EEGCAPS: Brain Activity Recognition using Modified Common Spatial Patterns and Capsule Network

Hamidreza Sadreazami [†], Marzieh Amini[*], M. Omair Ahmad [‡] and M. N. S. Swamy [‡]
[†]Bioengineering Department, McGill University, Montreal, Canada
[*]School of Information Technology, Carleton University, Ottawa, Canada
[‡]Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada

2021 IEEE International Symposium on Circuits and Systems (ISCAS) | 978-1-7281-9201-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISCAS51556.2021.9401332

***Abstract*—Brain computer interface is a developing technology that can provide enhanced quality of life to individuals suffering from various disabilities. In this work, a new binary electroencephalography (EEG) signal decoding algorithm is proposed using a modified common spatial pattern and capsule network. The proposed method is realized by extracting the spectral-temporal common spatial pattern features from the EEG signals while preserving the time resolution of the signal. The resulting features are fed into the capsule network for automatic feature extraction and classification. The capsule network is known to be superior to convolutional neural networks in requiring less training data, which makes it a promising candidate for EEG signals classification. The performance of the proposed method is evaluated and compared to that of the other methods by conducting several experiments. The results demonstrate that the proposed method provides recognition accuracy higher than that provided by other methods.**

***Index Terms*—Common spatial pattern, binary classification, EEG signals, capsule network, brain computer interface.**

I. Introduction

Motor imagery brain computer interface (MI-BCI) provides the potential for controlling external devices by translating the brain signals. This has been a breakthrough in many applications including wheelchair control for people with physical disabilities and realization of a new indirect communication path for amyotrophic lateral sclerosis patients [1]. To realize a patient-friendly neuro-rehabilitation system using MI-BCI [2]–[4], developing automatic, reliable and efficient algorithms to translate the brain activities into the device commands is of great importance.

The electroencephalogram (EEG) signals capture changes in brain's electrical waves and are used to analyze the brain activities. Common spatial patterns (CSP) and its variants have been widely applied in MI-BCI task recognition studies [2], [5]–[8]. These methods take into account spectral characteristics of the EEG signals to optimally distinguish different classes. This is realized by determining the directions in the pattern space through projecting the EEG signals onto a subspace. For instance, in [9], a subband common spatial pattern (SBCSP) was developed to decompose EEG signals by a filter bank and extract discriminative features. In [10], a filter bank-based method (FBCSP) was proposed to process EEG signals using CSP applied to different frequency bands. In [6], regularized CSP with Tikhonov regularization was presented. In [11], a method based on filter-bank regularized common spatial pattern (FBRCSP) was proposed and tested on small sets of samples. However, all the above-mentioned works were reliant on heavy pre-processing and extracting hand-crafted features.

Deep learning has made a revolution in brain activity recognition by providing promising tools to handle automatic feature extraction. In [12] and [13], high-level features of EEG signals were extracted with convolutional neural networks and used in classification of the EEG data. In [14], a deep belief network was used to model EEG waveforms for classification in a semi-supervised manner. A similar network was investigated in [15] for exploratory analysis and pre-training in neuroimaging applications. A deep learning scheme based on Boltzmann machines was proposed in [16], in which the EEG signals in the frequency domain are employed. The above studies have shown the benefits of adopting deep learning in systematically extracting features from EEG signals, even in the absence of large volume of data.

Leveraging the recent advances in deep neural networks, in this work, a new method, called EEGCAPS, for MI-BCI EEG signal classification based on modified common spatial patterns and deep capsule network is proposed. The goal is to improve the recognition accuracy of the EEG signals by taking into consideration spectral as well as temporal resolutions of the EEG signals. To this end, the spatially-filtered features are extracted from the EEG signals which are then converted to image representation. The obtained images are fed into the capsule network for automatic feature learning process. Experiments are conducted to evaluate the performance of the proposed method using the standard data set taken from the BCI competitions III.

II. Modified Common Spatial Patterns

In order for a brain activity recognition task to be realized, pre-processing is required. In the pre-processing step, the EEG signals are bandpass filtered using a 5th order Butterworth filter to extract the contents of the signal in the range of $(7-30)$ Hz. The signal is then smoothened using a weighted moving average filter of length 10 [1]. A supervised learning

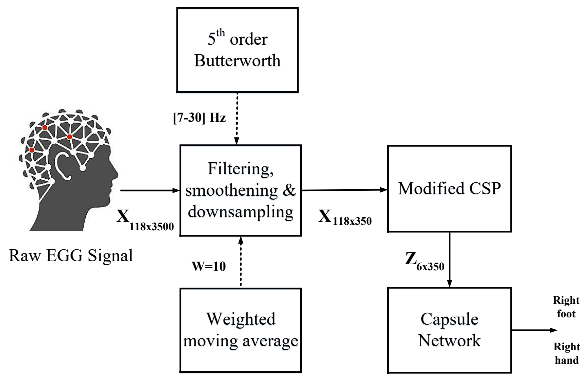

Fig. 1. EEGCAPS block diagram for MI-BCI EEG signal classification.

is considered from the EEG trials set, i.e., $\mathbf{X}_i \in \mathbb{R}^{N_{ch} \times N_t}$, for $(1 \leq i \leq N_{tr})$, where N_{ch}, N_{tr} and N_t, are the number of electrodes, trials and time samples collected from each electrode in a trial, respectively.

CSP filters are commonly used in the BCI domain [6], [17]. For a binary classification problem, CSP provides spatial filters which maximize the variance between two classes. As a result, the extracted signals optimally discriminate two different EEG classes, while revealing spatial patterns of different classes. More specifically, spatial filters are obtained by joint diagonalization of the two covariance matrices corresponding to the two classes. The normalized covariance matrix $\mathbf{R}$ for each trial of the EEG signal $\mathbf{X}_i$, is obtained as

$$\mathbf{R} = \frac{\mathbf{X}_i \mathbf{X}_i^T}{trace\left(\mathbf{X}_i \mathbf{X}_i^T\right)}, \tag{1}$$

where the *trace* of a matrix is the sum of the elements on the main diagonal and $\mathbf{X}^T$ denotes the transpose of the matrix $\mathbf{X}$. The covariance matrices $\mathbf{R}_a$ and $\mathbf{R}_b$ of the two classes a and b are summed up to a new covariance matrix $\mathbf{R}_c$. This matrix is then decomposed into a set of eigenvectors $\mathbf{B}_c$ and eigenvalues $\boldsymbol{\lambda}_c$, i.e., $\mathbf{R}_c = \mathbf{B}_c \boldsymbol{\lambda}_c \mathbf{B}^T$. This decomposition yields a whitening transform, given by $\mathbf{W} = \boldsymbol{\lambda}_c^{-1/2} \mathbf{B}_c^T$, and used to transform $\mathbf{R}_a$ and $\mathbf{R}_b$ into

$$\begin{aligned} \mathbf{S}_a &= \mathbf{W}\mathbf{R}_a\mathbf{W}^T \\ \mathbf{S}_b &= \mathbf{W}\mathbf{R}_b\mathbf{W}^T, \end{aligned} \tag{2}$$

where $\mathbf{S}_a = \mathbf{U}\psi_a\mathbf{U}^T$ and $\mathbf{S}_b = \mathbf{U}\psi_b\mathbf{U}^T$, $\mathbf{U}$ being the orthonormal eigenvectors of $\mathbf{S}_a$ and $\mathbf{S}_b$. In addition, ψ_a and ψ_b are the corresponding diagonal matrices of eigenvalues, which add up to 1. The eigenvectors $\mathbf{U} = \left(U_1, ..., U_{N_{ch}}\right)$ are sorted in a descending order with respect to the eigenvalues in $\psi_a = \left(\psi_{a,1}, \psi_{a,2}, ..., \psi_{a,N_{ch}}\right)$, $\psi_{a,1} \geq ... \geq \psi_{a,N_{ch}}$ and in a ascending order with respect to $\psi_b = \left(\psi_{b,1}, \psi_{b,2}, ..., \psi_{b,N_{ch}}\right)$, $\psi_{b,1} \leq ... \leq \psi_{b,N_{ch}}$. When classes a and b are both projected onto the first eigenvector U_1, the class a yields a maximum variance, whereas the class b yields a minimum variance value. On the other hand, when the classes are projected onto the last eigenvector $U_{N_{ch}}$, class a yields a minimum variance value, while the class b yields a maximum variance value. It is noted that only a few eigenvectors are selected for the discrimination analysis, i.e., $\mathbf{U}^* = \left(U_1, ..., U_m; U_{N_{ch}-m+1}, ..., U_{N_{ch}}\right)$, where $m << N_{ch}$. In other words, only a small number of channels are utilized. The final projection matrix is defined as

$$\mathbf{P} = \mathbf{U}^*\mathbf{W}, \tag{3}$$

The projection matrix $\mathbf{P}$ is further used to map each EEG trial $\mathbf{X}_i$ as $\mathbf{Z}_i = \mathbf{P}\mathbf{X}_i$. It is noted that the dimension of the trial signals is reduced to $2m \times N_t$, i.e., $\mathbf{Z}_i$ is constructed from the first and last $m = 3$ largest eigenvalues that maximizes the difference of the variances of the two classes. At this stage in the proposed method, unlike the commonly-used CSP technique, where the features are obtained by taking the logarithm of the variance of $\mathbf{Z}_i$, we keep the temporal resolution of the EEG signals unchanged and project the spatially-filtered signals onto a 2D representation, resulting in features having both the frequency and time resolutions. These 2D signals are used as input to the capsule network for automatic feature learning. Fig. 1 shows the block diagram of the proposed EEGCAPS.

III. Automatic Feature Learning Using Capsule Network

In recent years, deep learning has taken feature engineering to perception and meta-structure engineering by automating feature learning. Deep learning approaches may work independently or in tandem with other feature selection methods. The latter is the main focus in this work. Leveraging the recent deep learning advances, in the proposed EEGCAPS method, deep capsule network [18] is applied to the spatially-filtered EEG signals for automatic feature learning. Unlike the convolutional neural network (CNN) [19], capsule network do not need pooling layers, resulting in the same amount of information to be processed at different layers. The capsule network is formed of a primary capsule layer to capture low-level features, followed by a secondary capsule layer to explore patterns in spatially-filtered EEG signals.

It is noted that in the capsule network, the weight optimization is through the routing-by-agreement algorithm [18], [20], where features are sent from a primary capsule (children) to a secondary capsule (parent) that better agrees with the features. Thus, after some iterations, parents' outputs may converge to predictions of some children and diverge from those of others. Parent capsules s_j are obtained as the weighted sum of predictions from all the children capsules, as given by $\mathbf{s}_j = \sum c_{ij} \hat{\mathbf{u}}_{j|i}$, where $\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij}\mathbf{u}_i$ predicts the output of the parent capsule and c_{ij} is a coupling coefficient given by a routing softmax, as given by

$$c_{ij} = \frac{\exp\left(b_{ij}\right)}{\sum_k \exp\left(b_{ik}\right)}, \tag{4}$$

where b_{ij} is the prior log probability for the routing, i.e., the prior probability that capsule i should connect to capsule j in a higher level. The weights $\mathbf{W}_{ij}$ between the ith primary to

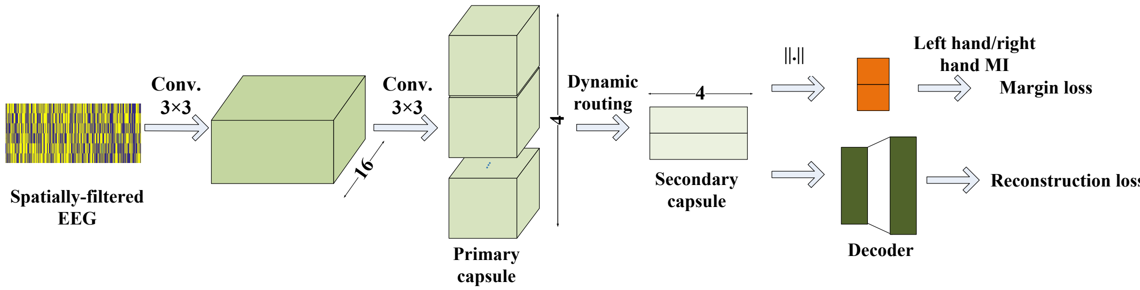

Fig. 2. Block diagram of the capsule network used in the proposed EEGCAPS method.

the jth secondary capsule filter the signal at different scales not only to detect independent features, but to identify a relationship between the learnt features.

In the output layer, feature channels are grouped by the capsule network through building a vector from the output of each neuron. The activation function used in the capsule network is a squashing function, as given by

$$\mathbf{z}_j = \frac{\left\|\mathbf{s}_j\right\|^2}{1+\left\|\mathbf{s}_j\right\|^2}\frac{\mathbf{s}_j}{\left\|\mathbf{s}_j\right\|} \tag{5}$$

The magnitude of the resulting vector is regarded as the probability of the presence of a capsule's entity, i.e., the patterns in the 2D spatially-filtered EEG signals. The prior log probabilities b_{ij} are then updated for a number of iterations as $b_{ij} = b_{ij} + \mathbf{z}_j \hat{\mathbf{u}}_{j|i}$. For the output layer, unlike CNN, which uses a softmax activation and cross-entropy loss function, the capsule network computes a margin loss for the secondary capsule layers as

$$\begin{aligned} L_r = {} & y_r \max\left(0, m^+ - \|\mathbf{z}_r\|\right)^2 \\ & + 0.5\left(1-y_r\right)\max\left(0, \|\mathbf{z}_r\| - m^-\right)^2 \end{aligned} \tag{6}$$

where y_r is the label for the class r. It is noted that the class r is predicted if $\|\mathbf{z}_r\| > \mu^+$, and is not predicted if $\|\mathbf{z}_r\| < \mu^-$. The parameters μ^+ and μ^- are experimentally found and set to 0.9 and 0.1, respectively [18]. The margin loss forces the class instances to be close to each other.

In addition, the capsule network uses a decoder for regularization, which is comprised of fully-connected layers. The loss value of this decoder is computed as $L = \left\|\mathbf{S} - \hat{\mathbf{S}}\right\|^2$, where $\mathbf{S}$ and $\hat{\mathbf{S}}$ are, respectively, the input signal and its reconstructed version. The total loss value of the capsule network is obtained as the weighted average of the margin loss and reconstruction loss. The binary cross-entropy loss function is used to measure the error. The error is then minimized through backpropagation using the Adam optimizer. The learning rate and batch size are tuned through grid searching. The number of epochs is set to 200 for all the subjects. The number of iterations used in the routing algorithm is set to 3.

In the following, a summary of the different layers of the proposed EEGCAPS is presented.

- The 2D spatially-filtered EEG signals of size 6×350 are used as input to the capsule model.
- The second layer is comprised of a convolutional layer having 16 filters of size 3×3 each with a stride of 1, resulting in 16 feature maps of size 4×348.
- The primary capsule layer includes 4 convolutional filters of size 3×3 each with a stride of 2. This layer has 4 capsules of dimension 4, leading to 64 feature maps each of size 2×346.
- The final capsule layer includes 2 capsules of length 4, i.e., class capsules, corresponding to right hand and right foot MIs.
- The decoder is composed of fully-connected layers having 64, 256 and 2100 neurons, respectively, where the input signal is reconstructed and the sum of squared differences between the input signal and its reconstructed version is minimized.

IV. Experimental Results

The experimental results are obtained using MI-BCI EEG signals taken from BCI competitions, i.e., dataset 4a, BCI competition III. The dataset is collected when the subjects perform MI tasks through imagination of limb movements. It is comprised of the EEG signals from five subjects, performing 280 trials of right hand and right foot MIs. This dataset includes the EEG recordings of 118 electrodes.

We first compare the performance of the EEGCAPS method to those using other classifiers such as linear discriminant analysis (LDA), support vector machine (SVM), decision tree (DT), multi-layer perceptron (MLP) and CNN, where 2D spatially-filtered EGG signals, obtained in Section II, are used as input to the classifiers. In the case of MLP, the signal is compressed into 200, 100 and 50 features using three hidden layers. In the case of CNN, two convolutional blocks having 4 and 8 filters of size 3×3, respectively, followed by batch normalization and a non-overlapping 2×2 max-pooling layer are used. The final feature map is passed through a global average pooling to reduce the dimensionality and avoid overfitting. For the sake of consistency, the number of epochs is set to 200 for MLP and CNN, and the learning rate is set to 0.001.

Table I gives cross-validated classification metrics of the proposed method as well as those of LDA, SVM, DT, MLP

TABLE I
ACCURACY, PRECISION AND SENSITIVITY (%) OBTAINED USING THE PROPOSED METHOD AS WELL AS THOSE OBTAINED USING SVM, DT, LDA, MLP AND CNN, ON SUBJECT *AW*.

	Accuracy	Precision	Sensitivity
TSCP+LDA	51.09$\mp$2.74	50.52$\mp$2.76	46.62$\mp$5.69
TCSP+SVM	55.93$\mp$2.35	55.70$\mp$2.05	60.40$\mp$3.37
TCSP+DT	73.25$\mp$3.12	73.57$\mp$3.21	72.36$\mp$3.41
TCSP+MLP	76.12$\mp$1.90	75.10$\mp$2.00	76.43$\mp$4.74
TCSP+CNN	90.87$\mp$1.50	90.60$\mp$3.21	91.78$\mp$2.35
EEGCAPS	**93.85$\mp$0.89**	**92.29$\mp$0.88**	**95.87$\mp$0.37**

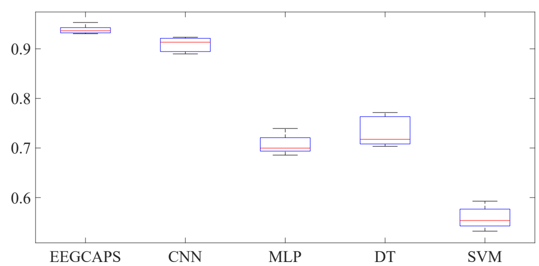

Fig. 3. Box plots for classification accuracy of various methods.

and CNN, for one of the subjects, i.e., *AW*. It is seen from this table that the proposed method provides a classification accuracy higher than those of the other methods. Similar results have also been observed for other subjects.

The statistical significance of the proposed method is also investigated. The p-value for a balanced one-way ANOVA between different methods, when comparing the classification accuracies, is smaller than 0.05 against LDA, SVM, DT and MLP, indicating the statistical significance of the proposed method. Yet, against CNN, p-value is equal to 0.21, indicating that the difference between the two methods is not statistically significant. Fig. 3 shows box plots of the classification accuracies for various methods. It is seen from this figure that the proposed method is more robust than the other methods, as observed from the small standard deviation value.

The classification accuracy of the proposed method is also compared to those obtained using other existing methods on the MI-BCI EEG signal classification. To this end and to evaluate the generalizability of the results, the dataset for each subject is split into five partitions, i.e., 5-fold cross validation. Tables II gives classification accuracies of the proposed method as well as those of other existing methods for various subjects. As observed in table II, the classification accuracy obtained using the proposed EEGCAPS method is higher than those yielded by the other methods. The superior performance of the proposed method in providing better class discrimination and higher recognition accuracies can be attributed to the fact that it takes into account both the spectral and temporal features, and to the reliance on hierarchical feature extraction using the capsule network. In particular, the proposed method outperforms its closest competitor, SBRCSP [11], with an improved

TABLE II
CLASSIFICATION ACCURACY OBTAINED USING THE PROPOSED METHOD AS WELL AS THOSE YIELDED BY THE OTHER METHODS.

	Dataset 4a, BCI III				
	AA	AL	AV	AW	AY
CSP [5]	66.07	96.43	43.47	71.88	49.60
R-CSP [2]	77.7	96.4	58.7	92.0	68.3
MSRCSP [21]	69.64	96.43	59.18	71.88	52.78
SBCSP [9]	83.03	98.21	52.04	89.05	58.33
FBCSP [10]	83.93	96.43	63.26	72.32	54.37
SSRCSP [6]	70.54	96.43	53.57	71.88	75.39
LRDS [8]	80.4	94.6	50.0	90.6	83.3
FBRCSP [11]	84.82	96.43	63.78	74.55	73.81
EEGCAPS	**85.50**	**97.52**	**62.15**	**94.70**	**83.57**

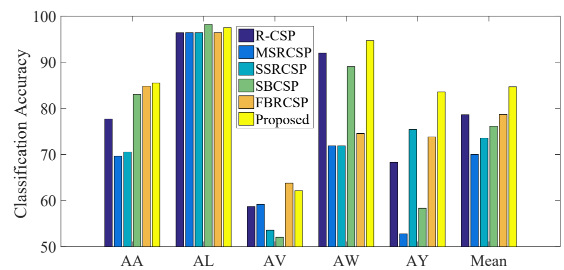

Fig. 4. Classification accuracy comparison of various methods.

classification accuracy of about 7.5% averaged over all the subjects. Fig. 4 shows EEG signal classification performance of the proposed method and those obtained using R-CSP [2], MSRCSP [21], SSRCSP [6], SBCSP [9] and FBRCSP [11], for different subjects in dataset 4a, BCI competition III, as well as their mean value. It is observed from this figure that the proposed method is superior to the other methods on average classification accuracy, especially when compared to the recent work in [11], where the proposed EEGCAPS method provides higher accuracies for all the subjects except for *AV*.

V. CONCLUSION

In this paper, a new method for MI-BCI EEG signal classification has been proposed based on spectral-temporal common spatial patterns and the capsule network. Discriminant spectral and temporal features are extracted from the EEG signals and projected onto image representations by preserving both the frequency and time resolutions of the EEG signals. These features are then fed into the capsule network for feature extraction. The recognition accuracy in the binary classification problem achieved by the proposed method was enhanced compared to some of the state-of-the-art methods. It is worthwhile to mention that the improved classification performance of the proposed method is thanks to the preservation of the temporal features of EEG signals, as well as, the automatic feature learning using the capsule network.

References

[1] G. Kalantar, H. Sadreazami, A. Mohammadi and A. Asif, "Adaptive dimensionality reduction method using graph-based spectral decomposition for motor imagery-based brain-computer interface," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP),* pp. 990-994, 2017.

[2] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting", *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936-2946, 2010.

[3] S.-H. Park, D. Lee and S.-G. Lee,"Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. on Neural Systems and Rehabilitation Engineering,* vol. 26, no. 2, pp. 498-505, 2018.

[4] T. Tanaka, T. Uehara and Y. Tanaka, "Dimensionality reduction of sample covariance matrices by graph Fourier transform for motor imagery brain-machine interface", *IEEE Statistical Signal Processing Workshop (SSP)*, pp. 1-5, 2016.

[5] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement", *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441-446, 2000.

[6] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering,* vol. 58, no. 2, pp. 355-362, 2011.

[7] S. Lemm, B. Blankertz, G. Curio and K.-R. Muller, "Spatio-spectral filters for improved classification of single trial EEG", *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541-1548, 2005.

[8] R. Tomioka and K. Aihara, "Classifying matrices with a spectral regularization," in *Proc. International Conference on Machine Learning,* pp. 895-902, 2007.

[9] Q. Novi, "Subband common spatial pattern for brain-computer interface", in *Proc. IEEE/EMBS International Conference on Neural Engineering,* pp. 204-207, 2007.

[10] K. K. Ang, Z. Y. Chin, H. Zhang and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE International Joint Conference on Neural Networks,* pp. 2390-2397, 2008.

[11] S.-H. Park and S.-G. Lee, "Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern," *IEEE Sensors Journal,* vol. 17, no. 10, pp. 2977-2983, 2017.

[12] S. Sakhavi, C. Guan and Sh. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Transactions on Biomedical Engineering,* vol. 29, no. 11 , pp. 5619-5629, 2018.

[13] P. Bashivan, I. Rish, M. Yeasin and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in *Proc. International Conference on Learning Representations (ICLR),* 2016.

[14] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco and B. Litt, "Modeling EEG waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *Journal of Neural Engineering,* vol. 8, no. 3, pp. 1-28, 2011.

[15] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner and V. D. Calhoun, "Deep learning for neuroimaging: a validation study," *Frontiers in Neuroscience,* vol. 8, pp. 1-11, 2014.

[16] N. Lu, T. Li, X. Ren and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Transctions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 6, pp. 566-576, 2017.

[17] H. Sadreazami and G. D. Mitsis, "Motor task learning in brain computer interfaces using time-dependent regularized common spatial patterns and residual networks", in *Proc. IEEE 18th International New Circuits and Systems Conference (NEWCAS),* pp. 190-193, 2020.

[18] S. Sabour, N. Frosst and G. E. Hinton, Dynamic routing between capsules; *Advances in Neural Information Processing Systems 30, Curran Associates, Inc.*, pp. 3856-3866, 2017.

[19] H. Sadreazami, M. Bolic and S. Rajan, "Contact-less fall detection using time-frequency analysis and deep neural network", *IEEE Transactions on Industrial Informatics,* 2021.

[20] H. Sadreazami, M. Bolic and S. Rajan, "CapsFall: Fall detection using ultra-wideband radar and capsule network", *IEEE Access,* vol. 7, pp. 55336-55343, 2019.

[21] X. Li and H. Wang, "Smooth spatial filter for common spatial patterns," in *Proc. International Conference on Neural Information Processing (ICONIP),* pp. 315-322, 2013.