بیستمین کنفرانس ملی دانشجویی مهندسی برق ایران
دانشگاه رازی کرمانشاه
دانشگاه رازی
انجمن علمی مهندسی برق دانشگاه رازی
20th IRANIAN STUDENT CONFERENCE ON ELECTRICAL ENGINEERING
بیستمین کنفرانس ملی دانشجویی مهندسی برق ایران
NSSOEE
سازمان علمی دانشجویی مهندسی برق کشور
تاسیس ۱۳۷۷
اتحادیه انجمن‌های علمی دانشجویی مهندسی برق
ISC
Islamic World Science Citation Center

Multi-Class Motor Imagery Classification

M. Jyannasab
Elec. Eng. Dept., Shahed University
Email: mohamadjeyannasab@yahoo.com

S. Seyedtabaii
Elec. Eng. Dept., Shahed University
Email: stabaii@shahed.ac.ir

Abstract
The Motor Imagery (MI) classification task is a high dimension multivariate and complicated subject. In this respect, the original signals are analyzed and minimal unique features of the classes are extracted to facilitate accurate classification of the actions performed. The fusion of common spatial pattern, Fisher discrimination ratio, and filter bank alongside the SVM and CNN-LSTM are incorporated to provide accurate clustering. As a result and after extensive simulations, it is shown that the CSP+ FDR + CNN-LSTM setup more accurately differentiates the classes.

Keywords: Motor Imagery Classification, SVM, LSTM, CSP

1 INTRODUCTION

People with neurological disease may find trouble in walking, speaking, and writing due to the lack of functioning of the motor control. Brain-computer interface (BCI) technology can help them to back to the quality of normal life.

The problem of systems action interpretation may be employed by model-based (Nemati and Seyedtabaii 2020, Shams and Seyedtabaii 2020) and/or signal-based (Qaedi and Seyedtabaii 2012, Seyedtabaii 2012, Fasihipour and Seyedtabaii 2020) approaches; wherein the case of MI the signal based is favorite. Low SNR, fewer data, and being multi-channels are among the difficulties related to this study. Signal-based BCI consists of two modules: feature extraction and command interpretation.

In this respect, a new CNN architecture to introduce the temporal representation of the data for MI classification is presented in (Saputra, Setiawan et al. 2019). Regarding multi-class MI signal analysis, the classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain-computer interfaces has been detailed in (Dong, Li et al. 2017). A novel hybrid deep learning scheme for four-class motor imagery classification has been investigated in (Zhang, Zong et al. 2019). Recurrent Deep Learning for EEG-based motor imagination recognition has been detailed in (Rammy,

Abrar et al. 2020). Adaptive transfer learning for EEG motor imagery with deep Convolution Neural Network is the subject of study in (Zhang, Robinson et al. 2021). An EEG Channel Selection Method for motor imagery-based brain-computer interface and neurofeedback using Granger causality has been introduced in (Varsehi and Firoozabadi 2021). In (Wang, Huang et al. 2020), temporal-spatial-frequency depth extraction of brain-computer interface based on mental tasks has been described. A CNN with a hybrid convolution scale for EEG motor imagery classification is reported in (Dai, Zhou et al. 2020). Recognizing single-trial motor imagery EEG based on interpretable clustering method is investigated in (Fu, Li et al. 2021). Adaptive spatiotemporal graph convolution nnetworks for MI classification has been suggested in (Sun, Zhang et al. 2021).

In this paper, four-class MI classification algorithms are studied. Various features from the common spatial pattern (CSP), filter bank idea, and Fisher discrimination ratio (FDR) are derived. Both SVM and CNN-LSTM deep learning are utilized for the classification where the best results obtained are the product of the CSP+FDR+CNN+LSTM algorithm.

The paper is organized as follows. In Section 2, the basic sub-algorithms used are briefly described. Section 3 introduces the setups of two algorithms and their variations for the MI classification and lastly, the conclusion comes in section 5.

2 BRIEF DESCRIPTION OF SUB-ALGORITHMS

Basic CSP

Consider two tasks, H and F where each has been tried M times and the recorded signals are $X_H(i) \in R^{N\times T}$ and $X_F(i) \in R^{N\times T}$ where N is the number of channels and T is the number of samples. The main idea is to use a linear transformation to project the two sets of the multi-channel EEG data into another space with maximum distance to be easily classified. In the multiclass cases, where one versus rest (OVR) is requited, the following average for the rests are employed,

$$X_F(i) = \overline{X_{F1} + X_{F2} + X_{F3}} \tag{1}$$

The normalized spatial covariance of the two matrixes and their averages are computed using the following equations (Wang, Gao et al. 2006),

$$C_H(i) = \frac{X_H(i)X_H^T(i)}{trace(X_H(i)X_H^T(i))}, \quad \bar{C}_H = \frac{1}{M}\sum_{i=1}^{M} C_H(i)$$

where C_* is an N by N matrix. Trace(A) computes the sum of the diagonal elements of A. Then, the following composite spatial covariance is calculated and decomposed as below,

$$C = \bar{C}_H + \bar{C}_F = URU^T$$

where U is the eigenvectors and R is the diagonal matrix of eigenvalues. The results are multiplied by the whitening transformation matrix as below,

$$P = R^{-\frac{1}{2}}U^T \Rightarrow \quad S_H = P\bar{C}_H P^T, \quad S_F = P\bar{C}_F P^T$$

The eigenvectors with the largest eigenvalues of S_H have the smallest eigenvalues in S_F and vice versa. Thus, by employing the projection matrix W, the original samples are converted to the uncorrelated Z counterparts,

$$W = U^T P \Rightarrow Z = WX \tag{2}$$

The columns of W^{-1} are the CSPs filters. For order reduction, *d* rows from the top and bottom of Z are selected for Z_R which often d=1 is considered to be adequate (Jamaloo and Mikaeili 2015).

$$Z_R = Z(k,t), \quad K = 1,...,d,N-d,...,N \tag{3}$$

Signal features

From Z_R, the scalar row variances are the features that are fed to the classifier section (Olivas-Padilla and Chacon-Murguia 2019),

$$f_i = \mathrm{var}(Z_{R,i}(j,t)),\quad i=1,...,M,\ t=1,...,T,\ j=1,...,2d \tag{5}$$

where j is the Z_R rows, M the total number of trials, and t is the signal samples.

Fisher's discriminant ratio

The Fisher criterion for optimal class separability aims at maximizing the between class-variance while minimizing the within class-variance, i.e., the distance between the feature vectors from the same class. This is done by the V vector mapping the feature from one space to another with better classification capacity (Veksler).

$$f_f = V'f \tag{6}$$

where f is the feature vector (5) and V is the FDR transformation matrix.

FILTER BANK (FB)

Physiologically, the content of the discriminative information in the different frequency ranges differs for individuals. Therefore, decomposing the signal in terms of filter bands is helpful in better MI signal classification (Park, Lee et al. 2017). By filtering the signals into K bands, the covariance matrix for each band is obtained,

$$C_i^k,\quad i=1,...,M \quad and \quad k=1,...,K$$

and then separately undergoes the CSP operation for each band and the feature vector f is computed. The best band is the one with the highest d computed below,

$$d(k) = mean\left(\sum_{i=1}^{M}\left(f_{H,1}(i,k)-f_{H,2}(i,k)\right)\right) - mean\left(\sum_{i=1}^{M}\left(f_{F,1}(i,k)-f_{F,2}(i,k)\right)\right) \tag{7}$$

LSTM (long short-term memory)

LSTM has three layers (Wiki). The first layer is defined by the following equations which receives the input pattern x_t and the past hidden layer output h_{t-1} to produces,

$$\bar{C}_t = \tanh(W_c[h_{t-1},x_t]+b_c)$$
$$i_t = sig(W_i[h_{t-1},x_t]+b_i)$$

where b_*, W_*, i_t, and $\bar{C}_t$ are the biases, weights, input gate, and the input tangent hyperbolic function output.

The second layer is described by,

$$f_t = sig(W_f[h_{t-1},x_t]+b_f)$$
$$C_t = f_t C_{t-1} + i_t \bar{C}_t$$

where f_t is the forgetting gate and C_t is the updated cell output.

Finally, the output of the current state and the hidden layer output are calculated as below,

$$o_t = sig(W_o[h_{t-1},x_t]+b_o)$$
$$h_t = o_t * \tanh(C_t)$$

Convolutional Neural Network (CNN)

CNNs are comprised of four layers (Wiki). These are input layers, convolutional layers, pooling layers, and fully connected network layers.

The input layer holds the input pattern.

The convolutional layer with *l* hidden layers operates as given by the following equations,

$$hc_l = R(conv(W_l, x_l) + b_l)$$

$$R(\alpha) = \max(0, \alpha)$$

where b_l, x_l, w_l, and R are the bias, input pattern, weight, and the rectified linear unit(RELU) of the l*th* hidden layer.

The pooling layer will then simply perform downsampling along with the spatial dimensionality of the given input, further reducing the number of parameters within that activation.

The fully connected layers will then perform the same as the conventional neural network for classification.

SUPPORT VECTOR MACHINE (SVM)

SVM is a binary classifier that linearly optimally maps the input x onto binary decision variable y∈ [±1]. In the case of linear separation, the problem is expressed by the following optimization problem,

$$\min \ \tfrac{1}{2}\|w\|^2$$

$$s.t \quad y_i \times (w.x_i + b) > M(1 - \varepsilon_i), \quad \varepsilon_i > 0, \sum \varepsilon_i < C$$

where y∈[1,-1] is the class label, x is the input, b is the bias, and ϵ is the margin which improves the performance in a case of not linearly separable problems. By minimizing w, the size of M gap between the two classes is maximized.

3 MI CLASSIFICATION METHODS

Data set

The dataset used in this study was taken from BCI competition IV-II-a provided by Graz University (BCI). It includes signals with the following specifications:

- 9 healthy subjects
- 4 operations: movement of the left hand, right hand, both feet, and tongue.
- 22 EEG channels: The signals were sampled at 250 Hz and band-pass filtered between 0.5 Hz and 100 Hz and notch filtered at 50Hz.
- 6 runs per subject per session.
- 12 trial * 4 task for each run per subject (72 trail per task per subject and overall of 288 trial per subject)

The μ rhythms (8–12 Hz) and β rhythms (14–30 Hz) of EEG signals amplitudes and powers vary due to the intention for performing actions. In addition, the high-frequency components in EEG signals were usually nebulous, so the raw EEG signals were filtered by a band-pass filter (3–34 Hz).

Therefore, there are 288 matrices of X(N, T) with 22 channels by 315 samples. From the 72 trials, 50 trails are randomly assigned for training and the remaining 22 trails are assigned for the test and validation. A sample of X signal has been shown in Fig. 1

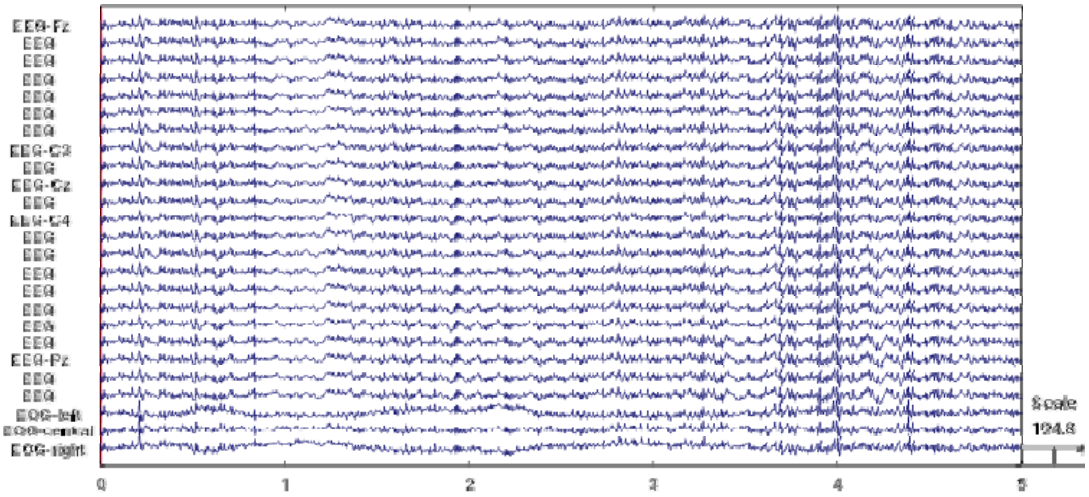

Fig. 1. One of the trials

3-1 SVM classification

In this section, several feature extraction approaches are incorporated and a set of SVMs are utilized for the MI classification.

CSP+SVM

In this approach, four similar parallel classifiers are developed where each takes care of one of the classes. The algorithm configuration for Class 1 is as below:

Preprocessing:

- Class 1 is introduced as X_H and the rest as X_F (1).
- X_H and X_F are used for computing $W \in R^{22\times22}$ using (2).

Training

- Randomly 50 samples of X_H and X_F are picked for training
- They are multiplied by W for Z (2)
- 2D $f \in R^{2\times100}$ is obtained using (3) and (5).
- f is applied to SVM for training.

Test

- 72 samples of both X_H and X_F are multiplied by W for Z (2)
- 2D $f \in R^{2\times144}$ is obtained using (3) and (5).
- f is applied to SVM for the test.
-

The test success rates are presented in the training, test, and overall groups as depicted in Table I.

CSPFDR+HSVM

In this test, 10 classifiers with almost similar configurations are arranged. 6 for one versus one (OVO) and 4 for one versus rest (OVR) using (1). The basic algorithm configuration for the distinction between Class one and the rest is as below.

Preprocessing:

- Class 1 is introduced as X_H and the rest as X_F (1). In the case of OVO, X_F is the second class.
- X_H and X_F are used for computing $W \in R^{22\times22}$ using (2).
- $Z_R(i) \in R^{2\times313}$ is computed using (3)
- $f \in R^{2\times144}$ is calculated for both classes and by using FDR (6) the vector *V* is derived.

Training

- Randomly 50 samples of X_H and X_F are picked for training
- They are multiplied by W for Z (2)
- 2D f$\in R^{2\times100}$ is obtained using (3) and (5).
- f is multiplied by V of FDR to generate the input pattern f_f for SVM training.

Test

- 72 samples of both X_H and X_F are multiplied by W for Z (2)
- 2D f$\in R^{2\times144}$ is obtained using (3) and (5).
- f is multiplied by *V* and applied to SVMs for the test.
- If there is no conflict in the results of OVR classifiers the success rate is 100% and no further action is required.
- If there is a conflict between 2, 3, and 4 of the classes, their OVO counterpart classifiers will judge the situation for discrepancy removal (Dong, Li et al. 2017).

The test success rates are presented in the training, test, and overall groups as depicted in Table I. Besides their confusion matrixes have also been provided in Fig. 2.

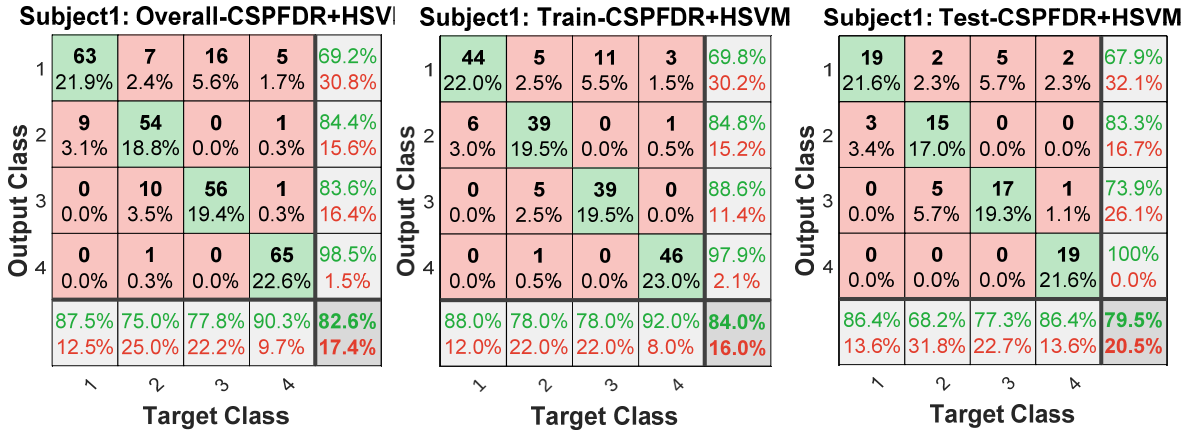

Fig. 2. The CSP+FDR+HSVM confusion matrixes

FB+CSP+HSVM

In this test also 10 classifiers with almost similar configurations are arranged. 6 for one versus one (OVO) and 4 for one versus rest (OVR) using (1). The basic algorithm configuration for the distinction between Class one and the rest is as below.

Preprocessing:

- Class 1 is introduced as X_H and the rest as X_F (1). In the case of OVO, X_F is the second class.
- X_H and X_F are partitioned into 5 filter banks, $X_{H,k,}$ and $X_{F,k}$, k=1,…,5.
- $X_{H,k,}$ and $X_{F,k,}$ are used for computing $W_k\in R^{22\times22}$ using (2).
- $Z_{R,k}(i)\in R^{2\times313}$ is computed using (3)
- $f_k\in R^{2\times144}$ is calculated for both classes.
- The best bank k0 is selected using (7).

Training

- Randomly 50 samples of X_H and X_F are picked for training
- By filter banks, $X_{H,k0}$ and $X_{F,k0}$ are picked
- They are multiplied by W_{k0} for Z (2)

- 2D f∈ $R^{2\times100}$ is obtained using (3) and (5).
- *f* is the input pattern for SVM training.

Test

- 72 samples of EEG signal go through the filter bank for $X_{H,k0,}$ and $X_{F,k0}$.
- $X_{H,k0,}$ and $X_{F,k0}$ are multiplied by W_{k0} for Z (2)
- 2D f∈ $R^{2\times144}$ is obtained using (3) and (5).
- f is applied to SVMs for the test.
- If there is no conflict in the results of OVR classifiers the success rate is 100% and no further action is required.
- If there is a conflict between 2, 3, and 4 of the classes, their OVO counterpart classifiers will judge the situation for discrepancy removal (Dong, Li et al. 2017).

The test success rates are presented in the training, test, and overall groups as depicted in Table I.

TABLE I. Algorithms' performances

		Class 1	class 2	class 3	class 4	Overall
CSP+SVM	train	68.0	78.0	52	86	71.0
	test	50.0	54.5	27.3	81.8	53.4
	overall	62.5	70.8	44.4	84.7	65.69
CSP+FDR+ HSVM	train	88.0	78.0	78.0	92.0	84.0
	test	86.4	68.2	77.3	86.4	79.5
	overall	87.5	75.0	77.8	90.3	82.6
FBCSP+ HSVM	train	100.0	100.0	100.0	100.0	100.0
	test	86.4	59.1	31.8	45.5	55.7
	overall	95.8	87.5	79.2	83.3	86.5

3-2 CSP+ Deep learning

The CNN-LSTM configuration shown in Fig. 3 is used for MI classification where the OVR features first derive separate CNNs and the result is applied to LSTM for final multiclass MI classification. Various tests using different features are conducted which are detailed in the sequel.

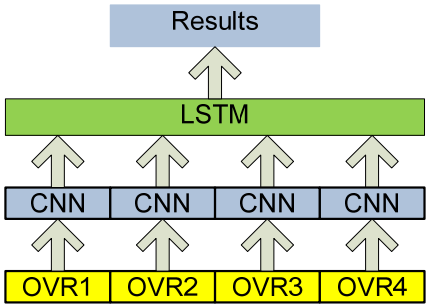

Fig. 3. The CNN-LSTM configuration

CSP+CNN-LSTM

In this test, four OVR features are calculated and applied to CNN-LSTM for test and training as elaborated below,

Preprocessing:

- Similar to CSP+SVM calculate W_m for the m_{th} OVO case.

Training

- Randomly pick 50 samples of $X_{H,}$ and X_F for training

- They are multiplied by W_m for Z (2)
- Overall $Z_R \in R^{(4*10)\times 100}$ is obtained using (3) and (5) which is applied to CNN-LSTM for training. Five from the top and five from the bottom of Z are selected.

Test

- X_H and X_F are multiplied by their corresponding W_m for Z (2)
- $Z_R \in R^{(4*10)\times 100}$ features are applied to the CNN-LSTM for the test.
-

The test success rates are presented in the training, test, and overall groups as depicted in Table II.

CSP+FDR+CNN-LSTM

In this test, four OVR features are computed and applied to CNN-LSTM as below,

Preprocessing:

- Class m is introduced as X_H and the rest as X_F (1).
- $X_{H,m,}$ and $X_{F,m}$ are used for computing $W_m \in R^{22\times 22}$ using (2).
- $Z_{R,m}(i) \in R^{4\times 313}$ is computed using (3)(2 from the top and 2 from the bottom)
- $f_m \in R^{4\times 1}$ is calculated for both classes and by using FDR (6) the vector *V* is derived

Training

- Randomly 50 samples of $X_{H,}$ and X_R are picked for training
- They are multiplied by W_m for Z (2)
- $f_m \in R^{(4*4)\times 100}$ is obtained using (3) and (5).
- f is multiplied by V of FDR to generate the input pattern $f_f \in R^{(4*4)\times 100}$ for CNN-LSTM training.

Test

- 72 samples of both $X_{F,m}$ and $X_{R,m}$ are multiplied by W_m for Z_m (2)
- $f_m \in R^{(4*4)\times 144}$ is obtained using (3) and (5).
- f_m is multiplied by V for $f_f \in R^{(4*4)\times 144}$ and applied to CNN-LSTMs for the test.
-

The test success rates are presented in groups of training, test, and overall as depicted in Table II.

FBCSP+CNN-LSTM

A filter bank is employed to decompose the EEG in every time window into 16 frequency passbands by using causal Chebyshev Type II filter

Preprocessing:

- Class m is introduced as $X_{F,m,}$ and the rest as $X_{R,m}$ (1).
- $X_{F,m}$ and $X_{R,m}$ are partitioned into 16 filter banks, $X_{F,m,k}$ and $X_{R,m,k}$, k=1,…,16.
- $X_{F,m,k,}$ and $X_{R,m,k}$ are used for computing $W_{m,k} \in R^{22\times 22}$ using (2).
- $Z_{R,mk}(i) \in R^{4\times 313}$ is computed using (3)
- $f_{m,k} \in R^{4\times 144}$ is calculated for both classes.

Training

- Randomly 50 samples of X_F and X_R are picked for training
- They are multiplied by $W_{k,m}$ for Z (2)
- Z_R is obtained using (3).
- $f \in R^{(16*4*10)*100}$ is calculated for all classes for training where 16, 4 and 10 stand for the number of filter banks, the number of classes and the numbers of rows of Z, picked for classification.

Test

- 72 samples of EEG signal go through the filter banks for $X_{F,k,m,}$ and $X_{R,k,m}$.

- $f \in R^{(16*4*10)*144}$ is obtained using (3) and (5) and is applied to the CNN-LSTM for test

The test success rates are presented in the training, test, and overall groups as depicted in Table II.

TABLE II. Table II The results obtained using CNN-LSTM classifier.

		Class 1	class 2	class 3	class 4	Overall
CSP+ CNN-LSTM	Train	100.0	100.0	100.0	100.0	100.0
	Test	45.54	59.09	31.82	63.64	50.0
	Overall	83.83	87.50	79.17	88.89	84.72
CSP+FDR+ CNN-LSTM	Train	100.0	100.0	100.0	100.0	100.0
	Test	70.0	80.45	70.0	86.45	76.73
	Overall	100.0	98.61	100.0	98.61	92.8
FBCSP+ CNN-LSTM	Train	100.0	100.0	100.0	100.0	100.0
	Test	27.27	77.27	40.91	59.09	51.14
	Overall	77.78	93.06	81.94	87.50	85.07

4 CONCLUSION

In this paper, a four-class MI classification is studied. Features from the common spatial pattern (CSP), Fisher discrimination ratio (FD) are integrated to provide distinctive features for SVM and deep learning classifiers. The results indicate that the algorithm consists of CSP+FDR+CNN+LSTM yields the best success rate as high as 92%. -

Acknowledgment

This work has been partially supported by the research department of the Shahed University, Tehran, IRAN.

References

BCI BCI dataset.

Dai, G., J. Zhou, J. Huang and N. Wang (2020). "HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification." Journal of neural engineering **17**(1): 016025.

Dong, E., C. Li, L. Li, S. Du, A. N. Belkacem and C. Chen (2017). "Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain–computer interfaces." Medical & biological engineering & computing **55**(10): 1809-1818.

Fasihipour, H. and S. Seyedtabaii (2020). "Fault detection and faulty phase (s) identification in TCSC compensated transmission lines." IET Generation, Transmission & Distribution **14**(6): 1042-1050.

Fu, R., W. Li, J. Chen and M. Han (2021). "Recognizing single-trial motor imagery EEG based on interpretable clustering method." Biomedical Signal Processing and Control **63**: 102171.

Jamaloo, F. and M. Mikaeili (2015). "Discriminative common spatial pattern sub-bands weighting based on distinction sensitive learning vector quantization method in motor imagery based brain-computer interface." Journal of medical signals and sensors **5**(3): 156.

Nemati, M. and S. Seyedtabaii (2020). "Fault severity estimation in a vehicle cooling system." Asian Journal of Control **22**(6): 2543-2548.

Olivas-Padilla, B. E. and M. I. Chacon-Murguia (2019). "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters." Applied Soft Computing **75**: 461-472.

Park, S.-H., D. Lee and S.-G. Lee (2017). "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification." IEEE Transactions on Neural Systems and Rehabilitation Engineering **26**(2): 498-505.

Qaedi, S. and S. Seyedtabaii (2012). "Improvement in power transformer intelligent dissolved gas analysis method." World Academy of Science, Engineering and Technology **61**: 1144-1147.

Rammy, S. A., M. Abrar, S. J. Anwar and W. Zhang (2020). Recurrent Deep Learning for EEG-based Motor Imagination Recognition. 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), IEEE.

Saputra, M. F., N. A. Setiawan and I. Ardiyanto (2019). "Deep Learning Methods for EEG Signals Classification of Motor Imagery in BCI." IJITEE (International Journal of Information Technology and Electrical Engineering) **3**(3): 80-84.

Seyedtabaii, S. (2012). "Performance evaluation of neural network based pulse-echo weld defect classifiers." Measurement Science Review **12**(5).

Shams, Z. and S. Seyedtabaii (2020). "NONLINEAR FLEXIBLE LINK ROBOT JOINT-FAULT ESTIMATION USING TS FUZZY OBSERVERS." International Journal of Robotics and Automation **35**(1).

Sun, B., H. Zhang, Z. Wu, Y. Zhang and T. Li (2021). "Adaptive Spatiotemporal Graph Convolutional Networks for Motor Imagery Classification." IEEE Signal Processing Letters **28**: 219-223.

Varsehi, H. and S. M. P. Firoozabadi (2021). "An EEG channel selection method for motor imagery based brain–computer interface and neurofeedback using Granger causality." Neural Networks **133**: 193-206.

Veksler, O. "Pattern Recognition." from https://www.csd.uwo.ca/~oveksler/Courses/CS434a_541a/Lecture8.pdf.

Wang, L., W. Huang, Z. Yang and C. Zhang (2020). "Temporal-spatial-frequency depth extraction of brain-computer interface based on mental tasks." Biomedical Signal Processing and Control **58**: 101845.

Wang, Y., S. Gao and X. Gao (2006). Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. 2005 IEEE engineering in medicine and biology 27th annual conference, IEEE.

Wiki, A. I. "A Beginner's Guide to LSTMs and Recurrent Neural Networks." from https://wiki.pathmind.com/lstm.

Zhang, K., N. Robinson, S.-W. Lee and C. Guan (2021). "Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network." Neural Networks **136**: 1-10.

Zhang, R., Q. Zong, L. Dou and X. Zhao (2019). "A novel hybrid deep learning scheme for four-class motor imagery classification." Journal of neural engineering **16**(6): 066004.