COGNITIVE ANALYSIS OF WORKING MEMORY LOAD FROM EEG, BY A DEEP RECURRENT NEURAL NETWORK

Shiba Kuanar[1], Vassilis Athitsos[2], Nityananda Pradhan[3], Arabinda Mishra[4], K.R.Rao[1]

Department of [1]Electrical and [2]Computer Science Engineering, University of Texas Arlington, USA
[3] Department of Psychopharmacology, NIMHANS, India
[4] Institute of Imaging science, Vanderbilt University, Nashville, Tennessee, USA

ABSTRACT

One of the common modalities for observing mental activity is electroencephalogram (EEG) signals. However, EEG recording is highly susceptible to various sources of noise and to inter subject differences. In order to solve these problems we present a deep recurrent neural network (RNN) architecture to learn robust features and predict the levels of cognitive load from EEG recordings. Using a deep learning approach, we first transform the EEG time series into a sequence of multispectral images which carries spatial information. Next, we train our recurrent hybrid network to learn robust representations from the sequence of frames. The proposed approach preserves spectral, spatial and temporal structures and extracts features which are less sensitive to variations along each dimension. Our results demonstrate cognitive memory load prediction across four different levels with an overall accuracy of 92.5% during the memory task execution and reduce classification error to 7.61% in comparison to other state-of-art techniques.

Index Terms— RNN, LSTM, Softmax, EEG, FFT.

1. INTRODUCTION

EEG is a noninvasive neuroimaging modality which measures the electrical signal changes on the scalp induced by cortical activity. Using the classical blind source separation analogy (ICA), EEG data can be considered similar to multi-channel speech signals obtained from several electrodes. These electrodes record signals and modulate the cortical activities. Recent EEG-based mental state recognition techniques used manual feature selection from time series and applied supervised machine learning techniques to learn discriminative manifolds between the states [2]. But the main challenge in correctly recognizing mental states has been to construct a model that is robust to signal noise and distortion. Variations occur due to the presence of inter-subject differences and signal acquisition constraints. However most of variations originate from differences in individual cortical mapping. Spatial variations in responses may also be caused by imperfect placing of caps at predetermined cortical regions and heads of different shapes. The source code for this paper is available on http://omega.uta.edu/~spk7522/Cognitive/EEG/

The proposed deep learning approach learns representations from EEG data and appears to be more robust to inter subject differences and unwanted acquisition noise. We transform EEG data into a multi-dimensional array tensor and obtain a sequence whose topology retains spatial information. Once such multi-spectral frames are obtained, we train those video frame sequences using our proposed recurrent architectures. We use a convolutional neural network (ConvNet) to extract the spatial and spectral invariant EEG representations, and an RNN to extract temporal patterns in sequential frames. Overall our proposed model is able to preserve the spectral, spatial and temporal structure of EEG data and extract more robust features for further analysis.

2. RELATED WORK

In recent years deep neural networks have achieved great success in classification [4, 5] and pattern recognition tasks [19] within a wide range of speech, text, video and image applications. ConvNets have demonstrated the ability to extract features that are invariant to translation, deformation (rigid/non-rigid) and rotation of input patterns [20]. In handwriting and speech recognition [7, 16], the RNN architecture has delivered state-of-the-art performance using the temporal sequence dynamics. A combination of ConvNet and RNN networks has been used for video classification [9, 1], and extracting representations from EEG series [15, 8] to evaluate medical diagnostic accuracy. ConvNets have already been used to learn features from Magnetic Resonance Imaging (resting state and stimulus driven fMRI) with moderate datasets [12]. Despite the successes, deep neural network applications remain relatively unexplored in neuroimaging area.

3. METHODS

The human brain contains many diverse networks which are responsible for many specialized tasks like working memory (WM).The WM retains information for a short duration and it is crucial for brain information manipulation. Working memory capacity can limit the individual's ability in a range of cognitive tasks [8]. Increasing the cognitive load over an individual's capacity can lead to a state of confusion and diminishes learning ability [21]. Therefore, recognizing

individual working memory loads is important for applications such as human computer interaction and brain computer interfaces.

3.1. Data Recording and Preprocessing

We collected our datasets from the EEG cognitive database of the Psychopharmacology Department, NIMHANS. Twenty five subjects (ten female) of age 16-28 performed a standard WM experiment. EEG signals were recorded from 64 electrodes placed over the scalp at standard 10-20 locations. The data were acquired at 256 Hz through each channel from Neurofax EEG-1200 (Nihon Kohden) machine. The raw EEG signals were then filtered through a band pass filter to remove unwanted signals. Three subjects' data were excluded because of noise and artifacts. The digitalized data were then ported to a computer workstation for further analysis.

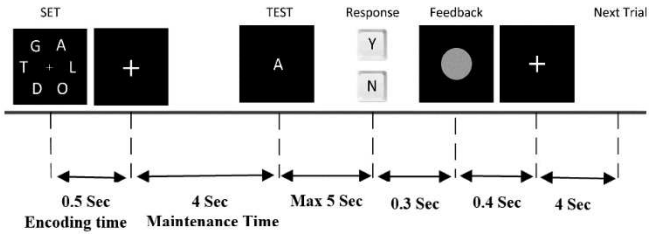

Figure 1: Working memory experiment diagram.

Figure 1 illustrates the experiment process. First, an array of English characters in SET was shown for 0.5 seconds and participants were instructed to memorize the characters. A TEST character was shown after 4 seconds and participants indicated whether the test character was in the SET array or not by pressing a button. Each participant repeated the experiment for 320 times. In each trial the number of characters were randomly chosen from the set {4, 6, 8, 10}. These characters determined the quantity of cognitive load introduced to the subject. We labeled each of the task conditions containing 4, 6, 8, 10 characters with loads 1- 4 respectively. The brain activity was recorded during the above 4.5 seconds trial in which patients kept information in their memory and recognized as a mental workload. A total of 6490 correctly responded samples were collected from 22 subjects and assigned to four different classes corresponding to loads from 1 to 4. The task of the classification was to recognize load levels corresponding to the character set size from recordings. EEG signals from each trial of 4.5 sec were sliced into 0.5 sec pieces through an offline windowing process, and an image was constructed over each time slice, to produce nine frames for training. We followed the leave-subject-out cross validation technique [13] by repeatedly splitting 22 fold dataset into test, validate, training datasets and evaluated the performance of classifier.

3.2. EEG Feature

On each subject trail the time intervals from SET to TEST were recorded for each electrode and these time spans contained the total encoding and maintenance stages of the WM operation. The power spectra for each time sliced window (0.5 sec) was estimated by applying Fast Fourier transform (FFT). In our EEG analysis the whole frequency spectrum were divided into three sub-bands: theta (4-7Hz), alpha (8-13Hz), beta (13-30 Hz). Based on numerous evidence the above three frequency bands were chosen for our cognitive experiment [3] and aggregated the feature vectors. The mean spectral power within the three sub-bands was calculated by averaging associated FFT magnitudes and considered as a feature. Finally the 192 features (64 channels x 3 bands) were combined to form a big feature vector.

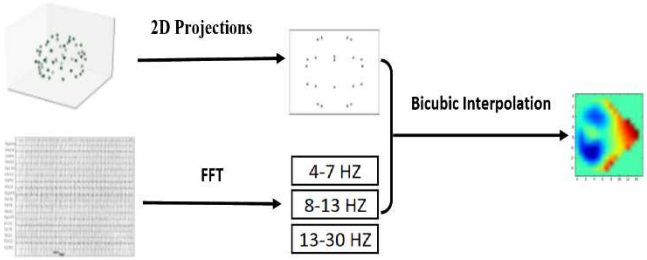

Figure 2: Image based representation of EEG signals

3.3. Images from Multichannel Time Series (EEG)

The EEG signal included multivariate time series which correspond to measurements across spatial cortex locations. We computed the sum of the squared absolute power values for each of the theta, alpha and beta frequency bands associated with each electrode. We then transformed the measurements into 2D images to preserve spatial structures and corresponding color channels to represent the spectral dimensions. Finally image frame sequences were derived from consecutive time windows and accounted for our temporal evolutions. In our experiment we projected scalp electrode locations from 3D space to 2D surface [8] and transformed spatially distributed activity maps as 2D frames. The Azimuthal Equidistant (Polar) Projection technique [10, 8] was used to preserve relative distance between neighboring electrodes. The x and y dimensions of the image represented the spatially distributed activities over the cortex. We applied Clough Tocher technique [11] to interpolate scattered power over scalp and estimated intermediate electrode values over a 32× 32 mesh. This procedure, repeated for each of the three sub-bands, resulted into three topographical activity maps. The spatial maps were merged together to form color images with 3 channels and was presented as input to ConvNet (Figure 2).

4. NEURAL NETWORK MODELS

We adopted a hybrid combination of ConvNet and RNN (Figure 3) to deal with the inherent structures of EEG data. The ConvNet was used to handle the variations in space and frequency domains because of its ability to learn 2D data representations. The extracted ConvNet feature vectors (FV) were fed into recurrent LSTM layers to learn the temporal variations. We evaluated the cognitive state classification problem using multi frame approach. Each trial was divided

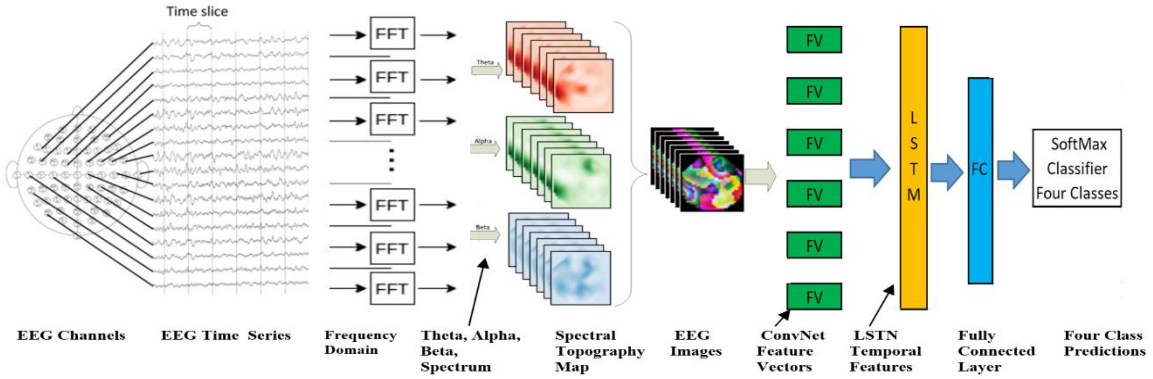

Figure 3: Our proposed framework overview: (i) EEG signals from multiple cortex locations (ii) FFT and topographical maps (iii) Spectral maps combined to form 3 channel images, (iv) ConvNet FV and LSTM for representation learning (v) Softmax classification.

into 0.5 sec time slices, images were constructed over each time window, and those images were used as input to our network.

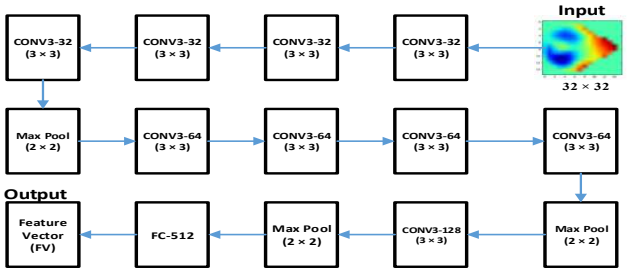

Figure 4: Convolutional neural network architecture configuration

4.1. Convolutional Neural Network Architecture (CNN)

Our ConvNet network is summarized in Figure 4. It contained nine conv layers and one fully connected layer. The input color image to ConvNet was a fixed size of 32×32. The image was passed through a stack of conv layers with a small 3×3 receptive field and stride 1. To restore the spatial resolution, intermediate conv layer inputs were zero padded of one pixel. All hidden layers were equipped with ReLU non-linearity. Multiple conv layers were stacked together and then followed by a 2 × 2 max-pool layer with stride 2. Finally the conv layer parameters were denoted as "conv <receptive field size>-<number of channels>".

4.2. Recurrent Neural Networks (RNN)

The ConvNet outputs were reshaped as sequences of frames and later used to investigate the temporal sequence in maps. Inspired by deep learning video classification techniques [1] we evaluated two models 1) Bidirectional LSTM (BiLSTM) and 2) Long Short-Term Memory (LSTM) to extract the temporal information (Figure 5). The RNN model [14] considered the sequence of CNN activations, processed forward inputs $x = \{x_1 \ldots x_T\}$, computed hidden vector $h = \{h_1 \ldots h_T\}$ and output responses $y = \{y_1 \ldots y_T\}$ by iterating equations from time t = 1 to T: $h_t = H(W_{xh} \times x_t + W_{hh} \times h_{t-1} + b_h)$; $y_t = W_{hy} \times h_t + b_y$. The W, b and h terms denotes weight, bias and hidden function respectively. The brain activity is a dynamic process which shows the temporal fluctuation over time. These temporal variations among frames might contain useful information about the underlying mental states. Given the dynamic nature of neural responses, RNN framework appeared to be reasonable modeling the temporal brain dynamics. The hidden function (h) for our LSTM network was computed by the below set of equations:

$$i_t = \sigma(W_{xi} \times x_t + W_{hi} \times h_{t-1} + W_{ci} \times c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf} \times x_t + W_{hf} \times h_{t-1} + W_{cf} \times c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(W_{xc} \times x_t + W_{hc} \times h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo} \times x_t + W_{ho} \times h_{t-1} + W_{co} \times c_t + b_o) \quad (4)$$

$$h_t = o_t \times \tanh(c_t) \quad (5)$$

where σ is the sigmoid function. The LSTM model components: input, forget, cell activation vector and output gate were denoted as i, f, c, and o respectively. According to our dataset limits we used two LSTM layers each with 64 memory cells. The complete LSTM sequence of frames were propagated to FC layer (Figure 5) and prediction was made by Softmax classifiers. Bidirectional LSTMs [6, 7] processed the EEG data in both forward and backward directions using two separate hidden layers and can access long frames in both input directions. As illustrated in Figure 5, BiLSTM computed backward hidden sequence $\overleftarrow{h}$, forward hidden sequence $\overrightarrow{h}$ and updated output y; by iterating backward layer from t = T to 1 and forward layer from t = 1 to T [7]. Hence at every point in a given time sequence, BiLSTM had the information about all points before and after it.

5. NETWORK TRAINING

Our ConvNet network was trained by optimizing the cross-entropy cost function using stochastic gradient decent (SGD) and backpropagation. We trained our RNN network with Adam parameter update and a learning factor of 1×10^{-4}. The first and second momentum decay rates were set to 0.90 and 0.99 respectively. The batch sizes were set to 30 and training was regulated by L_2 weight decay of 0.0001. To overcome the overfitting issue we adopted dropout method [5] with a probability of 0.5 in FC layer. The network parameters converged after around 900 iterations with six epochs. The data was augmented by adding Gaussian noise to the image. We experimented with various noise levels. Our implementation was derived from publicly available Python based Theano framework and performed 18 hours training on a NVIDIA K40 GPU machine. We compared our results

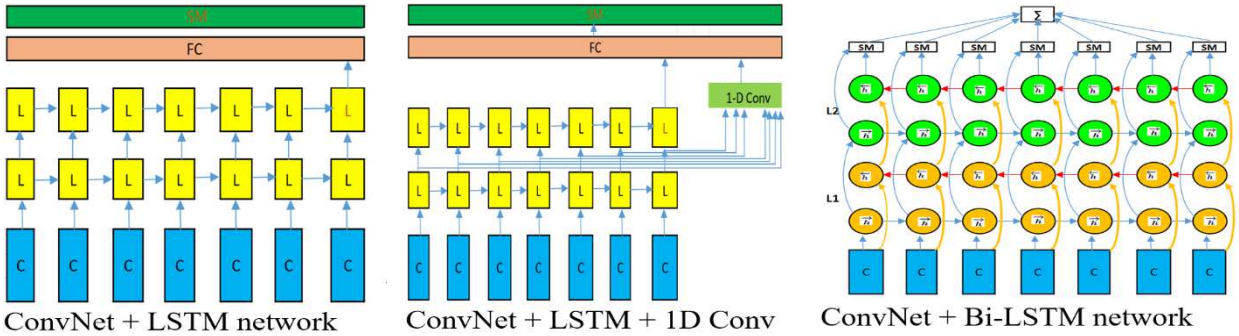

Figure 5: Different LSTM (L) models with ConvNet (C); BiLSTM (L1, L2); FC: Fully Connected Layer SM: Softmax

against the commonly used classifiers: Random Forest (RF), Support Vector Machines and Logistic Regression. The SVM parameters: regularization penalty (C) and RBF kernel $\gamma = 1/2\sigma$ were selected by a grid search through cross validation on a training set ($C = \{0.01, 0.1, 1, 10, 100, 1000\}$, $\gamma = \{0.1, 0.2... 1, 2... 10\}$). The number of trees for RF were varied within a set of {10, 20, 50, 100, 500}.Each decision tree output was computed form a random set of input features and final class was selected with majority voting. L1-regularization was introduced on our Logistic Regression classification and solved the unconstrained optimization.

Table 1: Classification results of different architectures

Architecture	Test Errors (%)	Validation Error	Number of Parameters
SVM	14.96	-	-
Logistic Regression (L1)	14.45	-	-
Random Forest	12.23	-	-
ConvNet + LSTM	9.87	6.13	1.29 Mil
ConvNet+ LSTM+1D-Conv	8.34	8.32	1.47 Mil
ConvNet + Bidirectional LSTM	7.61	8.11	1.66 Mil

6. RESULTS

We empirically chose the ConvNet described on Figure 4 and applied it on EEG image frames. We explored three different approaches and aggregated the temporal features from multiple frames (Figure 5). Using LSTM and BiLSTM structures, the classification accuracy improved significantly (Table 1). The accuracies on individual subjects show that our three models achieved a consistent improvement on classification accuracies except S3, S4, S5, S6, S15 and S20 (Table 2). The average accuracy of BiLSTM was 92.5%, which was higher than conventional methods.

Table 2: Classification accuracy results for subjects folds

Test Subjects	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
LSTM	88.45	71.27	93.22	97.43	98.2	81.1	94.5	93	86	85.25	87.4
LSTM + 1D Conv	89.9	75.3	92.5	96.4	95.4	94.5	96.4	95.8	91.8	93.45	90.5
BiLSTM	94.5	86.5	96.8	98.5	97.3	95.3	99.25	97.7	99.5	97.5	94.5

Test Subjects	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22
LSTM	80.5	46.7	81.45	92.53	89.3	100	91.4	90.5	82.4	80.5	47.5
LSTM + 1D Conv	81.7	50.62	92.5	87	96.5	100	93.5	95	87.2	81.65	51.4
BiLSTM (Mix)	89.8	78.5	95.2	92.5	98.45	97.3	94.34	96.34	75.4	88.6	71.3

Average Accuracy (%):	BiLSTM (Mix) = 92.5	LSTM + 1D Conv = 87.68	LSTM = 84.48

It highlights the role of the LSTM network in extracting features and demonstrates the effectiveness of our model in learning temporal dynamics. Table 1 also shows that classification test errors lowered significantly when the temporal LSTM models were added. The validation loss over number of training set epochs is shown on Figure 6. The ConvNet maxpooling operation created the invariant feature maps in deeper layers and this could hamper overall performance if map size was reduced to an extent where the regional activities cannot be distinguished. Our ConvNet learned a stack of filters which introduced nonlinearity on feature maps and maximized classification accuracy.

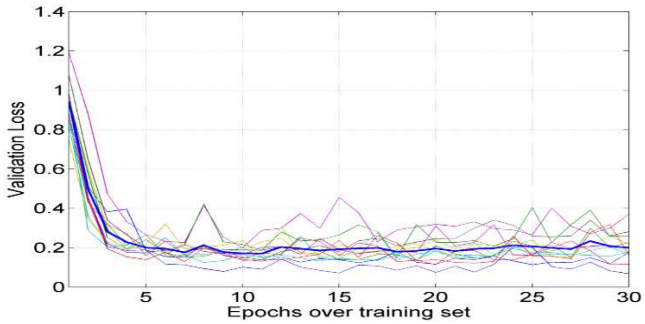

Figure 6: Validation Loss along the epochs. Average loss→ Blue line.

7. CONCLUSION

The objective of this work was to find robust representations from the EEG multi-channel time series that were invariant to inter-subject differences and data acquisition noise. We followed a methodology to learn spatial, spectral and temporal representations from the EEG datasets and demonstrated its advantages in the context of cognitive memory load classifications. Our implementation was different from the previous attempts and learned the robust representations from EEG image sequences using a ConvNet and BiLSTM hybrid network. Our proposed hybrid network demonstrated the significant improvements in finding better classification accuracy i.e. up to 92.5% over various existing LSTM models. In future, we would like to experiment on the unsupervised generative frameworks with larger labeled and unlabeled EEG datasets prior to training the network with task-specific data.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grants IIS 1565328 and IIP 1719031.

REFERENCES

[1] J. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, G. Toderici, "Beyond Short Snippets: Deep Networks for Video Classification", IEEE, CVPR, Boston, vol 2, June 2015.

[2] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, "A review of classification algorithms for EEG-based Brain-Computer Interfaces", Journal of Neural Engineering, IOP Publishing Ltd, vol 4, Number 2, Jan 2007.

[3] Jensen, O. and Tesche, C. "Frontal theta activity in humans increases with memory load in a working memory task", European Journal of Neuroscience, Blackwell Science, Ltd, vol. 15, pp. 1395-1399, 2002

[4] Karpathy, A. and Toderici, G. "Large scale video classification with convolutional neural networks", IEEE, CVPR, Columbus, Ohio, pp.1725–1732, June, 2014.

[5] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", NIPS, Lake Tahoe USA, pp. 1097–1105, Dec, 2012.

[6] M. Wöllmer, F. Eyben, A. Graves, B. Schuller, G. Rigoll, "Bidirectional LSTM networks for context-sensitive keyword detection in a cognitive virtual agent framework", Cognitive Computation, vol 2, issue 3, pp 180-190, Sep 2010

[7] A. Graves, N. Jaitly, A. R. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM", ASRU, IEEE Workshop, pp 273–278 2013.

[8] P. Bashivan, I. Rish, M. Yeasin, N. Codella, "Learning representations from EEG with deep recurrent convolutional neural networks", ICLR, Puerto Rico, May, 2016.

[9] Simonyan K and Zisserman A. "Very Deep Convolutional Networks for Large-Scale Image Recognition", ICLR, San Diego USA, pp. 1–14, May, 2015.

[10] J. P. Snyder, "Map projections–A working manual", US Government Printing Office, Washington USA, vol 1395, 1987.

[11] P. Alfeld, "A trivariate cloughtocher scheme for tetrahedral data. Computer Aided Geometric Design", Journal Computer Aided Geometric Design archive, Salt Lake City USA, pp 169-181, Nov, 1984.

[12] S. M. Plis, D. R. Hjelm, R. Slakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. Johnson, J. Paulsen, J. Turner, V. D. Calhoun, "Deep learning for neuroimaging: a validation study", Frontiers in Neuroscience, Article 229, Aug 2014.

[13] S. Arlot, A. Celisse, "A survey of cross-validation procedures for model selection", Statistics Surveys, vol 4, pp. 40-79, July 2010.

[14] Hochreiter, S. and Schmidhuber, J. "Long Short-Term Memory", Journal Neural Computation, MIT Press, vol 9 , pp 1735-1780, Nov, 1997

[15] N. F. Güler, E. D. Übeyli, İ. Güler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification", Expert Systems with Applications, Tarrytown, NY, vol 29, issue 3, pp 506-514, Oct 2005.

[16] A. Graves, M. Liwicki, H. Bunke, J. Schmidhuber, S. Fernández, "Unconstrained online handwriting recognition with recurrent neural networks", NIPS, Vancouver Canada, pp 577-584, Dec, 2007.

[17] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, "Greedy layer-wise training of deep networks. Advances in neural information processing systems", NIPS, MA USA, pp 153-160 Dec 2006.

[18] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors" arXiv preprint arXiv: 1207.0580, 2012.

[19] A. Graves, A.R. Mohamed, G. Hinton, "Speech recognition with deep recurrent neural networks", ICASSP, IEEE, Vancouver Canada, pp. 6645–6649, vol 1, May 2013.

[20] X. Zhang, J. Zhao, Y. LeCun, "Character-level Convolutional Networks for Text Classification", NIPS, Montreal Canada, vol. 3, Dec 2015

[21] J. Sweller,Jeroen J. G. van Merrienboer, Fred G. W. C. Paas, "Cognitive architecture and instructional design", Educational Psychology review, vol. 10, Issue. 3, pp 251–296, Sep, 1998.