# An Autoencoder-based Approach to Predict Subjective Pain Perception from High-density Evoked EEG Potentials

Jiahao Wang, Mengying Wei, Li Zhang, Gan Huang, Zhen Liang, Linling Li, and Zhiguo Zhang

*Abstract*— **Pain is a subjective experience and clinicians need to treat patients with accurate pain levels. EEG has emerged as a useful tool for objective pain assessment, but due to the low signal-to-noise ratio of pain-related EEG signals, the prediction accuracy of EEG-based pain prediction models is still unsatisfactory. In this paper, we proposed an autoencoder model based on convolutional neural networks for feature extraction of pain-related EEG signals. More precisely, we used EEGNet to build an autoencoder model to extract a small set of features from high-density pain-evoked EEG potentials and then establish a machine learning models to predict pain levels (high pain vs. low pain) from extracted features. Experimental results show that the new autoencoder-based approach can effectively identify pain-related features and can achieve better classification results than conventional methods.**

*Index Terms*— **EEG, pain, deep learning, autoencoder, laser-evoked potentials**

## I. Introduction

Pain is a subjective feeling, and self-report is the gold standard to assess pain [1]. Clinically, patients often rely on a pain scale (for example, "0" means no pain and "10" means unbearable pain) to communicate their degree of pain with doctors. However, self-report is subjective and could lead to some serious clinical problems. For example, some patients (such as dementia patients, infants, and patients with severe coma) are not able to report their pain and some others may deliberately provide false pain scores [2]. So, it is necessary to develop new objective and reliable pain assessment tools for accurate prediction of pain.

Various functional brain imaging techniques, such as electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI), have been widely applied to study the neural mechanisms of pain and to develop objective pain assessment tools [3-5]. Given the high time resolution and low cost of EEG, it is more common to develop an EEG-based pain prediction model [6, 7]. Normally, such pain prediction models are based on EEG potentials evoked by pain stimulation. For example, in laser-evoked pain experiments, a set of pain-related features can be extracted from laser-evoked potentials (LEP), such as N2 (180 to 300 ms), P2 (250 to 500 ms) and gamma-band event-related synchronization (180 to 260 ms, 60 to 85 Hz) [8, 9]. Subsequently, pain prediction models can be established based on these pain-related EEG features [10-12].

However, the signal-to-noise ratio (SNR) of pain-evoked potentials is very low, so it is difficult to accurately extract pain-related features of evoked EEG [13]. Also, conventional feature extraction methods are heavily dependent on prior knowledge of pain-related EEG markers. With the rapid development of deep learning techniques, data-driven and end-to-end EEG feature extraction methods have gained increasing popularity. Previous researchers have proposed an EEGNet model based on artificial neural networks for the classification of EEG data of brain-computer interfaces [14]. It has been shown that the convolutional neural network (CNN) layers in the EEGNet model and others can effectively extract accurate features from EEG data [15]. However, these deep learning methods have not been used in the application of pain prediction.

In the present study, we proposed to use a CNN-based autoencoder (AE) method to extract pain-related EEG features in a data-driven manner. The new AE method is based on the EEGNet, and it can explore the relationship between pain scores and EEG features in an unsupervised manner. By convolution and deconvolution to encode and decode EEG signals, the AE-based method is capable of reducing high-dimensional EEG data to low-dimensional features. This method can extract EEG features in relation to pain and further improve the accuracy of machine learning prediction.

## II. Materials and Methods

### A. EEG Data and Experimental Design

Twenty-nine participants (9 females and 20 males) were enrolled in this experiment, aged from 17-25 years (22.2 ± 1.9 years). All participants had no history of chronic pain. Meanwhile, all participants were informed with written agreement. The experiment program was approved by local ethics committee. The infrared neodymium yttrium aluminum perovskite (Nd:YAP) laser (Electronic Engineering, Italy) generated a radiant-heat stimuli of 1.34 μm that directly activates the nociceptive terminals in the most superficial skin layers on the back of the left hand. Each subject received laser stimulation with of 4 different energy levels (E1: 2.5 J; E2: 3 J; E3: 3.5 J; E4: 4 J), each with ten trials. Stimulation sequence is a pseudo-random process. Participants reported their pain ratings (ranged from 0 to 10, "0" means no feeling and "10" means unbearable pain) after each stimulus. Laser beams were randomly moved about 1 cm, to avoid nociceptor fatigue or sensitization after each stimulus.

*Research supported by the National Natural Science Foundation of China (Grant No. 81871443 ).

Jiahao Wang, Mengying Wei, Li Zhang, Gan Huang, Zhen Liang, Linling Li, and Zhiguo Zhang are with the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China. Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen University, Shenzhen 518060, China. Peng Cheng Laboratory, Shenzhen 518055, China. (phone: +86-755-26917125; Fax: +86-755-26534314; E-mail: zgzhang@szu.edu.cn)

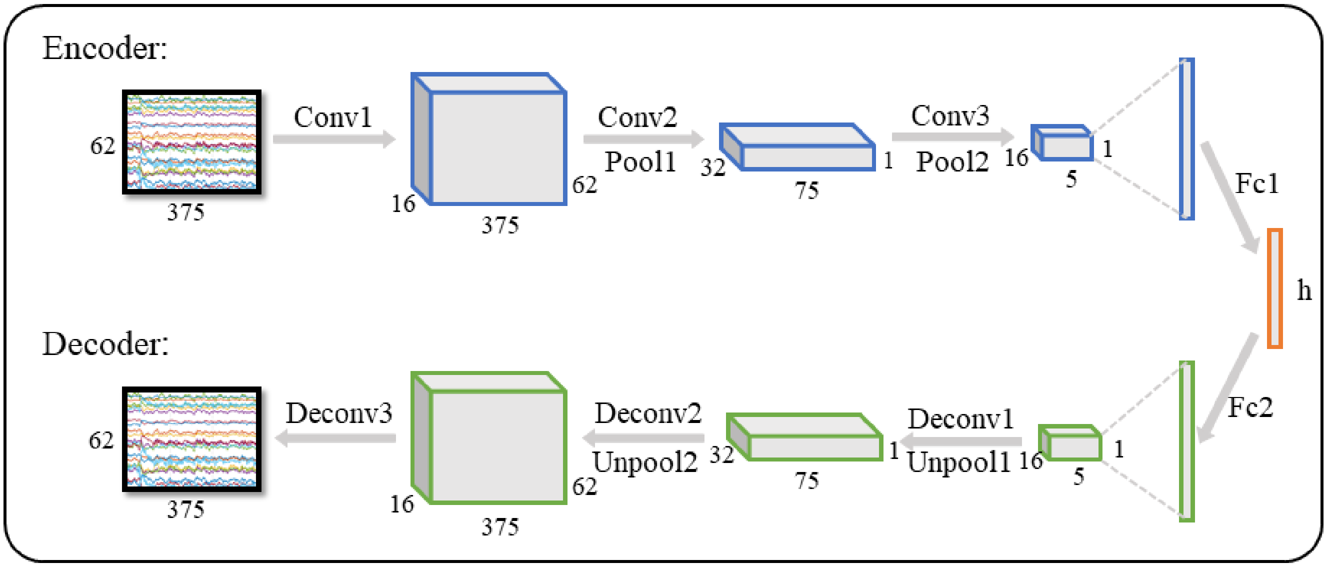


Fig. 1. The proposed AE model based on EEGNet [14].

All participants were seated in a quiet room, and they were asked to focus on laser stimulation and relax their muscles. EEG data were recorded using a 64 AgCl-channel Brain Products system, using the nose as reference (Brain Products GmbH, Munich, Germany; bandpass filter, 0.01 to 100Hz; sampling rate, 1000Hz), all channel impedances were kept lower than 10 kΩ. In order to monitor ocular movements and eye blinks, electrooculographic signals were simultaneously recorded from 4 bipolar electrodes: one pair placed over the upper and lower eyelids, and another pair placed 1 cm lateral to the outer corner of the left and right orbits.

### *B. Data preprocessing*

We use LETSWAVE toolbox to preprocess the data. Sixty-two-channel continuous EEG data were band-pass filtered between 1 and 30Hz. EEG epochs of laser-evoked trials were extracted from 0.5 s before stimulus to 1.0 s after stimulus, and each epoch was baseline-corrected by subtracting the mean of pre-stimulus data. Then, independent component analysis (ICA) was applied to remove eye artifacts [16]. And EEG data is down-sampled from 1000Hz to 250Hz to reduce the feature amount of data.

### *C. Feature extraction*

#### *1) Feature extraction based on Prior Knowledge*

We select the amplitude of the N2 wave of LEP as a feature for machine learning prediction, which is measured as the mean value of the post-stimulus period from 0.18s to 0.30s. N2 amplitudes of all 62 channels were extracted.

#### *2) Feature extraction based on AE*

The proposed autoencoder (AE) model has two parts. (1) Encoding: The input is an *i*-dimensional signal *x* which is processed by the hidden layer of the neural network to obtain *m*-dimensional data after intermediate dimensionality reduction. (2) Decoding: The original *i*-dimensional signal can be reconstructed through the hidden layer. Next, we minimize the error between the reconstructed signal and the input signal until the model converges. Last, we can obtain the intermediate parameter *h*.

The function of the encoder is defined as

$$h = \sigma(Wx + b) \qquad (1)$$

where $x$ is input data, $h$ is output of encoder, $W$ is the parameter of the hidden layer in the encoding process, and $b$ is a bias vector. In this work, $\sigma(\cdot)$ is an elu non-linear activation function.

The function of the decoder is defined as

$$y = \sigma(W'h + b') \qquad (2)$$

where $y$ is the output of decoder, and $W'$ and $b'$ are the hidden layer parameters and bias vector during the decoding process, respectively.

By continuously minimizing the error $L = |x - y|$ between the reconstructed signal and the original signal, the model is converged. We used the Adam algorithm to optimize the loss function $L$.

As shown in Figure 1, the proposed AE model is based on EEGNet [14], which is used to reduce the high-dimensional full-channel EEG data to a low-dimensional space without well-known prior knowledge to facilitate the prediction of machine learning models. In our AE model, based on the EEGNet model, a CNN is used to extract time and spatial domain information during the encoding process, and the use of a deep separable convolution layer can greatly reduce the number of network parameters and speed up the training process. Pooling and fully connected layers reduce the data feature dimension to a preset dimension. Then, in the decoder process, the extracted features after encoder are used to reconstruct EEG signal by up-sampling and deconvolution. By calculating the gradient of the difference between the reconstructed signal and the input signal, the model parameters can be updated iteratively until convergence. At this time, we use the encoded data $h$ for machine learning model training and prediction.

Table 1 shows the parameter settings for convolution and pooling of each layer in the AE model. In the model, we omit the activation function of each layer and the batchnorm layer in order to accelerate the model convergence. It should be noted that the Conv1 convolution layer relies on a convolution

TABLE 1
AE MODEL PARAMETERS

| Layer | Details of processing and parameters |
|---|---|
| Conv1 | Convolutional kernel size: (1,125) |
| Conv2 | Convolutional kernel size: (62,1)<br>Pooling: (1,5) |
| Conv3 | Convolutional kernel size: ((1,15), (16,1))<br>Pooling: (1,15) |
| Fc1 | Reshape<br>Fully connected layer |
| Fc2 | Fully connected layer<br>Reshape |
| Deconv1 | Unpooling: (1,15)<br>Deconvolutional kernel size: ((16,1), (1,15)) |
| Deconv2 | Unpooling: (1,5)<br>Deconvolutional kernel size: (62, 1) |
| Deconv3 | Deconvolutional kernel size: (1,125) |

kernel on a time scale with padding operation to extract time information and the Conv2 convolution layer relies on a convolution kernel on a spatial scale, and there is no padding operation during the convolution process. In addition, the Conv3 layer is a deeply separable convolution layer. We used two ordinary convolution layers instead of this special layer, which can significantly reduce the number of model parameters. The decoding process uses up-sampling and deconvolution corresponding to the encoding process to reconstruct the signal, respectively.

*3) Feature extraction based on PCA and MDS*

To compare the results obtained by the proposed AE model, we use the principal component analysis (PCA) method [15] and multidimensional scaling (MDS) method [17] to reduce the dimensionality of multi-channel EEG data and then use the features obtained by PCA and MDS to predict pain levels.

### *D. Classification*

For our 29 participants, we used the leave-one-subject-out cross-validation to make predictions, and calculated the accuracy (Acc) of each participant's prediction as an evaluation index. To examine the robustness of the proposed AE method, we used different machine learning models, including k-Nearest Neighbour (k-NN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Logistic Regression (LR), to train and predict the different types of features obtained.

All machine learning methods were trained based on the scikit-learn of python. The AE model proposed were realized based on the Pytorch deep learning framework, and were run on a GeForce GTX 1080Ti GPU.

TABLE 2
PREDICTION ACCURACY OF DIFFERENT FEATURES

| **AE (dimension of features)** | **k-NN** | **SVM** | **LDA** | **LR** |
|---|---|---|---|---|
| 2 | 63.5±11.7 | 71.8±20.2 | 71.9±20.3 | 72.0±20.3 |
| 4 | 66.7±12.7 | 71.8±20.2 | 69.7±15.7 | 69.6±15.7 |
| 8 | 67.0±11.5 | 71.8±20.2 | 70.3±14.5 | 70.7±14.4 |
| 16 | 67.0±12.5 | 70.3±19.6 | 68.2±16.5 | 69.5±14.9 |
| 32 | 65.9±11.0 | 68.5±16.0 | 71.5±12.5 | 72.1±12.4 |
| 64 | **68.4±10.3** | **73.4±11.3** | **73.9±10.4** | **74.6±11.2** |
| 128 | 68.1±10.7 | 71.6±13.0 | 72.0±11.9 | 71.8±13.2 |
| 256 | 67.9±9. 9 | 71.9±12.0 | 72.4±11.6 | 73.4±11.5 |
| N2 | 69.1±14.5 | 69.0±14.0 | 68.0±16.0 | 69.1±16.7 |

### *E. Statistical analysis*

We used paired t-test to test whether the performances of four classifiers based on features extracted by the proposed AE model or based on N2 is significant or not. Next, we compared the significance of the difference between the prediction results obtained by AE and the prediction results obtained by the PCA and MDS.

## III. Results

Table 2 shows that, when the AE model encoded the signal to a 64-dimensional feature vector, the best prediction performance can be achieved. As seen in Figure 2, the classification accuracies of AE-64 (64-dimensional features encoded by the AE model) in SVM, LDA and LR were all significantly higher than the conventional features N2 of all channels.

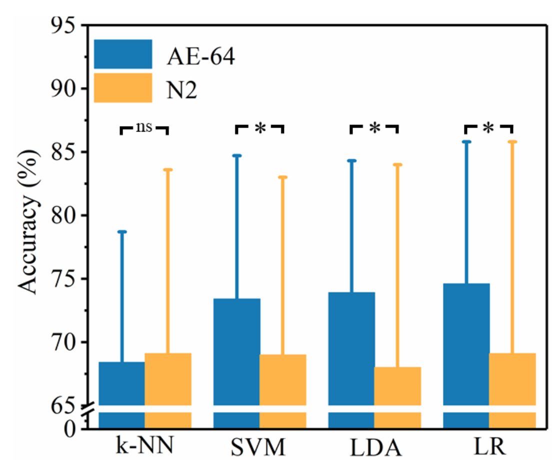


Fig. 2. Comparisons of classification accuracies between AE-64 and N2 for 4 different classifiers. “AE-64” represents 64-dimensional features encoded using the AE model, and “N2” represents N2 amplitude features of all 62 channels. Four classifiers, k-NN (k = 3), SVM (linear kernel), LDA, LR

(logistic regression) were used to predict pain in EEG features. In the figure, ns represents no significant difference, and * represents a significant difference of paired t-test results between two groups of data.

TABLE 3
CLASSIFICATION ACCURACY BASED ON FEATURES EXTRACTED BY THE PROPOSED AE MODEL, PCA, AND MDS

| Dimension of Features | PCA | MDS | AE |
|---|---|---|---|
| 2 | 66.8±19.9 | 67.5±19.0 | 72.0±20.3 |
| 4 | 67.0±20.2 | 67.8±19.7 | 69.6±15.7 |
| 8 | 66.6±19.4 | 66.5±19.1 | 70.7±14.4 |
| 16 | 65.8±19.6 | 65.6±18.4 | 69.5±14.9 |
| 32 | 63.5±19.8 | 64.2±18.7 | 72.1±12.4 |
| 64 | 58.9±19.0 | 63.2±17.8 | **74.6±11.2** |
| 128 | 50.9±18.6 | 56.8±17.3 | 71.8±13.2 |
| 256 | 42.7±18.0 | 53.6±13.4 | 73.4±11.5 |

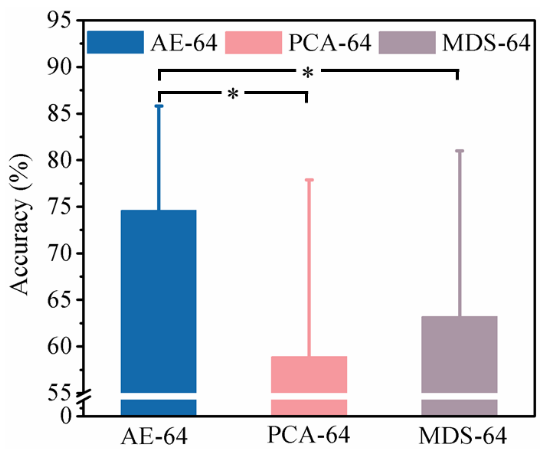


Fig. 3. Comparison of classification accuracies between the AE model, PCA and MDS, all of which encode or reduce the original EEG data to 64 dimensions. The logistic regression classifier was used to make predictions. * means a significant difference between the two groups of data.

Further, we used MDS and PCA to reduce the dimensionality of multi-channel EEG data for pain prediction. The results are shown in Table 3 and Figure 3. As shown in Figure 3, when the dimensionality of features was 64, the proposed AE method had significantly higher performance than PCA and MDS.

## IV. CONCLUSION

Due to the very low SNR as well as the remarkable inter-individual difference of pain-related EEG features, traditional pain prediction methods that mainly rely on well-known features from prior knowledge cannot achieve satisfactory results. In this study, we proposed a new data-driven EEGNet-based AE method to extract features from EEG data. As compared with traditional feature extraction and dimensionality reduction methods such as PCA, the proposed AE method can achieve significantly higher accuracy when classifying high-pain and low-pain EEG epochs. This study shows that the proposed AE model and other deep learning methods can be potentially used for EEG-based pain prediction.

## REFERENCES

[1] M. Ploner, C. Sorg, and J. Gross, "Brain Rhythms of Pain," *Trends Cogn Sci,* vol. 21, no. 2, pp. 100-110, Feb 2017.

[2] D. A. Engemann *et al.*, "Robust EEG-based cross-site and cross-protocol classification of states of consciousness," *Brain,* vol. 141, no. 11, pp. 3179-3192, Nov 1 2018.

[3] Y. Tu, B. Zhang, J. Cao, G. Wilson, Z. Zhang, and J. Kong, "Identifying inter-individual differences in pain threshold using brain connectome: a test-retest reproducible study," *Neuroimage,* vol. 202, p. 116049, Nov 15 2019.

[4] Y. Bai, G. Huang, Y. Tu, A. Tan, Y. S. Hung, and Z. Zhang, "Normalization of Pain-Evoked Neural Responses Using Spontaneous EEG Improves the Performance of EEG-Based Cross-Individual Pain Prediction," *Front Comput Neurosci*, vol. 10, p. 31, 2016.

[5] Q. Lin, L. Li, J. Liu, W. Liu, G. Huang, and Z. Zhang, "Influence of Individual Differences in fMRI-Based Pain Prediction Models on Between-Individual Prediction Performance," *Front Neurosci*, vol. 12, p. 569, 2018.

[6] Z. G. Zhang, L. Hu, Y. S. Hung, A. Mouraux, and G. D. Iannetti, "Gamma-band oscillations in the primary somatosensory cortex--a direct and obligatory correlate of subjective pain intensity," *J Neurosci*, vol. 32, no. 22, pp. 7429-38, May 30 2012.

[7] L. Hu, P. Xiao, Z. G. Zhang, A. Mouraux, and G. D. Iannetti, "Single-trial time-frequency analysis of electrocortical signals: baseline correction and beyond," *Neuroimage,* vol. 84, pp. 876-87, Jan 1 2014.

[8] L. Hu and G. D. Iannetti, "Neural indicators of perceptual variability of pain across species," *Proc Natl Acad Sci U S A,* vol. 116, no. 5, pp. 1782-1791, Jan 29 2019.

[9] A. Mouraux, J. M. Guérit, and L. Plaghki, "Non-phase locked electroencephalogram (EEG) responses to CO2 laser skin stimulations may reflect central interactions between A∂- and C-fibre afferent volleys," *Clinical Neurophysiology,* vol. 114, no. 4, pp. 710-722, 2003.

[10] Y. Tu *et al.*, "Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming nociceptive stimuli," *Hum Brain Mapp,* vol. 37, no. 2, pp. 501-14, Feb 2016.

[11] Z. G. Zhang, L. Hu, Y. S. Hung, A. Mouraux, and G. D. Iannetti, "Gamma-band oscillations in the primary somatosensory cortex--a direct and obligatory correlate of subjective pain intensity," *J Neurosci,* vol. 32, no. 22, pp. 7429-38, May 30 2012.

[12] L. Li, G. Huang, Q. Lin, J. Liu, S. Zhang, and Z. Zhang, "Magnitude and Temporal Variability of Inter-stimulus EEG Modulate the Linear Relationship Between Laser-Evoked Potentials and Fast-Pain Perception," *Front Neurosci,* vol. 12, p. 340, 2018.

[13] A. Mouraux and G. D. Iannetti, "Across-trial averaging of event-related EEG responses and beyond," *Magnetic Resonance Imaging*, vol. 26, no. 7, pp. 1041-1054.

[14] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *J Neural Eng,* vol. 15, no. 5, p. 056013, Oct 2018.

[15] T. Wen and Z. Zhang, "Deep Convolution Neural Network and Autoencoders-Based Unsupervised Feature Learning of EEG Signals," *IEEE Access,* vol. 6, pp. 25399-25410, 2018.

[16] G. Huang, P. Xiao, Y. S. Hung, G. D. Iannetti, Z. G. Zhang, and L. Hu, "A novel approach to predict subjective pain perception from single-trial laser-evoked potentials," *Neuroimage,* vol. 81, pp. 283-293, Nov 1 2013.

[17] P. Krauss et al., "Analysis of Multichannel EEG Patterns During Human Sleep: A Novel Approach," *Front Hum Neurosci*, vol. 12, p. 121, 2018.