35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Single-Trial Discrimination of EEG Signals for Stroke Patients: A General Multi-Way Analysis

Ye Liu, Mingfen Li, Hao Zhang, Junhua Li, Jie Jia, Yi Wu, Jianting Cao and Liqing Zhang*

Abstract— **It has been demonstrated that Brain-Computer Interface (BCI), combined with Functional Electrical Stimulation (FES), is an effective and efficient way for post-stroke patients to restore motor function. However, traditional feature extraction methods, such as Common Spatial Pattern (CSP), do not work well for post-stroke patients' EEG data due to its irregular patterns. In this study, we introduce a novel tensor-based feature extraction algorithm, which takes both spatial-spectral-temporal features of EEG data into consideration. EEG data recorded from post-stroke patients is used for simulation experiments to assess the effectiveness of the proposed algorithm. The results show that the the proposed algorithm outperforms some traditional algorithms.**

I. Introduction

Stroke, or cerebrovascular accident (CVA), is a medical emergency that results in multi-organ dysfunction and can cause permanent neurological damage and death [1]. For post-stroke patients, regular daily life and social activities are severely affected. Effective rehabilitation approaches should be applied immediately onto patients in order to help them return to normal life. However, traditional rehabilitation therapies, such as constraint-induced movement therapy and speech and language therapy, achieve disappointing results and poor rehabilitation performances. Poor enthusiasm are aroused for stroke patients by these boring therapies. Deeply, active motor control loop between brain and impaired limbs is not established, and thus it is difficult to reconstruct the neural circuits on impaired brain cortex.

Brain-Computer Interface (BCI), as a new pathway between brain and external devices, is an alternative way of rehabilitation training in the people with disabilities. It has been proved that BCI, combined with Functional Electrical Stimulation (FES), is an effective and efficient way for post-stroke patients to restore motor function [2]. Feature extraction and classification are crucial for designing a BCI-based rehabilitation training system. However, traditional feature extraction methods, such as Common Spatial Pattern (CSP) [3], can not ensure persuasive performance when directly applied on post-stroke patients because of irregular imagery patterns [4]. Additionally, another issue that few EEG data is recorded from real patients usually leads to limited available training dataset, arousing the small sample size (SSS) problem.

In this study, a robust tensor-based algorithm, namely general tensor-based nearest feature line (GTNFL), is introduced to overcome the above problems. Spatial-spectral-temporal patterns are taken into account, obtaining an optimal multi-way discriminative subspace for data projection. Moreover, the training dataset can be expanded due to the extra virtual training data, and thus the small sample size problem is better solved.

The rest of the paper is organized as follows: Section II gives a brief view of data acquisition based on the BCI-FES system. Section III presents a detail description of data processing, especially the GTNFL algorithm for feature extraction. Section IV lists the comparative results of the simulation experiments when applying the proposed algorithm and other traditional algorithms on EEG data recorded from real patients. Finally, Section V gives a discussion of the results.

II. Materials

A. Stimuli and Procedure

A motor imagery based BCI-FES rehabilitation system [5] is applied for data acquisition in this study. The whole rehabilitation process consists of three courses called training model course, virtual games rehabilitation course and post-training model course. In the whole experiment, patients sit still in a comfortable wheelchair with their hands rested on the armrest. In the training model course, patients have to participate in the basic motor imagery based tasks for eight sessions. In each trial, patients are required to imagine for two seconds based on the direction shown by a bold arrow. In the virtual games rehabilitation course, patients are asked to complete five virtual games involving different categories of daily life. In the post-training model course, patients continue to finish another two sessions of the basic motor imagery based tasks, as described in the training model course, for purpose of rehabilitation efficacy assessment.

B. Data Acquisition

Five post-stroke patients participate in our BCI-FES rehabilitation training per day and three days per week, along with the ordinary medical treatments in the hospital every day. It takes two months for the patients to complete the whole training process. Raw data is collected using Ag/AgCl electrodes from the 16-channel (namely FC3, FCZ, FC5,

This work was supported by the National Natural Science Foundation of China (Grant No. 90920014 and 91120305), the NSFC-JSPS International Cooperation Program (Grant No.61111140019) and the European Union Seventh Framework Programme (Grant No. 247619)

Y. Liu, H. Zhang, J. Li and L. Zhang are with MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

M. Li, Y. Wu and J. Jia are with Department of Rehabilitation, Huashan Hospital, Fudan University, Shanghai, 200240, China

J. Cao is with Saitama Institute of Technology

*Corresponding author (zhang-lq@cs.sjtu.edu.cn)

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

C1-C6, CZ, CP3, CPZ, CP4, P3, PZ and P4) g.USBamp amplifier (G.tecGuger Technologies, Graz, Austria) with a sampling rate of 256Hz.

III. METHODS

A. Data Pre-processing

Pre-processing for EEG data is required to improve the low signal-noise ratio (SNR) before feature extraction. Firstly, EEG data is spatial filtered through Common Average Reference (CAR) [6] method. Afterwards, we employ Independent Component Analysis (ICA) [7] method to remove artifacts, such as Electrocardiograph (ECG) and Electromyography (EMG), from EEG signals. Furthermore, EEG signals are bandpass filtered within the specific frequency band 4-45Hz, aiming at extracting the most active rhythms related to motor imagery for post-stroke patients. Finally, with time-frequency decomposition methods, EEG data is converted into the format channel*frequency*time (16 channels; 4-45 Hz; 1-2000 ms). In our study, wavelet transform method is chosen as the time-frequency decomposition method and the complex Morlet wavelet is utilized as the mother wavelet, as it is successful in the analysis of time-frequency information of EEG data.

B. Feature Extraction and Classification

General tensor-based nearest feature line (GTNFL) algorithm is proposed to extract features of EEG data recorded from post-stroke patients in this section. Support vector machine (SVM) [8] classifier is used as classification method. Some basic tensor operators are presented firstly in the following subsection.

(1) Tensor Operators

Tensors, as multi-dimensional arrays, present the linear relations between vectors, scalars and other tensors. The contracted product of two tensors $\underline{\mathbf{X}} \in R^{N_1 \times N_2 \times \cdots \times N_M \times Q_1 \times Q_2 \times \cdots \times Q_D}$ and $\underline{\mathbf{Y}} \in R^{N_1 \times N_2 \times \cdots \times N_M \times S_1 \times S_2 \times \cdots \times S_P}$ along the first M modes is denoted as $\underline{\mathbf{Z}} = [\![\underline{\mathbf{X}} \otimes \underline{\mathbf{Y}}; (1:M)(1:M)]\!] \in R^{Q_1 \times \ldots \times Q_D \times S_1 \times \ldots \times S_P}$. Especially, contracted product of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ on all indices except the k-th index is denoted as $[\![\underline{\mathbf{X}} \otimes \underline{\mathbf{Y}}; (\overline{k})(\overline{k})]\!]$. The mode-$d$ product of a tensor $\underline{\mathbf{X}} \in R^{N_1 \times N_2 \times \cdots \times N_M}$ and a matrix $\mathbf{P} \in R^{J_d \times N_d}$ is a tensor $\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_d \mathbf{P} \in R^{N_1 \times \ldots \times N_{d-1} \times J_d \times N_{d+1} \times \ldots \times N_M}$. Moreover, $mat_d(\underline{\mathbf{X}}) \in R^{I_d \times I_1 \ldots I_{d-1} I_{d+1} \ldots I_M}$ stands for the mode-d matricization of the tensor $\underline{\mathbf{X}}$.

(2) GTNFL for Feature Extraction

Suppose x_1 and x_2 are two 2-dimensional points, a line passing through x_1 and x_2 is defined as *Feature Line* (FL), as described in [9]. As a natural extension of the traditional definition, we define *Multi-dimensional Feature Line* (MFL) $\underline{\mathbf{X}}_1\widehat{\underline{\mathbf{X}}}_2$ as a line passing through two high-dimensional points $\underline{\mathbf{X}}_1 \in R^{I_1 \times I_2 \times \cdots \times I_M}$ and $\underline{\mathbf{X}}_2 \in R^{I_1 \times I_2 \times \cdots \times I_M}$ sharing the same label. Suppose $\underline{\mathbf{X}}^*$ is the projection of $\underline{\mathbf{X}}$ onto $\underline{\mathbf{X}}_1\widehat{\underline{\mathbf{X}}}_2$, *FL distance* between $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}_1\widehat{\underline{\mathbf{X}}}_2$ is expressed as the Euclidean distance between $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}^*$.

Given the training dataset $\left\{\underline{\mathbf{X}}_{i;j}^{I_1 \times I_2 \times \ldots \times I_M}\right\}$, where $\underline{\mathbf{X}}_{i;j}^{I_1 \times I_2 \times \ldots \times I_M}$ denotes the j-th point in the i-th class with $1 \le i \le C$ and $1 \le j \le N_i$, the mean point for the i-th class is $\underline{\mathbf{M}}_i = (1/N_i)\sum_{j=1}^{N_i} \underline{\mathbf{X}}_{i;j}$. For each point $\underline{\mathbf{X}}_{i;j}$, the number of MFLs, formed by the mean point $\underline{\mathbf{M}}_i$ and the other points sharing the same class with $\underline{\mathbf{X}}_{i;j}$, is $P_{i;j}$. Afterwards, we use $\{\underline{\mathbf{X}}_{i;j;1}^*, \underline{\mathbf{X}}_{i;j;2}^*, \ldots, \underline{\mathbf{X}}_{i;j;P_{i;j}}^*\}$ as all the projections of $\underline{\mathbf{X}}_{i;j}$ onto all the above MFLs. Extra virtual points, like these projections, are generated, and thus the training dataset is enlarged. We define *the within-class FL distance* between $\underline{\mathbf{X}}_{i;j}$ and all its projections as $\sum_{p=1}^{P_{i;j}} d^2(\underline{\mathbf{X}}_{i;j}, \underline{\mathbf{X}}_{i;j;p}^*)$, where $\underline{\mathbf{X}}_{i;j;p}^*$ is the p-th projection of $\underline{\mathbf{X}}_{i;j}$. Furthermore, the within-class FL distance between all the points in the training set and their projections is expressed as:

$$J_{within} = \sum_{i=1}^{C}\sum_{j=1}^{N_i}\sum_{p=1}^{P_{i;j}} d^2(\underline{\mathbf{X}}_{i;j}, \underline{\mathbf{X}}_{i;j;p}^*) \tag{1}$$

Similarly, the *between-class FL distance* between all the points and their projections is expressed as:

$$J_{between} = \sum_{i=1}^{C}\sum_{j=1}^{N_i}\sum_{q=1}^{Q_{i;j}} d^2(\underline{\mathbf{X}}_{i;j}, \underline{\mathbf{X}}_{i;j;q}^*) \tag{2}$$

where $Q_{i;j}$ is the number of the MFLs formed by the mean point $\underline{\mathbf{M}}_{\bar{i}}$ with $1 \le \bar{i} \ne i \le C$ and all the points $\underline{\mathbf{X}}_{\bar{i};j}$ with $1 \le j \le N_{\bar{i}}$. $\underline{\mathbf{X}}_{i;j;q}^*$ is the q-th projection of $\underline{\mathbf{X}}_{i;j}$ onto the corresponding MFL. Fig. 1 gives a brief view of the concept of within-class FL distance and between-class FL distance.

The aim of GTNFL algorithm is to find a set of projection matrices $\left\{\mathbf{W}^{(d)} \in R^{I_d \times U_d}, U_d \le I_d\right\}_{d=1}^{M}$, by which original EEG data $\underline{\mathbf{X}}_{i;j}$ can be projected to $\underline{\mathbf{Y}}_{i;j} = \underline{\mathbf{X}}_{i;j}\prod_{d=1}^{M} \times_d \mathbf{W}^{(d)^T} \in R^{U_1 \times U_2 \times \ldots \times U_M}$. These matrices, which are optimal for distinguishing different classes, can project original data in a low dimensional space in which the within-class FL distance is minimized and the between-class FL distance is maximized. Therefore, GTNFL algorithm can be converted to the following optimization problem:

$$f(\mathbf{W}^d|_{d=1}^M) = \arg\min \frac{1}{K_{in}}\sum_{i=1}^{C}\sum_{j=1}^{N_i}\sum_{p=1}^{P_{i;j}} d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}_{i;j;p}^*) - \frac{1}{K_{\text{bet}}}\sum_{i=1}^{C}\sum_{j=1}^{N_i}\sum_{q=1}^{Q_{i;j}} d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}_{i;j;q}^*) \tag{3}$$

where $K_{in} = \sum_{i=1}^{C}\sum_{j=1}^{N_i} P_{i;j}$ and $K_{\text{bet}} = \sum_{i=1}^{C}\sum_{j=1}^{N_i} Q_{i;j}$. $\underline{\mathbf{Y}}_{i;j}$ is the projection of $\underline{\mathbf{X}}_{i;j}$ by $\left\{\mathbf{W}^{(d)}\right\}_{d=1}^{M}$. $\underline{\mathbf{Y}}_{i;j;p}^*$ is the p-th projection of $\underline{\mathbf{Y}}_{i;j}$ onto the corresponding MFL. Similarly, $Q_{i;j}$ and $\mathbf{Y}_{i;j;q}^*$ represent corresponding parameters of the between-class FL distance.

However, optimization problem in equation (3) does not have a closed form as M parameters need to be determined in one function. Therefore, the alternative least square (ALS)

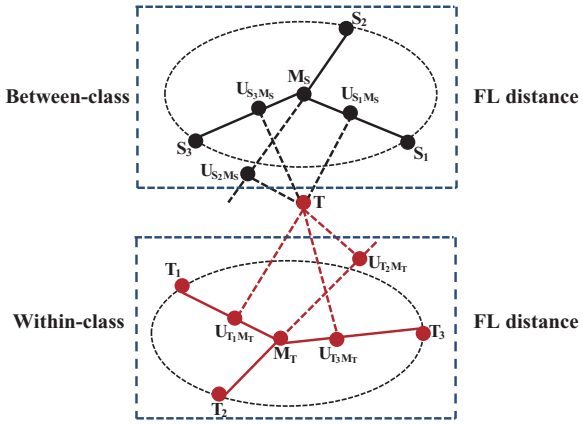

Fig. 1. A brief view of the concept of within-class FL distance and between-class FL distance. Different dots with different colors share different classes. Within-class FL distance is defined as the sum of the distance between data T and all its projections onto the MFLs formed by the mean point M_T and T_1, T_2, T_3. Similarly, the between-class FL distance is the sum of the distance between data T and all its projections onto the MFLs formed by the mean point M_S and S_1, S_2, S_3.

method is applied to decompose the optimization problem into several optimization subproblems, obtaining a numerical solution. As the optimization function defined in equation (3) is only determined by $d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;p})$ and $d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;q})$, we just reform $d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;p})$ as follows:

$$
\begin{aligned}
& d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;p}) \\
&= d^2(\underline{\mathbf{X}}_{i;j} \prod_{d=1}^{M} \times_d \mathbf{W}^{d^T}, \underline{\mathbf{X}}^*_{i;j;p} \prod_{d=1}^{M} \times_d \mathbf{W}^{d^T}) \\
&= [\![(\underline{\mathbf{X}}_{i;j} \prod_{d=1}^{M} \times_d \mathbf{W}^{d^T}) \otimes (\underline{\mathbf{X}}_{i;j;p} \prod_{d=1}^{M} \times_d \mathbf{W}^{d^T}); \\
&\quad (1:M)(1:M)]\!] \\
&= [\![(\underline{\mathbf{X}}_{i;j} \overline{\times}_m \mathbf{W}^{m^T} \times_m \mathbf{W}^{m^T}) \otimes (\underline{\mathbf{X}}^*_{i;j;p} \overline{\times}_m \mathbf{W}^{m^T} \times_m \\
&\quad \mathbf{W}^{m^T}); (1:M)(1:M)]\!] \\
&= tr\{\mathbf{W}^{m^T} [\![(\underline{\mathbf{X}}_{i;j} \overline{\times}_m \mathbf{W}^{m^T}) \otimes \\
&\quad (\underline{\mathbf{X}}^*_{i;j;p} \overline{\times}_m \mathbf{W}^{m^T}); (\overline{m})(\overline{m})]\!] \mathbf{W}^m\}
\end{aligned}
$$

Similarly, $d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;q})$ can be re-expressed in the same form as $d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;p})$. Therefore, new expressions of $d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;p})$ and $d^2(\underline{\mathbf{Y}}_{i;j}, \underline{\mathbf{Y}}^*_{i;j;q})$ are put into equation (3), we can get:

$$
f(\mathbf{W}^m) = \arg\min tr\left(\mathbf{W}^{m^T}(A_m - B_m)\mathbf{W}^m\right) \qquad (4)
$$

where A_m and B_m are defined as:

$$
A_m = \frac{1}{K_{in}} \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{p=1}^{P_{i;j}} [\![\begin{array}{c} (\underline{\mathbf{X}}_{i;j} \overline{\times}_m \mathbf{W}^{m^T}) \otimes \\ (\underline{\mathbf{X}}^*_{i;j;p} \overline{\times}_m \mathbf{W}^{m^T}); (\overline{m})(\overline{m}) \end{array}]\!]
$$

$$
B_m = \frac{1}{K_{\text{bet}}} \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{q=1}^{Q_{i;j}} [\![\begin{array}{c} (\underline{\mathbf{X}}_{i;j} \overline{\times}_m \mathbf{W}^{m^T}) \otimes \\ (\underline{\mathbf{X}}^*_{i;j;q} \overline{\times}_m \mathbf{W}^{m^T}); (\overline{m})(\overline{m}) \end{array}]\!]
$$

Thus, the optimization problem defined in equation (4) is only determined by one parameter $\mathbf{W}^m$. After setting partial derivative of $\mathbf{W}^m$ to zero, we get the optimal $\mathbf{W}^m$ as the U_m unit eigenvectors based on the smallest U_m eigenvalues of $(A_m - B_m)$. The algorithm will stop running if the convergence condition $\sum_{m=1}^{M} \|\mathbf{W}_t^m - \mathbf{W}_{t-1}^m\| < \varepsilon$ is satisfied, where ε is the threshold to test convergence.

IV. Results

200 EEG trials (100 trials for each class) are extracted from EEG dataset recorded from each patient for each day. The training dataset is consisted of the first 60 trials (30 trials for each class) and the other 140 trials are used as test dataset. GTNFL algorithm is applied for feature extraction, together with the obtained optimal discriminative subspace by which the original data $\underline{\mathbf{X}}_{i;j}$ is projected to $\underline{\mathbf{Y}}_{i;j} \in R^{U_1 \times U_2 \times \ldots \times U_M}$. Furthermore, we apply fisher score [10] as the redundant features elimination method. In our study, we just assess the performance of the proposed GTNFL method from two aspects of the obtained optimal discriminative subspace and recognition rate.

A. Optimal Discriminative Subspace

We just choose two days (day 30 and day 60) during the rehabilitation period to obtain the comparative results and assess GTNFL algorithm when compared with some traditional algorithms. Fig. 2 gives a brief illustration of the optimal subspace extracted by GTNFL algorithm on day 30 and day 60 for the second patient. In each spatial pattern on day 30, it is obvious that C3 is the most significant channel for imagining right while the left movement imagination is focused at channel CP4 and P4. In terms of spectral patterns, the two most remarkable bands are around 7-13 Hz and 20-30 Hz. Moreover, in each spatial pattern on day 60, C3 and C4 are the primary channels for imagining left and right. For spectral patterns on day 60, 6-14 Hz and 15-26 Hz are the two most reactive bands.

CSP, which relies on physiological phenomenon of event-related desynchronization (ERD), can calculate the most discriminative spatial patterns of bandpass filtered EEG data. In Fig. 3, for CSP, the first two most significant spatial patterns have no discrimination when distinguishing left imagination from right imagination. As the most active frequency bands contributing to ERD of post-stroke patients are not available in advance, CSP is not effective for extracting features of EEG signals from post-stroke patients. However, GTNFL algorithm has a better performance without the neurophysiologic knowledge, as it calculates discriminative spatial patterns and reactive frequency bands directly from multi-way EEG signals.

B. Recognition Rate

Five traditional algorithms of power spectrum density (PSD) [11], CSP, regularized CSP (RCSP) [12], wavelet transform (WT) [13] and nonnegative multiway factorization (NMWF) [14] are implemented on the dataset for comparison. Six weeks out of two months (three weeks per month) are chosen for each patient. Afterwards, classification accuracy on test dataset of each experiment in each week is

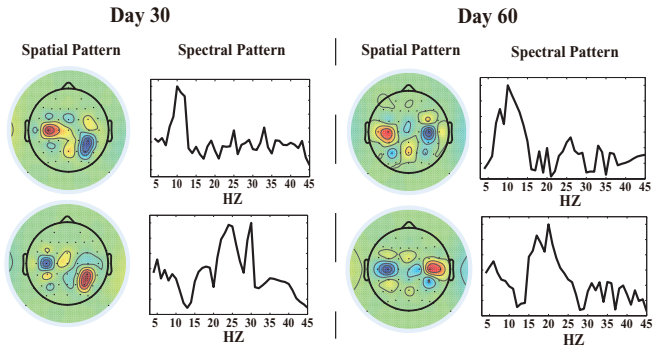

Fig. 2. The optimal subspace extracted by GTNFL for projection on day 30 and day 60 for the second patient .

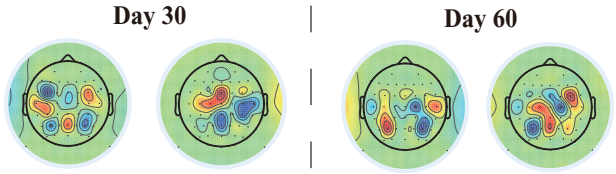

Fig. 3. The optimal subspace extracted by CSP for projection on day 30 and day 60 for the second patient.

computed using the trained SVM. Finally, mean accuracies of the six weeks are calculated for all the patients.

TABLE I

MEAN ACCURACIES (%) OF THE SIX CHOSEN WEEKS USING GTNFL AND THE FIVE OTHER TRADITIONAL METHODS FOR FEATURE EXTRACTION.

Subject	PSD	CSP	RCSP	WT	NMWF	GTNFL
1	48.35	59.70	60.46	61.93	65.71	71.64
2	60.05	57.60	59.10	63.85	67.95	76.52
3	45.83	57.20	59.30	65.68	65.56	75.36
4	50.11	53.48	55.36	51.88	59.78	68.17
5	56.92	57.90	58.51	59.81	64.45	66.64
Mean	52.25	57.18	58.55	60.63	64.69	71.67

Table I lists the results of mean accuracies of our proposed GTNFL algorithm and the five other methods. For each patient, mean accuracy of the six chosen weeks for GTNFL algorithm is obviously higher than that of the other algorithms. On average of all the patients, mean accuracies can not exceed 70% for the traditional five algorithms while GTNFL achieves an improvement to 71.67%. Meanwhile, the same accuracy improvement is found in each individual. These results verify that the performance of GTNFL outperforms the other traditional algorithms.

V. Discussion

The results show that GTNFL algorithm outperforms the five other traditional feature extraction algorithms. As GTNFL calculates the optimal subspace directly from multi-dimensional data, structural information and correlation are kept in the original data, and thus the diverse EEG data is well-represented by the the extracted features. Additionally, the small sample size problem is better solved due to the expanded training dataset through the generated virtual data. As a consequence, there is a significant improvement in performance for GTNFL algorithm.

VI. Conclusion

In this study, a robust feature extraction algorithm, namely general tensor-based nearest feature line (GTNFL), is proposed to calculate the optimal projection subspace based on multi-dimensional EEG data. We evaluate GTNFL algorithm based on the EEG datasets recorded from post-stroke patients, compared with five other classical algorithms. The results show that the proposed algorithm outperforms the other algorithms and is efficient in discrimination of EEG signals for post-stroke patients.

References

[1] N. R. Sims and H. Muyderman, "Mitochondria, oxidative metabolism and cell death in stroke," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1802, no. 1, pp. 80–91, 2010.

[2] K. A Rahman, B. Ibrahim, M. Jamil, and A. Leman, "The development of control system via Brain Computer Interface (BCI)-Functional Electrical Stimulation (FES) for paraplegic subject," *International Journal of Integrated Engineering*, vol. 4, no. 3, 2013.

[3] 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.

[4] D. Garrett, D. Peterson, C. Anderson, and M. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 141–144, 2003.

[5] H. Zhang, J. Liang, Y. Liu, H. Wang, and L. Zhang, "An iterative method for classifying stroke subjects' motor imagery EEG data in the BCI-FES rehabilitation training system," in *IEEE 1st Inter. Conf. on Cognitive Systems and Information Processing*. IEEE, 2012.

[6] D. McFarland, L. McCane, S. David, and J. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalography and clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, 1997.

[7] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4, pp. 411–430, 2000.

[8] V. Vapnik, *The nature of statistical learning theory*. springer, 1999.

[9] J. Chien and C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1644–1649, 2002.

[10] R. Duda, P. Hart, and D. Stork, "Pattern classification," *New York: John Wiley, Section*, vol. 10, p. l, 2001.

[11] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[12] 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.

[13] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1057–1061, 2004.

[14] M. Mørup, L. Hansen, J. Parnas, and S. Arnfred, "Decomposing the time-frequency representation of EEG using non-negative matrix and multi-way factorization," *Technical University of Denmark Technical Report*, 2006.