

Classification of motor imaginary in EEG using random forest classifier

Gurkan Kucukyildiz, Mechatronics Engineering Department, Kocaeli University–Umuttepe, Kocaeli 41380, Turkey.

Suat Karakaya*, Mechatronics Engineering Department, Kocaeli University–Umuttepe, Kocaeli 41380, Turkey.

Hasan Ocak, Mechatronics Engineering Department, Kocaeli University–Umuttepe, Kocaeli 41380, Turkey.

Suggested Citation:

Kucukyildiz, G., Karakaya, S. & Ocak, H. (2017). Classification of motor imaginary in EEG using random forest classifier. *Global Journal of Computer Sciences: Theory and Research*. 7(3), 129-135.

Received July 23, 2017; revised October 21, 2017; accepted December 9, 2017.

Selection and peer review under responsibility of Prof. Dr. Dogan Ibrahim, Near East University, North Cyprus.

©2017 Academic World Education & Research Center. All rights reserved.

Abstract

Although the motor-imagery-based brain computer interface (BCI) has become popular in recent years, its practical application is limited due to the classification accuracy of methods. In this study, a new classification scheme is proposed for the classification of multi-class motor imaginary in EEG using random forest (RF) classifier. In the proposed scheme, a four-stage binary classification tree is constructed. An RF model is trained for each stage of decision tree using features extracted from the EEG channels. The EEG band powers of each channel are the extracted features from the EEG signal. The proposed classification scheme is applied on the BCI competition IV dataset 2a recordings. The EEG data is acquired from nine subjects and the proposed scheme is performed for each subject independently. The kappa values of the proposed scheme are calculated to compare the results with the methods in the literature. It is demonstrated that the proposed classification scheme has higher kappa values than the methods in the literature.

Keywords: Brain computer interface, motor imaginary, random forest.

* ADDRESS FOR CORRESPONDENCE: **Suat Karakaya**, Mechatronics Engineering Department, Kocaeli University, – Umuttepe, Kocaeli 41380, Turkey. *E-mail address:* suatronik@gmail.com / Tel: +90 (262) 303 10 00

1. Introduction

People who suffer from tetraplegia or ALS can lose their motor ability functions. Therefore, these people cannot control external devices with their muscles. Brain-computer interfaces (BCI) are an alternate way to control electronic devices. BCIs are based on the understanding of the brain activity and converting them into control commands [1]. Generally, during the imaginary movements of different parts of their body, spatially localised brain activity arises [2]. Junhua *et al.* [3] developed a motor-imaginary-based BCI system. During the experiments, subjects were asked to accomplish three cognitive tasks: The imaginary left-hand, right-hand and right-foot movement. EEG data was collected by an amplifier, which has 118 EEG and 10 EOG channels. The sampling frequency of the amplifier is set to 1 KHz. EEG amplifier has an internal bandpass filter of 0.05 and 200 Hz. The delta frequency (0–4 Hz) band energy of the EEG data and common spatial pattern (CSP)-based features are extracted as a feature. The fisher discriminant classifier is used in order to classify the EEG data. Authors tested the developed classification scheme on the two subjects' EEG data. It was observed that the proposed classification scheme has 93.45% and 91.88% sensitivity for each cognitive task.

Izguerdo *et al.* [4] developed a BCI system that consists of three motor imaginary cognitive tasks. The cognitive tasks are determined as left-hand movement, right-hand movement and constructing sentences with a keyword. The developed classification scheme is applied on the BCI Competition III dataset V benchmark data. In the feature extraction step, the band energies of EEG data are extracted as feature. An ANN and fuzzy-logic-based hybrid classifier is used in the signal classification stage. The authors claimed that the developed classification scheme has 87.21%, 82.26% and 58.72% classification accuracy per cognitive task, respectively.

In this study, the benchmark data of BCI competition IV dataset 2a is used. EEG data is obtained from nine healthy subjects in two sections. The first 5 min of EEG data is neglected in order to reduce the EOG effects on the EEG signal. The first 5 min can be divided into three sections: 2 min eyes open, 2 min eyes closed and 1 min with eye movements. The EOG parts of the experiment are given in Figure 1.

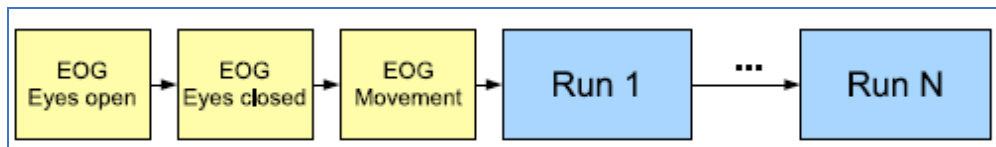


Figure 1. Sample scenario for trapping into local minimum

During the experiment, subjects were asked to accomplish four different cognitive tasks: left-hand, right-hand, both foot and tongue movement. Sampling frequency and resolution of the EEG amplifier was set to 250 Hz and 100 uv, respectively. The data consisted of the 22 channel EEG and 3 channel EOG data. A 50-Hz digital notch filter is applied on the EEG in order to suppress the power line frequency. The EEG data is also filtered with a 20th degree bandpass filter (8–30 Hz).

1.1. Path Planning Phase

The CSP method is developed by H. Ramoser [5] to classify multi-channel EEG signals. The main goal of the CSP is to get a linear transform matrix that maximises the variance of the two-class signal matrix. Rows of the gathered transform matrix are assumed as weights for the channels. The raw EEG data is represented as an $N \times M$ matrix \mathbf{X} . N is the number of the channels and M is the number of samples per channels. The normalised spatial covariance of \mathbf{X} can be calculated as follows:

$$C = \frac{\mathbf{X}\mathbf{X}^T}{\text{trace}(\mathbf{X}\mathbf{X}^T)} \quad (1)$$

The trace of a matrix is defined as the sum of diagonal elements. The spatial covariance is calculated for both the classes by using the EEG data matrices \mathbf{X}_r and \mathbf{X}_l (i.e., left and right motor imaginary)

$$\mathbf{C}_l = \frac{\mathbf{X}_l \mathbf{X}_l^T}{\text{trace}(\mathbf{X}_l \mathbf{X}_l^T)}, \mathbf{C}_r = \frac{\mathbf{X}_r \mathbf{X}_r^T}{\text{trace}(\mathbf{X}_r \mathbf{X}_r^T)} \quad (2)$$

The composite spatial covariance is then calculated as

$$\mathbf{C}_c = \mathbf{C}_l + \mathbf{C}_r \quad (3)$$

The \mathbf{C}_c matrix can be decomposed with the standard matrix decomposition

$$\mathbf{C}_c = \mathbf{U}_c \mathbf{\Lambda}_c \mathbf{U}_c^T \quad (4)$$

In Eq. (4), \mathbf{U}_c is the eigenvector matrix and $\mathbf{\Lambda}_c$ is the eigenvalue diagonal matrix. The whitening transformation matrix \mathbf{P} is calculated as follows:

$$\mathbf{P} = \sqrt{\frac{1}{\mathbf{\Lambda}_c}} \mathbf{U}_c^T \quad (5)$$

The variances in the space spanned by \mathbf{U}_c are equalised using \mathbf{P} matrix as follows:

$$\mathbf{S}_l = \mathbf{P} \mathbf{C}_l \mathbf{P}^T, \mathbf{S}_r = \mathbf{P} \mathbf{C}_r \mathbf{P}^T \quad (6)$$

\mathbf{S}_l and \mathbf{S}_r matrices share common eigenvectors. Therefore, the sum of the corresponding eigenvalues of two matrices is equal to I

$$\mathbf{S}_l = \mathbf{B} \mathbf{\Lambda}_l \mathbf{B}^T, \mathbf{S}_r = \mathbf{B} \mathbf{\Lambda}_r \mathbf{B}^T \text{ and } \mathbf{\Lambda}_l + \mathbf{\Lambda}_r = \mathbf{I} \quad (7)$$

The eigenvectors, which have the largest eigenvalues for \mathbf{S}_l , have the smallest eigenvalues for \mathbf{S}_r , and vice versa. Therefore, the eigenvectors of \mathbf{B} matrix can be used for two-class classification problems. The transformation of the whitened EEG data is used in the first and last eigenvectors in \mathbf{B} , providing the feature vectors that are optimal for discriminating two populations of odd EEG in the least squares sense.

The projection matrix (\mathbf{W}) can be defined as follows:

$$\mathbf{W} = (\mathbf{B}^T \mathbf{P})^T \quad (8)$$

The EEG data is decomposed with the \mathbf{W} as

$$\mathbf{Z} = \mathbf{W} \mathbf{X} \quad (9)$$

The columns of the \mathbf{W}^{-1} matrix are called the CSPs.

1.2. Random Forest

The decision tree technique was initially proposed by Morgan and Sonquist in 1963 [x]. The technique was used frequently for classification applications after a study published by Breiman in 1984 [6, 7]. The classification algorithm starts by splitting the training data into left and right nodes based on a threshold applied to the first feature. The threshold is determined as the boundary value, which minimises the residual of sum squares (regression criteria) and is determined as the threshold. The residual of the sum squares can be calculated as

$$RSS = \sum_{\text{left}} (y_i - y_l)^2 + \sum_{\text{right}} (y_i - y_r)^2 \quad (10)$$

The splitting operation is applied to each node repeatedly using the next feature at each level of the decision tree till pure class samples are obtained. Therefore, the size and computational cost of the decision tree depends on the number of features. A sample of the separation flowchart of random forest (RF) algorithm is shown in Figure 2.

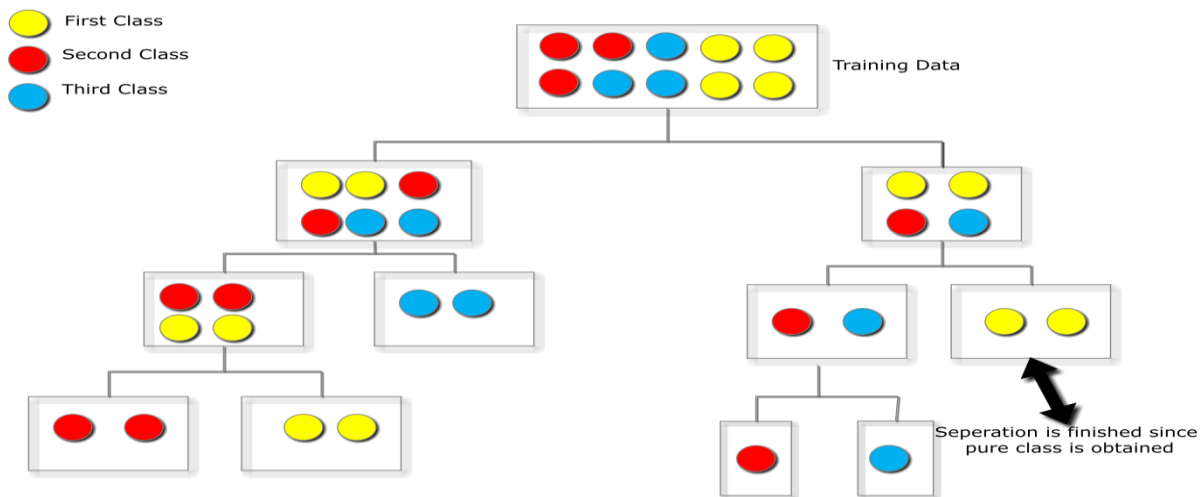


Figure 2. Sample flowchart for RF classifier

The size of the decision tree can be decreased by pruning operation. The pruning operation can be described as the elimination of useless features. The gini-index, which measures the statistical dispersion, can be used to decide whether the node is promising or not. Therefore, the gini-index of a node can be used for feature selection as follows:

$$G = N_l \sum_{k=1}^K p_{kl}(1 - p_{kl}) + N_r \sum_{k=1}^K p_{kl}(1 - p_{kl}) \quad (11)$$

The classification accuracy of a binary decision tree depends on the training data. To overcome this problem, the training data are segmented into randomly selected sub-datasets. A decision tree is built for each sub-dataset. The output of the RF classifier is set as the average of the decision trees' outputs

$$RF = \frac{1}{M} \sum_{i=1}^M T_i \quad (12)$$

where M is the number of decision trees and T_i is the output of the i th tree.

1.3. Feature Extraction

Feature extraction, which maps the measurement space into the feature space, is the initial step in any particular pattern recognition problem. In this study, a set of both time and frequency domain features were extracted from the related signals. In addition to CSP-based features, band energies (alfa, beta and theta) of EEG data signal are included in the feature vector as frequency domain-based features. During offline analysis of the EEG data, it is observed that the delta energy band decreases the classification performance. Therefore, the delta band is not included in the feature vector. The CSP is defined between two classes. Therefore, two methods, one versus rest and the other versus one, are used in the literature. In this study, one versus all method is used to extract feature.

1.4. Signal Classification

A multi-stage decision-making algorithm is developed in order to analyse motor imaginary EEG data in this study. The proposed motor imaginary classification system consists of three decision stages. In the first stage, the developed algorithm decides whether the EEG data sample is left hand or not. If the algorithm determines that the EEG data sample does not belong to the left hand, it goes to the second stage where it decides whether it is the right-hand or other classes. In the final stage of the algorithm, the EEG data is assigned to both the hands or tongue. RF classifiers are used at each stage of the proposed motor imaginary classification system. The number of the grown trees in each classifier is set as 100 by the trial-and-error method. The developed classification scheme is shown in Figure 3.

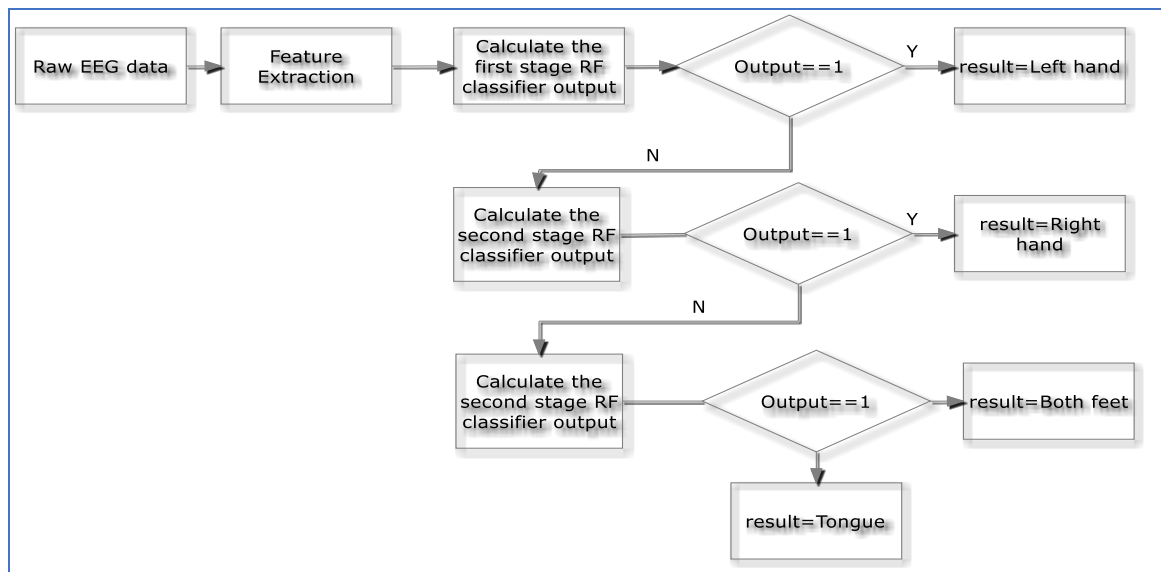


Figure 3. Multi-stage classification scheme

2. Results

In the literature, the performance of classifiers is compared by two performance criteria: sensitivity and specificity. The sensitivity and specificity of a classifier can be calculated as follows:

$$Sensitivity = TP / (TP + FN) \quad (13)$$

$$Specificity = TN / (TN + FP) \quad (14)$$

In Eqs. (13) and (14), TN, TP, FN and FP correspond to true negatives, true positives, false negatives and false positives, respectively. TP samples are assumed as the correctly classified data samples in a related class. Like TP samples, TN samples are the correctly classified data samples belonging to other classes. FN samples are the data samples in a class incorrectly classified as a related class to one of the other classes. Besides, FP samples are the data samples in other classes which are incorrectly classified as the related class. For analysis, the EEG data is divided into 1.25-sec time windows. Therefore, each class has 144 samples per subject. A total of 50% of the data are used for training and the remaining 50% is used for testing the proposed classifier. Totally, each class has 324 samples for testing. The confusion table of classifier is given in Table 1.

Table 1. Confusion table of RF classifier

Classifier	Left hand	Right hand	Both foot	Tongue
RF	200	44	48	32
	77	207	10	30
	51	32	212	29
	45	29	36	214

Of the 324 samples corresponding to the left-hand class, 200 (sensitivity: 0.617) samples are correctly classified by the proposed classification scheme. The classification sensitivities for the right hand, foot and tongue are calculated as 0.638, 0.654 and 0.660, respectively.

Table 2. Class specificities of RF classifier

Classifier	Left hand	Right hand	Both foot	Tongue	Mean
RF	0.822	0.887	0.907	0.923	0.880

Using the confusion table, class specificities of the proposed classification scheme are calculated and are given in Table 2. It is observed that the RF classifier has the highest classification specificity on the tongue class.

3. Discussion

In this study, a three-stage RF-based classification scheme is proposed for the motor imaginary in EEG signals. The developed scheme is tested on the four-class BCI Competition dataset 2a benchmark data. Both time (CSP) and frequency domain (EEG band energies)-based features are extracted from the EEG data. It is observed that the proposed classification scheme has the highest sensitivity and specificity on the tongue class.

References

- [1] Cano-Izquierdo, J. M., Ibarrola, J., & Almonacid, M. (2012). Improving motor imagery classification with a new BCI design using neuro-fuzzy S-dFasArt. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(1), 2-7.

- [2] Kucukyildiz, G., Ocak, H., Karakaya, S., & Sayli, O. (2017). Design and implementation of a multi sensor based brain computer interface for a robotic wheelchair. *Journal of Intelligent & Robotic Systems*, 87, 1-17.
- [3] Leo, B. (1984). *Classification and regression trees*. Pacific Grove, CA: Wadsworth & Brooks.
- [4] Li, J., Wang, Y., Zhang, L., Cichocki, A., & Jung, T. P. (2016). Decoding EEG in cognitive tasks with time-frequency and connectivity masks. *IEEE Transactions on Cognitive and Developmental Systems*, 8(4), 298-308.
- [5] Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58(302), 415-434.
- [6] Nezamfar, H., Orhan, U., Purwar, S., Hild, K., Oken, B., & Erdogmus, D. (2011). Decoding of multichannel EEG activity from the visual cortex in response to pseudorandom binary sequences of visual stimuli. *International Journal of Imaging Systems and Technology*, 21(2), 139-147.
- [7] Pfurtscheller, G., Guger, C., & Ramoser, H. (1999). EEG-based brain-computer interface using subject-specific spatial filters. *Engineering Applications of Bio-Inspired Artificial Neural Networks*, 1607, 248-254.