# A Multichannel Deep Belief Network for the Classification of EEG Data

Alaa M. Al-kaysi[a], Ahmed Al-Ani[a] and Tjeerd W. Boonstra[b,c,d]

[a]Faculty of Engineering and Information Technology, University of Technology, Sydney, Ultimo NSW 2007 Australia
[b]School of Psychiatry, University of New South Wales, Sydney, Australia
[c] Black Dog Institute, Sydney, Australia
[d]Research Institute MOVE,VU University Amsterdam, The Netherlands
Alaa.M.Dawood@student.uts.edu.au, Ahmed.Al-Ani@uts.edu.au,
t.boonstra@unsw.edu.au

**Abstract.** Deep learning, and in particular Deep Belief Network (DBN), has recently witnessed increased attention from researchers as a new classification platform. It has been successfully applied to a number of classification problems, such as image classification, speech recognition and natural language processing. However, deep learning has not been fully explored in electroencephalogram (EEG) classification. We propose in this paper three implementations of DBNs to classify multichannel EEG data based on different channel fusion levels. In order to evaluate the proposed method, we used EEG data that has been recorded to study the modulatory effect of transcranial direct current stimulation. One of the proposed DBNs produced very promising results when compared to three well-established classifiers; which are Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Extreme Learning Machine (ELM).

**Keywords** — Multichannel Deep Belief Network; EEG classification; transcranial direct current stimulation.

## 1 Introduction

Deep Belief Networks (DBNs) have emerged recently as new learning algorithms composed of multilayer neural networks that have the ability of training high dimensional data. A deep belief network is a probabilistic, generative model which is constructed of multiple layers of hidden units. The undirected layers in the DBNs are called Restricted Boltzmann Machines (RBM) which are trained layer by layer. RBMs have been used in machine learning as a generative model for different types of data [1].

DBNs have mainly been applied to image classification due to the hierarchical structure of images, where edges can be grouped to form segments, which when grouped form objects [2]. They have also been successfully applied to a number of other classification tasks, such as speech recognition and natural language processing [2]. In contrast, DBNs have not been fully explored for the

classification of electroencephalogram (EEG) data, which records the electrical activity of the human brain from multiple electrodes on the scalp. EEG plays a vital role in understanding the functional state of the brain and diagnosing disorders. For example, EEG has been used to study the modulatory effect of transcranial direct current stimulation (tDCS) on changes in cortical activities [3] and the treatment of brain disorders [4].

Wulsin et al. evaluated the classification and anomaly measurement of EEG data using an autoencoder produced by unsupervised DBN learning [5]. Zheng et al. introduced a deep belief network to search the neural signature associated with positive and negative emotional categories from EEG data. They trained the DBN using different entropy features that were extracted from multichannel EEG data. They showed that the performance of DBN was better in higher individual frequency bands [6].

In addition to DBNs that have been proposed for the classification of single-stream data, multi-modality DBNs have also been investigated by a number of researchers. Ngiam et al. proposed a deep network to learn unsupervised features for text, image or audio. They showed that the performance of a bimodal DBN based autoencoder was worse in cross modality and shared representation tasks [2]. Srivastava and Salakhutdinov proposed a DBN of multimodal system for learning a joint representation of data. They used images and text bi-model data to demonstrate that Multimodal DBN (MDBN) can learn good generative attributes. [7]. Cheng et al. proposed MDBN as a new deep learning model to fit large images (e.g. 1024×768). First, they used Scale-invariant feature transform (SIFT) descriptor to reprocess the images and sent these images to MDBN to extract features. Also, they adapted the Markov sub-layer to reflect the neighboring relationship between the inputs [8].

In this paper, three implementations of DBN are presented; one is similar to the single-stream DBN, the second implements a single stream DBN for each channel and then combines the classification outcomes of the different channels using another DBN, and the third case utilises the concept of MDBN in proposing a new multi-stream (or multichannel) DBN. In the third case a "partial" DBN with no decision layer is constructed for each channel, and the last hidden layers of the partial DBNs are combined using higher level hidden and decision layers.

## 2 Deep Belief Networks

Deep learning is a set of algorithms in machine learning that is based on distributed representations. Hinton et al. proposed a Deep Belief Network as an efficient unsupervised learning algorithm to overcome the complexity of training deep generative model [9][10]. The principal operation of DBN depends on using hierarchical structure of multiple layers of Restricted Boltzmann Machines (RBM). The process of training DBNs involves the individual training of each RBM one after another and then stack them on top of each other. Each two consecutive layers in DBNs are treated greedily as a Restricted Boltzmann Machine

[11]. RBM consists of a weight matrix $w_{ij}$, where $i$ represents a visible node and $j$ a hidden node. Gaussian or Bernoulli stochastic visible units are usually used in RBMs, while the hidden units are usually Bernoulli [10]. The energy function of a joint configuration $(v, h)$ of the visible and hidden units is described by [1]:

$$E(v,h) = \sum_{i \in visible} a_i v_i - \sum_{j \in hidden} b_j h_j - \sum_{ij} v_i h_j w_i j \qquad (1)$$

Where $v_i h_j$ are the binary states of visible and hidden units $i, j$ respectively, $w_{ij}$ is the weight between visible and hidden units and $a_i b_j$ are their biases. The model parameters are composed of $[w, v, h]$, $a = [a_1, a_2, ..., a_V]^T$, $b = [b_1, b_2, ..., b_H]^T$. The probability for every possible pair of a visible and a hidden victor was assigned by the following function:

$$p(v,h) = \frac{1}{Z} e^{E(v,h)} \qquad (2)$$

where $Z$ is the partition function that represents the summing of all possible pairs of visible and hidden vectors:

$$Z = \sum_{v,h} e^{E(v,h)} \qquad (3)$$

In each RBM, there are direct connections between units of the two layers, but there are no connections between units of the same layer, which leads to an unbiased sample of the state of a hidden and a visible units. Updating the visible and hidden units can be implemented using Gibbs sampling [1].

## 3 The Proposed Implementations of Multichannel DBN

Three DBN architectures have been implemented for the classification of multi-channel EEG data, as shown in Figure 1.

– The first architecture (Figure 1(a)), is the traditional single-stream DBN that trains features extracted from all channels using a single network that consists of two or more RBMs. The process of training the RBM models is as follows: the hidden layer of the first RBM is the visible layer of the second RBM, and the hidden layer of the second RBM is the visible layer of the third RBM, and so on until arriving to the top RBM. This architecture is used to evaluate the ability of a single-stream DBN to extract high level information from the features of the different channels.
– The second architecture (Figure 1(b)) consists of multiple DBNs, one for each EEG channel, each of which is trained in a way similar to that of the single-stream DBN. The labels of these DBNs are trained using another DBN to obtain the final decision. This architecture evaluates the ability of individual channels to correctly classify the data and the advantage of combining the classification results of individual channels.

– The third architecture, or the multi-stream DBN, (Figure 1(c)) is implemented using partial DBN that is used to process individual channels. The top hidden layer of each of those partial DBN are combined using unified hidden layer(s) that is(are) followed by the decision layer. The rationale behind proposing this architecture is to extract "local" information from each channel using the partial DBNs, while the higher level information is extracted using the combined hidden layers that fuse the local information of the different channels.
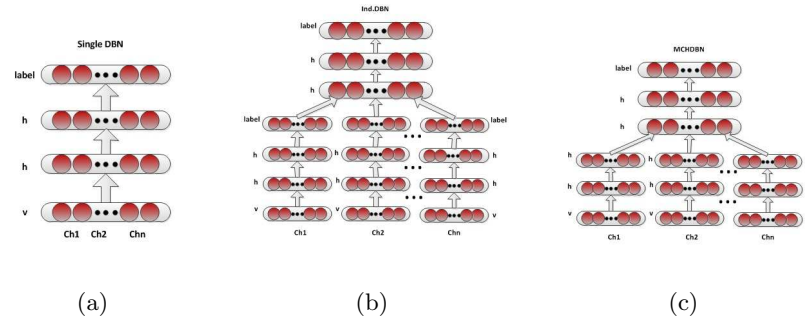


Fig. 1: ($a$) Single-stream DBN, ($b$) Individual DBN for each channel and a top DBN to combine the channels' classification results, and ($c$) Multi-stream (or multichannel) DBN

## 4  Data Set

In our experiments, we used data that was collected from patients suffering from depression. Depression is one of the significant mental disorders that affects many people around the world. According to the World Health Organization (WHO), depression is expected to be the major disability causing disease in the world by 2020 [12]. Transcranial direct current stimulation (tDCS) is a promising approach for treating depression [13][14]. tDCS treatment has only been recently established and it is important to study the effects of tDCS on brain functioning. In this study, we utilized data that was collected from twelve patients who were asked to go through three EEG recording sessions [4]. Each session lasted for 20 minutes. In the first session, baseline EEG was recorded for each patient. This was followed by one session of active and one session of sham tDCS that were randomly ordered among the twelve patients. Scalp EEG was recorded while patients were sitting still with their eyes closed using 62 electrodes placed according to the International 10/20 system, as shown in Figure 3(a). In this paper, we attempt to differentiate between the three session types using the recorded EEG data. The data of each EEG session was sampled at F = 2000 Hz. A Hanning window was applied and the power spectral density (PSD) was

estimated using the Fourier transform. The average PSD in the five conventional frequency bands, which are Delta $(0.5 \leq \delta < 4)Hz$, Theta $(4 \leq \theta < 8)Hz$, Alpha $(8 \leq \alpha < 13)Hz$, Beta $(13 \leq \beta < 30)Hz$, Gamma $(30 \leq \gamma < 100)Hz$), were used as features.
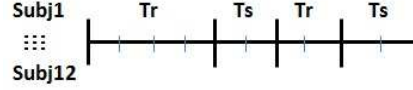


Fig. 2: Testing and training divisions

## 5    Experimental Results and Discussion

Two experiments were conducted, one aims at ranking channels based on their importance for this particular classification task, while the other aims at evaluating the three proposed multichannel DBN implementations and compare their performance with the benchmark classifiers of SVM, LDA and ELM. We divided the data of each session in to 10 segments, and used segments $[5, 6, 9, 10]$ for testing, while the remaining segments were used to train the classifiers, as shown in Figure 2. To increase the training data, we further divided each segment into 10 windows.
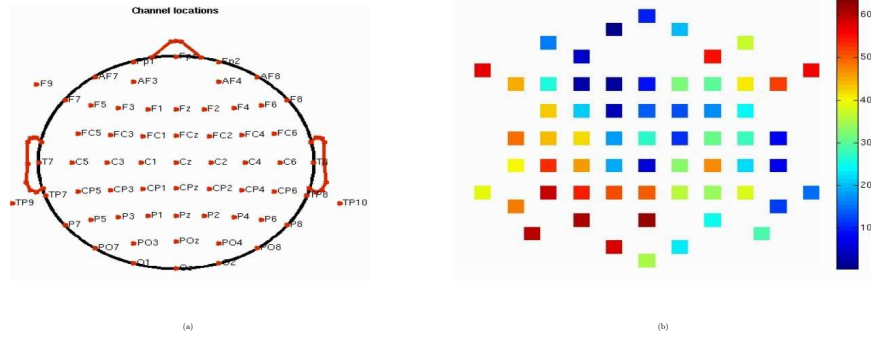


Fig. 3: (a) EEG electrode montage. (b) Classification accuracy for individual channels

We used a linear SVM, LDA and ELM classifiers to rank the performance of channels. The aim of using three classifiers is to identify channels that perform well for all three classifiers, and hence, reduce sensitivity to a particular classifier. We started by evaluating the performance of individual channels. In order

to reduce fluctuations in the performance of channels, we considered evaluating subsets of four neighboring channels, and then for each channel we averaged the accuracy of the subsets that consist of that particular channel. For example, the accuracy of channel $C4$ is obtained by averaging the accuracy of subsets ($\{C4, FC2, C2, FC4\}$, $\{C4, CP6, CP4, C6\}$, $\{C4, CP2, C2, CP4\}$, $\{C4, FC6, FC4, C6\}$). The 62 channels were ranked based on their classification performance, as shown in Figure 3(b). The color bar represents the rank order, where the best channels have dark blue color. The Figure indicates that channels located over the frontal midline perform relatively better than other regions of the brain, and hence, have the ability to differentiate between the three sessions of baseline, active tDCS and sham tDCS.

In the second experiment we constructed channel subsets of various sizes to have a better evaluation of the performance of the classifiers. These subsets are formed using frontal channels as well as channels of other regions of the brain. The channel combinations are listed in Table 1. Unlike the last experiment, each channel is used by itself in this experiment.

Table 1: Classification error rates of the six classification methods for each of the selected channel subsets

| Channels | Classification Methods | | | | | |
|---|---|---|---|---|---|---|
| | SVM | LDA | ELM | DBN1 | DBN2 | DBN3 |
| [F7,AF4] | 0.507 | 0.534 | 0.243 | 0.396 | 0.361 | 0.264 |
| [O1,AF3,FCz] | 0.319 | 0.333 | 0.298 | 0.389 | 0.340 | 0.257 |
| [TP9,CPz,POz] | 0.423 | 0.416 | 0.312 | 0.359 | 0.403 | 0.299 |
| [O1,TP9,P6,CPz] | 0.340 | 0.354 | 0.305 | 0.375 | 0.389 | 0.257 |
| [O1,O2,PO3,PO4,PO7,PO8] | 0.465 | 0.423 | 0.312 | 0.382 | 0.306 | 0.278 |
| [Fp2,TP9,AF3,AF4,CPz,POz] | 0.306 | 0.305 | 0.229 | 0.340 | 0.340 | 0.243 |
| [Fp1,Fp2,F7,F8,F9,F10,Fpz] | 0.444 | 0.326 | 0.229 | 0.292 | 0.292 | 0.236 |
| [Fp1,Fp2,AF3,AF4,AF7,AF8,Fpz] | 0.333 | 0.263 | 0.277 | 0.271 | 0.243 | 0.215 |
| [O1,O2,PO3,PO4,PO7,PO8,Oz,POz] | 0.347 | 0.284 | 0.326 | 0.333 | 0.285 | 0.243 |
| [P4,O2,P8,P2,PO4,P6,PO8,Oz] | 0.368 | 0.368 | 0.263 | 0.368 | 0.278 | 0.257 |
| [Cz,FC1,FC2,CP1,CP2,C1,C2,CPz] | 0.410 | 0.314 | 0.256 | 0.361 | 0.312 | 0.271 |
| [Fp1,F3,F7,F1,AF3,AF7,F5,F9,Fpz] | 0.333 | 0.298 | 0.256 | 0.278 | 0.243 | 0.215 |
| [Fp2,F4,F8,Fz,F2,AF4,AF8,F6,F10] | 0.431 | 0.291 | 0.326 | 0.306 | 0.292 | 0.222 |
| [P3,O1,P7,Pz,P1,PO3,P5,PO7,POz] | 0.486 | 0.430 | 0.291 | 0.368 | 0.361 | 0.285 |
| [C3,T7,FC5,CP5,TP9,FC3,CP3,C5,TP7] | 0.424 | 0.416 | 0.333 | 0.319 | 0.299 | 0.229 |
| [C4,T8,FC6,CP6,TP10,FC4,CP4,C6,TP8] | 0.389 | 0.305 | 0.333 | 0.347 | 0.333 | 0.181 |
| [Fp1,Fp2,F7,F8,AF7,AF8,F9,F10,Fpz] | 0.396 | 0.222 | 0.259 | 0.299 | 0.222 | 0.167 |
| [Fz,AF3,AF4,AF7,AF8,F5,F6,F9,F10,Fpz] | 0.299 | 0.229 | 0.270 | 0.257 | 0.188 | 0.111 |
| Average | 0.390 | 0.340 | 0.284 | 0.336 | 0.305 | 0.235 |

In order to evaluate the performance of the three DBN architectures, they were used to classify the subsets of selected channels. For the single-stream DBN shown in Figure 1(a), the two hidden layers consist of 200 and 150 units. For the other two DBNs, the four hidden layers consist of 200, 150, 200 and 150 units. The models were trained for 1200 iterations. The obtained results of the six classifiers shown in the table indicate that the ELM classifier performed noticeably better than both SVM and LDA. As for the DBN architectures, results obtained using the single-stream architecture was not found to produce very competitive results. The performance of the second DBN was slightly better than the first one, particularly for the large channel subsets. The multi-stream

DBN on the other hand achieved smaller error rates in most channel subsets compared to the other five classifiers, especially for the larger ones. Figure 4 shows a box-plot of the obtained classification error.

The obtained results indicate that DBN needs to be carefully designed when applied to the classification of EEG data, as traditional or single-stream architecture may not produced the expected outcomes. The same is also applicable to combining the classification results of individual channel DBN classifiers. The multi-stream DBN proved to be the most appropriate architecture, as it attempts to extract local attributes from each channel and combine those at a higher level. It is worth mentioning that we have not attempted to optimize the computational cost of the DBN algorithms, and hence, their execution time was noticeably higher than the other three classifiers.
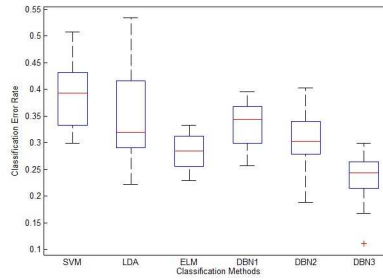


Fig. 4: Classification error of the six classification methods

## 6   Conclusion

In this paper, we evaluated the effectiveness of DBN in the classification of EEG data. Three DBN architectures were presented that fuse the channel information at different levels. In order to evaluate the proposed DBNs, we utilized 62-channel EEG data that was collected from patients suffering from major depressive disorder and started the tDCS treatment. The classification task involved differentiating between baseline EEG, active tDCS and sham tDCS sessions, and the data of each session was represented using the power spectral density in five frequency bands. We first identified the most relevant channels for classification and found the frontal midline channels to be more influential. We have then evaluated the performance of the three DBNs and found the one the multi-stream DBN to outperform the other two DBN implementations as well as the SVM, LDA and ELM classifiers. The multi-stream DBN attempts to extract local attributes from the individual channels in the lower network layers, which are processed and then combined in the higher unified layer(s) to extract high level information that facilitates more accurate labeling of the data. These findings demonstrate the capacity of DBNs as a classification platform for

multichannel EEG data. In future studies, we are planning to extend this work to differentiate patients that respond to the tDCS treatment from those who do not.

## References

1. Hinton, G.: A practical guide to training restricted boltzmann machines. Momentum **9** (2010) 926
2. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y.: Multimodal deep learning. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11). (2011) 689–696
3. Jacobson, L., Ezra, A., Berger, U., Lavidor, M.: Modulating oscillatory brain activity correlates of behavioral inhibition using transcranial direct current stimulation. Clinical neurophysiology **123** (2012) 979–984
4. Powell, T.Y., Boonstra, T.W., Martin, D.M., Loo, C.K., Breakspear, M.: Modulation of cortical activity by transcranial direct current stimulation in patients with affective disorder. PloS one **9** (2014) e98503
5. Wulsin, D., Gupta, J., Mani, R., Blanco, J., Litt, B.: Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. Journal of neural engineering **8** (2011) 036015
6. Zheng, W.L., Zhu, J.Y., Peng, Y., Lu, B.L.: Eeg-based emotion classification using deep belief networks. In: Multimedia and Expo (ICME), 2014 IEEE International Conference on, IEEE (2014) 1–6
7. Srivastava, N., Salakhutdinov, R.: Learning representations for multimodal data with deep belief nets. In: International Conference on Machine Learning Workshop. (2012)
8. Cheng, D., Sun, T., Jiang, X., Wang, S.: Unsupervised feature learning using markov deep belief network. In: Image Processing (ICIP), 2013 20th IEEE International Conference on, IEEE (2013) 260–264
9. Do, V., Xiao, X., Chng, E.: Comparison and combination of multilayer perceptrons and deep belief networks in hybrid automatic speech recognition systems. In: Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). (2011)
10. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313** (2006) 504–507
11. Wu, Y., Cai, H.: A simulation study of deep belief network combined with the self-organizing mechanism of adaptive resonance theory. In: Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on, IEEE (2010) 1–4
12. Hosseinifard, B., Moradi, M.H., Rostami, R.: Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from eeg signal. Computer methods and programs in biomedicine **109** (2013) 339–345
13. Kalu, U., Sexton, C., Loo, C., Ebmeier, K.: Transcranial direct current stimulation in the treatment of major depression: a meta-analysis. Psychological medicine **42** (2012) 1791–1800
14. Loo, C.K., Alonzo, A., Martin, D., Mitchell, P.B., Galvez, V., Sachdev, P.: Transcranial direct current stimulation for depression: 3-week, randomised, sham-controlled trial. The British Journal of Psychiatry **200** (2012) 52–59