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

Single feature spatio-temporal architecture for EEG Based cognitive load assessment

Akshaya Ramaswamy, Arpit Bal, Abhranila Das, Jayavardhana Gubbi,
Kartik Muralidharan, Ramesh Kumar Ramakrishnan, Arpan Pal, Balamuralidhar P
Embedded Devices and Intelligent Systems, TCS Research, India

***Abstract*— The study of electroencephalography (EEG) data for cognitive load analysis plays an important role in identification of stress-inducing tasks. This can be useful in applications such as optimal work allocation, increasing efficiency in the workplace and ensuring safety in difficult work environments. In order for such systems to be realistically deployable, easy acquisition and processing of the data on a wearable device is imperative. Current techniques primarily perform offline processing to analyse a multi-channel EEG to make a post facto assessment. This work focusses on building a new deep learning architecture that performs a single feature based spatio-temporal analysis of EEG data. This is achieved by creating a brain topographic map based on a single feature followed by spatio-temporal analysis using the developed network architecture. Data from two cognitive load experiments on the Physionet EEGMAT dataset were used to validate the performance. The network achieves an accuracy of 98.3% which is better than similar state-of-the-art approaches. Moreover, the proposed approach facilitates analysis of the spatial propagation of a signal, which is not possible through conventional EEG signal representations.**

I. INTRODUCTION

Memory in the human brain can be functionally segmented into two parts according to cognitive load theory — working memory and long-term memory [1]. Cognitive load is used as an important criterion for analyzing proficiency in various tasks like driving [2] and learning [3]. The load on the working memory while doing various mental tasks is the cognitive load on the person. It is a key attribute for determining cognitive abilities such as problem-solving and stress-endurance. Objective characterisation of the cognitive load in near real-time can help in preventing burnout, prolonged stress and ensuring safety in high mental load working environments.

EEG has been used to capture the activity of the brain for a while. The data is collected in a multi-channel setup, where 23 channel data is collected from the scalp. This can help in determining the cognitive load and the state of a person [4]. The analysis of brain EEG based cognitive load analysis is used in diverse applications starting from designing brain-human computer interfaces (BCI/HCI) to designing more retentive advertisements [5]. These methods use multiple channels to calculate time and frequency domain features for determining the cognitive state. Hence, the application of these methods are often restricted to post facto analysis of the data rather than continuous assessment of the cognitive state.

An EEG signal can be represented more intuitively in the two-dimensional spatio-temporal space using topographic maps. A topographic mapping establishes the relation between the multi-channel EEG signal and the spatial and temporal dimensions enabling a visualization of the activation in different regions of the brain. This representation helps in localizing activated regions and further assists in studying the responses that are evoked by different types of tasks. This work uses a topographic map as the representation for automatic assessment. The key advantage of this approach is the possibility of achieving the objective of cognitive load classification using a single or very few features with an acceptable level of accuracy. In this work, a deep learning approach for cognitive load assessment is proposed, with the following key contributions: (a) an approach for analysis of spatial propagation of brain signals is proposed, (b) Topographic maps (topomaps) are generated using entropy and PSD, and are converted into a continuous video of spatio-temporal maps, (c) a new deep learning architecture called EEG-TopoNet, based on spatio-temporal architecture [6] is proposed for EEG topomap analysis, (d) the developed architecture is validated using two levels of cognitive load assessment tasks, first to identify if the subject is at rest or active state and the second is to classify the subject based on count quality in the given arithmetic task.

II. RELATED WORKS

EEG signal analysis has been explored for various tasks ranging from mental workload estimation to seizure detection and emotion recognition [7]. Many classical machine learning techniques have been applied to EEG data for mental load analysis. This follows three main steps: removal of signal noise, extraction of hand-crafted features, followed by classification into different levels of load. Shivabalan *et al.* [8] used a novel machine learning classifier called the SMORASO-DT, which combines SMOte, Random forest and lASso- Decision Tree to separate subjects based on their count quality. Fatimah *et al.* [9] used quadratic discriminant classifier for the same task. In [10], subjects are assessed for mental cognitive load mismatching (CLMM) state (bad performer) or cognitive load matching (CLM) state (good performer), by analyzing multi-modal sensor data from both brain (EEG) and cardiac activities (HRV). An RBF-kernel SVM classifier was used in Kim *et al.* [11] with eleven features to segregate the quality of arithmetic tasks performed

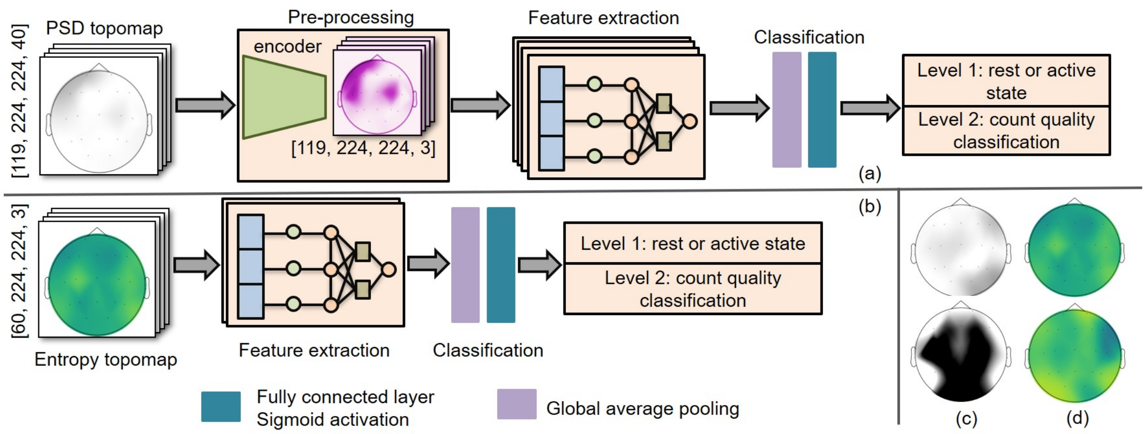

Fig. 1: Proposed framework using a) PSD-based topographic maps; b) entropy topographic maps. Topographic plots for a subject generated using c) PSD of baseline task (top) and PSD of mental calculation task (bottom), d) entropy of baseline task (top) and entropy of mental calculation task (bottom)

by the subjects. In the last few years the application of deep learning architectures has gained attention. Here deep learning techniques are combined with EEG topographic maps to understand the brain functionality with respect to different tasks. Deep recurrent convolutional networks are utilized in [12], [13] to learn spatial, spectral and temporal features from topographic representations for cognitive load classification. In these topomap-based approaches, the spatial and temporal features are learned sequentially by having either spatial or spatial-spectral layers followed by temporal layers in the network. This work investigates the three-dimensional convolution and long short-term memory (LSTM) networks that simultaneously look at the spatio-temporal and spectral planes to extract meaningful features for mental load classification. The use of feature based spatio-temporal topomaps followed by deep learning is shown to produce reliable accuracy and performance in near real-time.

III. MATERIALS AND METHODS

A. Dataset description

This work uses the publicly available EEG dataset for Mental Arithmetic Tasks (EEGMAT) from Physionet. This dataset contains EEG recordings of subjects before and during the performance of a mental arithmetic task [14]. The experiment was conducted in a controlled environment with thirty-six subjects who were given an arithmetic task. The task was carried out for four minutes with EEG and ECG data of every subject collected during the experiment. The subjects were separated into two categories — good and bad — based on the performance of the task. No further post-processing was performed on the signal data.

B. Topographic map generation

For the generation of topographic maps from the EEG signals, MNE toolbox was used [15]. Two sets of topographic maps are generated from the EEGMAT dataset, as shown in Figure 1(c) and Figure 1(d), one using a power spectral density (PSD) representation, and the second based on entropy.

For the first set, a method similar to that used by Zhang *et al.* [12] is employed. The topographic maps are computed by decomposing the downsampled 21-channel EEG data into frequency bands of 1 Hz each, from 1Hz to 40Hz and then computing their power spectral density using Morlet wavelet transform. The topographic maps are generated at an interval of 0.5 seconds for the total duration of 60 seconds resulting in 120 frames for each subject and for each class (baseline and mental calculation task). Each frame consists of forty topographic maps corresponding to the forty frequencies, stacked on top of each other. This results in a dimension of $[120, 232, 221, 40]$. The frames are resized to (224, 224), and the last frame is dropped since it is identified by the toolbox as being corrupt. For the second set of topographic maps, the sample entropy [16] is computed for each one-second interval of the downsampled EEG data, thereby generating 60 frames for each subject and each activity, and spatially resized to (224, 224). The PSD-based topographic maps are represented in grayscale with dimension $[119, 224, 224, 40]$, while the entropy based topographic maps are represented in the RGB space with dimension $[60, 224, 224, 3]$. These are given as input to the proposed spatio-temporal EEG network (EEG-TopoNet) explained in the subsequent sections.

C. Spatio-temporal EEG network (EEG-TopoNet)

The topographic maps capture brain activity across time, providing a visualization of activated regions during various tasks. This activity patterns can be extracted using a spatio-temporal network, similar to [6] with appropriate modifications to handle EEG topomaps instead of videos. Let T_{psd} and T_{ent} represent the generated PSD topomaps with dimension $[119, 224, 224, 40]$ and the generated entropy topomaps with dimension $[60, 224, 224, 3]$, respectively. These are fed as input to the proposed spatio-temporal deep network, referred to as EEG-TopoNet. Let T denote a generic topomap representation with dimensions $[30, 224, 224, 3]$. The feature extraction with T of a fixed dimension is described first, followed by the additional steps for T_{psd} and T_{ent}.

Feature Extraction: The core of the architecture is the feature extractor as shown in Figure 2, that forms the backbone for spatio-temporal feature representation. The goal is to extract descriptive features that capture the brain activation patterns over different time durations. This uses 3D convolution layers and long short-term memory (LSTM) layers as building blocks, and is inspired by the hierarchical arrangement in [6]. The architecture uses 3D pre-trained models to compute short-term spatio-temporal features, and these are aggregated using attention pooling and 2D convLSTM layers. As explained above, T is the input to the feature extractor. T is split into three parts across first dimension, such that $T_i, i = 1, 2, 3$, has dimension $[10, 224, 224, 3]$. I3D features, denoted by T_i^{i3d}, are computed for T_i by tapping the 'mixed-5c' layer output in the popular I3D architecture [17]. The three outputs T_i^{i3d}, $i = 1, 2, 3$, each of dimension $[14, 14, 832]$, are processed further to gather long-term variations. This is achieved by first adding a 2D convLSTM layer with 64 filters. This takes each T_i^{i3d} as input at every timestep. The intermediate cell states C_i^{l1}, $i = 1, 2, 3$ and hidden states H_i^{l1}, $i = 1, 2, 3$, are tapped from this layer and aggregated using attention-based pooling. Here, I3D features are used as context to compute two attention weighted feature sets C_{att}^{l1} and H_{att}^{l1}. These are finally combined using a second 2D convLSTM layer with 64 filters, to calculate the comprehensive features H^{l2} from the hidden state. The I3D features of T_2 are concatenated to this to form the final set of features F_T of dimension $[14, 14, 896]$ for a topomap representation T. As the generated topomaps have different number of frames as the input data, they are pre-processed, as explained below, before being sent to the feature extractor. This will keep the core EEG-TopoNet architecture intact and enables a variety of topomaps as inputs.

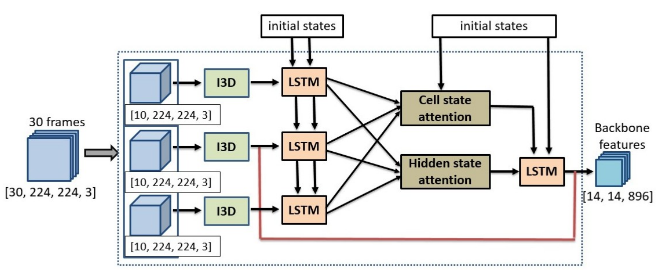

Fig. 2: The backbone network used for feature extraction from the topographic maps

Classification using PSD topomaps: The block diagram for PSD topomap classification is shown in Figure 1(a). Pre-processing block is an auto-encoder consisting of two 3D convolution layers with 32 filters and 3 filters respectively followed by two de-convolution layers to reconstruct the input. This is trained with T_psd as input to compress the data and remove redundancies. The trained encoder is used for further training. The encoder outputs a compressed representation E_{psd} of dimension $[119, 224, 224, 3]$. The feature extractor in the proposed network takes a fixed size of 30 frames as input. Therefore, E_{psd} is divided into four sets of 30 frames each E_{psd}^i, $i = 1, 2, 3, 4$, and provided to the feature extraction sequentially. For the last set of 29 frames, a blank frame is appended. The cell states and hidden states of the convLSTM layers are carried over from one set to the next and the final LSTM layer output is extracted for each set. The extracted feature maps F_{psd}^i, $i = 1, 2, 3, 4$ are concatenated and flattened using global average pooling and passed through a 3D convolution layer with sigmoid activation for binary classification.

Classification using entropy topomaps: Figure1(b) shows the steps for classification using entropy topomaps. The entropy input is directly fed to the feature extractor without any pre-processing. T_{ent} is provided as two sets of thirty frames sequentially. Further, the extracted features F_{ent}^i, $i = 1, 2$ are passed through the same set of layers as the PSD approach, for classification.

IV. EXPERIMENTS AND RESULTS

The proposed approach is evaluated on the EEGMAT dataset using the two generated topographic maps. Further, two cognitive load tasks are assessed: mental state qualification and count quality classification. For the count quality classification, only the active state topomaps are used. The performance using both inputs are compared with each other and with other methods in literature.

A. Evaluation setup

The network is implemented using Tensorflow libraries and the experiments are performed on an NVidia Tesla V100 GPU. The network is trained for 200 epochs using an Adam optimizer with learning rate 0.0001 and a binary cross-entropy loss. A ten fold cross-validation is performed for both levels of classification. The classification performance is evaluated by computing three metrics: accuracy, sensitivity and specificity. To the best of knowledge, there are no topomap-based approaches applied on the EEGMAT dataset. For performance comparison, the spatio-temporal approach in [13] is utilized, and three temporal aggregation techniques explored in [13] are implemented here for classification: a) maxpool; b) LSTM; c) temporal convolution. ResNet pre-trained model [18] is used for spatial feature extraction. For the proposed networks, the performance is evaluated with and without augmentation. For the entropy topomaps, augmentation of the topomaps is performed for both levels of classification, by randomly adding noise and varying brightness levels. For the PSD topomaps, augmentation is performed on the topomaps for the mental state classification and for the count quality classification, SMOTe is used to generate synthetic data to handle the class imbalance.

B. Results and discussion

The cross-validation performance for both levels of classification is summarized in Table I. Overall, EEG-TopoNet with PSD-topomaps achieves the best results with accuracy 98.3% for the mental state classification and 95% for count quality classification. In comparison, a) the entropy based

TABLE I: Quantitative evaluation on EEGMAT dataset. Evaluation of performance using EEG-TopoNet and two types of topomaps. Comparison with temporal aggregation techniques in [13].

Approach	Mental state classification			Count quality classification			Number of training parameters
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	
ResNet + maxpool	97.5	98.6	95.8	86.25	86.11	87.5	0.1 million
ResNet + temporal conv	90.1	90.3	90.0	88.75	87.5	90.3	1.8 million
ResNet + LSTM	88.3	87.5	88.9	81.25	81.9	80.5	10 million
EEG-TopoNet (entropy) without augmentation	87.5	84.7	90.2	87.3	85	88.46	3.5 million
EEG-TopoNet (entropy) with augmentation	90.3	88.89	91.67	86.1	83.3	88.9	3.5 million
EEG-TopoNet (PSD) without augmentation	94.2	93.05	94.5	93.75	94.5	93.05	3.7 million
EEG-TopoNet (PSD) with augmentation	**98.3**	**98.6**	**97.2**	**95**	**94.5**	**96.4**	3.7 million

topomaps do not perform well in both stages of classification; b) the ResNet + maxpool, with the least training parameters, achieves comparable performance of 97.5% on mental state classification, but does not perform well on count quality classification; c) The ResNet + LSTM has a lot of training parameters, however, does not impress at both the tasks; and 4) The convolution based temporal aggregation achieves the second best performance at the count quality assessment with 1.8 million trainable parameters. The proposed approach with PSD has twice the number of parameters but gives the highest performance for both the tasks and shows a considerable improvement of 6.8% over the state-of-the-art, for count quality classification.

A unique merit of the proposed approach is that, unlike the conventional approach which use signal analysis and are based on a number of features, the proposed approach operates on only one feature - either a PSD based topomap or entropy based topomap. Conventional techniques employ multiple hand-crafted features specific to each task. The baseline performance using the such techniques achieves an accuracy of 99.9% at mental task [10] and 94.2% at count quality task [9]. In contrast, the proposed approach achieves comparable performance with a single feature and an end-to-end training. Further, the same framework can be used with a variety of inputs and targeting multiple applications highlighting the horizontal nature of the proposed method.

V. CONCLUSION

This work explores the application of spatio-temporal attention-based feature learning using EEG topographic maps for cognitive load assessment. Topographic representations are generated by computing power spectral density and entropy from raw EEG signals. These are fed to a spatio-temporal deep network for cognitive load assessment at different levels. The method is evaluated on the EEGMAT dataset and the performance is compared with other state-of-art deep networks. The best results are achieved using the proposed network when applied to PSD-based topomaps. This shows great scope in localization of activations regions in the brain and facilitates longitudinal analysis by assessing spatial propagation of the signal.

REFERENCES

[1] P. Antonenko, F. Paas, R. Grabner, and T. Van Gog, "Using electroencephalography to measure cognitive load," *Educational Psychology Review*, vol. 22, no. 4, pp. 425–438, 2010.

[2] T. Taylor, A. Pradhan, G. Divekar, M. Romoser, J. Muttart, R. Gomez, A. Pollatsek, and D. L. Fisher, "The view from the road: The contribution of on-road glance-monitoring technologies to understanding driver behavior," *Accident Analysis & Prevention*, vol. 58, pp. 175–186, 2013.

[3] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive science*, vol. 12, no. 2, pp. 257–285, 1988.

[4] Y. Wang and S. Makeig, "Predicting intended movement direction using eeg from human posterior parietal cortex," in *International Conference on Foundations of Augmented Cognition*, pp. 437–446, Springer, 2009.

[5] J. García-Madariaga, I. Moya, N. Recuero, and M.-F. Blasco, "Revealing unconscious consumer reactions to advertisements that include visual metaphors. a neurophysiological experiment," *Frontiers in Psychology*, vol. 11, 2020.

[6] A. Ramaswamy, K. Seemakurthy, J. Gubbi, and B. Purushothaman, "Spatio-temporal action detection and localization using a hierarchical lstm," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 764–765, 2020.

[7] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: a review," *Journal of neural engineering*, vol. 16, no. 3, p. 031001, 2019.

[8] K. Shivabalan, B. Deb, A. Ramachandran, and S. Goel, "Smoraso-dt: A hybrid machine learning classification model to classify individuals based on working memory load in mental arithmetic task," *medRxiv*, 2020.

[9] B. Fatimah, D. Pramanick, and P. Shivashankaran, "Automatic detection of mental arithmetic task and its difficulty level using eeg signals," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–6, IEEE, 2020.

[10] R. Xiong, F. Kong, X. Yang, G. Liu, and W. Wen, "Pattern recognition of cognitive load using eeg and ecg signals," *Sensors*, vol. 20, no. 18, p. 5122, 2020.

[11] K. Kim, N. T. Duc, M. Choi, and B. Lee, "Eeg microstate features according to performance on a mental arithmetic task," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021.

[12] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning spatial–spectral–temporal eeg features with recurrent 3d convolutional neural networks for cross-task mental workload assessment," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 27, no. 1, pp. 31–42, 2018.

[13] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[14] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, p. 14, 2019.

[15] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in neuroscience*, vol. 7, p. 267, 2013.

[16] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, 2000.

[17] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," *CoRR*, vol. abs/1705.07750, 2017.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.