STOW-Net: Spatio-Temporal Operation Based Deep Learning Network for Classifying Wavelet Transformed Motor Imagery EEG Signals

Md. Shohel Rana[a], *Shaikh Anowarul Fattah*[a,b] *and Mohammad Saquib*[c]
[a]*Institute of Robotics and Automation, BUET, Dhaka, Bangladesh*
[b]*Department of EEE, BUET, Dhaka, Bangladesh*
[c]*Department of EE, The University of Texas at Dallas, Texas, USA*
Email: rana.shohel10@yahoo.com, fattah@eee.buet.ac.bd, saquib@utdallas.edu

***Abstract*—Brain-computer interface (BCI) systems rely on capturing characteristics of human brain activity from the electroencephalography (EEG) signals, especially for the reliable classification of motor imagery tasks. For multi-channel EEG signals, it is crucial to precisely capture the spatio-temporal variation along with the frequency characteristics. Hence, instead of directly operating on raw EEG data, in this paper, discrete wavelet transform (DWT) is first applied to the motor-imagery multi-channel EEG data and then a deep learning architecture is designed incorporating spatial-temporal operations, which operates on the DWT-transformed EEG signal. In the proposed architecture, temporal convolution followed by spatial convolution is performed on the DWT-operated MI-EEG signal, and this part is termed as SAT-net. Next, by considering all channels together convolutional operation is performed to reduce the number of channels and this part is termed as SOC-net. Finally, a fully connected layer is used to classify the MI-EEG data from the derived feature vector. Extensive experimentation is performed on multiple subjects taken from the MI-based EEG dataset BCI Competition IV 2a. It is found that the proposed model offers a classification accuracy of 84.65%, consistently providing better classification performance than that obtained by some state-of-the-art methods.**

***Index Terms*—Discrete Wavelet Transform, Convolutional Neural Network, Motor Imagery, EEG signals**

I. INTRODUCTION

Electroencephalography (EEG) signals are most widely used in brain-computer interface (BCI) systems to classify motor imagery (MI) tasks [1]. Brain activity related to various MI tasks needs to be precisely extracted from the EEG signals through the BCI systems for controlling external devices [2], [3]. Machine-learning based methods utilize various time-frequency operations or decomposition algorithms to extract hand-crafted features from the MI-EEG signals [4], [5]. In this case, getting the best set of features is always very difficult and strongly depends on the choice of feature extraction process.

On the contrary, deep neural network based methods are getting popularity, as they do not need any sort of feature extraction stage. In this case, most of the proposed methods for classifying the MI tasks utilize a 2D image generated from the corresponding given 1D multi-channel EEG signals or their transformed (or decomposed) versions.

Indeed, end-to-end deep learning algorithms have gained significant popularity in recent years for their remarkable performance in classifying motor imagery (MI) tasks from EEG signals [6], [7], [8], [9]. In [6], deep learning model is proposed consisting of a wide-band and a narrow-band MI-EEG signal. Here, it is demonstrated that incorporating beta wave motor imagery task cues resulted in an improvement in classification accuracy for MI-EEG signals. In [7], a pre-trained CNN algorithm is proposed for MI-based BCI systems to achieve improved classification accuracy where continuous wavelet transform (CWT) is utilized to obtain image from the EEG signals. In [8], a deep CNN (DCNN) is employed for the classification of MI-task EEG signals belonging to two different classes where short-term Fourier transform and CWT are utilized. Since these methods use the transformed images, they may not be able to exploit the information available in the 1D time-series data of each channel. Moreover, dealing with 2D images is computationally expensive in comparison to the 1D counterpart.

In this paper, a deep learning based efficient algorithm is proposed to improve the classification accuracy of discrete wavelet transform (DWT) operated MI-EEG signals. Here DWT is first used to obtain time-frequency decomposition of the MI-EEG signals. A deep learning network based on spatial-temporal operations on the wavelet transformed data, namely STOW-Net is proposed. In order to extract spatio-temporal features, a SAT-Net is first designed that involves temporal convolution followed by spatial convolution on the DWT-operated MI-EEG signal. Next, spatial convolutional operation is used to reduce the channels (SOC-Net). Finally, the classification performance of the proposed method is investigated on a publicly available MI-EEG dataset for multiple subjects.

II. BCI COMPETITION IV 2A DATASET

The BCI Competition IV 2a dataset provided by the University of Graz is a popular dataset used for investigating MI-EEG signal classification performance [10]. A band-pass filter (0.5 Hz - 100 Hz) and a 50 Hz notch filter are used to process the signal before sampling at a rate of 250 Hz. Motor imagery data were collected from 9 subjects in two separate

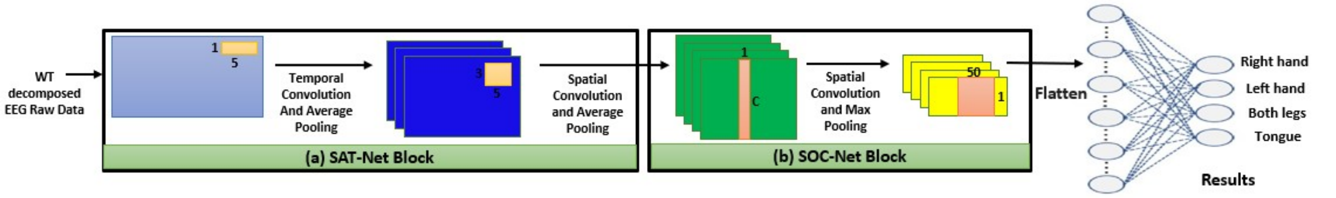

Fig. 1. Proposed STOW-Net Architecture: SAT-Net and SOC-Net

sessions conducted on different days. The dataset consists of four classes (left hand, right hand, foot and tongue), 12 trials per class, i.e. 48 total trials per session. In six sessions, there are total 288 trials for four classes.

III. PROPOSED METHODOLOGY

In this section, a deep learning method for MI task classification is proposed based on the DWT of MI-EEG signal. In the proposed network, instead of using the raw MI-EEG data, DWT-operated MI-EEG signals are utilized as the input. First, a spatial-temporal operation and then a spatial domain channel reduction is performed on the DWT-operated MI-EEG signals. As the proposed network performs spatial-temporal operations on the wavelet-transformed data, it is termed the STOW-Net.

A. Pre-processing

The dataset used in this study consists of 22 EEG channels with a sampling rate of 250 Hz. Here data is collected by applying a bandpass filter (0.5-100Hz). A duration of 4.5 seconds is taken for each recorded signal, which contains the motor imagery portion. Here 1125 samples are obtained for a sampling frequency of 250 Hz. For each trial a 2D representation of size 22×1125 is generated from the 22 EEG channels of data.

B. Discrete Wavelet Transform

Discrete wavelet transform is a multi-resolution time-frequency domain analysis technique that is suitable for extracting features of EEG signals of non-stationary nature [11], [12]. The DWT is widely used to decompose a given signal into low-pass and high-pass components, resulting in approximate and detailed coefficients, respectively. A single-stage filter bank architecture of the DWT is used, where the input signal $x(n)$ is passed through the low-pass filter $g(n)$ and the high-pass filter $h(n)$ and the corresponding output with a downsampling factor of 2.

C. Proposed STOW-Net Architecture

In this subsection, the proposed STOW-Net architecture is presented, which provides MI-tasks classification at the output by extracting the features from the DWT-operated MI-EEG signal. The STOW-Net architecture consists of spatial and temporal network (SAT-Net) block and spatially operated channel reduction network (SOC-Net) block. Basically the SAT-Net block is designed to extract temporal and spatial features from multi-channel wavelet transformed MI-EEG signals. The SOC-Net block is designed to extract the spatial features of one channel from multi-channel SAT-Net output feature vector. These blocks are discussed in the following subsections.

1) SAT-Net: A temporal-spatial convolutional network is designed to extract temporal and spatial features from the DWT decomposed EEG signal. As the network performs spatial and temporal operations, it is termed as the SAT-Net. In Fig. 1(a), the SAT-Net architecture is shown. Here at first a temporal convolution layer is used to extract the temporal characteristics of each channel. Next, an average pooling layer is added to reduce the size. Next, the output of the average pooling layer is fed to a spatial convolution layer. In the SAT-Net, temporal features of the DWT transformed input signal are extracted by twenty five filters with a kernel size of (1×5) in a convolutional layer. Next, batch-normalization and exponential linear units (ELU) are employed to the output of the temporal convolution layer. After using the temporal convolutional layer, an average-pooling layer is used to reduce the length of its outputs. Here the kernel size (1×2) and the stride size (1×2) are used. Next, the spatial convolutional layer is applied on the output of the average-pooling layer to extract the spatial features. Here kernel size of (3×5) and the padding size of (1, 0) are used along with 50 filters. Also batch-normalization and ELU operations are used after the spatial convolutional network block.

2) SOC-Net: Multi-channel to single-channel spatial convolutional network is applied to extract spatial features from the output of the SAT-Net stage. This spatially operated channel reduction convolutional network is collectively called SOC-Net. In Fig. 1(b), the SOC-Net architecture is shown. Here an average pooling layer is added after the spatial convolution layer. In the SOC-Net, an average-pooling layer is first used to reduce the length of the SAT-Net output. Here kernel size and stride size are kept the same as those used in the previous average pooling stage. Next, the multi-channel to single-channel spatial convolutional network layer is applied on the output of the average-pooling layer in this SOC-Net. Here the number of channels in the input signal is equal to the number of kernels in the spatial convolutional layer. In this operation, 50 filters are used to extract the temporal properties of each channel. Next, batch-normalization and exponential linear units (ELU) are employed to the output of the spatial convolution layer.

After using the SOC-Net convolutional network, a max-

TABLE I
LAYERS OF THE PROPOSED STOW-NET MODEL

Blocks	Layer	Filter	Kernel Size	STOW Output Shape	Options
-	Input	-	-	[4]	-
WT	dwt	-	-	[1, 22, 563]	-
SAT-Net	Conv2d	25	(1x5)	[25, 22, 559]	BatchNorm, ELU
-	AvgPool2d	-	(1x2), stride(1x2)	[25, 22, 279]	-
-	Conv2d	50	(3x5), padding(1x0)	[50, 22, 275]	BatchNorm, ELU
SAC-Net	AvgPool2d	-	(1x2), stride(1x2)	[50, 22, 137]	-
-	Conv2d	50	(22x1)	[50, 1, 137]	BatchNorm, ELU
-	MaxPool	-	(1x50), stride(1x5)	[50, 1, 18]	Dropout
-	Flatten	-	-	[900]	-
-	Linear	-	-	[4]	-

pooling layer is used to reduce the length of the output of multiple single-channel filters. Here the minimum number of output with the highest quality feature is selected. The kernel size of the max-pooling module is kept (1×50) and the stride size is chosen as (1×5). Finally, a flattening operation is performed to obtain the desired feature vector. A detailed step-by-step description of the proposed architecture is presented in Table I.

IV. EXPERIMENTAL RESULTS

In this section, a publicly available dataset is considered to analyze the performance of the proposed STOW-Net method. The dataset contains nine healthy subjects to calculate the classification accuracy. There are training and testing datasets separately available for each subject. For the purpose of achieving high training accuracy, 200 epochs are used in the training process. Here the results obtained from the dataset IV 2a using the proposed method are presented, which are compared with the results obtained by some state-of-the-art methods. An analysis considering different types of wavelet transformed MI-EEG signals is also presented for the proposed method. Moreover, an ablation study on two major steps involved in the proposed method is also presented.

TABLE II
CLASSIFICATION PERFORMANCE COMPARISON AMONG DIFFERENT METHODS

Subject	RME [13]	MEMDBFC [14]	DBCNN [6]	CWT-SCNN [15]	2L-CNN [16]	**STOW-Net**
01	82.00	90.28	99.44	74.70	82.14	**97.92**
02	86.00	65.28	75.69	81.30	75.05	**68.75**
03	84.00	93.75	86.11	68.10	90.10	**87.50**
04	78.00	74.31	84.72	96.30	94.61	**85.42**
05	79.00	68.06	60.42	92.50	92.30	**63.89**
06	86.00	78.47	74.41	86.90	86.75	**78.47**
07	75.00	79.86	88.19	73.40	76.60	**97.22**
08	80.00	97.22	84.72	91.60	80.50	**86.11**
09	75.00	93.75	94.44	84.40	79.55	**96.53**
Ave.	77.33	82.33	82.56	83.20	84.18	**84.65**
Std	11.59	11.86	10.22	12.09	16.64	**11.58**

In Table II, the classification accuracies achieved by different state-of-the-art methods on the same dataset are reported separately. To compare the proposed method in terms of subject-wise accuracy, reported overall average accuracy and standard deviation, we consider five methods. In [13], wavelet packet transform (WPT) is used to extract wavelet domain features and then various machine learning algorithms are applied on the selected features for EEG signal classification of motor imagery. Here the Rényi min-entropy (RME) method is applied for feature selection and the random forest classifier is found as the best performing method. In the DBCNN technique proposed in [6], signals from two band-limited MI-EEG signals are used as input raw data. Next, a deep learning network is used to extract the features from the MI-EEG data. In [14], the multivariate empirical mode decomposition (EMD) is used to decompose the EEG signal along with the common spatial pattern. Next, linear discriminant analysis (LDA) and support vector machine (SVM) classifiers are applied (two-class problem). A method combining CWT and a simplified CNN, namely CWT-SCNN, is proposed in [15]. In the CWT-SCNN method, the CWT is first used to decompose MI-EEG data and then the CNN is used to extract the features. A 2-layer CNN (2L-CNN) model is proposed in [16], where wavelet transform based 2D image is formed from the raw EEG data extracted from C3, Cz and C4 channels and a 2-layer CNN is applied as a classifier for 2D image classification.

TABLE III
EXPERIMENTAL RESULTS USING DIFFERENT DWT DECOMPOSED SIGNALS

Subject	DWT High-freq.	DWT Concatenated	**DWT Low-freq. Sig.**
S01	93.75	96.53	**97.92**
S02	70.14	65.97	**68.75**
S03	94.44	93.75	**87.50**
S04	87.50	84.03	**85.42**
S05	70.14	69.44	**63.89**
S06	74.31	78.47	**78.47**
S07	79.17	93.75	**97.22**
S08	82.64	79.86	**86.11**
S09	97.22	95.14	**96.53**
Ave.	83.26	84.10	**84.65**

TABLE IV
EFFECT OF DIFFERENT BLOCKS OF THE PROPOSED STOW-NET

Input Raw	Raw to DWT data	SAT-Net	SOC-Net	Avg.
✓	✓	×	✓	82.87
✓	×	✓	✓	83.79
✓	✓	✓	✓	**84.65**

It is observed from the Table II that the proposed method provides a very satisfactory classification accuracy with the lowest standard deviation compared to other methods. In particular, the use of wavelet decomposition based STOW-Net method results in consistently high accuracy with an average of 84.65% and standard deviation of 11.58 over nine subjects. For example, with respect to the average accuracy reported in [15], STOW-Net provides 2.32% higher accuracy.

In this part, our objective is to analyze the variation in classification performances in three cases based on the DWT operation: when only high-frequency (detail) DWT coefficients are used, when only low-frequency (approximate) DWT coefficients are used and the case when the features obtained from the above two cases are concatenated. Experimental results obtained for these three cases are shown in Table III. From these experimental results, it can be seen that the classification results using high-frequency wavelet decomposed signals or the concatenated features are slightly lower compared to that obtained for the low-frequency decomposed signals. The sampling frequency used is 250 Hz and thus the low frequency (less than 62.5 Hz) wavelet coefficients contain the major brain activities related to the motor-imagery tasks. Hence, the low-frequency DWT coefficients provide better results.

An ablation study is conducted on the proposed method to understand the effect of DWT operation and SAT-net on the overall classification performance. The results obtained for different cases where SAT-net and DWT are not used are shown in Table IV. From the table it is clearly observed that the use of the SAT-net module increases the classification accuracy significantly (2.22% improvement). It is because of the use of temporal convolution which extracts the temporal characteristics of individual channels and spatial convolution which considers inter-channel relationships in the small vicinity of a particular channel. Using wavelet decomposition data increases the classification accuracy by 0.86%. It is to be noted that in this case, the proposed SAT-net plays an important role even when the wavelet decomposition is not applied and helps offer satisfactory results.

V. Conclusion

In this paper, a deep-learning framework for MI-task classification from multi-channel EEG data is presented. In our proposed STOW-Net model, first discrete wavelet transform (DWT) is employed to decompose the MI-EEG signals into time-frequency band-limited signals. Then the DWT operated MI-EEG signals are passed through the proposed SOW-net architecture. In the SOW-net, SAT-Net and SOC-Net are used to extract features of motor imagery EEG tasks. Here SAT-Net explores the temporal convolution and spatial convolution on the decomposed band-limited MI-EEG signals and SAC-Net reduces the number of channels by using the is to extract spatial features from SAT-Net feature vectors for a single channel. Improved classification performance is achieved by using this band-limited decomposition of MI-EEG signals for motor imagery tasks.

It is clearly observed that the use of SAT-net improves the classification performance due to its capability of extracting temporal features (within an individual channel, temporal variation) as well as spatial features (inter-channel relationship in the small neighborhood). It is also found that the use of discrete wavelet transformed EEG data offers better performance, especially when low-frequency decomposed EEG data are used. Moreover, the incorporation of the SOC-net offers a channel reduction at the last stage via spatial convolution (inter-channel relationship among all channels at a time). Finally, the simulation results on the widely used public dataset BCI competition IV 2a exhibits a consistently better performance in comparison to some existing methods. The subject-wise classification performances are also very satisfactory for all subjects.

References

[1] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[2] P. Nuyujukian, J. M. Fan, J. C. Kao, S. I. Ryu, and K. V. Shenoy, "A high-performance keyboard neural prosthesis enabled by task optimization," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 1, pp. 21–29, 2014.

[3] F. Lotte, L. Bougrain, and M. Clerc, "Electroencephalography (EEG)-based brain-computer interfaces," 2015.

[4] S. Chaudhary, S. Taran, V. Bajaj, and S. Siuly, "A flexible analytic wavelet transform based approach for motor-imagery tasks classification in bci applications," *Computer methods and programs in biomedicine*, vol. 187, p. 105325, 2020.

[5] S. Phadikar, N. Sinha, and R. Ghosh, "Unsupervised feature extraction with autoencoders for eeg based multiclass motor imagery bci," *Expert Systems with Applications*, vol. 213, p. 118901, 2023.

[6] M. S. Rana and S. Anowarul, "An efficient dual-band cnn for motor imagery eeg signal classification," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, 2022, pp. 1127–1130.

[7] M. T. Sadiq, M. Z. Aziz, A. Almogren, A. Yousaf, S. Siuly, and A. U. Rehman, "Exploiting pretrained cnn models for the development of an eeg-based robust bci framework," *Computers in Biology and Medicine*, vol. 143, p. 105242, 2022.

[8] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional neural network based approach towards motor imagery tasks eeg signals classification," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4494–4500, 2019.

[9] H. K. Lee and Y.-S. Choi, "Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface," *Entropy*, vol. 21, no. 12, p. 1199, 2019.

[10] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–graz data set a," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[11] D. Rosca, "Wavelet bases on the sphere obtained by radial projection," *Journal of Fourier Analysis and Applications*, vol. 13, no. 4, pp. 421–434, 2007.

[12] Y. Wiaux, J. D. McEwen, P. Vandergheynst, and O. Blanc, "Exact reconstruction with directional wavelets on the sphere," *Monthly Notices of the Royal Astronomical Society*, vol. 388, no. 2, pp. 770–788, 2008.

[13] M. Rahman, F. Khanam, M. Ahmad, M. S. Uddin *et al.*, "Multiclass eeg signal classification utilizing rényi min-entropy-based feature selection from wavelet packet transformation," *Brain informatics*, vol. 7, no. 1, pp. 1–11, 2020.

[14] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "An automatic subject specific intrinsic mode function selection for enhancing two-class eeg-based motor imagery-brain computer interface," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6938–6947, 2019.

[15] F. Li, F. He, F. Wang, D. Zhang, Y. Xia, and X. Li, "A novel simplified convolutional neural network classification algorithm of motor imagery eeg signals based on deep learning," *Applied Sciences*, vol. 10, no. 5, p. 1605, 2020.

[16] B. Xu, L. Zhang, A. Song, C. Wu, W. Li, D. Zhang, G. Xu, H. Li, and H. Zeng, "Wavelet transform time-frequency image and convolutional network-based motor imagery eeg classification," *Ieee Access*, vol. 7, pp. 6084–6093, 2018.