EEG based mental state analysis

Krishnapriya Ajith, Dr.R Menaka and Shani S Kumar
Department of Electronics Engineering, Vellore Institute of Technology, Chennai, Tamilnadu, India

E-Mail: krishnapriyajith@gmail.com

Abstract. This work proposes a method for mental state analysis from the Electroencephalogram (EEG) signal data using Convolutional Neural Network (CNN). In recent years, deep learning techniques, particularly CNN have become a very popular topic. Emotions also play a very crucial role in our daily life. Emotions affect a person’s cognition, behaviour and decision-making. In this paper, we analyse the EEG signals and classify them into various emotions. The result would be a real-time EEG-based emotion recognition system. The DEAP dataset, which is a multimodal dataset, has been used in this study for analysis of human mental states.

1. Introduction
Emotions are psychological states that can impact a person’s cognition, behaviour, decision-making and attention. The past decades focused on various emotion recognition methods based on facial expression, speech pattern and body gestures. But recently, researchers have recognized that brain signals can be used in assessing the human emotions effectively. Electroencephalography is a non-invasive neuro-imaging medical technique which is used to evaluate the electrical activities in the brain. Due to ease of usage, EEG-based mental state analysis has become popular now. However, scalp EEG is a random dynamic signal from various cortex sources of the brain and hence cannot directly indicate any exact information. Hence the major challenge is to identify how the EEG could be mapped with the momentary mental-states. This paper focuses on identifying discriminative features and classifying the emotions appropriately.

The paper proceeds as follows. Related works are summarized in section II. Methodology is described in section III. Dataset details, implementation of model and results are presented in section IV.

2. Related Work
Guijun Chen et.al [1] proposed a method of emotion feature analysis and recognition based on reconstructed EEG sources. The paper focused on analysing significant differences between active cortex source regions of the brain and frequency bands for pairs of emotions-based reconstructed EEG sources. Emotions were classified using Support Vector Machines (SVM).

An EEG-based emotion recognition using multi-channel EEG via multi-level features guided capsule network was proposed by Yu Liu et.al [3]. Experiments were demonstrated on DEAP and DREAMER datasets. The work could achieve an average accuracy of 97.97%, 98.32% and 98.32% for the DEAP dataset and 94.59%, 95.26% and 95.13% for the DREAMER dataset from valence, arousal and dominance respectively.

Jatupaiboon et.al[4] proposed a mechanism for classifying emotions into happy and unhappy. EEG data were recorded from 10 subjects. The stimuli used were picture and classical music. Emotions were classified using Power Spectrum Density (PSD) and Support Vector Machine (SVM) classifier. The accuracy achieved was 75.62% for subject-dependent model and 65.12% for subject-independent model.
Automated emotion recognition system based on deep learning algorithm and higher order statistics was proposed by Rahul Sharma et.al [5]. The work focused on higher order statistics to explore the emotions. Signals were decomposed into sub-bands using Discrete wavelet transform (DWT). Non-linear dynamics of each sub-band signal were then explored with the use of Third-order cumulants (ToC). The deep learning technique of long short-term memory was used to classify the EEG signals. The study could achieve an accuracy of 82.01% on the DEAP dataset.

Hassouneh et. al [10] put forward a real-time emotion recognition system. The work focussed on classifying the emotions of physically disabled people using facial expressions and EEG signals in CNN and LSTM classifier respectively. The subjects were instructed to express six different emotions in front of the camera and these were recorded. Mathematical model of the facial markers which were placed on the face of the subjects was generated and analysed. This was classified with the help of CNN. Similarly, the EEG signals were used on the LSTM classifier. The work could achieve an accuracy of 99.81% on facial expressions while using CNN and 87.25% on EEG signals while using LSTM classifier.

In [13], Nitin Kumar et. al presented a bispectral analysis method for recognizing emotions using DEAP dataset. Bispectrum, which is the third order statistics of a signal and method of identifying the phase relationships between the various frequency components, was used. The emotions were classified on the basis of a valence-arousal model. The brain signals were filtered into theta, alpha and beta bands and bispectrum was calculated using 2 EEG channels. Using the various derived bispectrum features, emotion classification was performed by the Linear Kernal Least Square Support Vector Machine (LS-SVM) and ANN (Artificial Neural Network) back-propagation algorithm. Backward sequential search was employed as the feature vector and used for feature selection.

3. Proposed Methodology

Our methodology involves the below mentioned steps. In the first step, raw EEG signals are preprocessed. In the second step, wavelet transform is applied to the preprocessed EEG signals. Third step involves generating scalogram images. Next step is to train the CNN using scalogram images in order to classify the emotions.

3.1. Preprocessing
At first the EEG signals were recorded from 32 subjects using 32 active AgCl electrodes [2]. The sampling rate was 512 Hz. The raw EEG data is then down-sampled to 128Hz. EOG artefact removal was performed using a blind source separation technique in order to remove the noise elements caused by eye and muscle movements. A bandpass frequency filter from 4.0-45.0Hz was applied and the data was averaged to a common reference. The data was segmented into 60 second trials and a 3 second pre-trial baseline removed.

3.2. Dataset
The dataset we used is the DEAP dataset that analyses human emotions. This dataset is taken from 32 subjects. It includes multichannel EEG signals and peripheral physiological signals. A music video clip is shown to each of the subject and their feedback is recorded on a continuous 9-point scale in terms of valence, arousal, dominance, and liking. A total of 8064 samples were recorded from 32 individual subjects with 40 trials for each subject. The data is of the format 32×40×40×8064 that includes 32 subjects, 40 trials/videos, 40 channels and 8064 samples of data [2].

| Table 1. Dataset Summary |
|--------------------------|
| No. of participants      | 32 |
| Number of videos         | 40 |
| Video duration           | 1 minute |
| Total number of channels | 40 |
| Number of channels for EEG | 32 |
3.3. **Wavelet Transform**

A wavelet transform (WT) is the decomposition of a signal into a set of basic functions consisting of contractions, expansions, and translations of a mother function $\phi(t)$, called the wavelet [7]. The continuous wavelet transform (CWT) is defined by (1) in terms of dilations and translations of a prototype or mother function $\psi(t)$. In time and Fourier transform domains, the wavelet is defined as

$$
\Psi_u, s(t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t-u}{s} \right), \quad u \in \mathbb{R}, \ s > 0
$$

(1)

Given $f \in L^2(\mathbb{R})$, the continuous wavelet transform (CWT) of $f$ at time $u$ and scale $s$ is defined as

$$
W f(u, s) = (f, \Psi_{u,s}) = \int_{-\infty}^{+\infty} f(t) \psi^*_{u,s}(t) dt
$$

(2)

Scalogram is a function of frequency and time that is used to get better time localization for short-duration, high-frequency events, and better frequency localization for low-frequency, longer-duration events [6]. Scalograms are obtained by the Continuous Wavelet Transform. It is used to analyze non-stationary signals at multiple scales. The signals are decomposed into wavelets and time–frequency representation is created to distinguish noise from the signal.

3.4. **Scalogram**

Scalograms are used to get better time localization for short-duration, high-frequency events, and better frequency localization for low-frequency, longer-duration events. Here we need better time localization for short-duration, high-frequency events. It is a time – frequency plot obtained from the absolute value of the continuous wavelet transform (CWT) of a signal.

The scalogram of $f$ is defined by the function

$$
S(s) = \|W f(s, u)\|_2 = \left( \int_{-\infty}^{+\infty} |W f(s, u)|^2 du \right)^{1/2}
$$

(3)
3.5. Convolutional Neural Network (CNN)
Convolutional Neural Network (CNN) is a class of artificial neural network which is composed of multiple building blocks such as convolution layers, pooling layers, and fully connected layers. CNN is designed to automatically learn and extract features through a backpropagation algorithm. CNN has recently achieved expert-level performances in various fields, including the medical field.

3.6. Model Implementation
3.6.1. Architecture of AlexNet. AlexNet is a CNN architecture that consists of 8 layers, which includes 5 convolutional layers and 3 fully-connected layers. Some of the convolutional layers are followed by maxpooling layers and the fully connected layers are followed by a final 1000-way softmax. This is one of the simplest architectures compared to the many complicated ones introduced in the recent years. The input size of the images to the AlexNet is 224 x 224 x 3 without padding and 227 x 227 x 3 with padding. The architecture uses ReLU as the activation function for faster convergence and dropout layers to avoid over-fitting. Fully connected layers were employed for the classification purpose. The first convolutional layer has kernel size of 11 x 11 with stride 4 and the maxpooling uses a filter size of 3 x 3 with stride 2 [8].

3.6.2. Architecture of SqueezeNet. The SqueezeNet architecture consists of, as the name suggests, “expand” and “squeeze” layers. The squeeze convolutional layer has a filter size of 1 x 1, which are then fed into an expand layer that has a mix of 1 x 1 and 3 x 3 convolutional filters. The SqueezeNet uses 50x fewer parameters in order to achieve the AlexNet-level accuracy. The model also makes use of the fire modules. At first, the input image is fed into a standalone convolutional layer. This is then followed by 8 fire modules. The architecture also utilizes maxpooling with a stride of 2.
3.6.3. Architecture of GoogleNet. The GoogleNet architecture comprises of 22 layers and uses 3 different size of filters (i.e., 1 x 1, 3 x 3 and 5 x 5). The input size of the images to the GoogleNet is 224 x 224 x 3. There are 27 pooling layers included. 9 inception modules are stacked linearly in total and these are connected to the global average pooling layer. An average pooling layer has a filter size of 5 x 5 and stride 3. A 1 x 1 convolution consists of 128 layers. All the convolutions inside the architecture uses ReLU activation function.

![Architecture of GoogleNet](image)

**Figure 4.** Architecture of GoogleNet

3.6.4. Architecture of ShuffleNet. The ShuffleNet architecture was mainly designed for use in mobile devices with very less computing power. The architecture makes use of two operations, such as pointwise group convolution and channel shuffle. Using these two operations, feature map channels could encode more information. The usage of pointwise convolution helps in reducing the costly dense 1 x 1 convolutions. In order to overcome the issue of the group convolutions blocking the information and weakening the representation, channel shuffle operation is presented. The channel shuffle obtains the input data from different groups by splitting the channels in each group to several other groups. This is then fed into the next layer with different subgroups.

![Architecture of ShuffleNet](image)

**Figure 5.** Architecture of ShuffleNet

3.6.5. Architecture of ResNet-50. The ResNet-50 is a variant of ResNet model. It has 48 convolutional layers, 1 MaxPool layer and 1 Average Pool layer. ResNets used the concept of skip connection to mitigate the problem of vanishing gradient by allowing a shortcut path for the gradient flow.
3.6.6. **Architecture of MobileNetV2.** The MobileNetV2 architecture uses depthwise separable convolutions in order to reduce the complexity cost and model size of the network. The architecture is preferred for mobile devices or any other devices that uses low computational power. Linear bottleneck between layers have been introduced. The model also utilizes the inverted residual-structure and has two types of blocks- namely residual block and downsizing block. The basic intuition is that the bottlenecks in the model could encode the intermediate inputs and outputs while the inner layer could encapsulate the model’s capability to learn and classify.

4. **Experiment and Results**

This section describes about the dataset division, experiment and its setup as well as the results and performance analysis.

4.1. **Dataset Division**

In order to obtain a greater number of samples to train the CNN, the dataset was divided into 4 parts, i.e., the dataset is now in the format 4×40×40×2016 for each of the 32 subjects.

In our work, we classify the data into four different emotions: Happy, Scared, Sad and Calm based on the Valence and Arousal features of the DEAP dataset. Valence and arousal were rated on a scale of 1 to 9. The rating of valence greater than 5 indicates a high valence and less than or equal to 5 indicates a low valence. Similarly, the rating of arousal greater than 5 indicates a high arousal and less than or equal to 5 indicates a low arousal. The data is then reorganised into 4 categories on the basis of these emotions.
4.2. Experimental Setup

The experiments were performed in two different PCs with Intel(R) Core (TM) i5-10210U CPU at 1.6 GHz and AMD Ryzen 5 3500U CPU at 2.10 GHz. The tool used for generating the scalogram and training the CNN is MATLAB R2020b.

4.3. Experiment

The experiments are based on multi-channel EEG signals. Continuous Wavelet Transforms (CWT) is applied to the sampled data. We have chosen Morse Wavelets, which is a part of the analytic wavelets and are useful in analysing modulated signals. Morse wavelets are defined by two parameters—symmetry and time-bandwidth product. These parameters will define the behaviour of the transforms and also the shape of the wavelet. For a generalized Morse wavelet, the Fourier transform is expressed as

$$\Psi_{\gamma}(\omega) = U(\omega) a_{\gamma} \omega \gamma^{\frac{3}{2}} e^{-\omega \gamma}$$

(3)

Scalogram images are plotted for these wavelets as a function of time and frequency. These images are taken as inputs to the CNN model. CNN is trained to extract features from the scalograms in order to classify the emotions.

4.4. Results

The results of training performed are as shown below:
Figure 10. Training on AlexNet

Figure 11. Training on SqueezeNet

Figure 12. Training on GoogleNet

Figure 13. Training on ShuffleNet

Figure 14. Training on ResNet50

Figure 15. Training on MobileNetv2
4.5. Performance Analysis

Comparison on both architectures and their hyperparameters are listed in the table below:

| Name of Hyperparameter | AlexNet | SqueezeNet | GoogLeNet | ShuffleNet | ResNet50 | MobileNetV2 |
|------------------------|---------|------------|-----------|------------|----------|-------------|
| MiniBatchSize          | 10      | 10         | 10        | 10         | 10       | 10          |
| MaxEpochs              | 10      | 10         | 10        | 10         | 10       | 10          |
| Total Iterations       | 4090    | 3580       | 4090      | 3840       | 4090     | 3580        |
| Learning Rate          | 0.0001  | 0.0001     | 0.001     | 0.0001     | 0.0001   | 0.0001      |
| Dropout Probability    | 0.5     | 0.5        | 0.4       | 0.7        | 0.6      | 0.9         |
| Loss Function          | Cross entropy ex | Cross entropy ex | Cross entropy ex | Cross entropy ex | Cross entropy ex | Cross entropy ex |
| Activation Function in convolution layer | ReLU | ReLU | ReLU | ReLU | ReLU | Clipped ReLU |
| Pooling                | Max pooling | Max pooling | Max pooling | Max pooling | Max pooling | Global average pooling |
| Accuracy Obtained      | 34.31%  | 33.44%     | 34.31%    | 34.06%     | 34.31%   | 32.53%      |

5. Conclusion

In this research paper, we had focused on classifying the emotions into 4 categories based on the valence and arousal dimensions of the DEAP dataset. The proposed approach has not achieved the best accuracy and results. This could be due to the limitation in dataset size and hence the lesser number of scalogram images which are having high resemblance with each other. Future work aims at achieving better results with the limited datasets and hence classifying the emotions.

6. Acknowledgement

This work was supported by the Vellore Institute of Technology, Chennai, India. The authors would like to thank the authors (S. Koelstra et al. 2012), who proposed the multimodal dataset for emotion recognition and provided the dataset on request.

References

[1] G Chen, X Zhang, Y Sun and J Zhang 2020 IEEE Access Emotion feature analysis and recognition based on reconstructed EEG sources 8 11907-16.
[2] S Koelstra, C Muehl, M Soleymani, J S Lee, A Yazdani, T Ebrahimi, T Pun, A Nijholt and I Patras 2012 IEEE Transactions on Affective Computing DEAP: A database for emotion analysis using physiological signals 3 18-31.
[3] Yu Liu, Yufeng Ding, Chang Li, Juan Cheng, Rencheng Song, Feng Wan and Xun Chen 2020 Computers in Biology and Medicine Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network 123 103927.
[4] N Jatupaiboon, S Pan-ngum and P Israsena 2013 The Scientific World Journal Real-time eeg-based happiness detection system.
[5] Rahul Sharma, Ram Bilas Pachori and Pradip Sircar Biomedical Signal Processing and Control Automated emotion recognition based on higher order statistics and deep learning algorithm 58 101867.
[6] Divya Garg and Gyanendra K Verma 2020 Procedia Computer Science Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet based deep learning framework 171 857-67.

[7] E A B Da Silva and D G Sampson 1996 Advances in Imaging and Electron Physics.

[8] K K Sudha and P Sujatha 2019 International Journal of Recent Technology and Engineering (IJRTE) A qualitative analysis of Googlenet and Alexnet for fabric defect detection 8.

[9] Gopal Chandra Jana, Ratna Sharma and Anupam Agrawal 2019 International Conference on Computational Intelligence and Data Science (ICCIDS 2019) A 1D-CNN-Spectrogram Based Approach for Seizure Detection from EEG Signal 167 403-12.

[10] Aya Hassouneh, A M Mutawa and M Murugappan 2020 Informatics in Medicine Unlocked 20 Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods 20.

[11] Barjinder Kaur, Dinesh Singh and Partha Pratim Roy 2018 International Conference on Computational Intelligence and Data Science (ICCIDS 2018) EEG Based Emotion Classification Mechanism in BCI 132 752-58.

[12] Nitin Kumar, Kaushikee Khaund and Shyamanta M Hazarika 2016 Procedia Computer Science Bispectral Analysis of EEG for Emotion Recognition 84 31-35.

[13] Jordan J. Bird, Luis J. Manso, Eduardo P. Ribeiro, Anikó Ekárt, Diego R. Faria, “A Study on Mental State Classification using EEG-based Brain-Machine Interface”, 9th international Conference on Intelligent Systems, 2018.