EEG-based emotion recognition using LSTM-RNN machine learning algorithm

Jeevan Reddy Koya[1]
Associate Professor,ECE
SNIST.Senior Member IEEE.
Hyderabad,India
jeevanreddyk@sreenidhi.edu.in

Dr.Venu Madhava Rao S P[2]
Professor,ECE Dept
MVSR Engg College
Hyderabad,India
spvennu@yahoo.com

Shiva Kumar Pothunoori[3]
Student ,Dept of ECE
SNIST
Hyderabad,India
shiva0071998@gmail.com

Srivikas Malyala[4]
Student, Dept of ECE
SNIST
Hyderabad,India
srivikasnannu@gmail.com

***Abstract*— In recent days the knowledge in the Brain Machine Interface is manifesting emotion recognition and classification. There are many studies indicating potential evidence in identifying emotions using EEG brain waves. This paper investigates and proposes a new machine learning technology in identifying the emotions through the use of latest machine learning concepts using LSTN (Long short term memory) recurring neural networks. The acquired brain wave signals are processed for classification using discrete wavelet transform and then given to the proposed algorithm for specific emotion recognition.**

***Keywords*— EEG,LSTM,RNN,DWT.**

I. INTRODUCTION

A Brain Machine Interface (BCI) is a system that considers a biosignal, acquired from a subject, and predicts certain parameters of the person's cognitive state [1,2]. Brain computer interface is in its initial stage of research .Initially EEG stimuli recordings were proposed for the emotion recognition. Emotion recognition may be classified using discrete wavelet transform[3].Signal analysis is very important in emotion classification for good and efficient recognition. A multimodal classification supports classification of certain brain activities which area of interest and which are excited to stimuli[4]. Most of the emotion related signals are found in gamma frequencies .Parameters obtained in gamma show important information related to emotions.

Convolution neural networks are forward neural networks, generally including primary pattern extraction layer and feature mapping layer, and can learn local patterns in data by a method called convolution. A distinctive nature of CNN is that it is optimized for end-to-end learning without any a priori feature selection (Schirrmeister et al. 2017),which in general rejects any information loss, and is known for its good Signal to noise ratio, Task related and non related EEG raw data. Hence, lots of EEG-based researches and applications have emerged these years such as Emotiv EEG acquiring device and P300 feature detection (Cecotti and Graser 2011; Puanhvuan et al. 2017). The chemical composition of brain generates different electric activity at different potentials of brain workloads[5]. Momentary total workload calculation and clasification (Zhang et al. 2017,2017). Various mental workload states can be easily found with the EEG signal LSTM algorithms with a sufficient potential to depend for drivers mental workload.[6].

II. DEEP LEARNING AND LSTM

Distributed hidden state that allows them to store a lot of information about the past efficiently. Non-linear dynamics that allows them to update their hidden state in complicated ways. The early research in neural networks was not so successful because of many limitations of single layer linear networks. Multilayer networks were not found until much later but even then there were not suitable training algorithms.

Back propagation : This neural network used uses a i*j layer hidden nodes to estimate the pattern. The no of iterations required are quite high.

$$E = \frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(t_j^{(k)} - \underbrace{f_2\left(\sum_{s=1}^{h} W_{js}\underbrace{f_1\left(\sum_{i=1}^{n} w_{si}x_i^{(k)}\right)}_{z_s=\text{output of hidden node } s}\right)}_{y=\text{output of output node } j}\right)^2$$

LSTM networks introduce a new structure called a memory cell. Each memory cell contains four main elements: Input gate, Forget gate, Output gate, Neuron with a self-recurrent. These gates allow the cells to keep and access information over long periods of time.

The LSTM uses this idea of "Constant Error Flow" for RNNs to create a "Constant Error Carousel" (CEC) which ensures that gradients don't decay. The key component is a memory cell that acts like an accumulator (contains the identity relationship) over time Instead of computing new state as a matrix product with the old state, it rather computes the difference between them. Expressivity is the same, but gradients are better behaved.

Computing the hidden workA schematic of an LSTM unrolled in time to show how LSTM can preserve the information. The input, forget, and output gate activations are respectively displayed below, to the left and above the memory block. For simplicity, the gates are either entirely open (‘O’) or entirely closed (‘—’).

Traditional RNNs: a special case of LSTMs, If we set the input gate all ones (passing all of the new information), the forget gate all zeros (forgetting all of the previous memory) and the output gate to all ones (exposing the whole memory

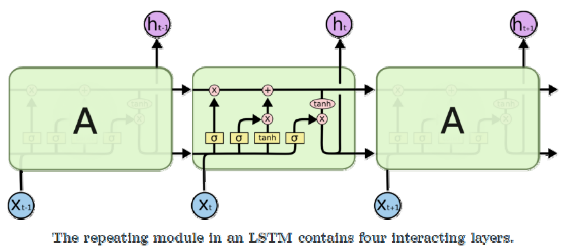

Fig 1 LSTM Network structure

Deep Learning (DL) is the recent method of excavating features present deep in complex data. In this paper, we introduce deep convolution neural networks to propose and design the EEG-based emotion recognizer primarily in to three modes. Differential Entropy (DE) is obtained as features at certain time interval for each channel. From a basic one-dimensional deep model, and from a SVM and KNN model . We found that HCNN (86.2% ± 2.5%) is better than SAE (83.2% ± 6.9%), and deep neural models are more preferable in emotion classification and estimation Machine computer interface system that models brain signals for emotions.

III. EXPERIMENT AND DATA AQUITION

The dataset is may be classified in to 2 groups . First, the subject is presented with a set of of stimuli for 1200ms:, two images of baby pictures with empathy activeness, or two images of scenes. Similarly a music will be given concurrently related to the scene seen by the subject.

This stimulus time period is generally considered as the initial presentation. After a certain delay, the subject will then presented with a stimuli for 1200ms representing a keeping a new condition, a not act case or an act case. In the refresh case, the subject is presented with an flashing light presented top bottom for identifying the one of the initial cues .

The subject will then be asked to think about the cue that was the topic of the interest for us in that particular location.Both the initial location present at that time of the stimuli and the refresh condition are used for classification in this research. The acquired data was from a 32 channel EEG at a 350 Hz sample rate Signals.

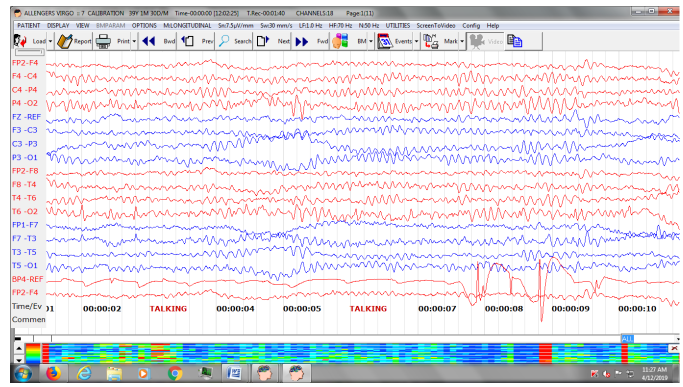

Fig 2 : EEG recording sample at the experimental setup using Virgo 24 channel speaking effort.

The data acquisition for a 2 level process before it is fed to the LSTM. A wavelet based filtering happens before it is classified as wavelet divides the data in to bands each band containing a particular spectral width. The spectrally separated data is then given to the classifier. The active classifier itself is a Long short term memory which posses a classified data spatially spectrally and then temporally.

Wavelet signal separation is an estimation technique here the EEG signal is represented as a sum of infinite series summation with weights. The purpose is to represent the signal as a linear weighted combination of spectral components. . The importance of wavelet transform is localization of time-frequency. Here the sub band energy of the wavelet is defined over a finite interval of time. Wavelet transform has its advantage of frequency localization at low frequencies and time localization at high frequencies

In this experimental setup we found to have concurrent study of fMRI and EEG was selected to see further clarification of activation to brain signal correlation.

To be able to clarify the activations few fMRI readings were obtained to get suspicious readings filtered. The correlation between the EEG and functional MRI were computed to place the nodes on the scalp and also to ensure the experimental justification. The activations from EEG and fMRI were similar and that shows a strong evidence for calibrating emotions from EEG. The concurrent study of fMRI and EEG made it possible to acquire the data with this relevance and actual data ,which is actually providing the best possible correlation for activation.

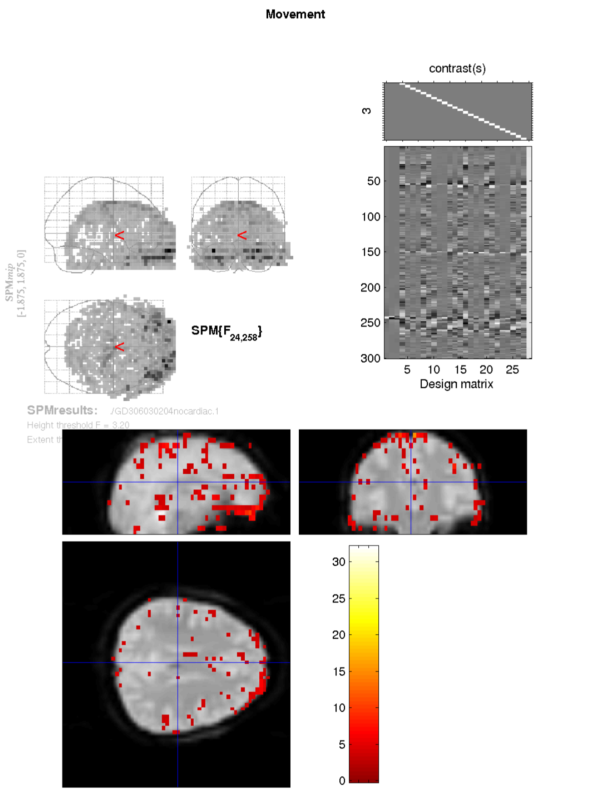

Fig3 : fMRI-EEG concurrently acquired brain activations.

In this section we first present the models determined by the model search and then the final classification results of these models in each task. The general preprocessing method was to scale the data by the 90 percentile value among all values in the training set. This regulates the important data to the 1 to 1 range at which deep learning methods function best, while reventing outliers from dominating the results.

using the fMRI time series modal first the artifacts from the acquired EEG signals will be removed.
The basic equation is

$$x_p(t) = m_p + b_p t + s_p(t) + v_p(t)$$

The above equation the parameters are baseline activation and the volume stimulus The mean will be removed to get clean EEG as the following data. The clean EEG data obtained from filtering the acquired data using hamming and Blackman window. The filtered data spatially separable which is then given for training to the LSTM based neural system.

This data which filtered is then given for artifact removal using Riemannian geometry based classifiers and other active classifiers. The data pre processing is done with the standard filtering approach and made adaptive for the LSTM layers.
The artifact removal plays a vital role in the performance of total emotion recognition as the quality of the EEG signal depends on the quality of artifact removal.

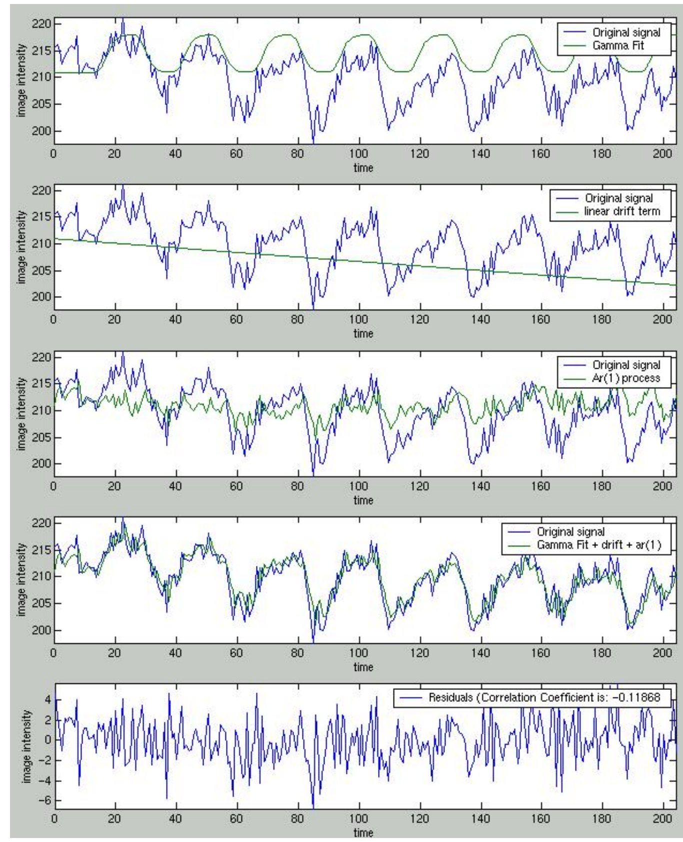

Fig 4 : Clean data after artifact removal through fMRI knowledge

The data is then given for a frequency translation and a mapping is done between the temporal and spatial parameters.

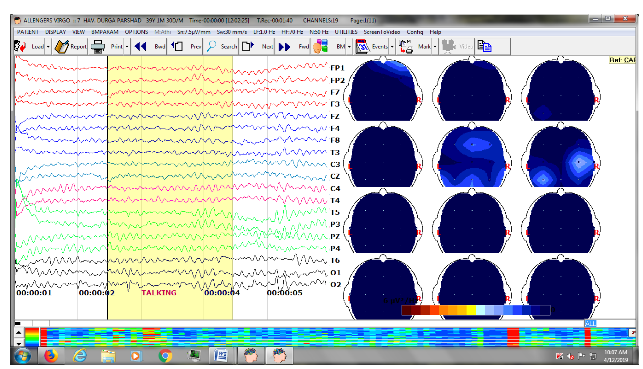

Fig 5 : Spatial temporal activations mapping for speaking effort.

While the LSTM(RNN) + CNN combination architecture generally has the most layers of any of the final model architectures, in this experimental setup less parameters than the LSTM architecture. The undertaken model has performed better with only a single LSTM layer, rather than using multiple layers as in the pure LSTM model. The structure of the convolution and max pooling layers serves to 57reduce the dimensionality of the network prior to the fully connected layer, thus leading to a drastic reduction in overall number of parameters.The convolutional architecture is partially based on Szegedy, Ioffe, & Vanhoucke, (2016)

where they have considered Two dimensional convolutions and were divided into split and pairs of one dimensional convolutions. This leads to a less number of parameters in the convolutional layers while simultaneously increasing the depth of the model. A 1x1 convolutional layer was used to separate the two convolutional passes . This will certainly provide benefits similar to a fully connected layer with a far lower increase in parameters, and led to an increase in performance over omitting the basic signal.

IV. RESULTS AND CONCLUSION

The deep learning models selected during finalization of the particular scheme is shown in this results. Since no model performed better in giving a reliable results on single subject data was converge reliably on single subject . In this present result the mean and variance over 10 subjects under similar conditions were manifested. And the single subject details are omitted.

All the results are presented checked and calculated through 15 fold cross validation check with a 70/18/18 training/validation/test split. Time binning significantly decreased the performance of all deep learning models. In the most extreme case, the LSTM + CNN model, time binning lowered the accuracy from 64.36% to 46.89%.Only the LSTM model was able to perform above chance and on-par with the traditional methods on the refresh dataset. Result obtained was 59%.

The MLP model performed poorly in all tasks –significantly lower than the traditional methods, MLP models are not invariant to temporal or spatial shifts in the data, and thus poor performance was expected when training across subjects.
The CNN model performed on par with the most successful traditional methods, with an initial presentation accuracy of 62.09%. Of the deep learning methods, it was by far the quickest to train at around 45 minutes to converge due to the small parameter amount.

The LSTM model performed slightly better than the traditional methods on the initial presentation data, with an accuracy of 63.61%in proper emotion recognition. It was the slowest model to train, taking 4 to 5 hours. Hence this proposed model having a higher potential of recognizing the emotional tasks. This may when data acquiring conditions improve and the computational devices speed improve. The traditionally performing devices generally perform quite slower than what an LSTM based device perform.

The LSTM + CNN model performed in a better way over traditional or deep learning model, with an accuracy of 64%.

V Acknowledgment

We are thankful for the Dept of ECE,SNIST and ECE dept MVSR for supporting the research. We are great full to UGC, India for funding the experimental setup.

REFERENCES

[1] Pun, T.; Alecu, T.L.; Chanel, G.; Kronegg, J.; Voloshynovskiy, S. Brain-computer interaction research at the computer vision and multimedia laboratory, university of geneva. IEEE Trans. Neural Syst. Rehabil. Eng. 2006, 14, 210–213. [CrossRef] [PubMed] Esfahani, E.T.; Sundararajan, V. Using brain-computer interfaces to detect human satisfaction in human-robot interaction. Int. J. Humanoid Robot. 2011, 8, 87–101. [CrossRef]I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, pp. 271–350. films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, pp. 271–350.

[2] Esfahani, E.T.; Sundararajan, V. Using brain-computer interfaces to detect human satisfaction in human-robot interaction. Int. J. Humanoid Robot. 2011, 8, 87–101. [CrossRef]I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, , pp. 271–350.

[3] Wavelet Transform for Classification of EEG Signal using SVM and ANN NITENDRA KUMAR, KHURSHEED ALAM AND ABUL Vol. 10(4), 2061-2069 (2017).

[4] Li, Jinpeng & Zhang, Zhaoxiang & He, Huiguang. (2016). Implementation of EEG Emotion Recognition System Based on Hierarchical Convolutional Neural Networks. 10023. 22-33. 10.1007/978-3-319-49685-6_3.

[5] Appl. Sci. 2017,7,1060; doi:10.3390/app7101060 Abdel-Hamid o,Mohammed Ar,Jiang H,Deng L,Penn G,Yu D(3014) Convolutional neural networks for speechrecognition .IEEE/ACM Trans Audio Speech Lang Process 22(10):1533-1545

[6] Ahn S, Nguyen T, Jang H, Kim JG, Jun SC (2016) Exploring neuro-physiological correlates of drivers’ mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and FNIRS data. Front Hum Neurosci 10:219. https://doi.org/10.3389/fnhum.2016.00219

[7] Bornas X, Fiolveny A, Balle M, Morillasromero A, Tortellafeliu M (2015) Long range temporal correlations in eeg oscillations of subclinically depressed individuals: their association with brooding and suppression. Cognit Neurodyn 9(1):53–62

[8] Brookhuis KA, De WD (1993) The use of psychophysiology to assess driver status. Ergonomics 36(9):1099

[9] Cecotti H, Graser A (2011) Convolutional neural networks for p300 detection with application to brain-computer interfaces. IEEE Trans Pattern Anal Mach Intell 33(3):433–445

[10] Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):1–27
Chen LL, Zhao Y, Ye PF, Zhang J, Zou JZ (2017) Detecting driving

[11] stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Syst Appl 85(C):279–291
Correa AG, Orosco L, Laciar E (2014) Automatic detection of drowsiness in eeg records based on multimodal analysis. Med Eng Phys 36(2):244