which suggested that the alpha and gamma bands were more strongly related with the depression state of individuals. A number of studies have reported the discovery that certain metrics on alpha band of EEG can distinguish between depression and healthy controls. Depressive patients display a greater frontal alpha power value than control groups [68]. Another functional connectivity study has shown that impaired functional connectivity at EEG alpha frequency band in depression [69]. In addition, after a great quantity of studies on gamma oscillations, they were widely regarded as a crucial part in integrating distributed neural processes into highly ordered cognitive functions, such as emotional processes [70]. Li *et al.* [71] reported abnormal functional connectivity of EEG gamma band in patients with depression during emotional face processing. A recent report was published that individuals with depression displayed sustained and increased gamma band EEG power [72]. Our result was in agreement with previous research findings.

C. COMPARISON OF PERFORMANCE IMPROVEMENT BY TWO FUSION STRATEGIES

Hidden Layer Fusion performed better in terms of improvement in the case of unimodal classification accuracy, whereas Feature Fusion did not perform as well, and instead, performed worse in the case of some EEG bands and classification algorithms. Moreover, the Hidden Layer Fusion strategy also showed a smaller standard deviation, which indicated that this model had better stability. Provost et al. recommended that when evaluating binary decision problems, instead of using the accuracy results directly, Receiver Operator Characteristic (ROC) curves should also be used as they can indicate the model's classification ability [73]. The area under an ROC curve (AUC), has a value between 0 and 1; the greater the AUC value, the better the classification ability of the model. Therefore, we compared the accuracy of the six classification algorithms and the ROC of the three better classification algorithms in the case of these four methods using the well performing alpha and gamma bands for the Neu_block and Emo_block, respectively. This comparison is shown in the Figure 8 and Figure 9. As can be seen from the figures, in the case of the Neu_block, the fusion of the alpha band and EM using the Hidden Layer Fusion strategy significantly outperformed the unimodal EEG and unimodal EM data with respect to classification accuracy and AUC value. However, the Feature Fusion strategy did not necessarily outperform the unimodal methods. In particular, in the case of the Emo_block, using the two feature fusion strategies for gamma band and EM led to a better improvement in classification performance than unimodal classification. Thus, the improvement in the classification on using the Hidden Layer Fusion is evidently better than that of the Feature Fusion approach. Therefore, in summary, the Hidden Layer Fusion of EEG and EM is more suitable for the identification of mild depression than the Feature Fusion approach. In a study by Ngiam *et al.* [19] that demonstrated

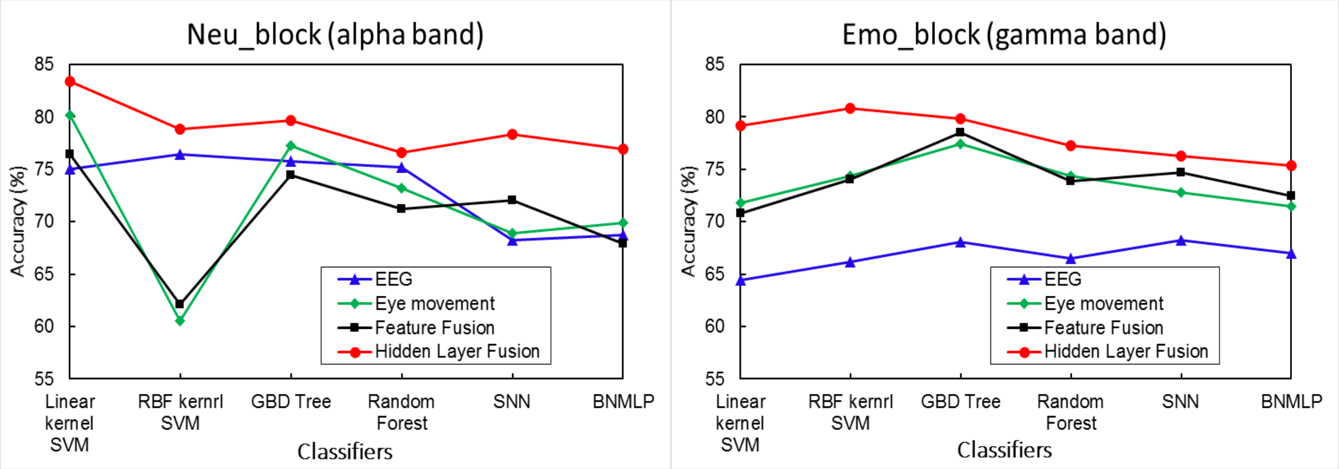
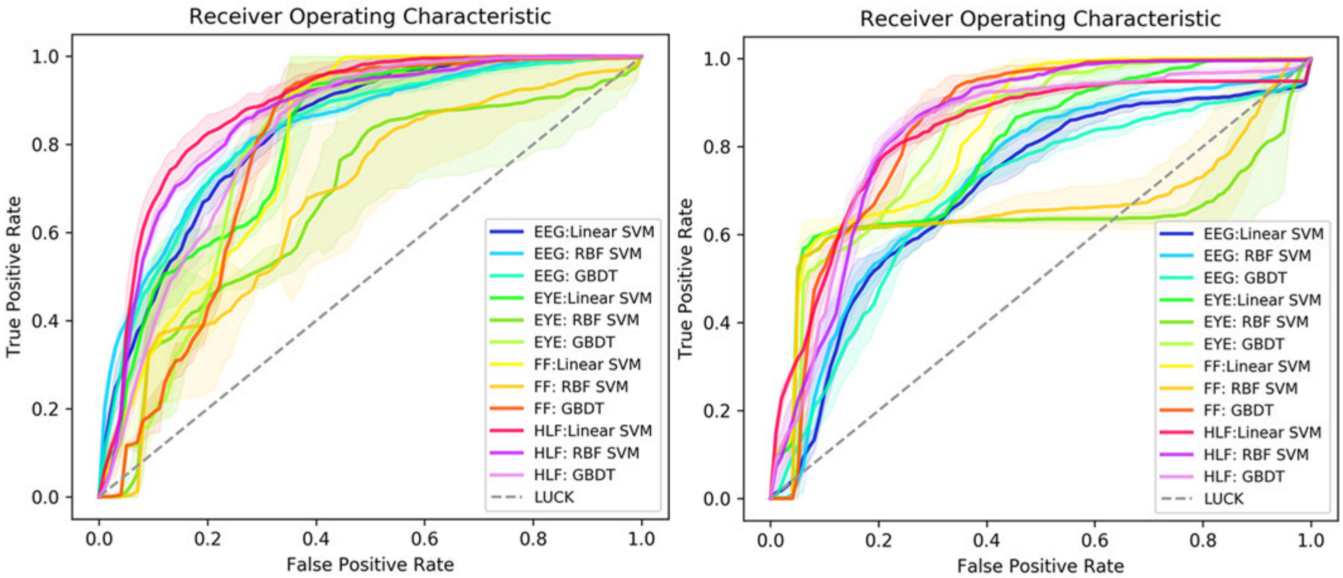


FIGURE 8. Comparison of the classification results of the four methods considered in the alpha band for the Neu_block and gamma band for the Emo_block.



(a) (b)

FIGURE 9. ROC of four methods in the alpha band for the Neu_block (a) and gamma band for the Emo_block (b). The shaded area in the graph is the result of each measure, and the solid line in the shadow is the average result of five measures.

best published visual speech classification on AVLetters using shared representation learning. Lu *et al.* [55] demonstrated that the shared representations are good features to discriminate different emotions by fusing EEG features and EM features with bimodal deep autoencoders (BDAE). For the SEED dataset, the BDAE model is better than other feature merging strategies.

D. COMPARISON OF CLASSIFICATION RESULTS OF THE TWO BLOCKS

In the case of the Neu_block and Emo_block, it can be observed that unimodal EEG led to classification accuracies of 79.69% and 76.27%, unimodal EM led to classification accuracies of 76.24% and 74.09%, Feature Fusion led to classification accuracies of 82.05% and 78.46%, and Hidden Layer Fusion led to classification accuracies of 83.42% and 80.85%, respectively. In addition, we found that the classification accuracy in the case of the Neu_block was always higher than that of the Emo_block among these four classification results, which suggests that, the neutral face picture stimulus had better discrimination ability between mild depression subjects and normal control subjects.

Other studies [74], [75] also found that depressed patients showed a clear impairment in the recognition of neutral facial expressions. In particular, depressed patients and controls performed nearly identically when happy or sad faces stimuli were presented; however, they performed differently in the case of neutral faces stimuli. Compared with controls subjects, depressed patients had lower awareness of neutral faces stimuli, which was characterized by lower judgment accuracy and longer response time. This observation was consistent with the conclusion that the Neu_block data could be used to better distinguish between mild depression subjects and control subjects.

VI. CONCLUSIONS

Mild depression is a complex psycho-physiological phenomenon; therefore, it is hard to establish a robust mild depression recognition model using only a single modality. Nevertheless, signals from different modalities can represent different aspects of mild depression and can integrate complementary information from these different modalities to build a more robust mild depression recognition model compared with the existing unimodal methods. Moreover, a major advantage of deep learning models over traditional models is the joint feature extraction and classification in a unified network. However, meaningful analyses of simultaneously recorded EM and EEG data requires that both data streams are synchronized with millisecond precision. In this study, we used an EEG-EM synchronization acquisition network that allowed us to simultaneously record both the EM and the EEG physiological signals of mild depression and normal controls during free viewing.

With a long-term aim to develop an objective multimodal system based on the EEG-EM synchronization acquisition network to assists doctors during the diagnosis and monitoring of mild depression, we investigated the mild depression recognition performance of EEG and EM individually as well as when fused. To develop a classification system-oriented approach, in this study, we considered feature selection, classification and fusion. We examined the performance of two fusion strategies using several feature selection methods as well as several classification algorithms. We found that the features selected using the BestFirst algorithm showed better results, and the Linear SVM classifier performed best. Furthermore, on fusing EEG and EM using the two fusion strategies, the ability to recognize mild depression is significantly improved in most bands and classification algorithms. Between the two fusion approaches used in our study, the more effective and robust fusion method was the Hidden Layer Fusion method, which led to an average accuracy of 83.42%.

VII. LIMITATIONS AND FUTURE WORK

Even though it is a common problem in similar studies, a known limitation is the relatively low number of both depressed and control subjects. Because the data collection work is ongoing, we anticipate on reporting on a larger dataset in the future. Another limitation of this study is that only two depression levels were used which leads to a binary classification. Future work will address these limitations, and we will try to classify more depression states such as severe depression (BDI scores ranging from 29-63). In particular, we hope to study multimodal depression recognition methods based on multiple depression states.

REFERENCES

[1] M. Marcus *et al.*, "Depression: A global public health concern," Dept. Mental Health Substance Abuse, Global Crisis World Fed. Mental Health, WHO, Geneva, Switzerland, Tech. Rep., 2012. [Online]. Available: https://www.who.int/mental_health/management/depression/wfmh_paper_depressionwmhd_2012.pdf

[2] G. H. Brundtland, "Mental health: New understanding, new hope," *Jama*, vol. 286, no. 19, p. 2391, 2001.

[3] X. Li, T. Cao, S. Sun, B. Hu, and M. Ratcliffe, "Classification study on eye movement data: Towards a new approach in depression detection," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 1227–1232. doi: 10.1109/CEC.2016.7743927.

[4] A. K. Ibrahim, S. J. Kelly, C. E. Adams, and C. Glazebrook, "A systematic review of studies of depression prevalence in university students," *J. Psychiatric Res.*, vol. 47, no. 3, pp. 391–400, 2013.

[5] S. D'Mello and J. Kory, "Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies," in *Proc. 14th ACM Int. Conf. Multimodal Interaction*, 2012, pp. 31–38.

[6] X. Li, Z. Jing, B. Hu, and S. Sun, "An EEG-based study on coherence and brain networks in mild depression cognitive process," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 1275–1282.

[7] X. Li, B. Hu, S. Sun, and H. Cai, "EEG-based mild depressive detection using feature selection methods and classifiers," *Comput. Methods Programs Biomed.*, vol. 136, pp. 151–161, Nov. 2016.

[8] H. Dibeklioğlu, Z. Hammal, and J. F. Cohn, "Dynamic multimodal measurement of depression severity using deep autoencoding," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 525–536, Mar. 2018.

[9] P. Giannakopoulos, P. Missonnier, G. Gold, and A. Michon, "Electrophysiological markers of rapid cognitive decline in mild cognitive impairment," in *Dementia in Clinical Practice*, vol. 24. Basel, Switzerland: Karger, 2009, pp. 39–46.

[10] M. Bachmann, J. Lass, and H. Hinrikus, "Single channel EEG analysis for detection of depression," *Biomed. Signal Process. Control*, vol. 31, pp. 391–397, Jan. 2017.

[11] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques," in *Proc. 19th Iranian Conf. Elect. Eng. (ICEE)*, May 2011, pp. 1–4.

[12] Y. Lu, W. L. Zheng, B. Li, and B. L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in *Proc. IJCAI*, 2015, pp. 1170–1176.

[13] M. Juhola, H. Aalto, and T. Hirvonen, "Using results of eye movement signal analysis in the neural network recognition of otoneurological patients," *Comput. Methods Programs Biomed.*, vol. 86, no. 3, pp. 216–226. 2007.

[14] G. J. Emslie, A. J. Rush, W. A. Weinberg, J. W. Rintelmann, and H. P. Roffwarg, "Children with major depression show reduced rapid eye movement latencies," *Arch. Gen. Psychiatry*, vol. 47, no. 2, pp. 119–124, 1990.

[15] A. Duque and C. Vázquez, "Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking study," *J. Behav. Therapy Exp. Psychiatry*, vol. 46, pp. 107–114, Mar. 2015.

[16] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, and M. Breakspear, "Eye movement analysis for depression detection," in *Proc. 20th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 4220–4224.

[17] R. Gupta *et al.*, "Multimodal prediction of affective dimensions and depression in human-computer interactions," in *Proc. 4th Int. Workshop Audio/Vis. Emotion Challenge*, 2014, pp. 33–40.

[18] A. B. Said, A. Mohamed, T. Elfouly, K. Harras, and Z. J. Wang, "Multimodal deep learning approach for joint EEG-EMG data compression and classification," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[19] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, 2011, pp. 689–696.

[20] N. Srivastava and R. Salakhutdinov, "Learning representations for multimodal data with deep belief nets," in *Proc. Int. Conf. Mach. Learn. Workshop*, 2012, pp. 1–8.

[21] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2222–2230.

[22] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, Jul. 2006.

[23] Y. Liu, X. Feng, and Z. Zhou, "Multimodal video classification with stacked contractive autoencoders," *Signal Process.*, vol. 120, pp. 761–766, Mar. 2016.

[24] J. Cohn, N. Cummins, J. Epps, R. Goecke, J. Joshi, and S. Scherer, "Multimodal assessment of depression and related disorders based on behavioural signals," in *The Handbook of Multimodal-Multisensor Interfaces*, vol. 2. New York, NY, USA: Morgan & Claypool, 2017.

[25] M. Valstar *et al.*, "Avec 2014: 3D dimensional affect and depression recognition challenge," in *Proc. 4th Int. Workshop Audio/Visual Emotion Challenge*, 2014, pp. 3–10.

[26] S. Scherer, Z. Hammal, Y. Yang, L.-P. Morency, and J. F. Cohn, "Dyadic behavior analysis in depression severity assessment interviews," in *Proc. 16th Int. Conf. Multimodal Interact.*, 2014, pp. 112–119.

[27] S. Scherer, G. Stratou, J. Gratch, and L.-P. Morency, "Investigating voice quality as a speaker-independent indicator of depression and PTSD," in *Proc. Interspeech*, 2013, pp. 847–851.

[28] H. Jiang *et al.*, "Investigation of different speech types and emotions for detecting depression using different classifiers," *Speech Commun.*, vol. 90, pp. 39–46, Jun. 2017.

[29] B. Hu *et al.*, "Emotion regulating attentional control abnormalities in major depressive disorder: An event-related potential study," *Sci. Rep.*, vol. 7, Oct. 2017, Art. no. 13530.

[30] N. C. Maddage, R. Senaratne, L.-S. A. Low, M. Lech, and N. Allen, "Video-based detection of the clinical depression in adolescents," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Sep. 2009, pp. 3723–3726.

[31] G. Stratou, S. Scherer, J. Gratch, and L.-P. Morency, "Automatic nonverbal behavior indicators of depression and PTSD: Exploring gender differences," in *Proc. Hum. Assoc. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2013, pp. 147–152.

[32] P. Li *et al.*, "Reduced sensitivity to neutral feedback versus negative feedback in subjects with mild depression: Evidence from event-related potentials study," *Brain Cogn.*, vol. 100, pp. 15–20, Nov. 2015.

[33] S. Lu *et al.*, "Attentional bias scores in patients with depression and effects of age: A controlled, eye-tracking study," *J. Int. Med. Res.*, vol. 45, no. 5, pp. 1518–1527, 2017.

[34] S. Alghowinem *et al.*, "Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 478–490, Oct./Dec. 2016.

[35] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proc. 4th Int. Workshop Audio/Vis. Emotion Challenge*, 2014, pp. 65–72.

[36] S. Scherer, G. Stratou, and L.-P. Morency, "Audiovisual behavior descriptors for depression assessment," in *Proc. 15th ACM Int. Conf. Multimodal Interaction*, 2013, pp. 135–140.

[37] H. Meng, D. Huang, H. Wang, H. Yang, M. AI-Shuraifi, and Y. Wang, "Depression recognition based on dynamic facial and vocal expression features using partial least square regression," in *Proc. 3rd ACM Int. Workshop Audio/Vis. Emotion Challenge*, 2013, pp. 21–30.

[38] S. A. Akar, S. Kara, S. Agambayev, and V. Bilgiç, "Nonlinear analysis of EEGs of patients with major depression during different emotional states," *Comput. Biol. Med.*, vol. 67, pp. 49–60, Dec. 2015.

[39] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, 2013.

[40] R. N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 395–404, 1997.

[41] B. Hu *et al.*, "EEG-based cognitive interfaces for ubiquitous applications: Developments and challenges," *IEEE Intell. Syst.*, vol. 26, no. 5, pp. 46–53, Sep. 2011.

[42] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.

[43] M. Bachmann, K. Kalev, A. Suhhova, J. Lass, and H. Hinrikus, "Lempel Ziv complexity of EEG in depression," in *Proc. 6th Eur. Conf. Int. Fed. Med. Biol. Eng.*, 2015, pp. 58–61.

[44] F.-Y. Fan, Y.-J. Li, Y.-H. Qiu, and Y.-S. Zhu, "Use of ANN and complexity measures in cognitive EEG discrimination," in *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBS)*, 2006, pp. 4638–4641.

[45] Z. Liang *et al.*, "EEG entropy measures in anesthesia," *Frontiers Comput. Neurosci.*, vol. 9, p. 16, Feb. 2015.

[46] N. Nicolaou and J. Georgiou, "Detection of epileptic electroencephalogram based on permutation entropy and support vector machines," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 202–209, 2012.

[47] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, p. R1, 2007.

[48] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997.

[49] D. E. Golberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.

[50] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 6, pp. 1437–1447, Nov./Dec. 2003.

[51] M. Gutlein, E. Frank, M. Hall, and A. Karwath, "Large-scale attribute selection using wrappers," in *Proc. IEEE Symp. Comput. Intell. Data Mining (CIDM)*, Mar. 2009, pp. 332–339.

[52] M. A. Hall and L. A. Smith, "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper," in *Proc. FLAIRS Conf.*, 1999, pp. 235–239.

[53] P. Poirson and H. Idrees, "Multimodal stacked denoising autoencoders," Center Res. Comput. Vis., Univ. Central Florida, Orlando, FL, USA, Tech. Rep., 2013. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.5211&rep=rep1&type=pdf

[54] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.

[55] W. Liu, W. L. Zheng, and B. L. Lu. (2016). "Multimodal emotion recognition using multimodal deep learning." [Online]. Available: https://arxiv.org/abs/160208225

[56] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[57] J. H. Friedman, "Stochastic gradient boosting," *Comput. Statist. Data Anal.*, vol. 38, no. 4, pp. 367–378, 2002.

[58] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, 1997.

[59] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 972–981.

[60] X. Han and Q. Dai, "Batch-normalized Mlpconv-wise supervised pre-training network in network," *Appl. Intell.*, vol. 48, no. 1, pp. 142–155, 2018.

[61] M. T. Arslan, S. G. Eraldemir, and E. Yıldırım, "Subject-dependent and subject-independent classification of mental arithmetic and silent reading tasks," *Uluslararasi Mühendislik Araştirma Geliştirme Dergisi*, vol. 9, pp. 186–195, Dec. 2017.

[62] A. T. Beck, R. A. Steer, R. Ball, and W. F. Ranieri, "Comparison of beck depression inventories-IA and-II in psychiatric outpatients," *J. Personality Assessment*, vol. 67, no. 3, pp. 588–597, 1996.

[63] X. Gong, Y.-X. Huang, Y. Wang, and Y.-J. Luo, "Revision of the Chinese facial affective picture system," *Chin. Mental Health J.*, vol. 25, no. 1, pp. 40–46, 2011.

[64] T. C. Ferree, P. Luu, G. S. Russell, and D. M. Tucker, "Scalp electrode impedance, infection risk, and EEG data quality," *Clin. Neurophysiol.*, vol. 112, no. 3, pp. 536–544, 2001.

[65] *EYE-EEG: Tutorial*. Accessed: Jan. 1, 2019. [Online]. Available: http://www2.hu-berlin.de/eyetracking-eeg/tutorial.html

[66] M. J. Patel, A. Khalaf, and H. J. Aizenstein, "Studying depression using imaging and machine learning methods," *NeuroImage, Clin.*, vol. 10, pp. 115–123, Jan. 2016.

[67] X. W. Wang, D. Nie, and B. L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, Apr. 2014.

[68] A. Kemp *et al.*, "Disorder specificity despite comorbidity: Resting EEG alpha asymmetry in major depressive disorder and post-traumatic stress disorder," *Biol. Psychol.*, vol. 85, pp. 350–354, Oct. 2010.

[69] A. A. Fingelkurts, A. A. Fingelkurts, H. Rytsälä, K. Suominen, E. Isometsä, and S. Kähkönen, "Impaired functional connectivity at EEG alpha and theta frequency bands in major depression," *Hum. Brain Mapping*, vol. 28, no. 3, pp. 247–261, 2007.

[70] E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann, "Gamma, alpha, delta, and theta oscillations govern cognitive processes," *Int. J. Psychophysiol.*, vol. 39, pp. 241–248, Jan. 2001.

[71] Y. Li, D. Cao, L. Wei, Y. Tang, and J. Wang, "Abnormal functional connectivity of EEG gamma band in patients with depression during emotional face processing," *Clin. Neurophysiol.*, vol. 126, pp. 2078–2089, Nov. 2015.

[72] G. J. Siegle, R. Condray, M. E. Thase, M. Keshavan, and S. R. Steinhauer, "Sustained gamma-band EEG following negative words in depression and schizophrenia," *Int. J. Psychophysiol.*, vol. 75, pp. 107–118, Feb. 2010.

[73] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240.

[74] J. M. Leppänen, M. Milders, J. S. Bell, E. Terriere, and J. K. Hietanen, "Depression biases the recognition of emotionally neutral faces," *Psychiatry Res.*, vol. 128, pp. 123–133, Sep. 2004.

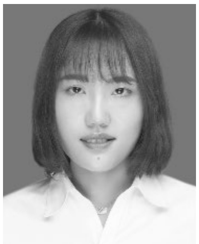
[75] K. M. Thomas *et al.*, "Amygdala response to facial expressions in children and adults," *Biol. Psychiatry*, vol. 49, no. 4, pp. 309–316, 2001.

JING ZHU received the M.Sc. degree in economics and finance from Xi'an Jiaotong University, in 2010. She is currently pursuing the Ph.D. degree in computer software and theory with Lanzhou University, where she is currently an Engineer. Her research interests include ubiquitous computing and data mining.

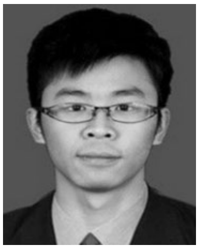
YING WANG received the bachelor's degree from Lanzhou University, in 2017, where she is currently pursuing the master's degree with the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering. Her research interests include analysis of affective learning, data mining, and deep learning.

RONG LA is currently pursuing the master's degree with the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University. Her research interests include analysis of affective learning, data mining, and deep learning.

JIAWEI ZHAN received the bachelor's degree in computer science from Lanzhou University, in 2018.



JUNHONG NIU received the bachelor's degree in computer science from Lanzhou University, in 2016, where she is currently pursuing the master's degree. Her research interests include the analysis of ubiquitous computing and data mining.



SHUAI ZENG received the bachelor's degree in software engineering from Chongqing University, in 2015. He is currently pursuing the master's degree with Lanzhou University. His research interests include the analysis of affective learning and data mining.

XIPING HU received the Ph.D. degree from The University of British Columbia, Vancouver, BC, Canada.

He was the Co-Founder and the CTO of Bravolol, Ltd., Hong Kong, a leading language learning mobile application company with over 100 million users and listed as the top-two language education platform globally. He is currently a Professor with the School of Information Science and Engineering, Lanzhou University, Lanzhou, China. He has authored or co-authored around 60 papers published and presented in prestigious conferences and journals, such as the IEEE Transactions on Emerging Topics in Computing, IEEE Internet of Things Journal, *ACM Transactions on Multimedia Computing, Communications, and Applications*, IEEE Communications Surveys and Tutorials, *IEEE Communications Magazine*, IEEE Network, ACM MobiCom, and WWW. His current research interests include mobile cyber-physical systems, crowdsensing, social networks, and cloud computing.

Dr. Hu has been serving as an Associate Editor for the IEEE Access and a Lead Guest Editor for the IEEE Transactions on Automation Science and Engineering and *Wireless Communications and Mobile Computing*.