EEG Based Analyzing Human Idiosyncrasies Within Unpredictable Circumstances Using 3D Convolutional Neural Networks

Tanzim Ahmed
Department of Computer Science And Engineering
BRAC University
Dhaka, Bangladesh
tanzim.ahmed@g.bracu.ac.bd

Sabbir Al Fahad
Department of Computer Science And Engineering
BRAC University
Dhaka, Bangladesh
sabbir.al.fahad@g.bracu.ac.bd

Joy Paul
Department of Computer Science And Engineering
BRAC University
Dhaka, Bangladesh
joy.paul@g.bracu.ac.bd

Rufaida Tasin
Department of Computer Science And Engineering
BRAC University
Dhaka, Bangladesh
rufaida.tasin@g.bracu.ac.bd

Abu Sayed Bin Abdullah
Department of Computer Science And Engineering
BRAC University
Dhaka, Bangladesh
abu.sayed.bin.abdullah@g.bracu.ac.bd

Muhammad Iqbal Hossain,PhD
Department of Computer Science And Engineering
BRAC University
Dhaka, Bangladesh
iqbal.hossain@bracu.ac.bd

***Abstract*—Automatic recognition of emotions is one of the most challenging of tasks. A sophisticated learning algorithm, which may represent high-level abstraction, is required to detect emotions from EEG signals. The detection of emotions from EEG signals requires a sophisticated learning algorithm that can reect high-level abstraction. Different techniques have been applied to reinforce the robustness of the emotion recognition systems using electroencephalogram (EEG) signals, in particular the learning of spatiotemporal traits. In this approach, the use of a model inspired from Siamese Network which deals with data through concatenation and 3D convolution. In this approach for training Adam optimizer was used. This model is a custom model, which can deal with a stack for a single time instant, which gives better efficiency than 3D Convolutional Network. This type of model is usually used for Face Recognition, but use of such a model for this purpose opens a new horizon of opportunity for Automatic recognition of emotion. Highest accuracy of our project is around 70% has been achieved by Dense Net and lowest accuracy was 30%.**

***Keywords*—3D Convolutional Neural Network, Seed IV, Siamese Network, Adam Optimizer,DenseNet, Deep Learning**

I. Introduction

Human emotions are isolated as the primary source of feelings from the external appearances. In any case, it realized that a few people might conceal their genuine emotions and use outward appearances to deceive. Analysts then keep fast to use specific data wellsprings, which are accurate and not defenseless to misrepresentation [1]. The brain-computer interface (BCI) has been one of the most interesting fields of research in biomedical engineering for decades [2]. It provides a promising technology that allows people to control outside devices by modulating their brain waves. Most BCI technologies have been developed for non-invasive brain signal processing that is feasible in real-world scenarios to implement. BCI can not only be used for mental control devices, but can also be implemented to understand our mental states. One such application is emotion recognition. Automatic emotion recognition algorithms could potentially bridge the gap between interactions between humans and machines. It varies from negative to positive. Electroencephalogram (EEG) is a record of brain electric potential oscillation likely to result from the ionic current between brain neurons. EEG signals are obtained by measuring the brain activity at the position of an electrode on the scalp [**?**] [**?**]. In addition, a Gaussian noise signal n with zero mean and unit variance generates randomly with N samples to generate the noisy EEG signals, so that N is the number of samples from the original EEG signals [?] [?]. Finally, it will get the noisy version of s by adding all samples of s and n signals together. The 3D-CNN is capable of learning spatial and temporal features [?] [?]. This requires that the EEG signals build up 3D input representations. To this end, the proposed work sets out a 3D representation procedure. The EEG data from each signal is usually recorded from various channels. Using a window size w, the data from each channel care segmented into small segments (frames) and each chunk is added with a label for better worked. There is an ever-growing need for computer applications that can detect the user's current emotional state [?]. Research has already been done in a report to copy human communication in recognizing emotion from the face and voice. With the increasing interest in brain-computer interaction of (BCI), the EEGs of the user (electroencephalograms) were also analyzed. If the EEG just displays a physiological reaction is still uncertain, or even provides insight into the emotion as to how it is psychologically perceived. Right EEG-recognition of arterially evoked Emotion is currently only about 60 percent, but much research is hopefully showing the EEG's suitability for our future implementation. EEG is one of the main diagnostic tests for epilepsy. A routine clinical EEG recording typically lasts 20-30 minutes (plus preparation time). It is a test that detects electrical activity in the brain using small, metal discs (electrodes) attached to the scalp. Routinely, EEG is used in clinical circumstances to determine changes in brain activity that might be useful in diagnosing brain disorders, especially epilepsy or another seizure disorder. An EEG might also be helpful for diagnosing or treating the following disorders.

II. Related Works

Emotional states are linked to a wide spectrum of human perceptions, thoughts and behaviors that influence our ability to behave rationally in situations like decision-making, perception and human intellect. Study into emotional awareness by using emotional signals as an important subject for therapeutic use and human social relations further strengthens the brain-computer interface (BCI) systems[8][6]. In recent years, EEG recognition systems have been developed into a popular subject of research among cognitive scientists. 3D input representation is produced in EEG segments [2] [?].This framework hopes that the 3D-CNN model framework proposed could be more specifically configured and simplifies job execution than was the case previously. The DEAP (Emotion Analysis Dataset) using EEG, Physiologic, and Video Signals is used to evaluate the process and it includes a multi-modal dataset named "SEED-IV" for human analysis of a given state. When there are many unlabeled data sets, the framework can work with an acceptable DLN algorithm and a successful one.

III. PROPOSED MODEL

Normally, the mechanism of automated emotional recognition may be done using one or more of the following modes: face, voice, body movements, and EEG signals. With the EEG signs, researchers aim at solving the issue of time-to - time emotion correlation. By default, feelings live for a brief or long time, not just a moment. Thus, the relationship between the time segments of emotions is extremely successful in improving the precision of identification. In order to accomplish this goal, the data increase step is first used to increase the amount of EEG samples available. The 3D representation of inputs is then generated from the EEG segments. Ultimately, the proposed 3D-CNN model framework is being developed. The method for the suggested system shown in Figure 1.

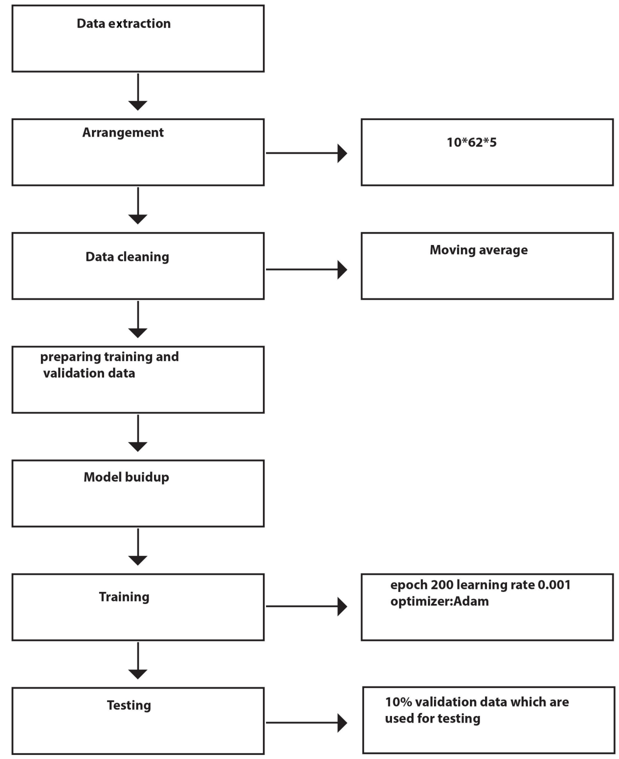

Fig. 1. Flowchart of the proposed system.

A. 3D Input Representation

This allows 3D input representation from EEG signals to be constructed. In addition, the proposed study introduces a 3D representation technique. EEG data is record on separate channels from each signal. The data of each channel is separate into small frames by a window size w.

Each channel has D frames. The i-th framework samples from all channels are joined together in a 2D matrix where the height of the i-th framework is the number of channels and the width of the I-th framework is the numbers of samples. Then, a set of consecutive frames, which also are called chunk dimensions, can be applied to the third temporal domain. If the chunk size is 10, a 3D matrix called B can add 10 consecutive frames in one chunk.

The majority rule is use to apply a label to each chunk to gain the corresponding ground truth label. This chunk is giving the majority label. Finally, the bits of frames and the required

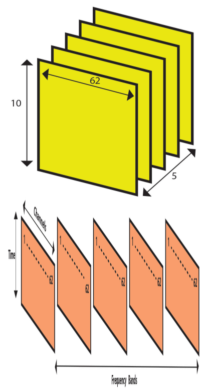

Fig. 2. 3D input representation.

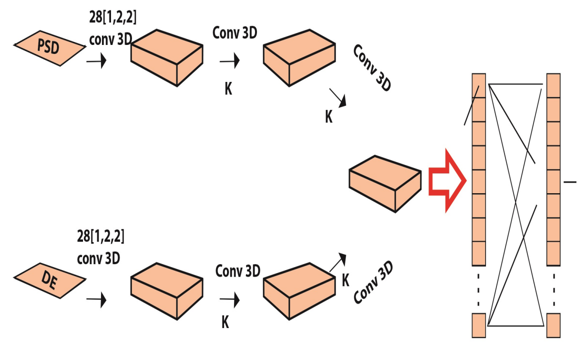

Fig. 3. Network Architecture of the Proposed 3D-CNN Model.

label is provide with a new 3D matrix C. Each 3D matrix C is an input to the 3D-CNN training model. Figure 2 shows the 3D volume input shape contains a chunk of 5 frequency bands and 62 channels is seen in the Figure 2.

B. The 3D-Convolutional Neural Network Model

The next move is to use the 3D-CNN proposed to classify emotions. It presents 3D-CNN and explains the proposed network architecture in the next chapters. 3D-CNN is the extension of standard CNN with updated convolution and pooling operations [4] [11]. Long lasting sequences such as voice, videos and EEG signals depend on their segments and negligence can affect the robustness of recognition systems. By using 3D transforming operations over the input segments, 3D-CNN models such temporal dependency. In addition, a 3D converting operational will imagine and modeless the spatial association between pixels in video frames or various EEG channel positions [11]. The 3D-CNN has been used to identify behavior. The convolution operation is inspire by visual neuroscience cell principles.

$$O(x, y, z) = \sum_{m\ np} f(m, n.p) * C(x - m, y - n, z - p)$$

Since O is the output of the convolution operation, f was its filter with the size m*n*p and C is the 3D input EEG chunk. Since O is the output of the convolution operation, f was its filter with the size m*n*p and C is the 3D input EEG chunk [11] [12]. In this proposed model, we used a small size of kernel, as the signal features are more important for this model, not the shape. The kernel size of our proposed model is (1, 3, 1).

1) Network Architecture: Choosing the right network configuration for a challenge provides a greater chance of more reliable outcomes. The 3D-CNN supports a wide range of connected layers. Due to a large number of different layer shapes, it is not optimal to find an optimal chain that fits the problem closely.

The adopted architecture consists of six layers. The first layer is the input volume. The middle layers are two 3D convolution layers. The last layer is one fully connected layer to extract the final features. A detailed illustration of the proposed network architecture is shown in Figure 3 . For the first layer, the input volume size is 10*62*5; 10 is the time stamp, 62 is the number of channels, and 5 is the number of frequency bands. The kernel shape of the first convolution layer is 1*3*1: where 1, 3, and 1 are the width, height, and depth respectively. The rectified linear unit (RELU) is used as activation function in both convolution layers since it is linear, drivable, and has a simple implementation, which can be expressed as: $RELU(x_i) =$

$$\begin{cases} x_i & x_i \geq 0 \\ 0 & x_i < 0 \end{cases}$$

x_i is the input to the new convolution list. The number of feature maps are set to 9. The same configurations of the first convolution layer is used for the second convolution layer, except for the number of feature maps set to 18. We have used RELU function in every layer. However, we have used Softmax activation function in the last layer. The continuous activation function that is widely used for multi-label problems that does not presume reliance on class labels.

IV. EXPERIMENTS AND RESULTS

A. Data Arrangement

Total number of 15 subjects participated the experiment. Each participant, 3 sessions are performed on different days, and each session contains 24 trials [16]. The raw EEG data are first down-sampled to a 200 Hz sampling rate. In order to filter the noise and remove the artifacts, the EEG data are processed with a band pass filter between 1 Hz to 75 Hz. After the extract power spectral density (PSD) and differential entropy (DE) features within each segment at 5 frequency bands: 1) delta: 1 4 Hz; 2) theta: 4-8 Hz; 3) alpha: 8-14 Hz; 4) beta: 14-31 Hz; and 5) gamma: 31-50 Hz [13]. That dataset “seed iv” had the noise clean using Moving average. A field is in the form of channel number*timestamp*frequency bands, i.e. 62*W*5, where W denotes the number of time windows in that trial (different trials have different W, since the video clips are not of the same length) [16].

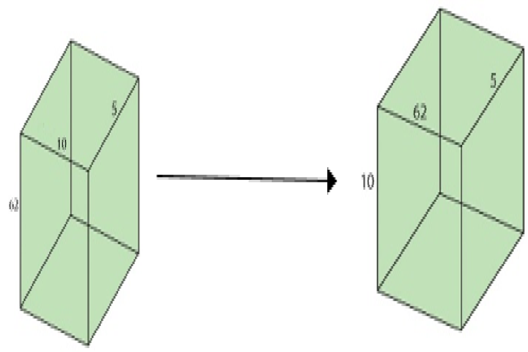

Fig. 4. Timestamp Transposition

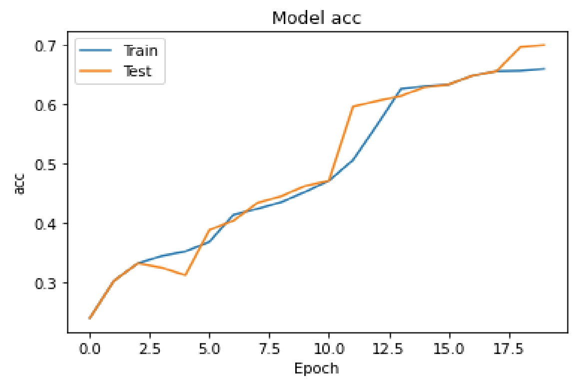

Fig. 5. Graph of average accuracy overall

We have transposed the field form of channel number*timestamp*frequency bands to timestamp*channel number*frequency for our convenience.

B. Training and validation

With Adam optimizer, the learning rate is set to 0.001. The batch size for training and testing shall be set at 200 samples. Final recognition accuracy is the average of all folds. The key priorities of the training process are integration and zero failure. If the loss approaches zero before passing the complete number of epochs, an early stop criterion is used to conserve time as further epochs are processed when the system is still converging. The proposed early stop criteria is satisfied by counting the number of times the failure is zero, the optimization is stopped after 20 epochs, as our proposed network model was overfitting.

C. Pre-Processing the EEG Signals

Various pre-processing operations have been used to improve the consistency of the EEG signal and as such the precision of the emotional detection task. The pre-processing requires a high-pass filter to suppress any signal that is less than 1 Hz or any ft. and a cut-off frequency filter of 50 to 60 Hz is often used to eliminate any unnecessary noise. Normalization is done between -1 and 1. for each channel. The EEG signal is 63 seconds for each video. In the first three seconds, the EEG signal is eliminated for preparation and monitoring only for 60 seconds. For a short amount of time, each EEG signal is stationary. In conclusion, every session is sliced into 4-second non-overlapping segments and this is regarded as one data sample during model training.

D. Results and Discussions

A variety of tests is carried out using SEED IV data to demonstrate the reliability of the proposed method. The best setup to date from their previous experiment is used for each experiment. This shows a comparison with and without an increase between the accuracy values. This experiment is performed by using a 1x62x5 frame chunk size and preparations as defined in section .The first x-axis element in the figure. , relates to the accuracy of our model. As is obvious, by increasing the amount of times that we conducted in our models the average accuracy increases. Since the method generalizes smaller data, findings display improved results by increasing the number of epochs.

We got around 70% accuracy from our model. Firstly we have divided our data into 2 segments .These segments are naming with PSD DE and, then take those data as an input into several Con3D layer from where we get many images as an result. After this, we use concatenate method to get a resultant Tensor as a single output. We have stopped the training as after 20 epoch testing accuracy started falling and training accuracy was increasing. So, we have to stop training because the network was overfitting. Finally, our proposed Model has gone through with DenseNet to get the expected outcomes with an efficient accuracy. Above the graph, the blue line indicates the training portion of our model and the red portion indicates with the testing part of our proposed model.

V. FUTURE WORK AND CONCLUSION

A. Future Improvement

If we could collect more data, the model is used might give better accuracy. This will give better outcomes if we can implement inter-layer-loss sharing. Adding more data on the dataset to get more accuracy in our work. As we know that machine, learning gives us more accuracy if we have more data on the dataset. In addition, we want to detect more emotions in future. Using Siamese Network, the model can be less complex than the present one. In short, the proposed model can be output better accuracy if we can train it with more amount of data and can make it less complex by upgrading it.

B. Conclusion

In this paper, the 3D-CNN emotion recognition approach is proposed to classify the emotions between the EEG signals. Since 3D-CNN needs 3D inputs, a new method has been created that represents EEG signals in a 3D format from multi-channel signals. The experimental work has shown that the proposed model can achieve very high accuracy of recognition compared to a limited number of data.

Acknowledgment

We sincerely thank all the people who helped us in developing our proposed model and we express special gratitude to our supervisor Assistant Professor Dr. Muhammad Iqbal Hossain for guiding, motivating and suggesting us at the progressing stages of the system

References

[1] N.N. Khatri, Z.H. Shah, and S.A. Patel, Facial expression recognition: A survey," International Journal of Computer Science and Information Technologies, vol. 5, no. 1, pp. 149,152, 2014.

[2] Q. Yao, Multi-sensory Emotion recognition with speech and facial expression," Ph.D. dissertation, Computing. University of Southern Mississippi, 2014.

[3] W. Heller and J. B. Nitscke, "Regional brain activity in emotion: A framework for understanding cogni-tion in depresion," Cognition Emotion, vol. 11, no. 5-6, pp. 637661, 1997.

[4] K. Nashiro, M. Sakaki, and M. Mather, Age differences in brain activity during emotion processing: Reections of age-related decline or increased emotion regulation," Gerontology, vol. 58, no. 2, pp. 156, 163, 2012.

[5] Anagnostopoulos, C.-N.; Iliou, T.; Giannoukos, I. Features and classiffiers for emotion recognition from-speech: A survey from 2000 to 2011.Artif. Intell. Rev.2015,43, 155177. [CrossRef]

[6] Tzirakis, P.; Trigeorgis, G.; Nicolaou, M.A.; Schuller, B.W.; Zafeiriou, S. End-to-End Multimodal Emo- tionRecognition Using Deep Neural Networks.IEEE J. Sel. Top. Signal Process.2017,11, 13011309. [CrossRef]

[7] H. Greenspan, B. Van Ginneken, and R. M. Summers, editorial deep learning in medical imaging: Overview and future promise of an exciting new technique", IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 11531159, 2016.

[8] H. J. Yoon and S. Y. Chung. EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm," Computers in Biology and Medicine, vol 43, no. 12, pp. 2230,2237. 2013.

[9] J. Atkinson and D. Campos. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," Expert Systems with Applications, vol. 47, pp. 35-6, 2015.

[10] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, Emotion recognition based on the EEG using LSTM recurrent neural network," International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, pp. 355-358, 2017.

[11] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, Spatial-temporal recurrent neural network for emotion recognition," Journal of Latex Class Files, vol. 13, no. 9, pp. 1,8, 2017

[12] Zheng, W.-L.; Lu, B.-L. Investigating Critical Frequency Bands and Channels for EEG-based Emo- tionRecognition with Deep Neural Networks.IEEE Trans. Auton. Ment. Dev. (IEEE TAMD)2015,7, 162,175

[13] Koelstra, S.; Muehl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I.DEAP: A Database for Emotion Analysis using Physiological Signals.IEEE Trans. Affect. Comput.2012,3,18,31.

[14] S. Koelstra, C. Muhl, and M. Soleymani, DEAP: A database for emotion analysis using physiological signals," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18,13, 2012.

[15] SEED IV Dataset "http://bcmi.sjtu.edu.cn/ seed/seed-iv.html"