Motor Imagery Classification for Brain Computer Interface using TSGL-EEGNet

Rajesh Bhambare[1],Manish Jain[2]
[1, 2] Department of Electrical & Electronics Engineering, Mandsaur University, Mandsaur, Madhya Pradesh, India.
[1]bhambare.rajesh@gmail.com, [2] manish.jain@meu.edu.in

Abstract

In spite of the fact that the accuracy of Motor Imagery (MI) Brain-Computer Interface (BCI) systems based on deep learning has been significantly increased in comparison with certain standard algorithms, it is still a significant challenge to comprehend the deep learning models accurately. This paper addresses the concerns by presenting a well-known deep learning model known as EEGNet, then contrasting that model with a more conventional approach known as Filter-Bank Common Spatial Pattern (FBCSP). After that, this work considers that a unique Discrete Wavelet Transform (DWT) can explain the 1-D convolution of EEGNet and that the depthwise convolution of EEGNet is comparable to the Common Spatial Pattern algorithm. In addition, the EEGNet has been made more effective by using the technique known as the Temporary Constrained Sparse Group Lasso (TSGL) developed for this study. The suggested model, a modified version of TSGL-EEGNet, was evaluated using the BCI Competition IV 2a dataset, consisting of MI tasks involving four classifications. The testing results indicate that the proposed model has achieved an average classification accuracy of 82.05 per cent on the dataset BCI Competition IV 2a. These results are higher than those achieved by TSGL-EEGNet, EEGNet, C2CM, MB3DCNN, SS-MEMDBF, using FBCSP, particularly on insensitive subjects. . Also, the testing results show 0.78 kappa value for subject dependent classification.

Index TermsBCI, CNN, Motor Imagery, Temporary constrained spares group lasso

DOINumber:10.14704/nq.2022.20.8.NQ44171 **NeuroQuantology2022;20(8):1582-1591**

INTRODUCTION

One of the most significant difficulties for the Brain Computer Interface (BCI) system is effectively decoding the Motor Imagery (MI) Electroencephalograph (EEG). In general, the BCI system consists of three components. The first is signal processing and data enrichment. The second step is feature extraction, which includes selecting the proper features and synthesis, and the final step is categorization and detection. In general, the first part is signal pre-purification, whereas the next two are related to signal decoding [1] [8]. The second operation is also known as detection, and the final procedure is referred to as classification. In comparison with classic machine learning algorithms, MI EEG based Brain Computer Interface systems principally inspect sensorimotor rhythm (SMR). The SMR is an oscillating pattern in electrical brain signals that initiates in brain areas concerned with voluntary movement planning, control, and execution [9], [10]. Event-related

*e*ISSN1303-5150

synchronization (ERS) is related to an enhanced activity in a specific frequency band, while event-related desynchronization (ERD) refers to a decrement in activity within a particular frequency range [9]. Motor imagery, motor action, and sensory input may cause the ERSs and ERDs to fire [11], [12]. The α frequency (8-12 Hz) obtained from inside the sensorimotor section of the cerebral cortex is referred to as μ frequency [4], [13]. The μ and β (12-30 Hz) frequency alterations are utilized to distinguish the task in process for the left-hand and right-hand categorization of motor imagery tasks. The γ frequency is utilized in invasive MI BCI. However, it is occasionally efficiently adapted in scalp EEG. Common class types for the multi-object MI BCI system consist of left hand, right hand, foot, and tongue movements [14], [16]. These incidences generate significant distinct changes in the EEG signal compared to the background EEG signal. Besides, foot movements are frequently merged, with no difference between left and right foot motions and finger movements. The cortical portions related with these various movements are too small to create discrete ERD and ERS signals [9]. However, as per our knowledge, just one study [17] reveals that β rhythm may be utilized to differentiate visual signals of variations between the left and right foot activity.

It is difficult to analyse EEG signals as they are continuously changing, easily affected due to external noise, and prone to signal changes. Furthermore, EEG extraction is challenging, the signal-to-noise ratio (SNR) is poor, and the size of the EEG dataset is many times moderate. There have been many studies on MI BCI in recent years. Filter Bank Common Spatial Pattern (FBCSP) is a traditional method proposed by Ang et al. [18]. It should be emphasized that the FBCSP has a mathematical basis and is easy to comprehend. It was the winner of the BCI Competition IV 2a/b dataset championship in 2008; however, it is proved insufficient to give results in reality. W.Burgard et al. [6] recommended (CNN) convolutional neural network-based Deep Conv. and Shallow Conv. (CNN). The Deep Conv. and Shallow Conv. extract useful information using one-dimensional (1-D) convolution, and they investigate the convolution structure and the efficacy of deep learning in the MI BCI system. EEGNet, a global deep learning framework for EEG tasks, was suggested by A.J. Solon et al. [3]. The EEGNet is a common Deep Learning model for numerous EEG paradigms that improves on the Deep and Shallow Convolution and gives improved interpretation accuracy and lower training time. J. Ren et al. [7] classified MI EEG data using a multi-branch 3D CNN approach. Their approach was similar to the widely used feature pyramid networks for object identification in videos. However, as EEG signals are time-varying and of continuously varying nature, the 3DCNN unable to perform similar to the object detection task. K.W. Ha et al. [8] created a MI Brain Computer Interface system using CapsuleNet. Some conventional algorithms, in general, offer strong interpretability and rapid computing rates. Although neural network methods often achieve higher accuracy rates, training times are sluggish. It isn't easy to comprehend how these neural networks acquire their outcomes. As a result, the primary goal of this study is to bring the neural network method's interpretability closer to that of the old approach while maintaining or even improving its accuracy.

The regularisation technique is used in this study to present the Temporal-constrained Group Lasso EEGNet (TSGL-EEGNet) algorithm, which is a CNN -based solution for the motor imagery Brain Computer Interface system. Furthermore, this technique is built on the FBCSP and EEGNet and enhances both using the TCSGL in [19],[21]. The main contribution of the

eISSN1303-5150

approach described in this analysis is about merging the conventional approach and deep learning methods with MI BCI techniques and to explore the interpretation ability of neural network techniques in the BCI field. The remainder of the paper is organized as follows. TheSection II gives insight on the dataset and pre-processing technique. Section III provides an overview of comparable work and the mathematical foundation of proposed TSGL-EEGNet. Section IV thoroughly assesses our model's performance and compares with other methods.

DATA

In this study we have referred public dataset the BCI Competition IV 2a dataset (2008) [25], which includes four categories of motor imagery samples from Nine participants' left hand, right hand, foot, and tongue motions. The data was collected using Twenty Two Ag/AgCl electrodes, captured at Two Hundred Fifty Hz, bandpass filtered in the range of 0.5 and 100 Hz, and 50 Hz notch filter is applied after it. One data trial lasted 4 seconds and was gathered following a 2-second fixation cross. There is no necessity of any type of specific data preprocessing methods for the EEG data given in dataset. The several pre-processing methods, such as standardization, filtering, and the technique of removing artifact signals, appear from previous research [3], [6], [7].

METHODOLOGY

A. Filter Bank Common Spatial Pattern

The FBCSP method is the outstanding conventional algorithm, and its former, the CSP algorithm, is a data-based algorithm that learns a spatial filter to optimize the 2 types of variance while training of data during classification. The spatial filter is primarily concerned with the EEG data channel because the left and right-hand motor imagery responses have distinct channel responses. In two class categorisation CSP algorithm gives good results. The FBCSP gives the frequency domain characteristic to the CSP algorithm, which has been tested for a prolonged time period. The FBCSP algorithm's core is explained as follows [5], [18], [23]: responses; hence the CSP method performs better in 2-category classification problems. The FBCSP adds the frequency domain characteristic to the CSP algorithm, which has been tested for an extended period. The FBCSP algorithm's core is interpreted as follows [5, 18, and 23]:

1) The EEG signal data is filtered with 9 filter banks of bandwidth 4 Hz. The frequencies of filter banks are of the order of 4-8 Hz, 8-12 Hz up to 32-36 Hz.
2) Filtering process divides the data into 9 bands. The Common Spatial Pattern (CSP) method is used to compute the data on each band. This is accomplished by maximizing the objective function, denoted by Eq. (1)

$$W^* = \underset{W}{\operatorname{argmax}} \frac{w^T \sum_{c1} w}{w^T \sum_{c1} + w^T \sum_{c1} w} \quad (1)$$

Where $\sum_{C1}$ and $\sum_{C2}$ are related with respective channel covariance matrices. W is the spatial filter. The analytical solution to this objective function, commonly known as the Rayleigh Quotient, is analogous to solving a generalized eigenvalue decomposition problem. If x represents the sample taken, i gives class number and j is corresponds to frequency bands ordinal, then it gives the feature F by following equation (2):

$$F = w\, x_{ij} \quad (2)$$

3) Spatial filters are chosen that correlate to the *2 X NW* extreme eigenvalues (NW most significant and *NW* lowest eigenvalues). Each

extreme spatial filter is coupled with its appropriate counterpart (spatially filtered channel pairs).

4) The energy (variance) of the spatially filtered channels (EC) is determined and normalised to the total energy of the channels in a specific frequency band. The energy logarithm calculated generates the final features.
5) All nine filter bands features are concatenated, and a mutual information-based feature selection is made on *2X NW X 9* spatially filtered channels, where *NS* filtered channels, and their pairings are selected. Depending on whether or not the selected characteristics already form pairs, a maximum of *2 X NS* features may be chosen.
6) Since CSP is created for a two-class issue, a one-versus-rest or one-versus-one method must be used for multiclass problems. In FBCSP, the first option is selected, which results in a maximum of *class* X 2 X *NS* features.
7) SVM and other maximum interval classification methods are used to categorise the features successfully.

B. EEGNET

EEGNet imitates FBCSP's feature engineering in some manner. It includes a 1-D convolution, a depth-wise convolution, and a separable convolution. The final classification is accomplished using the complete connection layer and the Softmax function.

The structure of EEGNet is shown in Fig.1 as consisting of three parts: feature extraction, feature selection, and classification. There are various grounds to believe that it imitates the engineering of the FBCSP's features. The EEGNet comprises a 1-D convolution and a depth-wise convolution for feature extraction. Calculating the similarity between the convolution and the convolution kernel is the core of convolution. This resembles the DWT method, and the convolution kernel resembles a fixed-scale wavelet basis function. The filter bank may implement the DWT. Therefore, 1-D convolution without bias is a signal filter bank, and the convolution kernel is the signal filter. Consequently, when a 1-D convolution kernel convolves along the time dimension for a single channel, it may be seen as frequency filtering. The depth-wise convolution convolutes the data on the dimension of the channel to produce a collection of spatial Filters. Each spatial filter represents a linear mapping of all channels to a single feature. The whole process is comparable to the FBCSP algorithm.

The following equation describes EEGNet's feature extraction:

$$F_{i,j} = w(f_j x_i) = w\, x_{i,j} \qquad (3)$$

Here ***F*** represents the features, ***x*** represents the samples, ***w*** represents the spatial filters, and f represents the frequency filters ***i*** and ***j*** denote, respectively, the category number and the frequency filter ordinal. In general, it can be seen that "(2)" is identical to "(3)", with the latter possessing more variable parameters and a greater number of degrees of freedom. The feature learning portion of the EEGNet comprises a filter bank (1-D convolution) and a spatial filter (depth-wise convolution), which constitutes the major part of the FBCSP.

At feature selection, time-spatial features in various frequencies are convoluted by ***1X16*** depth-wise convolutions. First, they are combined using ***1 X 1*** point-wise convolution, which is incomprehensible. In reality, the separable convolution is not selecting features but combining them with varying weights to simulate selection. In practical applications, the feature selection portion enhances the EEGNet's resilience and decoding and classification precision.

Therefore, feature selection is a crucial component in CSP algorithms. The commonly employed feature selection methods in MI BCI [23] include the Mutual Information based Best Individual Feature (MIBIF), the Mutual Information based Nave Bayesian Parzen Window (MINBPW), the Mutual Information based feature selection (MIFS), the Fuzzy-Rough set-based Feature Selection (FRFS), and the Mutual Information Rough Set Reduction (MIRSR). In addition, Zhang et al. [19] has recently used theregularization algorithms as feature selection techniques. Choosing the proper feature selection approach is an effective way to enhance the model's performance.

In the last phase of categorization, the input covers time, frequency, and spatial domains. The time dimension generates the time-domain characteristics. The frequency-domain characteristics are obtained using 1-D and separable convolution. The spatial domain characteristics are obtained by depth-wise convolution of frequency-domain characteristics. After selecting all of these characteristics, four probabilities are generated via a fully connected layer using the Softmax activation function to determine the category. In conclusion, the fundamental concepts of the EEGNet and FBCSP algorithms are identical. Hence the

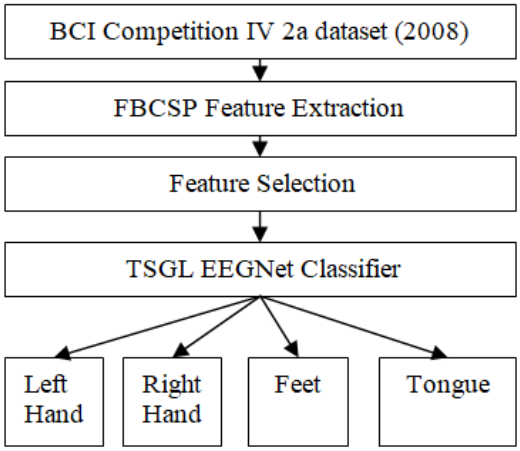

Fig.1 The Structure of TSGL-EEGNet

EEGNet is a deep learning algorithm derived from the traditional approach.

C. TSGL EEGNet

In the realm of BCI deep learning technology, EEGNet is a complex method with a straightforward explanation. However, it is still insufficient since its feature selection component is difficult to comprehend and expanding the feature space would result in over fitting. Keeping the main body of EEGNet untouched, enhancing the interpretability of the feature selection section, and minimizing the number of examples of over fitting would be a solid starting point for resolving the issue. Regularization is an effective strategy for resolving over-fitting issues and has strong interpretability. In addition, regularisation is one of the ways of selecting features. This research offers a regularization-based Temporal-constrained Sparse Group Lasso EEGNet (TSGL-EEGNet). The TSGL-EEGNet architecture is shown in Fig.1 and Table 1. The primary distinction from EEGNet is the Features section. Contrary to the EEGNet's separable convolution, only 2-D convolution is used here to improve interpretability. The TSGL penalty is introduced to this convolution, causing the layer's weight matrix to alter

*e*ISSN1303-5150

throughout training in the direction of penalty reduction.

TSGL regularization is seen in Fig.2. The black squares indicate parameter inhibition; the white squares denote parameter activation, and the grey squares represent parameter activation with small weights. The Group Lasso (GL) can inhibit groups corresponding to outgoing vectors. The Sparse Group Lasso (SGL) may suppress some active group parameters. Temporal-constrained Lasso (TL) can maintain a smooth temporal domain. The GL will inhibit or activate an entire group when it is impacted, the SGL will suppress specific parameters of the active groups depending on the GL, and the TL will reduce or raise specific parameters to maintain a modest difference between groups in the temporal domain. The TL and SGL make up the TSGL regularisation, as seen in Fig.3. To mathematically describe these Lassos, the following symbols are defined. In the N classification problem, the chosen feature may be written as ***wF*** for a single sample, where the feature produced by the feature learning component is a 2-D matrix denoted F and the 1-D convolution weight matrix is a 3-D matrix denoted ***w.***

Temporary Constraint aims to maintain the smoothness of the time domain characteristics and decrease the distortion induced by other regularisation techniques to be closer to the actual EEG signals. Temporary Constraint utilises the characteristics of the later period minus the characteristics of the earlier time. When the temporal domain is sufficiently

Table1: Structure of Modified TSGL EEGNet

Layer (Type)	Output Shape	Parameters
Input Layer	(None, 64, 128, 1)	0
Convolution 2D	(None, 64, 128, 8)	512
Batch Normalization	(None, 64, 128, 8)	32
Depthwise Convolution 2D	(None, 1, 128, 16)	1024
Batch Normalization	(None, 1, 128, 16)	64
Activation (elu)	(None, 1, 128, 16)	0
Average Pooling2D	(None, 1, 32, 16)	0
Dropout	(None, 1, 32, 16)	0
Batch Normalization	(None, 1, 32, 16)	64
Activation (elu)	(None, 1, 32, 16)	0
Average Pooling2D	(None, 1, 4, 16)	0
Dropout	(None, 1, 4, 16)	0
Flatten	(None, 64)	0
Dense	(None, 4)	260
Softmax (Activation)	(None, 4)	0

smooth, the temporary loss should thus be near 0. Assuming that EEG data contains T time steps, the characteristics of 1 to T – 1 time steps are ***wF1*** and those of 2 to T time steps are ***wF2***; therefore, the Temporal-constrained Lasso may be represented as $\|wF2 - wF1\|_1$.

The Sparse Group Lasso, comprising a Group Lasso and an L1 Norm, may create sparse groups. In the Sparse Group Lasso

allows the active group components to have sparsity as well. In this study, the frequency and spatial domain features are gathered to pick a suitable frequency and spatial features, and the time domain features of these may also be chosen using L1Norm, which can be represented as

$$\|w\|2,1 + \|w\|1\cdot\|w\|2,1 = \sum_{g\in w}\sqrt{|g|}\,\|g\|^2$$

$$= \sum_{g\in w}\sqrt{\sum g^2\,|g|}, \quad \|w\|_1 = \sum_{w}|w|,$$

Where 'g' is a group vector that is the length of a dimension of ***w. |g|.*** In this study, the features estimated by the feature selection convolution are treated as a group, with each feature represented as an element of the group vector. Consequently, |g| is the length of the dimension that affects the number of output features in the matrix ***w***.

By solving the following equation, the final weight matrix ***w**** is produced. Eq. (4):

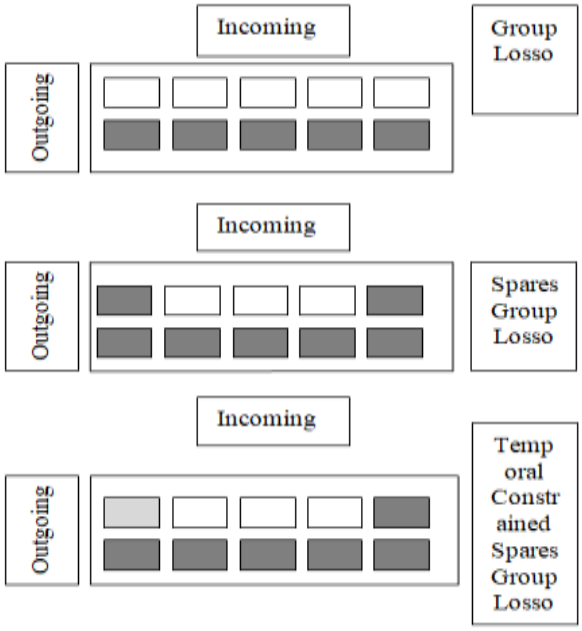

Fig.2 The GL, SGL and TSGL regularization.

The black squares correspond to the suppression of parameters, and the gray squares denote the activation related with tiny weights. Group Losso suppresses groups corresponding to outgoing vectors. Spares Group Losso suppresses few parameters from activated groups. Temporal-constrained Lasso makes the temporal domain smooth.

$$W^* = \underset{w,a,b}{argmin} -\frac{1}{n}\sum_{x}\sum_{i=1}^{N}[Softmax(awF+b)iy_i$$

+ (1 – ***yi***) ln(1- Softmax (a^{wF} + b)$_i$)] +

$$\frac{\beta_1}{n}\|w\|_{2,1} + \frac{\beta_{12}}{n}\|w\|_1 + \frac{\beta_3}{n}\|wF2 - wF1\|_1 \quad (4)$$

where β_1, β_2, β_3 are the regularisation coefficients, x represents the sample, and y represents the ground truth. . Here a, b represent the weight and bias of the Layer FC, respectively. n represents the number of samples. Let σ represent the activation function (here, Softmax) and z represents and ***awF+b;*** the loss function C of the issue may be stated as follows: The gradient descent technique is used to generate an approximation of the solution to the problem of loss function. Once the final weight matrix ***W**** has been obtained the selected features can be estimated by Eq. (3).

RESULTS AND DISCUSSION

The 5-fold average-validation training using the regularisation coefficients chosen by grid search is compared to various baseline models. This study proposes the modified TSGL-EEGNet, Table 2 provides the findings. Our design achieves an average accuracy of 82.05 per cent and an average kappa of 0.7806, which is greater than previous methods. Comparing the proposed TSGL-EEGNet to the EEGNet reveals that it has successfully minimised over-fitting and enhanced accuracy. Comparing the modified TSGL-EEGNet against other neural network approaches such

Table 2: Accuracy (kappa) results of various models using dataset BC IV 2a. The TSGL-EEGNet and EEGNet results are with Five Fold average-validation, and the remaining are from given references.

Subject	Proposed TSGL-EEGNet	TSGL-EEGNet [24]	EEGNet [24]	SS-MEMDBF [21]	MB3DCNN [7]	C2CM [5]	TSMM + SVM [22]	FBCSP [23]	FBCSP [20]
1	**86.48 (0.8001)**	83.77 (0.7836)	84.52 (0.7936)	(0.86)	77.40 (0.699)	87.50 (0.833)	(0.77)	76.00	(0.68)
2	73.22 (0.7658)	70.18 (0.6023)	61.17 (0.4821)	(0.24)	60.14 (0.459)	65.28 (0.537)	(0.33)	56.50	(0.42)
3	**97.39 (0.9393)**	94.36 (0.9248)	95.90 (0.9453)	(0.70)	82.93 (0.788)	90.28 (0.870)	(0.77)	81.25	(0.75)
4	78.91 (0.7864)	75.88 (0.6777)	66.58 (0.5532)	(0.68)	72.29 (0.594)	66.67 (0.556)	(0.51)	61.00	(0.48)
5	67.95 (0.7296)	64.35 (0.5249)	60.22 (0.4695)	(0.36)	75.84 (0.647)	75.84 (0.647)	(0.35)	55.00	(0.40)
6	68.98 (0.7495)	65.67 (0.5421)	56.65 (0.4217)	(0.34)	68.99 (0.538)	68.99 (0.538)	(0.36)	45.25	(0.27)
7	**91.95 (0.7254)**	88.95 (0.8528)	85.78 (0.8104)	(0.66)	76.04 (0.653)	76.04 (0.653)	(0.71)	82.75	(0.77)
8	86.97 (0.7394)	83.84 (0.7845)	84.83 (0.7978)	(0.75)	76.86 (0.702)	76.86 (0.702)	(0.72)	81.75	(0.75)
9	86.62 (0.7899)	83.64 (0.7818)	78.90 (0.7185)	(0.82)	84.67 (0.713)	84.67 (0.713)	(0.83)	70.75	(0.61)
Average	**82.05 (0.7806)**	78.96 (0.7194)	74.95 (0.6658)	(0.60)	75.02 (0.644)	75.02 (0.644)	(0.594)	67.75	(0.57)

Table3: Accuracy (kappa)'s p- value of Proposed model modified TSGL-EEGNet comparison with other methods using dataset BC IV 2a.

Test	Proposed TSGL-EEGNet	TSGL-EEGNet	EEGNet	SS-MEMDBF [26]	MB3DCNN [7]	C2CM [5]	TSMM + SVM [27	FBCSP [28]	FBCSP [23]
T Test	**0.003 (0.003)**	0.004 (0.004)	0.006 (0.006)	(0.014)	0.024 (0.003)	0.024 (0.023)	(0.002)	< 0.001	(0.001)
KS Test	<0.001 (<0.001)	<0.001 (<0.001)	(0.006)	(0.24)	0.034 (<0.001)	<0.001 (<0.001)	(<0.001)	(<0.001)	(<0.001)

As the MB3DCNN reveals that it has a more significant advantage in kappa than in accuracy, which explains why it performs well in all classes. Comparing the proposed TSGL-EEGNet to other classic techniques like as the SS-MEMDBF reveals that the

proposed TSGL-EEGNet has a more significant advantage on insensitive patients, which explains why it is more suited for the majority of people. The two-sided p-values of the proposed approach and the others are all less than 0.05, as shown in Table 3. The T-Test demonstrates that the two findings do not have the same mean, and the KS Test demonstrates that their distributions are distinct. This indicates that the approach suggested in this study is a substantial advance over its predecessor since it is more efficient and resilient. By combining classical machine learning with deep learning methods, it is possible to enhance classification accuracy, as shown by the findings. Moreover, the model given is relatively interpretable. After a mathematical explanation, it is possible to demonstrate that the proposed deep learning model and CSP methods have similar feature extraction and selection components. Consequently, the interpretability of deep learning models might be improved by adding conventional mathematical procedures. However, this does not imply that we should be satisfied with conventional algorithms. Deep learning, on the other hand, provides benefits that classical algorithms cannot reach, such as end-to-end models, adaptive hyper parameter learning, and high classification accuracy. This means that we must re-examine the conventional fields in order to develop neural network models that are better, quicker, and more interpretable.

Reference

[1] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, “Learning representation from EEG with deep recurrent – convolutional neural networks,” arXiv: 1511.06448[Online]. Available: http://arxiv.org/abs/1511.06448, 2015.

[2] M.-A. Li, Y.-F.Wang,S.-M. Jia, Y.-J.Sun, and J.-F.Yang ,”Decoding of Motor Imagery EEG based on brain source estimation, ”Neurocomputing, vol.339, pp. 182-193, Apr. 2019

[3] V. J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung and B.J. Lance , “ EEGNet: A compact convolutional network for EEG based brain-computer interfaces ,” J. Neural Eng. , vol. 15, Art. No. 056013, Oct. 2018 1993.

[4] N. Padfield, J. Zabalza, H. Zao, V. Masero, and J. Ren , “ EEG-based brain-computer interfaces using motor imagery: Techniques and challenges ,” Sensors, vol 19,no.6, pp. 1423, Mar. 2019

[5] S. Sakhavi, C. Guan, and S. Yan, “Learning temporal information for brain-computer interfaces using convolutional neural networks,” IEEE Trans. Neural NeuralNetw. Learn. Syst. , vol. 20, no. 11., pp. 5619-5629, Nov. 2018

[6] R.T. Schirrmeister , J.T. Springenberg , L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M.Tangermann, F. Hutter, W. Burgard and, T. Ball , “Deep learning with convolutional neural networks for EEG decoding and visualization,” Hum. Brain Mapping, vol. 38, no. 11, pp. 5391-5420, Nov. 2017

[7] X.Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, “ A multibranch 3D convolutional neural network for EEG based motor imagery classification, “IEEE Trans. Neural Syst. Rehabil. Eng., vol. 27, no.10, pp. 2164-2177, Oct. 2019

[8] K.-W. Ha and J.-W. Jeong, “Motor imagery classification using capsule networks,” Sensors, vol.19, no.13, p.2854, Jun. 2019

[9] B. Graimann, B. Allison, and G. Pfurtscheller, “Brain-computer interfaces : Springer, pp.1-27, 2009

[10] L.F. Nicolas-Alonso and J. Gomez-Gil, “Brain-computer interfaces a review,”

eISSN1303-5150

Sensor, vol. 12, no.2, pp. 1211-1279, 2012

[11] G. Pfurtscheller and A. Aranibar, " Event-related cortical desynchronization detected by power measurements of scalp EEG ," Electroencephalogr. Clin.Neuorophysiol. Vol. 42, no.6, pp. 817-826, Jun. 1977

[12] G. Pfurtscheller, C. Brunner, A. Schlogl, , and F.H.L. da Silva, " Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks," Neuroimage, vol.31, no.1, pp. 153-159, May 2006

[13] X. Deng, D. Li., J. Mi., F. Gao., Q. Chen, J. Wang, and R. Liu, "Motor imagery ECoG signal classification using sparse representation with elastic net constraint ," in proc. IEEE 7th Data Driven Control Learn. Syst. Conf. (DDLS), May2018, pp. 44-49

[14] G. Rodriguez-Bermudez and P. J. Garcia-laencina, "Automatic and Adaptive classification of electroencephalographic signals for brain computer interfaces," J. Med. Syst., vol.36, no. S1, pp. 51-63, Nov. 2012

[15] A. Scholgel, F.Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery data for the BCI-competition 2005," J. Neural Eng. , vol. 2, no.4, pp.L14-L22, Dec. 2005

[16] J. Z. Zhou, M. Meng, Y. Gao, Y. Ma, and Q. Zhang, "Classification of motor imagery EEG using wavelet envelope analysis and LSTM networks," in proc. Chin. Control Decis. Conf (CCDC), pp. 5600-5605, Jun. 2018

[17] Y. Hashimoto and J. Ushiba, "EEG-based classification of imaginary left and right foot movements using beta rebound , "Clin. Neurophysiol., vol. 124, no.11, pp. 2153-2160, Nov. 2013

[18] K.K. Ang, Z.Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in proc. IEEE Int. Joint Conf. Neural NeuralNetw. IEEE World Congr. Compute. Intell, Jun. 2008

[19] Y. Zhang, C.S. Nam, G.Zhou, J.Jin, X. Wang, and A.Cichoki, " Temporally constrained spares group spatial patterns for motor imagery BCI," "IEEE Trans. Cybern., vol.49, no 9, pp.3322-3332, Sep.2019

[20] K.K. Ang, Z. Y. Chin, C. Wang, C. Guan,and H. Zhang., "Filter bank common spatial pattern algorithm on BCI Competition IV dataset 2a and 2b," Frontiers Neurosci., vol.6, pp39, Mar. 2012

[21] P. Gaur, R.B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," Expert Syst. Appl., vol.95, pp. 201-211, Apr. 2018

[22] X. Xie, Z. L. Yu, H. Lu, Z. Gu, and Y. L," motor imagery classification based on bilinear sub-manifold learning of symmetric positive definite matrices," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 25, no.6, pp. 504-516, Jun. 2017

[23] H. Bashashati, R.K. Ward, and A. Bashashati, "User-Customized brain in brain-computer interfaces using Bayesian optimization," J. Neural Eng. Vol. 13, no.2, Art. No. 026001, Apr. 2016

[24] Xin Deng, Boxian Zhang, Nian Y.U., Ke Liu, and Sun, "Advanced TSGL-EEGNet for Motor Imagery EEG-Based Brain-Computer Interfaces," IEEE Access, vol.9, pp. 25118-25130, Feb. 202

[25] C. Brunner, R. Leeb, G. Mleer-Putz, A. Schlogl, "BCI competition 2008-graz data set A," Inst. Know.Disc., Lab. Berlin –Comp. Interfaces, Graz University, Austria, 2008 pp. 136-142, vol. 16 [Online]. Available: http://www.bbci.de/competition/iv/desc_2a.pdf