2020 25th International Conference on Pattern Recognition (ICPR)
Milan, Italy, Jan 10-15, 2021

Exploring Spatial-Temporal Representations for fNIRS-based Intimacy Detection via an Attention-enhanced Cascade Convolutional Recurrent Neural Network

Chao Li*, Qian Zhang*, Ziping Zhao*, Li Gu†‡, Björn Schuller§¶
*College of Computer and Information Engineering, Tianjin Normal University, Tianjin, China
†School of Humanities and Management, Guangdong Medical University, Dongguan, China
‡Faculty of Psychology, Tianjin Normal University, Tianjin, China
§Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany
¶GLAM, Group on Language, Audio & Music, Imperial College London, London, UK

***Abstract*—The detection of intimacy plays a crucial role in the improvement of intimate relationship, which contributes to promote the family and social harmony. Previous studies have shown that different degrees of intimacy have significant differences in brain imaging. Recently, work has emerged to recognise intimacy automatically by using machine learning techniques. Moreover, considering the temporal dynamic characteristics of intimacy relationship on neural mechanism, how to model spatio-temporal dynamics for intimacy prediction effectively is still a challenge. In this paper, we propose a novel method to explore deep spatial-temporal representations for intimacy prediction by an *Attention-enhanced Cascade Convolutional Recurrent Neural Network* (ACCRNN). Given the advantages of time-frequency resolution in complex neuronal activities analysis, this paper utilizes *functional near-infrared spectroscopy* (fNIRS) to analyse and infer intimate relationship. We collected fNIRS-based dataset for the analysis of intimate relationship. Forty-two-channel fNIRS signals are recorded from the 44 subjects' prefrontal cortex when they watched a total of 18 photos of lovers, friends and strangers for 30 seconds per photo. The experimental results show that our proposed method outperforms the others in terms of accuracy with the precision of 96.5%. To the best of our knowledge, this is the first time that such a hybrid deep architecture has been employed for fNIRS-based intimacy prediction.**

I. Introduction

The concept of intimacy has long permeated theories of social life [1] particularly in the study of close relationships like friendships, and love relationship [2]. Intimacy generally refers to the feeling of being in a close personal association and belonging together. It emphasizes the degree of interdependence between the two sides [3], which can be the romantic relationship of lovers, the marriage relationship between husband and wife, or the intimate friendship. Whether it is to build a happy family or a happy organization, the intimate relationship is important. Many studies show that intimate relationships can effectively alleviate people's negative emotions. For example, holding hands with a lover will reduce self-anxiety.

Previous work mainly used questionnaires to analyse intimate relationships in interpersonal communication, which leads to a strong subjectivity in the analysis of intimate relationships. Additionally, in the field of neuroscience, many studies show that different degrees of intimate relationships are associated with the activation of specific brain regions [3], [4], [5], [6]. When subjects viewed pictures of their lovers vs. friends, the activation of dopamine-rich brain regions and the midbrain marginal dopamine circuit was significantly enhanced. Therefore, can we automatically detect the categories of intimacy by analysing the activity of certain brain regions?

Recently, multiple modalities, such as electroencephalography (EEG), functional magnetic resonance imaging (fMRI) and fNIRS are applied in various brain-computer interface (BCI) analysis tasks, including lie detection [7], intention recognition [8], or emotion recognition [9], [10]. However, the utilization of fMRI has some limitations because of its bulky size, and high cost of its scanners, which leads to it being suitable for the most resting-state tasks. EEG [11] and fNIRS are both flexible, scalp located procedures that can be employed for monitoring multiple populations in ecological conditions. The EEG experimental preparation process is complicated, and the collected signals are more objective and sensitive[12]. It is often utilized for the diagnosis and measurement of brain diseases such as epilepsy and sleep. In addition, it is greatly disturbed by external physiological signals and device electrodes. Recently, fNIRS [13], [14] has been recognized as a promising noninvasive optical imaging technique for monitoring the hemodynamic response of the brain using neurovascular coupling. Neurovascular coupling in the cerebral cortex captures the increases in oxygenated hemoglobin (HbO) and reductions in deoxygenated hemoglobin (HbR) that occur during brain activity. The functional near-infrared spectroscopy has been performed well in the field of laboratory advanced cognitive neuroscience research [15], brain-computer interface research [16] and other cognitive activities [17]. Due to the

advantages of low cost, good portability, non-invasiveness and no excessive sensitivity to the test action during the experiment, fNIRS-based analysis is gaining widespread attention from researchers.

To date, however, work on exploiting a predictive model for intimacy detection has been very limited [18]. Li et al. [18] use hand-crafted features based on a General Linear Model (GLM) and Complex Brain Network Analysis (CBNA) methods to build a predictive model for intimacy. In the field of BCI, most of the existing studies [14], [15] have relied on extracting the statistical features from the time-domain signal. However, reaching the highest classification accuracy depends on multiple factors, such as selecting the best set of combined features [16], [17] and the size of the time window [19]. And hand-crafted features always require domain knowledge for the specific task, and designing the proper features for a new task may be more time consuming than designing the model itself. To overcome the limitations of these conventional methods, an appropriate technique for feature extraction needs to be determined. The previous studies have demonstrated that convolutional neural networks (CNNs) can successfully achieve high classification accuracy in many applications, including image recognition [20], speech detection [21], and multiple time-series processing [22]. Considering CNNs' ability to extract important spatial features from a signal, it may be suitable for fNIRS-based intimate relationship as well. Meanwhile, as a popular RNN architecture specialized in sequence learning, LSTM-RNN has built-in memory gates to retain long-term information, which has the ability of learning the temporal features from sequences.

Based on the above considerations, we propose a novel framework to use an attention-enhanced cascade convolutional recurrent neural network (ACCRNN) in intimacy detection by capturing spatial-temporal feature representation. A cascade convolutional recurrent neural network is utilized to automatically learn the high-level spatial-temporal representation in terms of intimacy. And an attention model is used to capture the key information in a sequence. Fig. I shows the proposed framework for intimacy prediction using fNIRS signals.

The main contributions of this paper are as follows: a) A fNIRS-based database is collected to analyze human's complex brain response pattern corresponding to intimacy. Forty-two-channel fNIRS signals are recorded from 44 subjects when they watch the pictures from their lover, friend and strangers; b) a fNIRS-based cascade deep learning architecture is utilized to detect three different intimacy classes, including lover, friend and stranger. Compared with hand-craft features, the proposed method can automatically extract spatial and temporal features from fNIRS signals by a cascade convolutional recurrent neural network, which is capable of learning feature representations and modeling the spatial-temporal dependencies between their activation; c) we also investigate the usage of attention-based architectures to improve fNIRS-based intimacy prediction. The attention mechanism allows the network to focus on the salient parts of a sequence.

The reminder of this paper is organized as follows: Section II reviews some related work in the field of intimacy prediction. Section III presents the details of the proposed framework. Section IV shows the performance of our proposed method on an existing high quality laboratory dataset and our proposed method is also compared with other existing methods. Section V summarizes this paper and outlines the future work.

II. Related work

Brain-computer interface (BCI) is a highly active research field, with many novel approaches being proposed and investigated over the past decade. As the clue to explore brain activity, several modalities have been used for brain signal acquisition, which include EEG, MEG, fMRI and fNIRS. Among them, fNIRS is relatively new, which uses near-infrared-range light (usually of 650-1000 nm wave length) to measure the concentration changes of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) [23]. Its main advantages are relatively low cost, portability, safety, low noise (compared to fMRI), and easiness to use [24].

Recently, brain mechanism for intimate relationship between humans and their interaction has attracted extensive attention from researchers. Eisenberger et al [25] explored the brain mechanism of social pain caused by rejection. The results from fMRI found that the subjects had a significant activation of the anterior cingulate and the right ventral prefrontal lobe when they were rejected compared to the acceptance. By using fNIRS, Reindl et al. [26] found that brain-to-brain synchrony may represent an underlying neural mechanism of the intimate connection between parent and child, which is linked to the child's development of adaptive emotion regulation. Thus, exploring and modeling users' brain nerve activity patterns and accurately predicting their intimate relationship status is the key to improve intimate relationships.

With the increase of available data and computational power, deep learning methods have been successfully applied in various BCI tasks to learn robust feature representations, such as intention recognition [8], mental workload [27], emotion recognition [10], motor imagery [24], or neurorehabilitation [24]. Zhang et al [8] introduce both cascade and parallel convolutional recurrent neural network models for precisely identifying human intended movements and instructions by effectively learning the compositional spatio-temporal representations of raw EEG streams. Chiarelli et al. [28] design a hybrid EEG-fNIRS brain-computer interface for motor imagery classification by using DNN. While there is a range of work in the literature focusing on analysis and prediction of intimate relationship, very little research has been undertaken to explore complex spatial-temporal feature representation.

III. Methodology

A. Problem Definition

Given a period T representing a trial during which a subject watches a photo from a certain intimate relationship, we aim to recognize the intimacy by analyzing the fNIRS signals. There

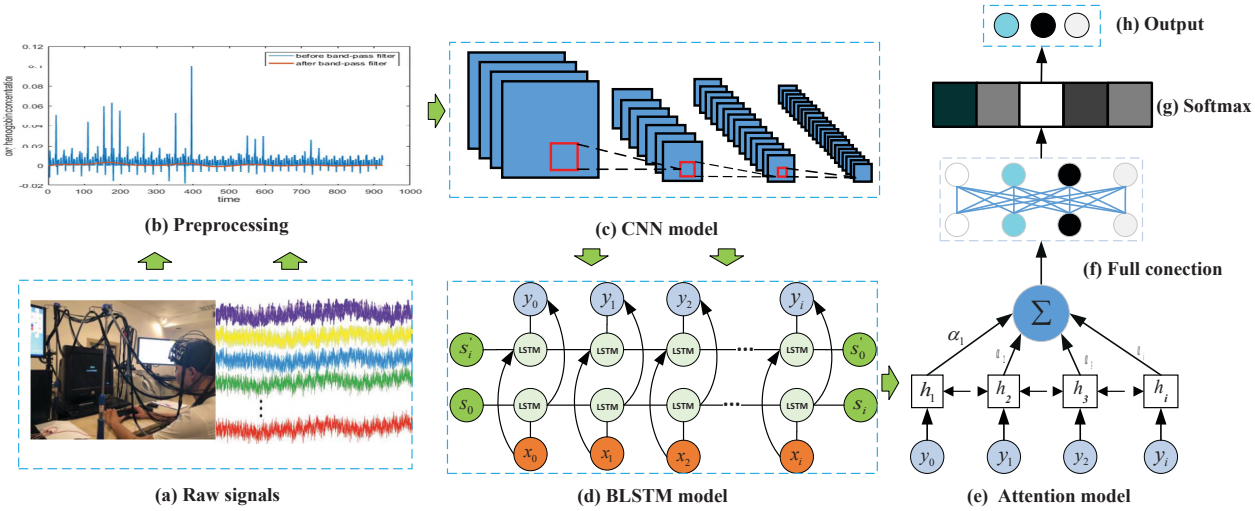


Fig. 1. The framework of the attention-enhanced cascade convolutional recurrent neural network based on fNIRS signals for intimacy prediction. (a) collects intimacy-induced fNIRS signals; (b) removes fNIRS signal noise; (c) captures high-level spatial representation by a CNN model; (d) captures high-level temporal representation by the LSTM model; (e) finds the salient parts of a sequence for intimacy by an attention mechanism; (f) learns the final representation for intimate relationship category; (g) predicts the final intimate relationship category by a softmax classifier; (h) outputs predicted results.

are n fNIRS sensor nodes, each of which has a k-time-point reading during T, constructing a two-dimensional (2D) tensor $\mathbf{X}_T = [\mathbf{r}_1, ..., \mathbf{r}_n] \in \Re^{n \times k}$ as raw fNIRS features of the trial T. Each fNIRS node reading is a one-dimensional (1D) tensor $\mathbf{r}_i = [s_1^i, ..., s_k^i] \in \Re^k$, where s_t^i is the sensor measurement of the ith fNIRS sensor at the time point t. Our goal is to predict the intimacy performed during one trial T by analyzing the fNIRS recording $\mathbf{X}_T$.

B. Spatial-temporal Feature Representation

To capture a better spatial-temporal representation, we design a cascade convolutional recurrent neural network (CCRNN) framework by combining CNN and LSTM. The input to the model is the preprocessed multi-channel fNIRS signals. We extract the spatial features of preprocessed fNIRS signals, and then feed the sequence of the extracted spatial features into the LSTM to extract temporal features. The CNN applied to each fNIRS signals is only responsible for spatial feature extraction, while the following LSTM network explores the relationships among multiple time steps.

The CNN has shown that it is highly capable of automatically learning appropriate features from the input data by optimizing the weight parameters of each filter, using forward and backward propagation to minimize classification errors. In convolutional layers, a convolutional filter whose width is equal to the dimension of the input and kernel size (height) of h is convolved with the input data, where the output of the ith filter is

$$o_i = \vec{w} \cdot \vec{x}[i : i + h - 1], \tag{1}$$

where $\vec{w}$ is the weight matrix, $x[\vec{i} : j]$ is the submatrix of input from row i to j, and o_i is the result value.

For learning a temporal representation, an LSTM unit can determine whether to retain existing memory or to overwrite it with new information. Thus, an LSTM-RNN has the ability to model long-range dynamic dependencies so the problem of vanishing or exploding gradients can be avoided during training [29]. According to the input of the previous unit, the input gate determines which information in the unit needs to be updated at time t. The forget gate calculates the importance of the information, discarding the useless information. The output unit controls and affects the final output state at time t.

$$\tilde{c}_t^j = \tanh\left(U_c \mathrm{x}_t + W_c h_{t-1} + b_c\right)^j \tag{2}$$

$$c_t^j = f_t^j c_{t-1}^j + i_t^j \tilde{c}_t^j. \tag{3}$$

In formula 2, $\tilde{c}_t^j$ represents c_{t-1} new memory gate unit, b is the memory content of the previous unit, and the new memory content is calculated by using the forget gate unit and the input gate unit, and c_t^j represents the memory content after the forget unit and the input unit are updated using the time t. c_t^j can be calculated by formula 3.

In formula 4 and 5, o_t^j represents h_t^j is the final LSTM output unit activated at time t.

$$o_t^j = \sigma(U_o x_t + W_o h_{t-1} + b_o)^j \tag{4}$$

$$h_t^j = o_t^j \tanh(c_t^j). \tag{5}$$

C. Attention model

Attention-based models have been successfully used in plenty of sequence-to-sequence learning tasks, such as speech recognition [30], part-of-speech tagging [31] and machine translation [32]. In fact, the attention mechanism is to select relevant encoded hidden vectors via attention weights (an informative sequence of weights) during the decoding phase. The architecture affords the possibility to construct an end-to-end system. In this paper, an attention mechanism aims to find the key part whose are more informative than others for intimacy prediction.

We calculate the attention weights α_i for each vector x_i in a sequence of inputs x, as follows:

$$\alpha_i = \frac{\exp(f(x_i))}{\sum_j \exp(f(x_j))}, \tag{6}$$

where $f(x)$ is the scoring function. Here, we use a linear function $f(x) = W^T X$. W in the linear function is a trainable parameter. The output of the attention layer is the weighted sum of the input sequence, which is denoted by $attention\, e_x$:

$$attention\, e_x = \sum_i \alpha_i x_i. \tag{7}$$

D. Intimacy classification

Vector $attention\, e_x$ represents a learned feature vector of an fNIRS sequence, which includes more discriminative and robust intimacy representation. And, it can be classified by:

$$p = soft\max(W_a ttention\, e_x v + b_c). \tag{8}$$

Our model is trained by minimizing the cross-entropy between the predicted label and the real label.

E. Attention-enhanced cascade convolutional recurrent neural network (ACCRNN) for fNIRS-based Intimacy Detection

The motivation of the proposed model is illustrated by three requirements of fNIRS-based intimacy prediction: a) Inspired by their performance in visual and speech recognition tasks, CNNs have been incorporated to extract features from raw signals. And CNNs are exceptionally good at capturing high-level representations in a spatial domain. b) The data of each sequence contains specific part of the complete intimate-induced brain activity. Thus, temporal information can be detected from the fNIRS signal. The final prediction result of intimate relationship is decided by sufficiently considering these contextual relationships. c) In addition to learning useful spatio-temporal features, it is also important to select the salient sections of an input signal to improve fNIRS-based intimacy prediction performance further. The use of attention mechanisms in RNN and CNN-based models has frequently been demonstrated as a useful tool to encourage a model to more heavily weight specific regions of an input sequence. While LSTMs are capable of modeling temporal dependencies in sequences, it is difficult for them to learn long temporal dependencies in long sequences. With the help of the attention mechanism [4], the LSTM-RNN can tackle this problem.

As shown in Fig. I, in our proposed model, we first extract the spatial features from preprocessed fNIRS signals, and then feed the sequence of the extracted spatial features into the LSTM-RNN to extract temporal features. An attention layer is designed for extracting the salient parts of a sequence. One fully connected layer receives the output of the attention layer, and feeds the softmax layer for final intimacy prediction.

IV. Experiments and Analysis

A. Data acquisition

1) Participants: In order to effectively analyse and infer to intimacy, forty-four healthy subjects were recruited for the experiment, 25 males and 19 females with an average age of 22.12 ± 2.51 years old and 20.4 ± 2.11 years old. All of the subjects are right-handed, with normal vision or corrected vision, no history of mental illness, and no major conflict with lovers during the week before the visit. Before the experiment, the principle of the instrument was introduced to ensure that it was harmless to the human body and does not involve any ethical issues. The participants were asked to sign the experimental informed consent form.

2) Stimuli: Previous works on intimacy, especially passionate relationships [33], [34], [35], usually obtain brain imaging in intimate relationships by allowing subject to recall the events from intimate relationship while watching lovers' photos. Inspired by these research efforts, we adopt the photos from the subject's lover, friends, and strangers to induce his/her brain imaging in different relationships. Each participant was asked to provide 20 photos (10 for lovers and 10 for friends). Thirty volunteers (15 male, 15 female, unrelated to the experiment) provided 60 photos (2 per person) as the induction of stranger photos. During the experiment, each subject views the photos from his/her lover, friends and strangers by random selection.

3) Instrumentation: In this paper, a near-infrared spectroscopy brain imager (LABNISR, Shimadzu Corporation, Kyoto, Japan) is used to record subject's fNIRS signals from 42 channels when he/she watch intimacy photos. The instrument monitors the cerebral cortex during the experiment by three kinds of semiconductor lasers with wavelengths of $780 \pm 5nm$, $805 \pm 5nm$, and $830 \pm 5nm$, and converts it into cortical hemoglobin concentration changes using the modified Beer-Lambert law. The sampling rate of the instrument used in the experiment is 11 Hz.

4) Experimental protocol: To study the intimacy elicited by the selected photos, we have designed an experiment of about 17 min length for each subject, which has been implemented using the EPrime 2.0 (Psychology Software Tools, Pittsburgh, PA). During the experiment, the subject is comfortably installed on a chair in front of a 21" computer screen, which is used for the presentation of the stimuli and which is placed at about 1m distance from the subject. The experiment starts with some general information on the experimental protocol and instructions for the subject to relax and to stay still as much as possible during the experiment. The experiment for each subject starts with a black screen shown for 120 s. A subject watching a group of photos including his/her lover, friend and stranger is denoted as a *block*. Each block is composed of 3 phases: watching friend's photo phase, watching lover's photo phase and watching stranger's photo phase. For each phase, there are three tasks for the subject as follow: a) gazing cross: a white fixation cross then appears for 300 ms in order to alert the subject to the beginning of the next photo;

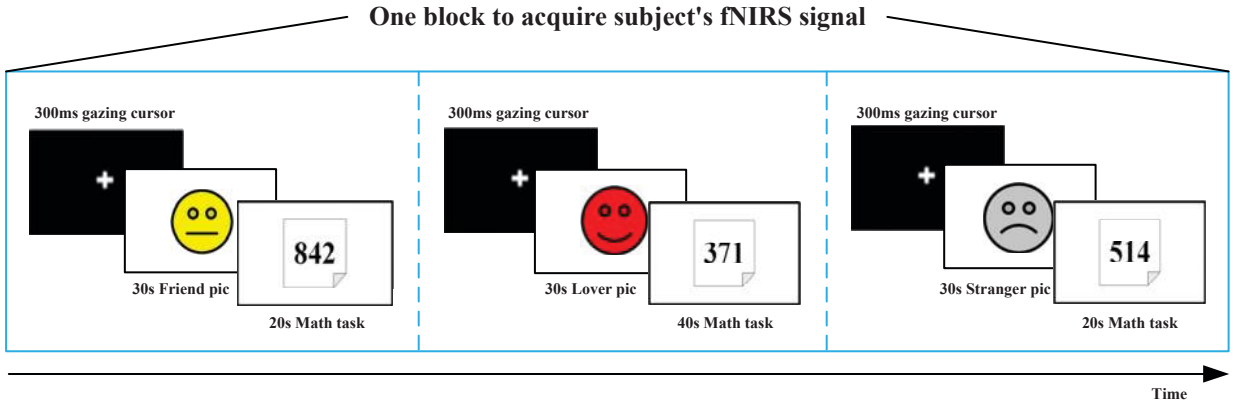


Fig. 2. Experimental paradigm for inducing a subject's fNIRS response with different intimate relationships

b) watching photo: the screen presents randomly a selected photo from his/her friend, lover, or stranger to the subject for 30 s, and the subject is asked to recall the happy events they experienced together for the corresponding people; c) math task: a random number (such as 842) is presented in the gap between pictures, with the subject being asked to cycle minus 7 until the number disappears. The purpose of this step is to allow the subject adequate time to, cognitively, eliminate the emotional stimulation after viewing the different pictures, and restore their neurophysiological baseline. The time length of the random number display is set to: 20 s after the picture of the friend and stranger, and 40 s after the picture of the lover. The purpose of different settings for the time of displayed random number is that a high-level arousal stimuli (e.g. lover's photo) requires more time to eliminate the emotional fluctuations and return to normal neurophysiological levels [34]. The experiment for each subject is composed of 6 blocks. An overview of the experimental paradigm associated with the database is shown in Fig. 2.

5) Preprocessing and Dataset Creation: For collected fNIRS data, the processing is implemented to reduce noise from a subject's head movements and instrument interference. In this experiment, a band-pass filter is used for preprocessing with a frequency range from 0.01 to 0.2 Hz. The processing result of the first subject viewing a lover's picture using the band-pass filter is given in Fig. 3; the blue line represents the original data, and the red line represents the data after bandpass filtering. It can be seen hat after band-pass filtering, the signal is smoother, and that the high-frequency noise is reduced.

Time stamps are recorded at the start of each experiment for both the stimuli and the fNIRS signals. The category to which the picture belongs to is used as label for induced intimacy. The subject's induced intimacy annotations are synchronized and paired with their respective fNIRS signals. For each subject, we obtain 18 segments of fNIRS signals induced by intimate relationship pictures (the window size is 30 seconds for each photo). Considering that deep learning methods require a large amount of training data to build a robust model, we obey the sample partitioning method in Zhang et al [36] to increase the number of samples. Smaller sliding windows (time=1, 2, 5 and 10 s) without overlap are set to crop samples. The final number of samples for each intimacy class with different time window lengths are given in Table I.

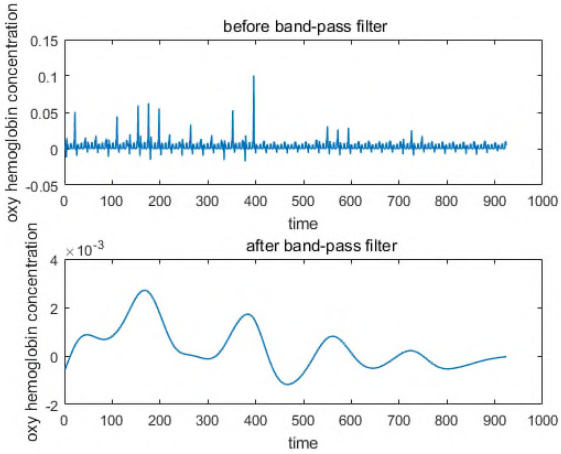


Fig. 3. Oxy hemoglobin concentration before and after using the bandpass filter when the first subject is viewing the lover's photo

TABLE I
INSTANCE DISTRIBUTION OVER THREE INTIMACY CLASSES-**F**RIEND, **L**OVER AND **S**TRANGER WITH DIFFERENT TIME WINDOW LENGTHS

Time window length	**#F**	**#L**	**#S**	Total
1 second	7,920	7,920	7,920	23,760
2 seconds	3,960	3,960	3,960	11,880
5 seconds	1,584	1,584	1,584	4,752
10 seconds	792	792	792	2,376
30 seconds	264	264	264	792

B. Network structure of our proposed method

The network structure and hyper-parameters of our proposed method in this paper are shown in Fig 4, which includes two

CNN layers, one LSTM layer, one attention layer, a fully connected layer and a softmax layer for intimacy detection. In the proposed model, the batch size and epoch are set to 32 and 200 respectively. For the two convolutional layers, the kernel size are 3*3 and the number of 32 and 16 respectively. The max-pooling layers are alternated between the CNN layer, which can increase the robustness of the features and reduce the dimensionality of the fNIRS signals vector. The size of the max pooling layer is 2*1 in order to preserve the information from each channel. After extracting spatial domain features by CNN layers, an LSTM layer is used to capture the temporal feature, and the number is set to 128. In order to prevent overfitting during training, a regularization term is applied and the dropout parameter is set to 0.2. The attention model is utilized to selectively learn these inputs by preserving the intermediate output of the input sequence by the LSTM encoder, and then training a model to selectively learn the inputs and correlate the output sequences with the model output. All learned features are fed into fully connected layer, then a softmax layer to achieve the output of intimacy prediction.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 110, 42, 32)	320
conv2d_2 (Conv2D)	(None, 110, 42, 16)	4624
max_pooling2d_1 (MaxPooling2	(None, 55, 42, 16)	0
reshape_1 (Reshape)	(None, 55, 672)	0
lstm_1 (LSTM)	(None, 55, 128)	410112
dropout_1 (Dropout)	(None, 55, 128)	0
attention_1 (Attention)	(None, 128)	183
dense_1 (Dense)	(None, 32)	4128
dropout_2 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 3)	99

Fig. 4. The network structure and hyper-parameters of our proposed method

In order to evaluate the intimacy detection, the cross-entropy loss function is set, which determines the degree of correspondence between the target output vector and the predicted output vector. In this paper, ReLU is employed as activation function. It can effectively avoid a vanishing gradient and in practice converges to the optimum point much faster. Consequently, it improves the training process of deep neural network architectures on large scale and complex data sets. In addition, the hyper-parameters for training all the proposed structures, including learning rate, number of epochs, and batch size, were chosen for each individual subject using Grid search. Adam is applied as a gradient descent optimization algorithm, whose parameters β_1, β_2, and ε are set to 0.8, 0.9, and 10^{-4}, respectively.

C. Experimental setting

In order to evaluate the performance, 10-fold cross validation is used to estimate the classification performance of the predictive model. In this paper, we compare the performance of the proposed method with the state-of-art approaches for intimacy prediction, including support vector machines (SVM), linear discriminant analysis (LDA), random forest (RF), CNN-based, LSTM-based and cascade convolutional recurrent neural network (CCRNN) methods. For shallow machine learning algorithms (e.g. SVM, LDA, RF), 168 intimacy-related features are extracted from preprocessed fNIRS signals, including mean, variance, kurtosis, skewness ($4\ features \times 42\ channels$). Since high-dimensional features usually suffer from performance degradation in classifiers, principle component analysis (PCA) is utilized to decrease the dimensions of features. Grid search is used to determine the number of principle components and the model parameters, which yields better performance. For deep learning algorithms, all neural networks are implemented with the Tensorflow library. For each method, we manually tune its parameters to achieve optimal performance.

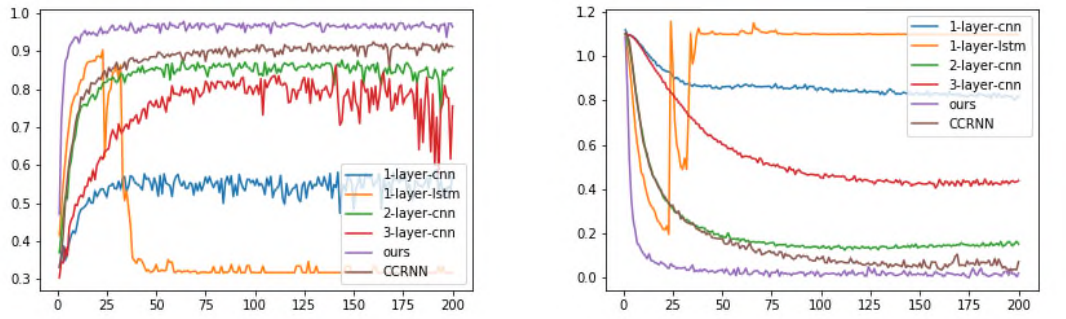
D. Experimental results

Table II shows that the performance results from SVM, RF, LDA, CNN-based, and LSTM-based methods with different instance lengths. Among shallow learning methods, RF achieves the best performance with the recognition precision of 34.4% when the instance length is 30 seconds. The SVM method achieves 69.2% and 60.5% recognition rate when the sample length is 1 second and 2 seconds respectively, which is the best performing classifier in the shallow learning method. For the CNN classifier, the 3-layer CNN achieves the best results with the sample length of 5 seconds and 10 seconds, respectively, and the recognition accuracies are 65.2% and 54.6%. We also see that the 2-layer CNN obtains the the recognition accuracy of 86.3% and 83.0% with the instance length of 1 second and 2 seconds, which is significantly higher than other CNN methods. For LSTM-based methods, a 1-layer LSTM is superior to 2-layer LSTM in recognition performance, especially with the instance length of 1 second. Due to the combination of the advantages of CNN and LSTM, CCRNN is superior to these two individual classifiers, when the instance length is 1 second. In general, deep learning methods are significantly better than shallow methods. With the increase of instances, the recognition performance based on deep learning methods generally shows an upward trend. Since our proposed method is a hybrid architecture based on CNN and LSTM, the number of CNN and LSTM layers is critical to the performance of the proposed method. Considering the promising performance of a 2-layer CNN and the 1-layer LSTM in a large number of instances, we choose a 2-layer CNN and the 1-layer LSTM to build our model. Note that our proposed method achieves the best performance with the accuracy of 97.4% and 94.8% with the instance length of 1 second and 2 seconds, respectively.

Fig. 5 depicts the experimental accuracy and loss curves during the testing for 200 epochs under different methods. As the number of epochs increases, the recognition performance of all methods except LSTM increases rapidly and tends to

TABLE II
PERFORMANCE ON DIFFERENT METHODS FOR FNIRS-BASED INTIMACY PREDICTION WITH DIFFERENT INSTANCE LENGTH

Classifer	Instance Length									
	1 second		2 second		5 second		10 second		30 second	
	ACC(%)	Loss	ACC(%)	Loss	ACC(%)	Loss	ACC(%)	Loss	ACC(%)	Loss
SVM	69.2	-	60.5	-	45.7	-	34.6	-	33.7	-
RF	41.5	-	40.8	-	38.6	-	37.1	-	**34.4**	-
LDA	37.8	-	37.1	-	35.0	-	34.3	-	33.6	-
1-layer CNN	73.7	0.656	53.8	0.974	38.0	1.135	35.6	1.572	33.1	1.971
2-layer CNN	86.3	0.401	83.0	0.496	54.2	1.837	40.1	2.854	30.7	3.248
3-layer CNN	67.4	0.821	60.9	0.866	**65.2**	**0.845**	**54.6**	2.312	31.8	7.572
1-layer LSTM	86.9	0.572	32.9	4.958	34.9	6.251	35.7	2.875	27.8	1.101
2-layer LSTM	50.1	0.877	33.8	9.465	34.7	7.263	32.4	3.821	28.4	**0.991**
CCRNN	95.7	0.178	80.1	0.539	47.1	1.013	38.6	4.567	32.9	2.802
Ours	**97.4**	**0.130**	**94.8**	**0.247**	38.6	3.903	36.2	**1.126**	27.9	1.167



(a) Accuracy curves

(b) Loss curves

Fig. 5. Accuracy and loss curves during the testing for 200 epochs under different methods

be stable. Relatively, the value of the loss rapidly decrease and then stabilizes. As can be seen from the figure, the recognition accuracy of our proposed ACCRNN approach is higher than other methods. For the LSTM model, when the number of epochs exceeds 25, its trend of recognition performance exhibits irregularity.

E. Discussions

From an overall experimental view point, the presented results demonstrate that our proposed model achieves the best performance for accuracy. However, we also notice that when the instance length is set to 5 to 30 seconds, the recognition performance seems to be not satisfactory. The main reason is that the deep learning method needs to be fed by a large amount of training data to obtain a robust predictive model. Not only for the method we proposed, but for all deep learning methods, we can clearly see that there is a trend of recognition performance increasing with the number of sample, specially, when the sample length is 1 second and 2 seconds, compared with the other lengths of instances, the performance has a huge improvement. In addition, due to the limitations of hand-crafted features on spatial and temporal representation, shallow learning methods, such as SVM, RF, LDA that rely on feature engineering are difficult to achieve satisfactory prediction results. Compared with the shallow learning method, deep learning methods have significantly improved overall recognition accuracy. The results from CNN- and LSTM-based methods imply that it is crucial to use either spatial or temporal information to boost intimacy prediction and analysis respectively. And the CCRNN method also provides an important evidence that the fusion of both temporal and spatial characteristics from brain activity is benefical to intimacy detection. In terms of improved performance, it is clear that an attention mechanism can improve the prediction accuracy of the cascade convolutional recurrent neural network modules.

V. CONCLUSION

The proposed cascade model in the paper is motivated by the existing progress on deep models, and takes advantage of CNN, LSTM, and the attention mechanism for intimacy prediction. With the proposed model, we achieved a potent improvement in the current state-of-the-art for the task of intimacy prediction on the fNIRS-based dataset. The increase in performance in comparison to other existing models shows that an attention mechanism can improve the performance of a cascade convolutional recurrent neural network for intimacy prediction. An overall analysis of the performance of our pro-

posed method was provided and compared to other techniques. In the future, we will expend our dataset by increasing the number of subjects to make it publicly available to the research community, and a multi-modal fusion method will also be investigated to further boost the performance of the intimacy prediction task.

ACKNOWLEDGEMENT

The work presented in this paper was substantially supported by the National Natural Science Foundation of China (Grant No: 61702370), the Key Program of the Natural Science Foundation of Tianjin (Grant No. 18JCZDJC36300), the Open Projects Program of the National Laboratory of Pattern Recognition, and the Senior Visiting Scholar Program of Tianjin Normal University. Chao Li and Qian Zhang share joint first authorship. Ziping Zhao and Li Gu are the corresponding authors (ztianjin@126.com, gulimail@gdmu.edu.cn) .

REFERENCES

[1] C. B. Burgoyne, "Conflict and decision-making in close relationships: Love, money and daily routines," *Journal of Socio-Economics*, vol. 31, no. 2, pp. 175–177, 2002.

[2] A. Bartels and S. Zeki, "The neural basis of romantic love," *Neuroreport*, vol. 11, no. 17, pp. 3829–3834, 2000.

[3] B. P. Acevedo, A. Aron, H. E. Fisher, and L. L. Brown, "Neural correlates of long-term intense romantic love," *Social cognitive and affective neuroscience*, vol. 7, no. 2, pp. 145–159, 2012.

[4] A. De Boer, E. Van Buel, and G. Ter Horst, "Love is more than just a kiss: a neurobiological perspective on love and affection," *Neuroscience*, vol. 201, pp. 114–124, 2012.

[5] L. M. Diamond and J. A. Dickenson, "The neuroimaging of love and desire: review and future directions." *Clinical Neuropsychiatry*, vol. 9, no. 1, pp. 1–8, 2012.

[6] S. Ortigue, F. Bianchi-Demicheli, A. de C. Hamilton, and S. T. Grafton, "The neural basis of love as a subliminal prime: an event-related functional magnetic resonance imaging study," *Journal of cognitive neuroscience*, vol. 19, no. 7, pp. 1218–1230, 2007.

[7] M. R. Bhutta, M. J. Hong, Y.-H. Kim, and K.-S. Hong, "Single-trial lie detection using a combined fnirs-polygraph system," *Frontiers in Psychology*, vol. 6, pp. 709: 1–9, 2015.

[8] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, R. Boots, and B. Benatallah, "Cascade and parallel convolutional recurrent neural networks on eeg-based intention recognition for brain computer interface," in *Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA, United States, 2018, pp. 1703–1710.

[9] S. K. Piper, A. Krueger, S. P. Koch, J. Mehnert, C. Habermehl, J. Steinbrink, H. Obrig, and C. H. Schmitz, "A wearable multi-channel fnirs system for brain imaging in freely moving subjects," *Neuroimage*, vol. 85, no. 2, pp. 64–71, 2014.

[10] C. Li, Z. Bao, L. Li, and Z. Zhao, "Exploring temporal representations by leveraging attention-based bidirectional lstm-rnns for multi-modal emotion recognition," *Information Processing & Management*, vol. 57, no. 3, pp. 102 185: 1–9, 2020.

[11] K. Ioulia, M. H. Shalinsky, M. S. Berens, and P. Laura-Ann, "Shining new light on the brain's 'bilingual signature': a functional near infrared spectroscopy investigation of semantic processing." *Neuroimage*, vol. 39, no. 3, pp. 1457–1471, 2008.

[12] G. Aranyi, F. Pecune, F. Charles, C. Pelachaud, and M. Cavazza, "Affective interaction with a virtual character through an fnirs brain-computer interface," *Frontiers in Computational Neuroscience*, vol. 10, no. 70, pp. 70: 1–14, 2016.

[13] F. Marco and Q. Valentina, "A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application," *Neuroimage*, vol. 63, no. 2, pp. 921–935, 2012.

[14] T. Trakoolwilaiwan, B. Behboodi, J. Lee, K. Kim, and J.-W. Choi, "Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a braincomputer interface: three-class classification of rest, right-, and left-hand motor execution," *Neurophotonics*, vol. 5, no. 1, pp. 1–15, 2017.

[15] U. Chaudhary, B. Xia, S. Silvoni, L. G. Cohen, and N. Birbaumer, "Brain-computer interface-based communication in the completely locked-in state," *PLOS Biology*, vol. 15, no. 1, pp. 1–25, 2017.

[16] S. Weyand, L. Schudlo, K. Takehara-Nishiuchi, and T. Chau, "Usability and performance-informed selection of personalized mental tasks for an online near-infrared spectroscopy brain-computer interface," *Neurophotonics*, vol. 2, no. 2, pp. 1–14, 2015.

[17] B. Abibullaev and J. An, "Classification of frontal cortex haemodynamic responses during cognitive tasks using wavelet transforms and machine learning algorithms," *Medical Engineering & Physics*, vol. 34, no. 10, pp. 1394–1410, 2012.

[18] C. Li, Q. Zhang, Z. Zhao, L. Gu, N. Cummins, and B. Schuller, "Analysing and inferring of intimacy based on fnirs signals and peripheral physiological signals," in *2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 2019, pp. 1–8.

[19] K. S. Hong, N. Naseer, and Y. H. Kim, "Classification of prefrontal and motor cortex signals for three-class fnirsbci," *Neuroscience Letters*, vol. 587, pp. 87–92, 2015.

[20] P. Y. Simard, D. Steinkraus, and J. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *the Seventh International Conference on Document Analysis and Recognition*, Washington, DC, United States, August 2003, pp. 1–6.

[21] S. Sukittanon, A. C. Surendran, J. C. Platt, and C. J. Burges, "Convolutional networks for speech detection," in *Interspeech / 8th International Conference on Spoken Language Processing*, Korea, 2004, pp. 1–4.

[22] Y. Bengio, "Learning deep architectures for ai," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[23] A. Villringer, J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl, "Near infrared spectroscopy (nirs): a new tool to study hemodynamic changes during activation of brain function in human adults," *Neuroscience letters*, vol. 154, no. 1-2, pp. 101–104, 1993.

[24] N. Naseer and K.-S. Hong, "fnirs-based brain-computer interfaces: a review," *Frontiers in human neuroscience*, vol. 9, p. 3, 2015.

[25] N. I. Eisenberger, M. D. Lieberman, and K. D. Williams, "Does rejection hurt? an fmri study of social exclusion," *Science*, vol. 302, no. 5643, pp. 290–292, 2003.

[26] V. Reindl, C. Gerloff, W. Scharke, and K. Konrad, "Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fnirs-based hyperscanning," *NeuroImage*, vol. 178, pp. 493–502, 2018.

[27] J. Benerradi, H. A. Maior, A. Marinescu, J. Clos, and M. L. Wilson, "Exploring machine learning approaches for classifying mental workload using fnirs data from hci tasks," in *Proceedings of the Halfway to the Future Symposium 2019*, 2019, pp. 1–11.

[28] A. M. Chiarelli, P. Croce, A. Merla, and F. Zappasodi, "Deep learning for hybrid eeg-fnirs brain–computer interface: application to motor imagery classification," *Journal of neural engineering*, vol. 15, no. 3, p. 036028, 2018.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *Advances in neural information processing systems*, Montral, Canada, 2015, pp. 577–585.

[31] O. Vinyals, Ł. Kaiser, T. Koo, S. Petrov, I. Sutskever, and G. Hinton, "Grammar as a foreign language," in *Advances in neural information processing systems*, Montral, Canada, 2015, pp. 2773–2781.

[32] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, vol. 1, pp. 1–15, 2014.

[33] A. Bartels and S. Zeki, "The neural correlates of maternal and romantic love," *Neuroimage*, vol. 21, no. 3, pp. 1155–1166, 2004.

[34] A. Aron, H. Fisher, D. J. Mashek, G. Strong, H. Li, and L. L. Brown, "Reward, motivation, and emotion systems associated with early-stage intense romantic love," *Journal of neurophysiology*, vol. 94, no. 1, pp. 327–337, 2005.

[35] X. Xu, A. Aron, L. Brown, G. Cao, T. Feng, and X. Weng, "Reward and motivation systems: A brain mapping study of early-stage intense romantic love in chinese participants," *Human brain mapping*, vol. 32, no. 2, pp. 249–257, 2011.

[36] J. Zhang, M. Chen, S. Hu, Y. Cao, and R. Kozma, "Pnn for eeg-based emotion recognition," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Budapest, Hungary, 2016, pp. 2319–2323.