Hybrid RNN-ANN Based Deep Physiological Network for Pain Recognition

Run Wang[1], Ke Xu[1], Hui Feng[2], and Wei Chen[1]

Abstract— **Quantitative assessment of pain is vital progress in treatment choosing and distress relief for patients. However, previous approaches based on self-report fail to provide objective and accurate assessments. For impartial pain classification based on physiological signals, a number of methods have been introduced using elaborately designed handcrafted features. In this study, we enriched the methods of physiological-signal-based pain classification by introducing deep Recurrent Neural Network (RNN) based hybrid classifiers which combines auto-extracted features with human-experience enabled handcrafted features. A bidirectional Long Short-Term Memory network (biLSTM) was applied on time series of pre-processed signals to automatically learn temporal dynamic characteristics from them. The handcrafted features were extracted to fuse with RNN-generated features. Finely selected features from biLSTM layer output and handcrafted features trained an Artificial Neural Network (ANN) to classify the pain intensity. The handcrafted features enhance the RNN classification performance by complementing RNN-generated features. With our accuracy reaching 83.3%, comparison results on an open dataset with other methods show that the proposed algorithm outperforms all of the previous researches with higher classification accuracy. Therefore, this research is a good demonstration of introducing hybrid features for pain assessment.**

I. INTRODUCTION

Pain is defined as a distressing experience complicated with tissue damage and cognitive suffering. Uncontrolled pain not only deteriorates the quality of life but also endangers the immune system, impedes healing after surgery and even exacerbates tumor growth [1]. Moreover, wrong pain assessment leads to disproportionate treatment, which may cause risks for the patients [2]. Therefore, valid and reliable pain assessment is necessary for choosing adequate treatment and relieving physical and psychological distress.

To date, researchers still fail to classify pain in an ordinate way. As a personal experience, pain is uttered in different ways because of the subjects' discrepancy in pain sensitivity so that it can be exaggerated or attenuated easily. Meanwhile, the multidimensional nature of pain also hinders pain classification development [3].

The predominant pain assessment guideline is self-report, i.e. the patient quantifies the pain level by himself. Nevertheless, this empirical method fails when respondents lose the cognitive or verbal ability, e.g., demented patients. In contrast, physiological signals, as spontaneous responses towards pain, exhibit the potential reliability to be the manifestation of pain.

Regarding the physiological signals, different classifiers have been developed to do the bio-signal classification practices. In terms of the classic classifiers, such as Support Vector Machine (SVM), Latent Dirichlet Allocation (LDA), and K-Nearest Neighbors (KNN), rely on empirical experiments and expert experience to extract useful features, and the classification performance is limited by the finite accepted handcrafted features. Nowadays, with the extensive application of deep learning algorithms in bio-signal feature extraction, suitable representations can be generated automatically; thereby, deep learning methods can complement and even substitute the classic methods due to convenience and effectiveness. Based on existing research, deep neural network, e.g., Convolutional Neural Network (CNN), outperforms the majority of classifiers in certain bio-signal domains [4]. Nevertheless, regarding the drawback of CNN, CNN ignores the temporal characteristics of the bio-signal series, which are of tremendous significance for physiological signals.

Given the problem above, deep Recurrent Neural Network (RNN) is introduced for bio-signal classification. Thanks to the structure that connections between nodes form a directed graph along a temporal sequence, RNN exhibits the temporal dynamic characteristic reliably and automatically.

Collectively, a Hybrid RNN classifier was employed for pain intensity classification in our research. RNN, which involves a bidirectional LSTM network, can generate the abstract time representations of the pain data stream; meanwhile, accepted handcrafted features, which were included in our research, have been considered powerful indexes for pain classification in clinical pain relevant research. By fusing the handcrafted features and RNN generated features, the expert experience and the temporal characteristics can be utilized and the classification performance can be improved. To be specific, the best representations were selected from all RNN generated features and finely selected handcrafted features, and trained another shallow Artificial Neural Network (ANN) to classify the pain intensity.

It should be noted that few work relevant to RNN is conducted in the pain binary classification field [5]. Due to the feature fusion strategy, the handcrafted features were combined with RNN generated characteristics so that the reliability and performance of the algorithm are enhanced. Our Leave-One-Subject-Out accuracy results were conducted

This work is supported by National Key RD Program of China (Grant No. 2017YFE0112000) and Shanghai Municipal Science and Technology Major Project (Grant No. 2017SHZDZX01), and Tengfei Undergraduate Research Opportunities Program of Fudan.

[1]Run Wang, Ke Xu and Wei Chen are with the Center for Intelligent Medical Electronics, School of Information Science and Technology, Fudan University, Shanghai samanthawangdl@gmail.com

[2]Hui Feng is with the Department of Electronic Engineering, Fudan University, Shanghai hfeng@fudan.edu.cn

on an open database, which exceed the majority of other results in previous researches and are comparable to the state-of-art outcome, without introducing non-physiological observations such as video stream.

The remainder of this work is organized as follows. The details of the dataset are illustrated in the second section. The preprocessing procedure, RNN, feature selection, and fusion methods are included in the third section. Subsequently, a detailed description of the results corresponding to the specific methods is exhibited in the fourth section.

II. DATASET

The experiments were conducted with the BioVid Heat Pain database [6]. Because the part B of the database contains abnormal data, the majority of the relevant researches choose part A as a database, including us. Prior to data collection, each subject's pain threshold and tolerance threshold were determined. Four temperatures between the two thresholds were chosen as stimulation's temperatures. Each subject was stimulated for 20 times for every stimulation intensity (P1 to P4). Together with 20 baseline measurements, a total number of 100 experiments were carried out for each subject. The following bio-signals were collected: (1) skin conductance (SC), (2) electrocardiogram (ECG), (3) electromyogram (EMG). The preprocessed data were cut into 5.5 seconds for every stimulus. In total, the dataset contained 8700 samples from 87 subjects, equally distributed in 5 classes: no pain (BLN), and pain levels P1 to P4. With the sampling rate of 512Hz, every data series have 2816 points.

The Leave-One-Subject-Out (LOSO) validation was carried out on the PartA of the Biovid Heat database, and the accuracy is compared with other LOSO results on the same database.

III. METHODS

A. Data Preprocessing

According to prior research, skin conductance (SC), electromyography (EMG), and electrocardiography (ECG) were chosen as target signals. Noise and artefacts within the recorded data were ruled out by applying appropriate signal preprocessing techniques. A third-order lowpass Butterworth filter with a cut-off frequency of 4 Hz was applied to the skin conductance signals. The EMG signals were filtered by applying a fourth-order bandpass Butterworth filter with a frequency range of [20, 50] Hz. Finally, a third-order bandpass Butterworth filter with a frequency range of [0.1, 50] Hz was applied to the ECG signals. Because of the group delay of the SC filter, the first 118 points were deleted. Aiming at synchronizing the EMG and ECG signals with the SC signals, the last 118 points of EMG and ECG were removed. In other words, every signal had 2703 points after preprocessing.

B. Deep Recurrent Neural Network

Recurrent Neural Networks comprise a dominant class of temporal predictors and classifiers. To allow information to be stored across arbitrary time lags, Long Short-Term Memory networks adopt 'forget gates' to fade the out-of-date information slowly and reset the memory blocks with prior and immediate information [7]. Bidirectional LSTM reinforces LSTM by learning time characteristics bidirectionally. Fig.1. shows the structure within a bidirectional LSTM network, where x_i is inputted backwardly and forwardly into the networks and the state of W_b is updated by forget gate and input gate's products based on its current state, input x_i, prior node's output C_i and a_i. To date, RNN has been applied widely in the domain of physiology signal, e.g., biLSTM was designed to extract pulse rate variability from photoplethysmography data streams automatically [8].

TABLE I
TRAINING LAYERS AND DETAILS

Layer	Details	
Input Layer	[SC, ECG, EMG]	
biLSTM Layer	Hidden Unit Size:	200
	State Activation Function:	tanh
	Gate Activation Function:	Sigmoid
Fully-connected Layer		
Tanh Layer		
Output Layer	Softmax	

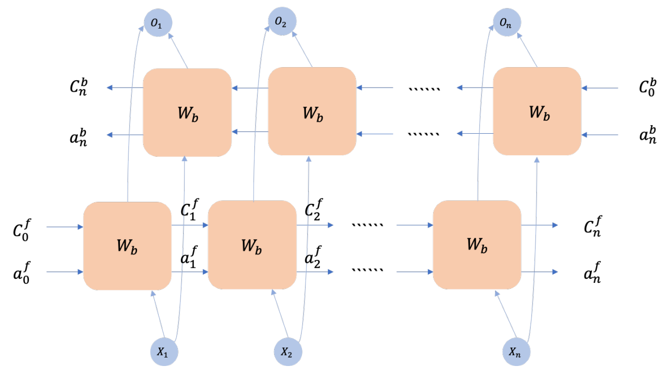

Fig. 1. Bidirectional LSTM Structure

In this study, the preprocessed SC, ECG, and EMG signals are connected in parallel and inputted into the network. Subsequently, a bidirectional LSTM is used for temporal representation generation. The softmax layer outputs the possibility of each category, and the basic RNN classification results are obtained. Then, the outputs of the biLSTM layer are extracted to fuse with handcrafted features.

To train this subject-independent network, data from the Biovid Heat Database were used, which include 87 subjects' responses towards four pain levels' stimulus and no pain stimulus for 20 times per level. The accuracy was tested on one left person, and the network was trained on the other 86 people. Hyper-parameters of RNN can be found in Table I.

C. Feature Extraction

This research has mainly focused on two RNN models: basic RNN model, and hybrid RNN-ANN model. The second model required handcrafted features to fuse with biLSTM layer output. The SC, ECG, EMG handcrafted features that widely used in prior pain researches were included in this research.

1) Skin Conductance Feature: The following features were extracted.

- Maximum
- Range
- Standard deviation
- Inter-quartile range
- Root mean square
- Mean
- Mean absolute value of the first differences

$$\frac{1}{N-1}\sum_{i=1}^{N-1}|x_{i+1}-x_i| \tag{1}$$

  where N is the number of the point in a single signal stream
- mean absolute value of the second differences

$$\frac{1}{N-2}\sum_{i=1}^{N-2}|x_{i+2}-x_i| \tag{2}$$

  where N is the number of the point in a single signal stream
- Mean absolute value of the first differences of the standardized signal
- Mean absolute value of the second differences of the standardized signal
- Skewness
- Kurtosis

2) Heart Rate Feature: Pan-Tompkins algorithm for QRS complex detection was employed to detect the R wave [9]. Then the following features were extracted:

- Mean of the interbeat intervals (IBI)
- Root mean square of the successive differences (RMSSD)
- Mean of the standard deviations of the IBIs (SDNN)

3) Electromyography Feature: The following features were extracted as follows.

- Maximum
- Standard deviation
- Mean absolute value of the first differences
- Mean absolute value of the second differences
- Mean absolute value of the first differences of the standardized signal
- Mean absolute value of the second differences of the standardized signal

D. Feature Selection and Fusion

The handcrafted features and RNN-generated features were fused to expect the final results. Because the biLSTM layer contains 200 units, the number of the output of the biLSTM layer is 400, which is much larger than the handcrafted features'. Besides, the significance of the handcrafted features and 400 RNN-generated features had not been weighed. Therefore, Minimum Redundancy Maximum Relevance method (MRMR) was employed to rank the significance of every feature to select the representations.

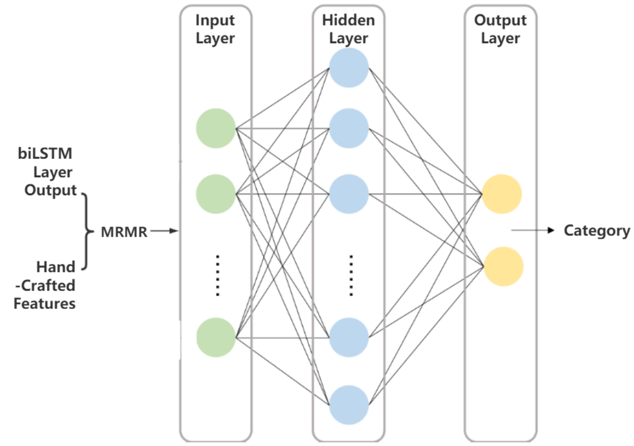

Fig. 2. ANN Fusion Structure

Minimum Redundancy Maximum Relevance method (MRMR) uses mutual information as a proxy for computing relevance and redundancy among characteristics and ranks the significance of features. The relevance of the set S and the redundancy of S are defined with mutual information I. The MRMR algorithm ranks features through the forward addition scheme by using the mutual information quotient (MIQ) value.

$$\max_{x\in S^c}\mathrm{MIQ}_x=\max_{x\in S^c}\frac{I(x,y)}{|S|}\sum_{z\in S}I(x,z) \tag{3}$$

MRMR was employed to select 50 features whose significance ranked highly among 400 biLSTM layer outputs and 21 handcrafted features. In this way, reliable representations were chosen, and adverse representations were excluded.

As is shown in Fig.2., 400 biLSTM layer outputs and 21 handcrafted features went through the feature selection procedure, and 50 features were chosen to be the input of the ANN. The ANN included a hidden layer with 100 units, and the predicted probabilities of each category were outputted in the output layer.

IV. RESULTS

Four methods are chosen in this study for comparison, which are based on random tree with physiological signals and video signals [11], random tree and SVM with the video signals [10], ANN [12] and random tree with physiological signals and video signals [13]. These methods are chosen because their methods represent the most existing methods in pain classification field, and they were conducted on the same database as our research; thus, it is reasonable to compare with these results.

The comparison results between methods and the proposed approach are illustrated in Table II. Our hybrid RNN method outperforms all methods because our results surpass theirs in every pain level. On the one hand, this research shares almost the same handcrafted features with other classic classification works; therefore, the limited ability of handcrafted features is proved by comparing the hybrid RNN result and classic

TABLE II
ACCURACY EVALUATION:COMPARE WITH OTHER METHODS

Approach	SIGNALS	BLN vs. P4	BLN vs. P3	BLN vs. P3	BLN vs. P1
Werner et al. [10]	Video	72.4	64.0	56.0	53.3
Werner et al. [11]	Physiological Signal and Video	80.6	72.0	60.5	49.6
Daniel et al. [12]	Physiological Signal	82.75	70.04	59.71	54.22
Kächele et al. [13]	Physiological Signal and Video	83.1	-	-	-
Our Approach (RNN)	**Physiological Signal**	**82.7**	**73.7**	**64.2**	**58.3**
Our Approach (Hybrid RNN-ANN)	**Physiological Signal**	**83.3**	**75.1**	**64.2**	**58.5**

results. Besides, Werner [11] has used video signals, which should have complemented the physiological indexes and overtaken other methods but failed. Kächele [13] used random forest methods and achieved a higher result, 83.1%, than our basic RNN result. However, he introduced video signals into his model. Thus it is understandable why his result surpasses our basic RNN's performance by a little margin. On the other hand, compared to Daniel's ANN method [12], our hybrid RNN results are superior to his results, and proved that the hybrid RNN-ANN method also reinforces the ANN performance considering that we shared almost the same handcrafted features.

Regarding our two classification methods, the hybrid RNN improves the basic RNN's performance in BLN vs. P4 and BLN vs. P3, and the other two results do not demonstrate a significant discrepancy. The RNN and handcrafted feature selection results can explain this. Among all of the 400 biLSTM layer outputs and 21 handcrafted features, two or three handcrafted features, including the skewness of SC and IBI, always ranked very high based on MRMR's significance score when classifying the BLN vs. P4 and BLN vs. P3. The handcrafted features complement the RNN generated features effectively and thus contribute to the hybrid RNN's performance. However, in the BLN vs. P2 and BLN vs. P1 experiments, far less handcrafted features were chosen after MRMR. In other words, when classifying low pain intensity, handcrafted features are not as reliable as RNN generated features are. Without the handcrafted features' influence, the majority of the ANN's inputs are RNN generated features; therefore, the hybrid RNN model exhibits quite similar results to the basic RNN's.

V. CONCLUSIONS

In this study, we first introduced deep Recurrent Neural Network (RNN) to extract temporal dynamic characteristics from the time series of pre-processed signals automatically. At the same time, 21 handcrafted features were obtained from skin conductance (SC), electromyography (EMG) and electrocardiography (ECG) signals. Features finely selected from the biLSTM layer and handcrafted features trained an ANN to classify pain intensity. Comparison results on an open dataset with other methods show that the proposed algorithm outperforms most of the previous researches with much higher classification accuracy, and the hybrid RNN model improves the basic RNN's high pain intensity classification performance by virtue of handcrafted features. Our strategy also achieves comparable results with state-of-the-art without involving additional video streams, which are required by the latter. To our best knowledge, this is the first time that RNN is conducted in the pain binary classification field only with physiological signal, and is a good example of introducing hybrid features for pain assessment.

REFERENCES

[1] M. E. Lynch, K. D. Craig, and P. W. Peng, *Clinical pain management: a practical guide*. Wiley Online Library, 2011.

[2] H. McQuay, A. Moore, and D. Justins, "Fortnightly review: treating acute pain in hospital," *BMJ*, vol. 314, no. 7093, p. 1531, 1997.

[3] M. H. Pitcher, M. Von Korff, M. C. Bushnell, and L. Porter, "Prevalence and profile of high-impact chronic pain in the united states," *The Journal of Pain*, vol. 20, no. 2, pp. 146–160, 2019.

[4] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2015.

[5] P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, and R. Picard, "Automatic recognition methods supporting pain assessment: A survey," *IEEE Transactions on Affective Computing*, pp. 1–1, 10 2019.

[6] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H. C. Traue, P. Werner, A. Al-Hamadi, S. Crawcour, A. O. Andrade, and G. M. da Silva, "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system," in *2013 IEEE international conference on cybernetics (CYBCO)*. IEEE, 2013, pp. 128–131.

[7] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[8] K. Xu, X. Jiang, H. Ren, X. Liu, and W. Chen, "Deep recurrent neural network for extracting pulse rate variability from photoplethysmography during strenuous physical exercise," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2019, pp. 1–4.

[9] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Trans. Biomed. Eng*, vol. 32, no. 3, pp. 230–236, 1985.

[10] P. Werner, A. Al-Hamadi, K. Limbrecht-Ecklundt, S. Walter, S. Gruss, and H. C. Traue, "Automatic pain assessment with facial activity descriptors," *IEEE Transactions on Affective Computing*, vol. 8, no. 3, pp. 286–299, 2016.

[11] P. Werner, A. Al-Hamadi, R. Niese, S. Walter, S. Gruss, and H. C. Traue, "Automatic pain recognition from video and biomedical signals," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 4582–4587.

[12] D. Lopez-Martinez and R. Picard, "Multi-task neural networks for personalized pain recognition from physiological signals," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2017, pp. 181–184.

[13] M. Kächele, M. Amirian, P. Thiam, P. Werner, S. Walter, G. Palm, and F. Schwenker, "Adaptive confidence learning for the personalization of pain intensity estimation systems," *Evolving Systems*, vol. 8, no. 1, pp. 71–83, 2017.