EEG Classification based on Image Configuration in Social Anxiety Disorder

Lubna Shibly Mokatren[1], Rashid Ansari[1], Ahmet Enis Cetin[1]
Alex D. Leow[2], Olusola Ajilore[2], Heide Klumpp[2] and Fatos T.Yarman Vural [3]

Abstract— The problem of detecting the presence of Social Anxiety Disorder (SAD) using Electroencephalography (EEG) for classification has seen limited study and is addressed with a new approach that seeks to exploit the knowledge of EEG sensor spatial configuration. Two classification models, one which ignores the configuration (model 1) and one that exploits it with different interpolation methods (model 2), are studied. Performance of these two models is examined for analyzing 34 EEG data channels each consisting of five frequency bands and further decomposed with a filter bank. The data are collected from 64 subjects consisting of healthy controls and patients with SAD. Validity of our hypothesis that model 2 will significantly outperform model 1 is borne out in the results, with accuracy 6–7% higher for model 2 for each machine learning algorithm we investigated. Convolutional Neural Networks (CNN) were found to provide much better performance than SVM and kNNs.

Index Terms— EEG, deep learning, classification

I. INTRODUCTION

Social Anxiety Disorder (SAD), world's third largest mental health care problem, affects 7% of the population [Richards(2017)]. It is characterized by extreme fear and avoidance of social situations and the fear of negative evaluations from others [Leichsenring and Leweke(2017)]. The diagnosis process of SAD was first characterized in 1980 by Diagnosis and Statistical Manual for Mental Disorders (DSM-III). However, the criteria evolved and the most recent description appears in the fifth edition (DSM-5) [Hofmann and Otto(2017)]. In the field of psychiatry, there is the need to pay considerable attention to the reliability and quality of the diagnostic process of SAD DSM-5 to give an accurate assessment of the disorder [Kraemer et al.(2012)Kraemer, Kupfer, Clarke, Narrow, and Regier].

Electroencephalography (EEG) is a useful mechanism for diagnosing mental disorders. EEG provides measurements of brain activities aquired using electrodes placed over the scalp. While other brain imaging techniques, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), are used in diagnosis, EEG has important attributes in that it captures the temporal activity of the brain and is affordable compared with other methods [Sanei and Chambers(2013)]. The EEG waveform is usually divided into five main frequency bands [Abo-Zahhad et al.(2015)Abo-Zahhad, Ahmed, and Seha]:

Delta (δ: up to 4 Hz) waves are generated during drowsiness and are the slowest, Theta (θ: 4-8 Hz) waves are observed during quiet focus or sleep, Alpha (α: 8 - 15 Hz) waves are observed during relaxation with closed eyes, Beta (β: 15-32 Hz) waves observed during normal consciousness and active thinking, and Gamma (γ: 32 Hz) associated with strong electrical signals caused by visual stimulation or information processing, learning, and perception. All the mentioned five frequency bands of the EEG signals are used in the analysis process for this study.

The use of EEG in SAD diagnosis has seen only limited study. The visual detection of differences in the EEG signals between SAD patient and control groups is impractical since the EEG activity appears to be similar in both. Therefore, automated detection using techniques such as machine learning is usually employed, which could lead to more precise diagnosis results and can be the first step for better connectivity analysis, pattern recognition process, and understanding treatment responses in SAD [Moscovitch et al.(2011)Moscovitch, Santesso, Miskovic, McCabe, Anton].

In the past, many classification algorithms were devised for using EEG data [Lotte et al.(2007)Lotte, Congedo, Lécuyer, Lamarche, and Arnaldi], such as, linear discriminant analysis, SVM, neural networks, nonlinear bayesian classifiers, kNN, hidden markov model, combination of classifiers, and others. However, none considered the spatial locations and configuration of the EEG channel sensors as a means to possibly achieving better accuracy in analysis or classification tasks. This was a key driving factor in our research. Two classification models, one which ignores the configuration (model 1) and one that exploits it with different interpolation methods (model 2), are considered in this study. We hypothesized that model 2 will significantly outperform model 1. Validity of our hypothesis is borne out in the results, with average accuracy 7% higher for model 2 for each machine learning algorithm we investigated. Convolutional Neural Networks (CNN) were found to provide much better performance than SVM and kNNs.

II. METHOD

A. EEG recording

The EEG data-set used in this paper was acquired in the Department of Psychiatry at University of Illinois at Chicago (UIC). The acquisition of multi-channel EEG is done using an Electro-Cap with electrodes positioned at 34 different

[1]Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago
[2]Department of Psychiatry, University of Illinois at Chicago, Chicago
[3]Department of Computer Engineering, Middle East Technical University, Ankara, Turkey

locations i.e. channels. The data are gathered from a total of 64 subjects divided into control and SAD patients groups. For this study, the brain activity being analyzed is at resting state, without the introduction of stimuli or task instructions. The duration of the EEG recording varies from 2-7 minutes. The signals are sampled at 1024Hz sampling frequency.

B. Data Preprocessing

EEG data are complex, as multiple processes take place simultaneously causing various artifacts and noise to appear, such as eye movements and EMG artifacts. Hence, it needs to be processed and cleaned for better interpolation during analysis. EEG data preprocessing stages are implemented using EEGLAB software. The frequencies of interest are in the range of 1-50 Hz, covering the five different frequency bands.

C. Data Analysis

Nonstationary phenomena are present in EEG data due to the constant switching of the meta-stable states of neurons assembling during brain functioning, causing signal changes in the form of spikes and momentary events. In our data analysis in time and frequency domains, each channel signal is divided into windows in which the data are assumed to be stationary. After testing multiple sizes of window segments based on the sampling frequency and the statistical properties, a window size of 5120 samples was found to yield the best results for detection.

The analysis of the signal content of each of the five main EEG frequency bands can be utilized to estimate subjects' cognition and emotional states. A dyadic wavelet packet transformation [Fliege(1994)] is used for the decomposition of subbands corresponding to the five frequency bands. [0-4] Hz for δ band, [4-8] Hz for θ, [8-16] Hz for α, [16-32] Hz for β, and [32-52] Hz for γ.

The energy of the content of each windowed segment is computed for the five frequency bands separately. The analysis is based on the energy content of these signals represented in two different ways: concatenation of the channels of the five frequency bands and image-like 2D representation of the EEG channel locations. The latter method is discussed in Section II.D. It should be noted that the outputs of the filter bank are ordered from highest frequency subband to the lowest.

D. Image representation of the EEG data

The data are acquired using 34 electrodes placed over different areas on the scalp, Frontal (F), Central (C) Temporal (T), Parietal (P) and Occipital (O) as shown in Fig. 1. It is hypothesized that the location of the channels can provide improved detection accuracy in the analysis of the data. To examine this hypothesis, two main data models are examined. First, the 34 channels of the five frequency bands are concatenated by creating a 34×5 energy matrix over each window, without accounting for the location of the channel electrodes. Second, a 3-D array of size $15 \times 15 \times 5$ is created, where the first two dimensions represent an image of 15×15
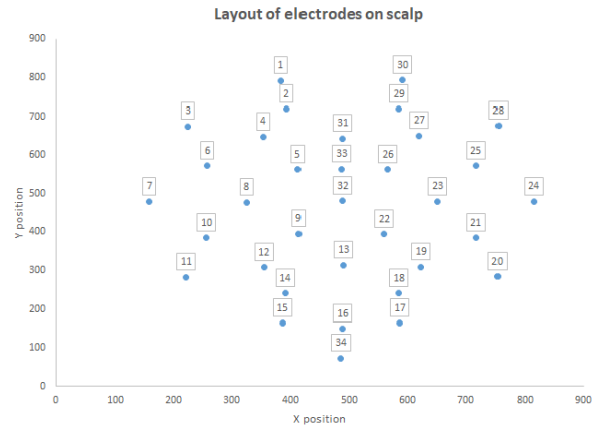

Fig. 1.   Layout of 34 electrodes on scalp

pixels corresponding with the channels positioning over the scalp while the third dimension represents the five frequency bands. In the latter method, the locations that not exactly correspond to any of the 34 channels are filled using different interpolation techniques.

The layout of the channels' location is given in Fig. 1. To construct an image-like representation of the electrodes layout, an image of size 15×15 was created. To fill in the missing pixel energy values, the following interpolation method was used [Eckstein(1989)]: an interpolated value e at point x, only the samples $u_i = u(x_i)$ for $i = 1, ...n_i$, which lay within a distance less than d_{max} from point x, are used to calculate interpolated value using the following weighted average.

$$u(x) = \frac{\sum_{i=1}^{n_i} w_i(x)u_i}{\sum_{i=1}^{n_i} w_i(x)} \quad (1)$$

where $w_i(x) = \frac{1}{d(x,x_i)}$ and $d(x,x_i)$ represents the distance between points x and x_i. This method is called Inverse Distance Weighting (IDW). In areas where no energy values exist within a distance of d_{max} i.e. "Border Points" (BP) for abbreviation, the nearest value is simply repeated. d_{max} is chosen empirically from a set of different distances.

It should be noted that other interpolation methods were examined such as nearest neighbor interpolation, bilinear interpolation, cubic spline interpolation [Lehmann et al.(1999)Lehmann, Gonner, and Spitzer], and zero insertion at the border points (BP) with zero. However, they all result in inferior performance compared with the method mentioned above. Table IV in Section IV summarizes the average performance of the suggested methods. The construction of the image in the second model is tested on the main deep network structure (CNN) described in Section III.B.

III. EXPERIMENTS

After the data are collected and preprocessed, two different models are used as previously discussed. In the first model, the energy for the five frequency bands is calculated separately in each window of size 5120 samples. Hence,
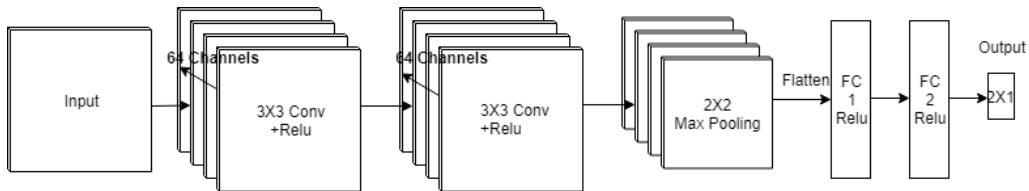
Fig. 2. Convolutional neural network structure

for each window an energy matrix of dimensions 34×5 is constructed. There are thirty four channels or electrodes and five frequency bands. Each row in the matrix represents one channel. The second model is based on the image representation of the EEG data, as discussed in Section II.D. For each window a 3D energy array of $15 \times 15 \times 5$ is built, where the third dimension represents the five frequency bands. For both models, each matrix is considered as a single sample for the training or testing data-set.

A. Acquisition of training and testing EEG data

To train the network, a stratified 8-fold cross-validation is applied, the data are shuffled randomly and each fold is made by preserving the percentage of subjects from the two classes. 7 folds (56 subjects) are used for training and validation (about 10% validation, 90% training) and the remaining fold (8 subjects) is used for testing. The classifier is trained 8 separate times. In each trial a different fold is used for testing. Early stopping is applied by monitoring the validation loss to avoid overfitting. In both models, for every window of size N, an energy matrix is constructed and considered as single sample, where $N = 5120$. The samples are gathered for each of the training, validation and testing sets by sliding a moving window of size N with no overlap of windows. The choice of sliding N samples is made empirically as it yields better results when compared with other overlapping windows of shifts of $N/4$, $N/2$ and $3N/2$. In summary, the data set of each subject was divided into multiple samples, and the samples collected for the training, validation and testing sets will never overlap. Samples are labeled 0 (negative) if collected from control group, and 1 (positive) if collected from SAD patients. The classification accuracy is the percentage of subjects classified correctly.

For every trial, each testing subject is evaluated as follows:

$$prediction = \begin{cases} patient & \text{if } x_i = \frac{p_i}{t_i} \geq Th \\ control & \text{else} \end{cases} \quad (2)$$

where $Th = 0.45$, p_i is number of samples classified as positive for subject i, t_i is total number of samples for subject i. Since some of the EEG data contain unwanted signals, the threshold was lowered from 0.5 to 0.45.

B. CNN Network Structure

Deep learning has significantly improved the performance for many problems compared with other machine learning algorithms [Deng et al.(2014)Deng, Yu, et al.]. Developing deep network-based solutions has been found to be effective in many applications. CNNs

have turned out to be the most powerful deep learning architectures for image-related problems [Krizhevsky et al.(2012)Krizhevsky, Sutskever, and Hinton]. The input to the network is arrays of data containing energy values of subband signals, in windows of size 5120. A sequential model is built for this classification problem as shown in Fig. 2. The first layer is a 2D convolution with kernel size (convolution window) 3×3, 64 output filters, and ReLu activation over the outputs, another 2D convolution layer with the same parameters followed. This is followed by a max-pooling layer with pool size of 2, followed by dropout with rate=0.25. The input is then flattened which is followed by a fully connected layer with 128 output dimension and ReLu activation. Another dropout is done with rate=0.2 followed by final fully connected layer with Softmax activation and output dimension that equals to 2. Dropout is a regularization method that is used to reduce over-fitting. Other networks with one or more convolutional layers are also tested. However, all produced an inferior performance compared with the proposed network. The parameters of this network were chosen using several different trials and the parameters that yielded the best performance were selected. It should be noted that batch normalization was applied to the input and the CNN layers to reduce the internal covariate shift [Ioffe and Szegedy(2015)].

IV. RESULTS

In this study, inputs constructed by the two main models are both tested, by feeding them to different machine learning algorithms or deep neural networks. Model 1 is built by concatenating the channels of the five frequency bands with no consideration for the spatial configuration of EEG electrodes. Model 2 takes the electrode configuration over the scalp into account. The two models are investigated for their classification performance using SVM, kNN, and a proposed CNN in Section III.B.

In the proposed CNN, the confusion matrices for Model 1 and Model 2 can be seen in Table I and Table II, respectively. The confusion matrix summarizes the prediction results. The first entry of the first column has the number of actual positive (i.e., SAD patients) predicted as patients, the second entry of the first column has the number of actual patients predicted as negative (i.e., healthy subjects), etc. The accuracy (ability to correctly classify a subject), sensitivity (ability to correctly identify patients when SAD is present), and specificity (ability to correctly identify control within a healthy group) are all higher for the proposed approach

TABLE I

CONFUSION MATRIX - MODEL I

Model I		Actual Condition		
		Positive	Negative	Accuracy = 80%
Predicted	Positive	26	7	Sensitivity = 81%
Condition	Negitive	6	25	Specifitty = 78%

TABLE II

CONFUSION MATRIX - MODEL II

Model II		Actual Condition		
		Positive	Negative	Accuracy = 87%
Predicted	Positive	29	5	Sensitivity = 90%
Condition	Negitive	3	27	Specificity = 84%

TABLE III

AVERAGE CLASSIFICATION PERFORMANCE

Classification Accuracy		
Model	Model I	Model II
Proposed CNN	80%	87%
SVM	72%	79%
kNN	69%	75%

TABLE IV

AVERAGE PERFORMANCE OF DIFFERENT INTERPOLATION METHODS

Interpolation Comparison	
Interpolation method	Average recognition perf.
Nearest Neighbor	73%
Bilinear	75%
IDW with 0 BP	78%
Cubic spline	83%
IDW with NN BP	87%

(Model 2), with overall classification accuracy of 87% for Model 2 and 80% for Model 1. Thus, Model 2 that takes advantage of the location of the channels is found to provide superior performance.

In this classification task the EEG input data are also fed to SVM with radial basis kernel function (RBF) with kernel parameter $\sigma = 0.4$ using LIBSVM tool. The parameters were chosen by estimating the performance of SVM with different kernels and hyper-parameters tuning using the validation set. The data are also applied to k-NN with k=3, and to other deep networks structures but the proposed network structure stood out as superior in terms of overall classification accuracy, as seen in Table III. It should be noted that, in all cases, Model 2 gave significantly higher accuracy than Model 1, establishing the importance of 2D representation of the spatial configuration of EEG sensors.

To represent the different channels as an image over the scalp in the second model, a few interpolation methods are tested, as mentioned in Section II.D. The average performance of these methods is presented in Table IV. Showing that Inverse Average Weighted interpolation yielded the best results among the methods tried.

V. CONCLUSION

In this study, it is found experimentally that SAD patients can be identified based on their EEG signals. Furthermore, the spatial configuration of the EEG electrodes is used for the first time in a classification task. When taking advantage of the positions of the channels, the performance of the CNN network shows 7% improvement in accuracy compared with the same CNN in which configuration is ignored.

REFERENCES

[Richards(2017)] Thomas A. Richards. The social anxiety association, 2017. URL http://socialphobia.org/.

[Leichsenring and Leweke(2017)] Falk Leichsenring and Frank Leweke. Social anxiety disorder. *New England Journal of Medicine*, 376(23): 2255–2264, 2017.

[Hofmann and Otto(2017)] Stefan G Hofmann and Michael W Otto. *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder Specific Treatment Techniques*. Routledge, 2017.

[Kraemer et al.(2012)Kraemer, Kupfer, Clarke, Narrow, and Regier] Helena Chmura Kraemer, David J Kupfer, Diana E Clarke, William E Narrow, and Darrel A Regier. Dsm-5: how reliable is reliable enough? *American Journal of Psychiatry*, 169(1):13–15, 2012.

[Sanei and Chambers(2013)] Saeid Sanei and Jonathon A Chambers. *EEG signal processing*. John Wiley & Sons, 2013.

[Abo-Zahhad et al.(2015)Abo-Zahhad, Ahmed, and Seha] M Abo-Zahhad, Sabah Ahmed, and Sherif Nagib Seha. A new eeg acquisition protocol for biometric identification using eye blinking signals. 07:48–54, 05 2015.

[Moscovitch et al.(2011)Moscovitch, Santesso, Miskovic, McCabe, Antony, and Schmidt] David A Moscovitch, Diane L Santesso, Vladimir Miskovic, Randi E McCabe, Martin M Antony, and Louis A Schmidt. Frontal eeg asymmetry and symptom response to cognitive behavioral therapy in patients with social anxiety disorder. *Biological Psychology*, 87(3): 379–385, 2011.

[Lotte et al.(2007)Lotte, Congedo, Lécuyer, Lamarche, and Arnaldi] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1, 2007.

[Fliege(1994)] Norbert J Fliege. *Multirate digital signal processing*, volume 994. John Wiley New York, 1994.

[Eckstein(1989)] Barbara Ann Eckstein. Evaluation of spline and weighted average interpolation algorithms. *Computers & Geosciences*, 15(1): 79–94, 1989.

[Lehmann et al.(1999)Lehmann, Gonner, and Spitzer] Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. Survey: Interpolation methods in medical image processing. *IEEE transactions on medical imaging*, 18(11):1049–1075, 1999.

[Deng et al.(2014)Deng, Yu, et al.] Li Deng, Dong Yu, et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387, 2014.

[Krizhevsky et al.(2012)Krizhevsky, Sutskever, and Hinton] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[Ioffe and Szegedy(2015)] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.