2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

3D CNN to Estimate Reaction Time from Multi-Channel EEG

Mohammad Samin Nur Chowdhury[1*], Arindam Dutta[2], Matthew K. Robison[3], Chris Blais[4], Gene Brewer[4], and Daniel W. Bliss[2]

***Abstract*— The study of human reaction time (RT) is invaluable not only to understand the sensory-motor functions but also to translate brain signals into machine comprehensible commands that can facilitate augmentative and alternative communication using brain-computer interfaces (BCI). Recent developments in sensor technologies, hardware computational capabilities, and neural network models have significantly helped advance biomedical signal processing research. This study is an attempt to utilize state-of-the-art resources to explore the relationship between human behavioral responses during perceptual decision-making and corresponding brain signals in the form of electroencephalograms (EEG). In this paper, a generalized 3D convolutional neural network (CNN) architecture is introduced to estimate RT for a simple visual task using single-trial multi-channel EEG. Earlier comparable studies have also employed a number of machine learning and deep learning-based models, but none of them considered inter-channel relationships while estimating RT. On the contrary, the use of 3D convolutional layers enabled us to consider the spatial relationship among adjacent channels while simultaneously utilizing spectral information from individual channels. Our model can predict RT with a root mean square error of 91.5 ms and a correlation coefficient of 0.83. These results surpass all the previous results attained from different studies.**

***Clinical relevance*— Novel approaches to decode brain signals can facilitate research on brain-computer interfaces (BCIs), psychology, and neuroscience, enabling people to utilize assistive devices by root-causing psychological or neuromuscular disorders.**

I. INTRODUCTION

Accurate prediction of human reaction time (RT) from electroencephalogram (EEG) signals can facilitate brain-computer interface (BCI) by providing meaningful information of the underlying neurophysiological phenomena. It can also help assess the performance of people with a wide range of speech or motor impairments. These include congenital impairments like autism and cerebral palsy, as well as acquired conditions such as spinal muscular atrophy (SMA) and amyotrophic lateral sclerosis (ALS) [1]. Besides that, a precise prediction can help determine attention or vigilance lapses, which is essential for critical tasks like air-traffic control or long-haul driving [2]. EEG-based BCI has been pursued extensively, and recent advancements suggest its potential in assisting, augmenting, or repairing human cognitive or sensory-motor functions [3]. With an accurate prediction of RT from EEG, the mental state can be more accurately assessed, yielding a better BCI accuracy [4]. However, not many studies have been conducted yet to estimate RT based on EEG.

One of the recent studies extracted Riemannian tangent space features to estimate RT across 16 participants [5] with a predicted root mean square error (RMSE) of 132.5 ms and a correlation coefficient (CC) of 0.61. Another similar study [6] used a rapid serial visual presentation-based experiment involving six subjects and predicted RT with an average RMSE of 119.5 ms. These studies portrayed subject-specific prediction methods. In our previous works [7], [8], [9] we created generalized models using machine learning and deep learning-based techniques to solve this problem and achieved a maximum CC 0f 0.80 and a minimum RMSE of 108.6 ms. This study tried to go a step further, using a more meaningful periodogram-based EEG data structure and developing a novel 3D neural network architecture to predict RT.

II. METHODS

A. Experiment Details

We conducted a straightforward experiment on 31 male and 17 female (ratio of 20:11) participants where multi-channel EEG signals and corresponding response times were recorded for a simple visual stimulus-based reaction task. In a digital screen, a plus symbol (+) appeared in the midpoint, and following a variable monitoring period, it transformed into a cross ($\times$). The participants were supposed to press the space bar with the minimum delay possible as soon as they observe the change in the symbol (+ to $\times$). This procedure was repeated several times for all the subjects, and each of

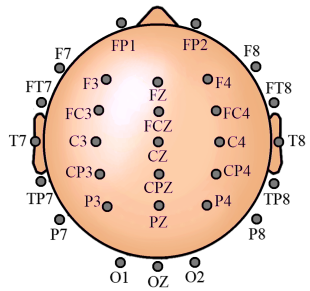

Fig. 1. Thirty-channel electroencephalogram (EEG) scalp locations.

This work was not supported by any organization

[1] M. S. N. Chowdhury is with School of Electrical & Computer Engineering, Purdue University

[2] A. Dutta, and D. W. Bliss are with School of Electrical, Computer & Energy Engineering, Arizona State University

[3] M. K. Robison is with Dpt. of Psychology, University of Texas at Arlington

[4] C. Blais, and G. Brewer are with Dpt. of Psychology, Arizona State University

* Corresponding author email: chowdh31@purdue.edu

the repetitions represented a single trial. The experiment's duration was 30-minutes for each of the 48 participants, providing 320 trials (range: 219–442) per subject on average. Noninvasive EEG data were recorded from 30 scalp locations for each trial. EEG scalp locations are shown in figure 1. All the participants possessed normal or corrected-to-normal visions with no previous record of neurological problems. The experiment was conducted with informed consent from the subjects, and the Institutional Review Board of Arizona State University approved all the procedures involved.

B. Data Preprocessing

The recorded EEG signals were sampled at a rate of 1 kilohertz. These recordings were then filtered from 0 to 400 Hz, preserving the useful information in the low-frequency region and getting rid of the high-frequency noise. To eliminate ocular artifacts, we first downsampled the data to 250 Hz, and the customized data were then forwarded to a GPU-optimized infomax independent component analysis procedure in EEGLAB. From the ICA, the ocular components were identified by visual inspection. We back-projected the remaining components to the original dataset and obtained artifacts-free unfiltered data. Afterward, the data were normalized by mean and variance. There were starting and ending point indicators for each subject, and between these two points, there were several trials. Each trial consisted of the following sequential events:

1) beginning of the trial (emergence of +)
2) change of the symbol (+ to $\times$)
3) response of the subject (space bar press)

We used the EEG signals from all 30 channels recorded between events 1 and 2 as the observation sequences. The differences in time between events 2 and 3 were also recorded as the corresponding RTs. In our experiment, only 2.6% of the RTs were over 1000 ms and so discarded as exceptions. The final dataset was comprised of 15,324 trials from all 48 subjects. The average RT over all the subjects was around 400 ms [10], with 66% of RTs lying between 315-515 ms. As a large number of RTs were in the same region, we needed to balance the dataset such that it had a close to even distribution of RTs across the whole range. We ended up with 6,000 trials out of 15,324 after balancing. Then we randomized the dataset and split it into training and testing sets. This was repeated until both sets had achieved comparable mean, variance, and range of RTs. We tried different training and testing ratios and settled for an 80:20 split, which provided us with low variance for both training parameter estimates and testing performance statistics.

C. Periodogram

The periodogram has been proven to be an efficient yet effective tool to represent a signal's spectral characteristics [11]. A periodogram is an estimation of a signal's power spectral density, which can be computed by taking the Fourier transform of the signal's autocorrelation function. As the data is discrete in our case, the periodogram of discrete signal $x[n]$ can be calculated from the following equation:

$$S\left(\frac{k}{NT}\right) = \left|\sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi kn}{N}}\right|^2 \quad (1)$$

where S is the periodogram of $x[n]$, k is the spectral-domain ordinal, and n is the time-domain ordinal. N and T are the sequence length and the sampling period, respectively. In compliance with earlier findings, we isolated data from *delta* to *beta* band (1–35 Hz) using a rectangular window to be used as features. A total of 72 feature points were obtained from each channel and trial, representing power at 72 frequency values between 1 and 35 Hz.

D. Data Restructuring

We restructured each of the single-trial, multi-channel periodogram data-subset in a particular way to create a 3D cuboid representation where both inter-channel and inter-frequency relationships for the particular trial were preserved. In our 3D cuboid, each of the XY planes contained the single-trial, single-frequency periodogram values from all the channels. It was essential to store the periodogram values in an order that reflects the corresponding channel locations. To do so, we organized the periodogram values from all 30 channels inside a 7x5 matrix considering the relevant channel positions. There were 5 (7x5-30 = 5) empty indices in the matrix, which were filled by computing the averages of the neighboring values. Figure 2 (a) shows how the 30 channel locations were restructured to create the 7x5 matrix. The empty indices are denoted by 'e' in the figure. Figure 2 (b) illustrates a symbolic representation of the topographical plot containing periodogram values before and after the transformation. On the other hand, the Z-axis represents all 72 frequency values ranging from 1 to 35 Hz. So, the XY plane preserved the inter-channel relationship, and the Z-axis preserved the inter-frequency relationship. After the transformation, we obtained a cuboid of dimension

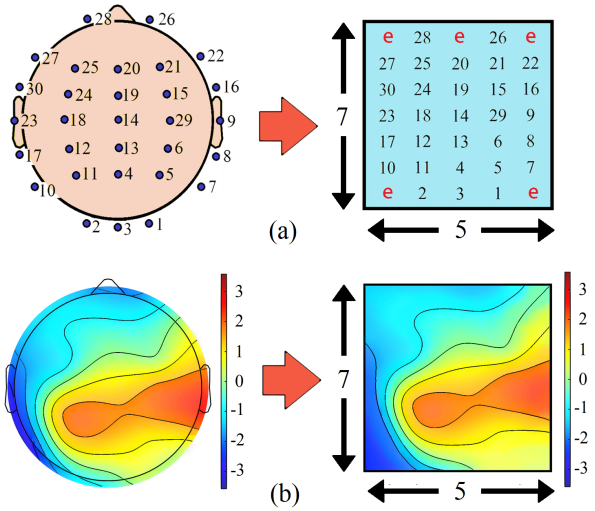

Fig. 2. (a) Channel distributions and (b) corresponding topographical plots, before and after the transformation

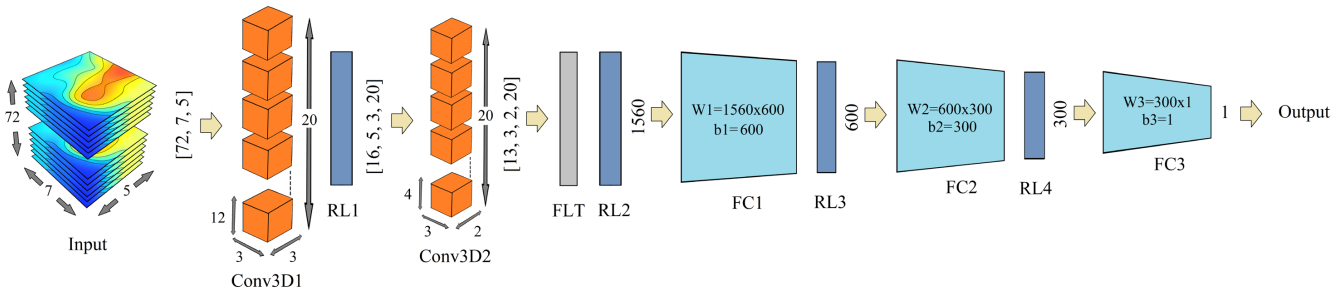

Fig. 3. 3D CNN model architecture and data flow

72×7×5 for each trial. Considering all the trials, we obtained a final dataset of dimension 72×7×5×6,000.

E. Model Architecture

In light of prior studies [9], [12], [13], we designed a simple 3D convolutional neural network (CNN) to predict RTs from the cuboids computed earlier. The model consists of 2 3D convolutional (Conv3D) layers and 3 fully connected (FC) layers. We flattened (FLT) the output of the second Conv3D layer to feed the first FC layer. In between these layers, we inserted 4 rectified linear unit (ReLU) layers to introduce non-linearity. Figure 3 shows the model.

1) 3D Convolutional Layers: Convolution operations preserve and propagate complex inter-component information of a given matrix at the cost of higher computational requirements. As the computational capabilities of present hardware devices are observing tremendous growth, convolutional layers are incorporated more often in neural network architectures designed for a number of applications along with biomedical signal processing. A 3D convolutional layer performs the following operation using several randomly initialized filters, optimized by back-propagation:

$$(h*f)[x,y,z] = \sum_{i=-m}^{m} \sum_{j=-n}^{n} \sum_{k=-l}^{l} h[i,j,k] f[x-i, y-j, z-k] \tag{2}$$

where h is the filter, f is the input, and m,n,l are the finite support sets.

We adopted 3D Convolutional layers in our study to observe how much we can improve our model's performance utilizing rich contextual information extracted from the 3D input layers. Figure 4 shows a sample 3D filter layer performing 3D convolution by moving across all 3 directions. For our study, we used 2 Conv3D layers with the following dimensions:

- Conv3D1: 12×3×3×20
- Conv3D2: 4×3×2×20

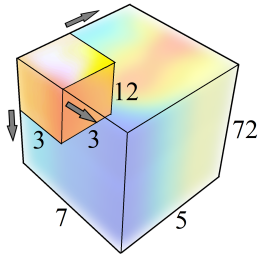

Fig. 4. Sample 3D convolution operation

2) Fully Connected Layers: Fully Connected layers are the layers where all the inputs from one layer are connected to each of the activation units of the following layer. It performs the affine transformation using matrix multiplication followed by bias offset. The following equation shows the fundamental operation undertaken by a fully connected layer:

$$\mathbf{Z} = \mathbf{W^T}.\mathbf{X} + \mathbf{b} \tag{3}$$

where **X** is an input matrix of size N × 1 and **Z** is the output matrix of size M × 1. The transpose of a weight matrix **W** of dimension N × M is first multiplied by the input matrix. Furthermore, an M × 1 bias matrix **b** is also offset. All the associated weight and bias values are optimized through back-propagation, starting from random initialization. In this model, we have 3 fully connected layers with the following dimensions:

- FC1:
  1) **W1** = 1560 × 600
  2) **b1** = 600 × 1
- FC2:
  1) **W2** = 600 × 300
  2) **b2** = 300 × 1
- FC3:
  1) **W3** = 300 × 1
  2) **b3** = 1 × 1

3) Rectified Linear Unit Layers: ReLU is one type of activation layer, and the main purpose of using this layer is to introduce non-linearities in the linear outputs of a layer. It is a very simple operation that is critical for the generalization of the model. The function of the layer is as shown in the following equation:

$$y = max(0, x) \tag{4}$$

where x is the input value, and y is the output value.

4) Loss Function: To train the model, we used a hybrid loss function (LF) that maximizes the correlation coefficient (CC) and minimizes the mean square error. The following equation illustrates the loss function:

$$LF = (1 - \frac{\sum_{i=0}^{n-1}(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}) + \frac{\sum_{i=0}^{n-1}(x_i - y_i)^2}{\sum_{i=0}^{n-1} y_i^2} \quad (5)$$

where n is the total number of predictions made. x_i is the predicted RT and y_i is the actual RT for $i \in [0, n-1]$. $\bar{x}$ and $\bar{y}$ are the means and σ_x and σ_y are the standard deviations of predicted and actual RTs, respectively. To maximize CC, the first part of the loss function minimizes (1-CC). The second part minimizes the mean square error.

III. RESULTS

For this analysis, predictions higher than 1000ms were clipped to 1000ms before computing the results. Using the CNN model, we achieved an RMSE of 91.5 ms and a CC of 0.83. The actual vs. predicted RTs for the testing set using our model are shown in figure 5. Also, a comparison of obtained RMSE and CC values for RT estimation among previous models [7], [8], [9] and the 3D CNN architecture is shown in Table I. Clearly, the 3D CNN model exhibited the best performance with improved results.

IV. DISCUSSION

This study has introduced a novel method of processing multi-channel EEG signals to predict RT. One of the most significant contributions of this work was the generation of the 3D data structure comprising the spectro-spatial EEG features. Single-trial RT was predicted from these EEG features using a 3D CNN architecture containing two 3D convolutional layers and 3 fully connected layers. The proposed model utilized around 6,000 trials from 48 subjects and performed better than all the methods from the previous studies. We were able to predict RT with a CC of 0.83 and an RMSE of 91.5 ms. It is one of the few studies where human RT was predicted using features from EEG in sustained attention-based visual response tasks. We believe our work will contribute to the further advancements of EEG-based brain-computer interface design and help accurate assessment of the mental states, which will benefit researchers in the field of psychology and neuroscience. Moreover, such a generalized model can help monitor people with all kinds of neuromuscular disorders and help them in rehabilitation.

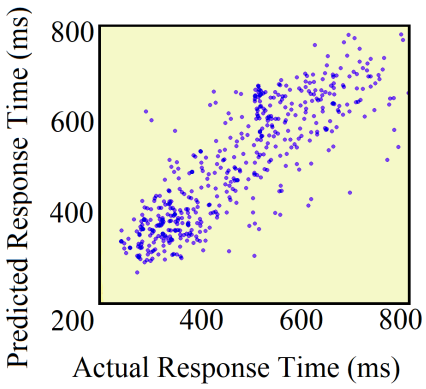

Fig. 5. Actual vs. predicted RT for the 3D CNN model

TABLE I
REGRESSION-BASED ESTIMATION RESULTS FOR DIFFERENT MODELS

Algorithm	CC	RMSE (ms)
Linear Regression	0.56	158.7
Ridge Regression	0.56	157.6
Support Vector Regression (SVR)	0.60	136.7
Extra Tree Regression	0.73	114.4
Random Forest Regression	0.74	111.2
FCNN + Random Forest	0.78	110.4
CNN+ Random Forest	0.80	108.6
3D CNN	0.83	91.5

REFERENCES

[1] A. Geronimo, Z. Simmons, and S. J. Schiff, "Performance Predictors of Brain–computer Interfaces in Patients with Amyotrophic Lateral Sclerosis," *J. Neural Eng.*, vol. 13, 2016, p. 26002.

[2] B. Binias, D. Myszor, H. Palus, and K. A. Cyran, "Prediction of Pilot's Reaction Time Based on EEG Signals," *Frontiers in Neuroinformatics*, vol. 14, 2020, p. 437.

[3] J. D. R. Millán, R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb, C. Neuper, K.-R. Müller, and D. Mattia, "Combining Brain–computer Interfaces and Assistive Technologies: State-of-the-art and Challenges," *Front. Neurosci.*, vol. 4, 2010, p. 161.

[4] A. Myrden and T. Chau, "Effects of User Mental State on EEG-BCI Performance," *Frontiers in Human Neuroscience*, vol. 9, 2015, p. 308.

[5] D. Wu, B. J. Lance, V. J. Lawhern, S. Gordon, T.-P. Jung, and C.-T. Lin, "EEG-Based User Reaction Time Estimation Using Riemannian Geometry Features," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25(11), November 2017, pp. 2157–2168.

[6] A. Luo and P. Sajda, "Using Single-Trial EEG to Estimate the Timing of Target Onset During Rapid Serial Visual Presentation," in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 79–82.

[7] M. S. N. Chowdhury, "EEG-Based Estimation of Human Reaction Time Corresponding to Change of Visual Event," M.S. Thesis, School of Elec., Comp. & Energy Eng., Arizona State University, Tempe, Arizona, USA, 2019.

[8] M. S. N. Chowdhury, A. Dutta, M. K. Robison, C. Blais, G. A. Brewer, and D. W. Bliss, "A Generalized Model to Estimate Reaction Time Corresponding to Visual Stimulus Using Single-Trial EEG," in *42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2020, pp. 3011-3014.

[9] M. S. N. Chowdhury, A. Dutta, M. K. Robison, C. Blais, G. A. Brewer, and D. W. Bliss, "Deep Neural Network for Visual Stimulus-Based Reaction Time Estimation Using the Periodogram of Single-Trial EEG," *Sensors*, vol. 20, no. 21, 2020, p. 6090.

[10] R. Whelan, "Effective Analysis of Reaction Time Data," *The Psychological Record*, 58, 2008, pp. 475–482.

[11] M. Akin and M. Kiymik, "Application of Periodogram and AR Spectral Analysis to EEG Signals," *J. Med. Syst.*, vol. 24, 2000, pp. 247-56.

[12] T. John, C. Luca, J. Jing, D. Justin, C. Sydney, and W. M. Brandon, "EEG CLassification Via Convolutional Neural Network-Based Interictal Epileptiform Event Detection," in *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2018, pp. 3148-3151.

[13] A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, and H. Takahashi, "Seizure Detection by Convolutional Neural Network-Based Analysis of Scalp Electroencephalography Plot Images," *NeuroImage Clin.*, vol. 22, 2019, p. 101684.