International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Robust unspoken Speech (Electroencephalogram) recognition algorithm using Long Short Term Memory –Deep Neural Networks approach

K Jeevan Reddy[1], Dr S P VenuMadhava Rao[2]

[1]*Associate Professor, ECE Department, SNIST, Hyderabad, (India)*

[2]*Principal, KPRIT, Hyderabad, (India)*

ABSTRACT

The brain signal recognition speech generally referred as unspoken speech is one of the most challenging task. EEG signals are highly non stationary and require complex processing in analyzing the signals.This paper proposes the application of a deep neural network (DNN) to discover unknown feature correlation between input signals that is crucial for the learning task. The DNN is implemented with a stacked autoencoder using hierarchical feature learning approach. Input features of the network are power spectral densities of 32-channel EEG signals from 32 subjects. To alleviate overfitting problem, principal component analysis (PCA) is applied to extract the most important components of initial input features. Long short term memory provides behavioral information on lapse events with good temporal resolution. We propose an automated behavior grading system and trained it to estimate the mean opinion of 3 human raters on the likelihood of a lapse. We then trained an LSTM neural network to estimate the output of the lapse rating system given only EEG spectral data through discrete arithmetic discrete wavelet transform. The detection system was designed to operate in real-time without calibration for individual subjects.

I.INTRODUCTION

Human-computer interface (BCI) is an emerging and complex biomedical engineering research fields for years. It provides a promising technology allowing humans to control external devices by modulating their brain waves. Most BCI applications have been developed for noninvasive brain signal processing which is practical to implement in real-world scenarios. There are plenty of successful EEG-based BCI applications such as word speller programs [1] and wheelchair controllers [2]. Not only can BCI be employed to mentally control devices, but also it can be implemented for understanding our mental states. Emotion recognition is one of such applications. Automatic emotion recognition algorithms potentially bridge the gap between human and machine interactions.A model of emotion can be characterized by two main dimensions called valence and arousal. The valence is the degree of attraction or aversion that an individual feels toward a specific object or event. It ranges from negative to positive. The arousal is a physiological and psychological state of being awake or reactive to stimuli, ranging from passive to active.

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

The valence-arousal dimensional model, represented in Figure 1, of emotion is widely used in many research studies.

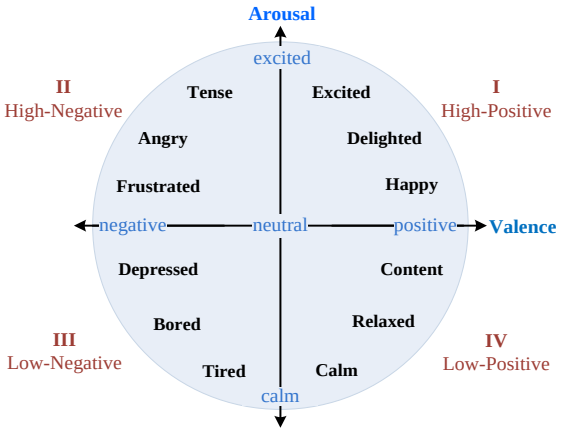

Figure 1 : Valence-Arousal modal.

The electroencephalogram (EEG) measures the activity of large numbers (populations) of neurons.EEG recordings are noninvasive, painless, do not interfere much with a human subject's ability to move or perceive stimuli, are relatively low-cost.Electrodes measure voltage-differences at the scalp in the microvolt (μV) range.Voltage-traces are recorded with millisecond resolution – great advantage over brain imaging (fMRI or PET).

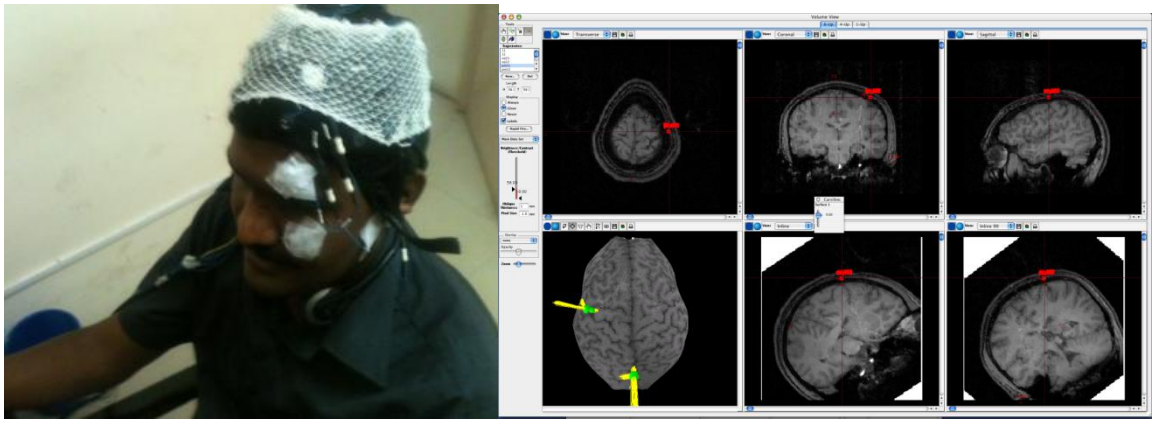

Figure 2 : standard EEG with fMRI analysis

Standard placements of electrodes on the human scalp: A, auricle; C, central; F, frontal; Fp, frontal pole; O, occipital; P, parietal; T, temporal. The additional analysis is done using fMRI on active subjects while concurrently recording both voice and EEG. EEG rhythms correlate with patterns of behavior (level of attentiveness, sleeping, waking, seizures, coma).

Rhythms occur in distinct frequency ranges:

Gamma: 20-60 Hz ("cognitive" frequency band)

Beta: 14-20 Hz (activated cortex)

Alpha: 8-13 Hz (quiet waking)

Theta: 4-7 Hz (sleep stages)

Delta: less than 4 Hz (sleep stages, especially "deep sleep")

Higher frequencies: Active processing, relatively de-synchronized activity (alert wakefulness, dream sleep).

Lower frequencies: Strongly synchronized activity (nondreaming *sleep, coma).*

II. LONG SHORT TERM MEMORY

An LSTM is a special kind of RNN architecture, capable of learning long-term dependencies.An LSTM can learn to bridge time intervals in excess of 1000 steps. LSTM networks outperform RNNs and Hidden Markov Models (HMM): Speech Recognition: 17.7% phoneme error rate on TIMIT acoustic phonetic corpus. Winner of the ICDAR handwriting competition for the best known results in handwriting recognition. This is achieved by multiplicative gate units that learn to open and close access to the constant error flow.

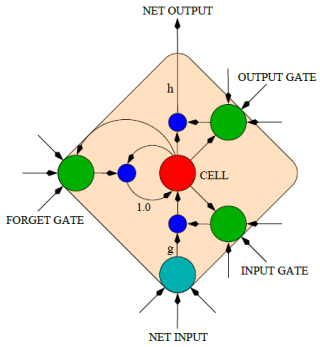

Figure 3 :LSTM Memory cell

Traditional RNNs are a special case of LSTMs: Set the input gate to all ones (passing all new information), the forget gate to all zeros (forgetting all of the previous memory) and the output gate to all ones (exposing the entire memory).

2.1 LSTM Equations

- $i = \sigma(x_t U^i + s_{t-1} W^i)$
- $f = \sigma(x_t U^f + s_{t-1} W^f)$
- $o = \sigma(x_t U^o + s_{t-1} W^o)$
- $g = \tanh(x_t U^g + s_{t-1} W^g)$

- $c_t = c_{t-1} \circ f + g \circ i$
- $s_t = \tanh(c_t) \circ o$

- i: input gate, how much of the new information will be let through the memory cell.
- f: forget gate, responsible for information should be thrown away from memory cell.
- o: output gate, how much of the information will be passed to expose to the next time step.
- g: self-recurrent which is equal to standard RNN
- c_t: internal memory of the memory cell
- s_t: hidden state
- v: final output

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

2.2 DBLSTM training and regularization

End-to-end training methods :

- Connectionist Temporal

Classification (CTC);

- RNN Transducer.

Regularization:

Early stopping: monitoring the model's performance on a validation set.

weight noise: adding Gaussian noise to the network weights during training. Weight noise was added once per training sequence, rather than at every time step.

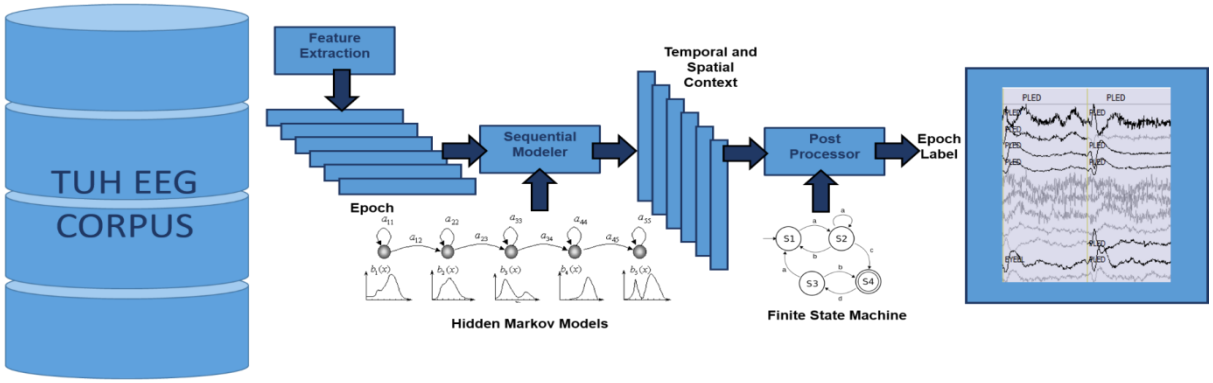

Figure 4: Training EEG Signals

III. METHODOLOGY

We set the number of convolutional filters as 30 in the first convolutional layer to extract 30 different kinds of correlation information, namely 30 different features. At the same time, to extract the multiple scale spatial characteristics of MFI, we use different size receptive fields in the first convolutional layer. The field sizes are 2 × 2 pixels, 5 × 5 pixels and 10 × 10 pixels, respectively. Corresponding to the different sizes of the field, the strides are 2, 5 and 10 pixels, respectively, without overlap between the strides. The activation function is ReLU. Following the first convolutional layer is a max pooling layer with pooling size of 2 × 2, and the strides are 2. The second convolutional layer is set as 10 different filters with a size of 2 × 2 without overlap between strides. This setting helps to further fuse the information of a specific scale range from the prior features. The structure of the hybrid deep neural networks used for emotion classification.

The Construction of Convolutional Neural Networks The inputting MFI size of the networks is 200×200 pixels, and it contains three color channels. We set the number of convolutional filters as 30 in the first convolutional layer to extract 30 different kinds of correlation information ,namely 30differentfeatures. At the same time, to extract the multiple scale spatial characteristics of MFI, we use different size receptive fields in the first convolutional layer. The field sizes are 2×2 pixels, 5×5 pixels and 10×10 pixels, respectively. Corresponding to the different sizes of the field, the strides are 2, 5 and 10 pixels, respectively, without overlap between the

strides. The activation function is ReLU. Following the first convolutional layer is a max pooling layer with pooling size of 2×2, and the strides are 2. The second convolutional layer is set as 10 different filters with a size of 2 × 2 without overlap between strides. This setting helps to further fuse the information of a specific scale range from the prior features.

IV. EXPERIMENTAL SETUP AND RESULTS

The proposed EEG-based emotion recognition system is implemented with a stack of three autoencoders with two softmax layer,.The system performs emotion classification by estimating valence and arousal states separately. Two softmax classifiers, one for valence and another for arousal, can share the outcome of unsupervised pretraining procedure because they both use the same set of unlabeled raw data. However, two softmax classifiers need to use different stacked autoencoders during fine-tuning back propagation.The DLN utilizes the unsupervised pretraining technique with greedy layerwise training, starting from the input layer to the softmax layer. The first sparse autoencoder (1st hidden layer) is trained on the inputs' features (230 power spectral features) to learn the primary features on these input features. We use L-BFGS to optimize the cost function, squared error between input features and outputs. Subsequently, the algorithm performs forward propagation by using the input features into this trained sparse autoencoder to obtain the primary feature activations. The features, deriving from feedforward propagation of the 1st hidden layer, must be used to perform unsupervised pretraining in the second hidden layer. The algorithm computes its features in the same procedure from the learned features from the previous hidden layers.

The weight and bias parameters of the softmax layer are trained by using a supervised learning approach. The output features of the last hidden layer are used as the input features of both softmax layers. We use a set of self-assessment emotion states (valence and arousal) of subjects as a ground truth. These softmax layers can be trained as the parameters concurrently. After the network finishes learning weight and bias parameters in both softmax classifiers, the algorithm has to perform fine-tuning of all weight and bias parameters in the whole network simultaneously. However, we are not able to use the same network parameters for two classifiers. We need to save the learned parameter outcomes of unsupervised pretraining and load the parameters for fine-tuning process of another softmax classifier. The fine-tuning process treats all layers of a stacked autoencoder and softmax layer as a single model and improves all the weights of all layers in the network by using backpropagation technique with supervised approach. The backpropagation process is used to learn the network weights and biases based on labeled training examples to minimize the classification errors.The algorithm performs a greedy layerwise unsupervised pertaining process, starting from the first hidden layer to the last hidden layer. Initial weights and biases of the trained hidden layer are assigned for parameter optimizations. Next, the features from feed forward propagation of the hidden layer must be used to perform unsupervised pretraining in the next hidden layer. After finishing unsupervised pretraining in the last hidden layer, softmax training and fine-tuning procedures are required.

Covariate Shift Adaptation of Principal Components

Deep learning networks implemented with stacked autoencoders have capability of representing a highly expressive abstraction. Therefore, we are confronted with overfitting problems, especially with the tremendous number of input features and hidden nodes. Moreover, a nonstationary effect of EEG signal is still challenging to develop a reliable EEG-based emotion recognition. The proposed system employs the concept of principal component based covariate shift adaptation [22] to handle both overfitting problems and nonstationary effects simultaneously. Principal component analysis (PCA) [23] is to extract the most important principal components and normalize these components individually by shifting a window over the data to alleviate the effect of nonstationarity.

PCA is a statistical method that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance. The proposed system reduces the number of input features from 230 to 50 features.
To minimize the nonstationary effects of input features, the proposed system normalizes the input features with the average of previous feature values within a rectangular window of length . We performed this normalization for each input feature individually. Figure 11 illustrates the shifting window during input feature normalization for covariate shift adaptation in each video trial.

In our experiments, the efficiency of our proposed EEG-based emotion recognition system was evaluated by four experiment setups. In the first setup, we implemented the emotion recognition by using a deep learning network with 100 hidden nodes in each layer (DLN-100). We employed the feature extraction process to calculate all of input features of the DLN from 32-channel EEG signals. At each epoch, the system learned 230 input features consisting of power spectral density of 5 frequency bands and the differences of power spectral densities of 14 asymmetry pairs. Next, the second experiment reduced the number of hidden nodes to 50 (DLN-50) for investigating the effect of hidden node size in the DLN.The PCA extracted the 50 most important components from initial 230 input features. The extracted features were fed into the DLN with 50 hidden nodes in each layer.The last experimental setup enhanced the efficiency of the emotion recognition system by applying covariate shift adaptation (CSA) concept to solve the problem of nonstationarity in EEG signals. The system normalized the input features with the average of previous feature values within a rectangular window of length. This normalization was processed for each input feature individually.

The classification accuracy of valence and arousal states in four experiment setups was measured with a leave-one-out cross validation scheme. The full leave-one-out cross validation of 32 subject acquisitions was performed. A training dataset was a composition of all input features from the other 31 subjects. A test dataset was the subject's input features under evaluation. Each individual dataset consisted of power spectral features from EEG signal records while the subject was watching 40 one-minute music videos. The DLN performed its

weight and bias optimization based on gradient descent approach. Therefore, the classification accuracy was occasionally affected by its initial weight and bias parameter. In our experiment, we repeated the classification accuracy measurement five times and used the average of the accuracy for further analysis.

The average accuracy and standard deviation of 32 subjects in four experiments are depicted in Figure 13. The DLN-100 provides the accuracy of 49.52% for valence and 46.03% for arousal. The DLN-50 accuracy slightly decreases into 47.87% and 45.50%. The number of hidden nodes in the DLN affects accuracy performance of affective state classification. The greater the number of hidden nodes is, the higher accuracy the DLN provides. In experiments, the number of hidden nodes in each layer was reduced from 100 to 50 nodes. The accuracy decreased 1.62% and 0.53% for valence and arousal classifications, respectively.

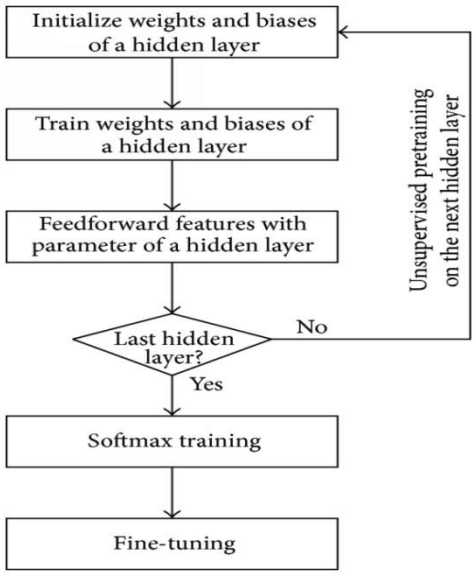

Figure 5:DNN Training Procedure Flow Chart

V. DISCUSSION AND CONCLUSION

The primary purpose of this research is to explore how well the deep learning network in the version of stacked autoencoder performs EEG-based affective computing algorithm. From our experimental results, the average of emotion classification accuracy from the deep learning network with a stack of autoencoders is better than existing algorithms. Consequently, the DLN is a promising alternative as EEG-based emotion classifier. However, one of the most challenging limitations for performing EEG-based emotion recognition algorithm is coping with the problem of intersubject variations in their EEG signals.

There are several promising methods to handle the inter subject variations. Lotte and Guan [27] proposed an algorithm for learning features from other subjects by performing regularization of common spatial patterns (CSP) and linear discriminant analysis (LDA). The method regularized the estimated covariance matrix toward the average covariance matrix of other subjects. Samek et al. [28] studied transferring information about nonstationarities in data, instead of learning the task-relevant part from others. These principal nonstationarities

are similar between subjects and can be transferable. Also they have an adverse effect on classification performance, and thus removing them is favorable. We plan to implement one of these two methods, depending on the nonstationary characteristics of the dataset, for alleviating the intersubject variations in our next version of EEG-based emotion recognition system.

One of the major limitations of the DLN is its tremendous amount of computational time requirement during unsupervised pretraining and supervised fine-tuning procedures. In our experiment setup, the DLN for EEG-based emotion recognition is constituted of three stacks of hidden layers and each hidden layer has 100 hidden nodes. At each epoch, the algorithm learned 230 input features. To estimate an individual subject's classification accuracy, there were in total 31 subjects watching 40 videos, each of 60 seconds (74,400) epochs. They are used to adjust the weight and bias parameters of the DLN. Table 1 shows other DLN's parameter settings. The approximated time used to train the DLN is 20–25 minutes on a laptop computer (Core i5-3320M 2.6 GHz, RAM 8 GB, SSD 128 GB, Windows 7 64-bit Professional).

To speed up training time of the DLN, we are able to exploit some parallelism between two softmax classifiers. However, we need to duplicate the stack of autoencoder implementation for valence and arousal states. Both stacks of autoencoders can be used for separated fine-tuning process of valence and arousal simultaneously. During unsupervised pretraining, two softmax classifiers can share the outcome of unsupervised pretraining procedure because they both use the same set of unlabeled raw data. After completing all sequences of DLN training procedure, shown in Figure 10, the DLN can be used to classify emotion states in real time. Even though the DLN requires tremendous amount of training time, it is able to perform EEG-based emotion classification in real time. During classification phase, the DLN simply feeds the input features through all layers of the network. To give better response, we are able to decrease the window size of covariate shift adaptation but we may trade off with lower classification accuracy.

The proposed EEG-based emotion recognition is implemented with a deep learning network and then enhanced with covariate shift adaptation of the principal components. The deep learning network is constituted of a stack of three autoencoders and two softmax classifiers for valence and arousal state classifications. The purpose of PCA is to reduce dimension of input features. The CSA handles the nonstationary effect of EEG signals. The classification accuracy of the DLN with PCA+CSA is 53.42% and 52.05% to classify three levels of valence states and three levels of arousal states. The DLN provides better accuracy performance compared to SVM and naive Bayes classifier. One of the major limitations for performing EEG-based emotion recognition algorithm is dealing with the problem of intersubject variations in their EEG signals. The common features of transferable nonstationary information can be investigated to alleviate the intersubject variation problems.

REFERENCES

1. F. Akram, M. K. Metwally, H. Han, H. Jeon, and T. Kim, "A novel P300-based BCI system for words typing," in Proceedings of the International Winter Workshop on Brain-Computer Interface (BCI '13), pp. 24–25, February 2013. View at Publisher · View at Google Scholar · View at Scopus

2. R. S. Naveen and A. Julian, "Brain computing interface for wheel chair control," in Proceedings of the 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT '13), pp. 1–5, Tiruchengode, India, July 2013. View at Publisher · View at Google Scholar

3. F. Sharbrough, G. E. Chatrian, R. P. Lesser, H. Luders, M. Nuwer, and T. W. Picton, "American Electroencephalographic Society guidelines for standard electrode position nomenclature," Journal of Clinical Neurophysiology, vol. 8, no. 2, pp. 200–202, 1991. View at Google Scholar

4. Wikipedia, "Electroencephalography," March 2014, .http://en.wikipedia.org/wiki/Electroencephalography.

5. S. Koelstra, A. Yazdani, M. Soleymani et al., "Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos," in Proceedings of the .International Conference on Brain Informatics, Toronto, Canada, 2010.

6. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42–55, 2012. View at Publisher · View at Google Scholar · View at Scopus

7. D. Huang, C. Guan, K. K. Ang, H. Zhang, and Y. Pan, "Asymmetric spatial pattern for EEG-based emotion detection," in Proceeding of the International Joint Conference on Neural Networks (IJCNN '12), pp. 1–7, Brisbane, Australia, June 2012. View at Publisher · View at Google Scholar · View at Scopus

8. G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," International Journal of Human Computer Studies, vol. 67, no. 8, pp. 607–627, 2009. View at Publisher · View at Google Scholar · View at Scopus

9. D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in Proceedings of the 5th International IEEE/EMBS Conference on Neural Engineering (NER '11), pp. 667–670, Cancun, Mexico, May 2011. View at Publisher · View at Google Scholar · View at Scopus

10. X.-W. Wang, D. Nie, and B.-L. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in Neural Information Processing, B.-L. Lu, L. Zhang, and J. Kwok, Eds., vol. 7062, pp. 734–743, Springer, Berlin, Germany, 2011. View at Google Scholar

11. N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Real-time EEG-based happiness detection system," The Scientific World Journal, vol. 2013, Article ID 618649, 12 pages, 2013. View at Publisher · View at Google Scholar

12.G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: arousal evaluation using EEG's and peripheral physiological signals," in Multimedia Content Representation, Classification and Security, B. Gunsel, A. Jain, A. M. Tekalp, and B. Sankur, Eds., vol. 4105, pp. 530–537, Springer, Berlin, Germany, 2006. View at Google Scholar

13.O. AlZoubi, R. A. Calvo, and R. H. Stevens, "Classification of EEG for affect recognition: an adaptive approach," in AI 2009: Advances in Artificial Intelligence, A. Nicholson and X. Li, Eds., vol. 5866 of Lecture Notes in Computer Science, pp. 52–61, Springer, Berlin, Germany, 2009. View at Google Scholar

14 .G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," IEEE Transactions on Systems, Man, and Cybernetics A Systems and Humans, vol. 41, no. 6, pp. 1052–1063, 2011. View at Publisher · View at Google Scholar · View at Scopus

15 .U. Wijeratne and U. Perera, "Intelligent emotion recognition system using electroencephalography and active shape models," in Proceedings of the 2nd IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES '12), pp. 636–641, December 2012. View at Publisher · View at Google Scholar · View at Scopus

16 .S. Y. Chung and H. J. Yoon, "Affective classification using Bayesian classifier and supervised learning," in Proceedings of the 12th International Conference on Control, Automation and Systems (ICCAS '12), pp. 1768–1771, October 2012. View at Scopus

17 .G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol. 18, no. 7, pp. 1527–1554, 2006. View at Publisher · View at Google Scholar · View at Zentralblatt MATH · View at MathSciNet · View at Scopus

18 .D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," Journal of Neural Engineering, vol. 8, no. 3, Article ID 036015, 2011. View at Publisher · View at Google Scholar · View at Scopus

19 .M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," Advances in Artificial Neural Systems, vol. 2012, Article ID 107046, 9 pages, 2012. View at Publisher · View at Google Scholar

20 .S. Koelstra, C. Mühl, M. Soleymani et al., "DEAP: a database for emotion analysis; using physiological signals," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18–31, 2012. View at Publisher · View at Google Scholar · View at Scopus

21 .J. D. Morris, "SAM: the self-assessment manikin. An efficient cross-cultural measurement of emotion response," Jounal of Advertising Research, vol. 35, no. 8, pp. 63–68, 1995. View at Google Scholar

22 .M. Spüler, W. Rosenstiel, and M. Bogdan, "Principal component based covariate shift adaption to reduce non-stationarity in a MEG-based brain-computer interface," EURASIP Journal on Advances in Signal Processing, vol. 2012, article 129, 2012. View at Publisher · View at Google Scholar · View at Scopus

23 .I. T. Jolliffe, Principal Component Analysis, Springer, New York, NY. USA, 1986. View at Publisher · View at Google Scholar · View at MathSciNet

24 .P. Baldi and K. Hornik, "Neural networks and principal component analysis: learning from examples without local minima," Neural Networks, vol. 2, no. 1, pp. 53–56, 1989. View at Publisher · View at Google Scholar · View at Scopus

25 .C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, article 27, no. 3, 2011. View at Publisher · View at Google Scholar · View at Scopus

26 .K. Li, X. Li, Y. Zhang et al., "Affective state recognition from EEG with deep belief networks," in Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, 2013.