The existing emotion recognition techniques based on the analysis of the tone of voice or facial expressions do not possess sufficient specificity and accuracy. These parameters can be significantly improved by employing physiological signals that escape the filters of human consciousness. The aim of this work was to carry out an EEG-based binary classification of emotional valence using a convolutional neural network and to compare its performance to that of a random forest algorithm. A healthy 30-year-old male was recruited for the experiment. The experiment included 10 two-hour-long sessions of watching videos that the participant had selected according to his personal preferences. During the sessions, an electroencephalogram was recorded. Then, the signal was cleared of artifacts, segmented and fed to the model. Using a neural network, we were able to achieve a F1 score of 87%, which is significantly higher than the F1 score for a random forest model (67%). The results of our experiment suggest that convolutional neural networks in general and the proposed architecture in particular hold great promise for emotion recognition based on electrophysiological signals. Further refinement of the proposed approach may involve optimization of the network architecture to include more classes of emotions and improvement of the network’s generalization capacity when working with a large number of participants.

Keywords: machine learning, artificial neural network, electroencephalogram, emotional state, valence, deep learning, convolutional networks

Author contribution: Savinov VB, Botman SA, Sapunov VV, Petrov VA — data acquisition and processing, manuscript preparation; Samusev IG — manuscript preparation and revision; Shusharina NN — project supervision and manuscript revision.

Compliance with ethical standards: this study was approved by the Ethics Committee of Immanuel Kant Baltic Federal University (Protocol № 7 dated March 26, 2019). The participants gave informed consent. The data was collected between March 26 and June 20, 2019. The collected data was stored in a database and deleted no later than 6 months after the study’s completion. The obtained data was used to perform scientific analysis only.

Correspondence should be addressed: Natalya N. Shusharina
Universitetskaya 2, Kaliningrad, 236006; nnshusharina@gmail.com

Received: 21.03.2019 Accepted: 16.05.2019 Published online: 29.06.2019

DOI: 10.24075/vrgmu.2019.037

Emotions play a crucial role in our daily lives, affecting our perception, decision-making and social interactions. Emotions can have an outward manifestation, showing in the tone of voice or a facial expression, as well as evoke physiological changes invisible to the naked eye. Although self-assessment studies do yield useful information, there are certain issues with the reliability and validity of the obtained data [1]. Because both voice and facial expression can be mimicked, they can hardly serve as reliable indicators of a person’s emotional state [2]. In contrast, the analysis of physiological signals fosters our understanding of basic emotional responses and the underlying biological mechanisms [3].
The following biosignals are commonly used to analyze a person’s emotional state: galvanic skin response (GSR), electromyogram (EMG), heart rate (HR), respiratory rate (RR), and electroencephalogram (EEG). Of them, EEG is the most interesting; it reflects the activity of the cerebral cortex that shapes a number of emotional responses. Although EEG has low spatial resolution, its temporal resolution is quite high, meaning that changes to the phase and frequency of the signal can be conveniently measured following exposure to an external emotional stimulus. Besides, electroencephalography is advantageous in terms of noninvasiveness, fast and inexpensive, in comparison with other techniques for the acquisition of bioclinical data.

As a rule, the EEG-based research into emotion recognition involves classification of a relatively small number of discrete states evoked by a specific stimulus. Usually, a raw EEG signal is passed through a filter first, and then features are extracted from it; classification is performed using a machine learning algorithm. The efficacy of this approach is largely determined by how features, which are the mathematically calculated signal attributes, are constructed and selected. Normally, feature construction accounts for the experimental and theoretical data on brain biology and the processes that an emotional stimulus induces in the brain.

The selected features are used to form feature vectors for machine learning models, such as random forests, multilayer perceptrons, support vector machines, k-nearest neighbors, etc. [4–6]. Classification accuracy of the listed models varies depending on the quality of input data, task criteria and the choice of a learning algorithm. In the case of an EEG signal, classification accuracy can be as high as 77% [7], whereas for multimodal signals it increases up to 83% [8]. The outcome is largely determined by the choice of a model and its parameters, signal features, and the techniques used to reduce data dimensionality.

Convolutional neural networks and deep learning are an alternative to the aforementioned algorithms for EEG-based emotion recognition. Neural networks are successfully used to process electrophysiological signals [9]. At present, the analysis of EEG signals by convolutional networks is employed to solve a variety of medical tasks and aid brain-computer interactions, including seizure prediction [10], detection of P300 waves [11], and recognition of emotions [12, 13] using DEAP datasets [14]. Such wide range of tasks proves that convolutional neural networks are a versatile and robust tool. Deep learning allows the optimal features to be automatically generated during the training process and, therefore, cancels the need for manual feature selection. Still, some authors use externally computed features [15] and Fourier and wavelet transforms [16, 17] as input data for a neural network.

The aim of this work was to carry out an EEG-based binary classification of emotional valence by creating and training a convolutional neural network and to compare its performance to that of a random forest algorithm with explicitly formed feature vectors.

METHODS

One of the authors of this work, a healthy male aged 30 years without a history of psychiatric disorders, volunteered to be an experimental subject. He underwent multiple sessions that involved exposure to different stimuli and lasted a few weeks. The design of our experiment was an adaptation of a well-known method [18]. It is common to recruit more than one participant and conduct a single session with each of the subjects. In our case, although the data were collected from one subject, the total data amount was quite substantial because the sessions were repeated multiple times. People differ in their emotional response to the same stimulus, and this diversity complicates the identification of physiological patterns that correspond to certain emotional states. When only one participant is engaged in a series of experiments, the interpretability of data improves significantly because the perception style remains unchanged.

The initial set of videos used to elicit a positive or negative emotional state was compiled based on the personal preferences of the study participant. The participant was familiar with the selected videos, therefore his emotional state during the sessions was largely determined by his inner psychic life, memories, etc. and not by the external stimulus as such. Later, the set was expanded to include additional videos that were categorized as emotionally positive or negative based on their resemblance to the original films. Data acquisition took 2 weeks. Ten two-hour long sessions were conducted. During a session that also included breaks, the participant watched six 15-min long videos (two from each category). Each video was played only once. A positive video was always followed by a negative.

EEG signals were recorded using a previously developed neurodevice [19], Ag/AgCl electrodes with a conductive gel were attached to a special cap for EEG at positions F3, F4, C3, C4, P3, P4, O1, O2 (the 10–20 system), as required by the monopolar recording technique. The electrode at position Fpz was used as a ground and reference. The sampling rate was 250 Hz. Second-order Butterworth filters were applied to cut off 1 and 50 Hz frequencies; additionally, a notch filter was used to remove a 50 Hz noise.

Once filtered, the data were standardized per channel and segmented into the sliding windows of 2 s with a 0.2 s overlap. Then, the resulting dataset was fed to the neural network. A filtered nonstandardized segmented signal was used to calculate features. The selected features included those recommended in the literature [20]: high-order crossings up to the 6-th order, band power for delta, theta, alpha and beta and gamma frequency ranges and power asymmetry ratio.

All models were trained as follows: the EEG data obtained during each of the sessions were assigned to one of the two classes depending on what type of stimulus was used to evoke the signal. The assignment was binary: positive stimuli were assigned to class 1, whereas negative, to class 2. The network was trained using a minibatch stochastic gradient descent method (the learning rate optimization algorithm Adam [21], the minibatch size of 64, the learning rate of 0.001, 30 epochs) and categorical cross-entropy as a loss function. Thus, the input tensor had 3 dimensions [64, 8, 500], where 64 was the minibatch size, 8 was the number of EEG channels, 500 was the length of an EEG segment with an overlap of 50 time points.

All stages of the experiment, including signal recording and processing, as well as model training, involved the use of Python and scikit-learn and Keras libraries.

RESULTS

We have created a neural network with the following architecture: 2 convolutional layers of 64 kernels each, a batch normalization layer, an ELU (exponential linear unit) layer, an average pooling layer (window size 4, stride 1), 2 convolutional layers of 64 kernels each, a batch normalization layer, a ReLU (rectifier linear unit) layer, a max-pooling layer (window size 2, stride 1), 2 convolutional layers of 128 kernels each, a batch normalization layer, a ReLU layer, a max-pooling layer (window size 2, stride 1),
a fully connected layer consisting of 256 neurons with a ReLU activation function. The following parameters were used in all convolutional layers: size 3, stride 1, padding 0. Softmax (normalized exponential function) was used as an activation function for the output layer.

The use of a neural network allowed us to achieve a F1 score of 87% on a validation sample, which is significantly higher than an F1 score for a random forest model (67%). Confusion matrices (see the Figure) demonstrate that the random forest model successfully identifies positive states, outperforming the neural network slightly, but has low specificity to negative states. By contrast, the performance of the deep learning model in identifying and differentiating between the two studied states is equally accurate.

When testing the performance of our convolutional neural network using a dataset cleared of noise (electrooculography, EOG), we did not observe any significant changes to classification accuracy. The analysis of neural network activations demonstrated that EOG artifacts did not significantly skew classification results.

**REFERENCES**

11. Cecotti H, Graser A. Convolutional neural networks for P300
литература


