

# The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on EEG Signal

Nurul E'zzati Md Isa<sup>1,\*</sup>, Amiza Amir<sup>1</sup>, Mohd Zaizu Ilyas<sup>1</sup>, and Mohammad Shahrazel Razalli<sup>1</sup>

<sup>1</sup>School of Computer and Communication Engineering, UniMAP, 02600 Arau, Perlis

**Abstract.** Most EEG-based motor imagery classification research focuses on the feature extraction phase of machine learning, neglecting the crucial part for accurate classification which is the classification. In contrast, this paper concentrates on the classifier development where it thoroughly studies the performance analysis of k-Nearest Neighbour (k-NN) classifier on EEG data. In the literature, the Euclidean distance metric is routinely applied for EEG data classification. However, no thorough study has been conducted to evaluate the effect of other distance metrics to the classification accuracy. Therefore, this paper studies the effectiveness of five distance metrics of k-NN: Manhattan, Euclidean, Minkowski, Chebychev and Hamming. The experiment shows that the distance computations that provides the highest classification accuracy is the Minkowski distance with 70.08%. Hence, this demonstrates the significant effect of distance metrics to the k-NN accuracy where the Minkowski distance gives higher accuracy compared to the Euclidean. Our result also shows that the accuracy of k-NN is comparable to Support Vector Machine (SVM) with lower complexity for EEG classification.

## 1 Introduction

Motor imagery classification is used to classify the imaginary task such as hand and foot movements by using brain signal which usually used in Brain-Computer Interfaces (BCI). BCI is an alternative communication pathway between the brain and computer which does not requires muscular movement or control [1] which has potential application value in wide range of the fields such as medical, military and entertainment [2].

The classification of brain signals is challenging as the signals are low signal to noise ratio, non-linearity and limited training data due to difficulties in collecting signals [3]. Electroencephalogram (EEG) is one of the techniques that usually been chosen by researchers to collect the signals instead of others such as computer tomography (CT), magnetic resonance imaging (MRI) and functional MRI (fMRI). The collection of signal using EEG is done by placing the electrode on the scalp to records electrical activity and brain waves.

Two main phases in data classification is feature extraction phase and training/ classification phase (classifier development). Unlike the existing researches [4-8] which focus on feature extraction, this paper focuses on the classifier development.

Classification task of EEG data can be done with many algorithms, one of them is k-Nearest Neighbor (k-NN). K-NN classifier is used in [9] to discriminate between seizure and non-seizure events for automated seizure detection using EEG signals. In [10], the k-NN is

used as classifier for its classification task to compare three different signal decomposition methods for motor imagery BCI systems. K-NN classifier provides better classification accuracy and requires less training and testing times compared with the Multiple Layer Perceptron (MLP) and Support Vector Machine (SVM) [11]. In [12], the k-NN classifier obtained the best accuracy in classifying EEG signals to identify the engagement, enjoyment, frustration and difficulty compared to Bayes Network, Naïve Bayes, SVM, Multilayer Perceptron, Random Forest and J48. Besides that, [13] shows that the k-NN also provides better accuracy in eye state classification compared to Multilayer Perceptron Neural Networks.

This research study the classification technique which using k-NN algorithm. K-NN is simple supervised learning algorithm [14] that classified a sample by a majority vote of its neighbors which means the sample is allocated to the class based on the most common class among its k closest neighbors. In k-NN, distance metrics is used to calculate distance between a new samples and existing samples in dataset. The literature is strongly influenced by a commonplace in using the Euclidean distance metric [10-13][15-18]. In fact, we are not aware of any studies with a focus on a performance comparison among various different distance metrics for k-NN. This paper aims to thoroughly investigate the effect of distance metrics and parameter k to the k-NN performance. In this paper, we study five distance metrics of k-NN: Manhattan, Euclidean, Minkowski, Chebychev and Hamming and vary k value in the

\* Corresponding author: nurulezzati@studentmail.unimap.edu.my

experiments. Additionally, the paper also verifies whether the k-NN algorithm is suitable or not to use for classification method in BCI system as in [15] states that k-NN is one of the most important non-parametric algorithms in BCI implementation and compare it with the performance of SVM which commonly used for EEG data classification.

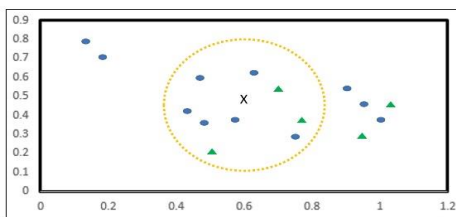
This paper is study the performance analysis of k-NN algorithms by using EEG data for BCI motor imaginary classification. The results helps in choosing better k value and a proper metric that makes a good classification performance. The rest of this paper is organized as follows: Section 2 gives the introduction of k-NN and its parameter. Section 3 contains the EEG processing using k-NN. In section 4, the results and discussion is presented. Lastly, section 5 is conclusion.

## 2 K-Nearest Neighbors

K-NN classification is a non-parametric model that is described as instance-based learning which the model are characterized by memorizing the training dataset. K-NN also a typical example of a lazy learner. It is called lazy not because of its apparent simplicity, but because it does not learn a discriminative function from the training data but memorized the training dataset instead. Lazy learning is a special case of instance-based learning that is associated with zero cost during the learning process [19].

The k-NN algorithm is suitable to classify EEG data as it is a robust technique for large noisy data [20]. The samples which is the data is classified by the majority vote of its neighbor's class. In order to determine the class, this algorithm requires training data and predefined k value as it will search through the training sample space for the k-most similar samples based on a similarity measure a distance function [21]. The value of k and distance metric will affect the result of classification [20].

Fig. 1 illustrates the concept of k-NN algorithm when applied the distance metric to determine the appropriate class of new data with k = 9. The data to be classified is at point (0.6, 0.45), which is shown with "X". The big circle with dot line is represented the distance metric using Euclidean distance computation. It has two possible classes which is circle class with six instances and triangle class with three instances. The algorithm will classify mark "X" to the circle class as the circle class have the majority of data within the radius.



**Fig. 1.** The classification of K-NN algorithm with using Euclidean distance and k=9

### 2.1. Value of k

The value of k is usually small and integer with positive value. If k = 5, class allocation of the sample is based on the nearest five neighbors within a certain distance. In this paper, the value of k is use from 1 to 15.

### 2.2. Distance metric

Distance metrics are a method to find distance between a new data point and existing training dataset [22]. There are 5 distance metrics used in this research which can be explained as follows:

#### 2.2.1 Manhattan / City Block Distance

The Manhattan distance or also be called as city block distance is the distance between two points  $(x, y)$  which the sum of the absolute difference of Cartesian coordinates, defined by:

$$d(x, y) = \sum_{i=1}^k |x_i - y_i| \quad (1)$$

#### 2.2.2 Euclidean Distance

The Euclidean distance is a measure to find distance between two points. In Cartesian coordinates, if  $x$  and  $y$  are two points in Euclidean k-space, then the distance (d) from  $x$  to  $y$  or from  $y$  to  $x$  is defined by the Pythagoras's theorem. The formula is as follow:

$$d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (2)$$

#### 2.2.3 Minkowski Distance

The Minkowski distance is a method to find distance between two points  $(x, y)$  in a normed vector space with various value of p. The distance is defined as follow:

$$d(x, y) = \left( \sum_{i=1}^k (|x_i - y_i|)^p \right)^{1/p} \quad (3)$$

For the special case of Minkowski distance: When p = 1, the Minkowski metric gives the Manhattan distance; when p = 2, the Minkowski metric gives the Euclidean distance; and when p = ∞, the Minkowski metric gives the Chebychev distance.

#### 2.2.4 Chebychev Distance

The Chebychev distance is a measure to find the distance between two vectors or points  $(x, y)$  which have the greatest of their differences along any coordinates dimension, defined by:

$$d(x, y) = \lim_{p \rightarrow \infty} \left( \sum_{i=1}^k |x_i - y_i|^p \right)^{1/p} \quad (4)$$

### 2.2.5 Hamming Distance

The Hamming distance, which is the percentage of coordinates that differ, can be defined by:

$$d(x, y) = \left( \frac{\#(x_i \neq y_i)}{n} \right) \quad (5)$$

## 3 K-NN for EEG Classification

### 3.1. Data description

The dataset used is Dataset 1 from BCI Competition IV [23] which contain recorded EEG data for motor imagery task perform by four healthy human subjects (a, b, f and g). In the experiment, each subject was asked to select two mental tasks to perform out of three tasks which are left hand, right hand or foot movements. The arrows pointing left, right or down is presented as visual cues on a computer screen as represented for left, right and foot movement respectively. The cues were displayed for a period of 4s which interleaved with 2s of blank screen and 2s with a fixation cross shown in the center of the screen. As in each run have 50 trials of each of the chosen two classes, give the total of 200 trials for each subject [24]. Eleven channels are used in this research, named as FC5, FCz, FC6, C5, C3, Cz, C4, C6, CP5, CPz and CP6.

### 3.2. Feature Extraction

The extraction of the wanted signals from the raw EEG signals and removing the irrelevant signals is called as feature extraction process as shown in Fig.2. The EEG data should be process in term of frequency as it is been classified based on frequency bands which are delta, theta, alpha, beta and gamma band. Therefore, the Fast Fourier Transform (FFT) is used to transform the time domain to frequency domain. The FFT features are then feed into k-NN classifier for classification process.

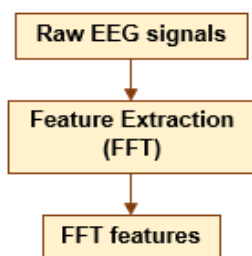


Fig. 2 Block diagram of the feature extraction process

As signals in time domain can be split into a group of sinusoids, the EEG signals which are lengthy and noisy can be easily convert into frequency domain. The hidden features can become visible after the conversion. The original signal can be restored by adding all the sinusoids up after FFT. Therefore, no information is lost.

### 3.3. Classification

In machine learning, classification is a prediction process to predict the categories or classes of the samples in the dataset by applying classification algorithm. The classifier used in this paper is k-NN classifier as shown in Fig. 3. This process is performed in python programming.

Five distance metrics: Manhattan distance, Euclidean distance, Minkowski distance, Chebyshev distance and Hamming distance are used to determine the best value of k to maximize the classification performance. The k-value is searched from 1 to 15. The classification is performed for all subjects and the result are expressed in percentage of accuracy.

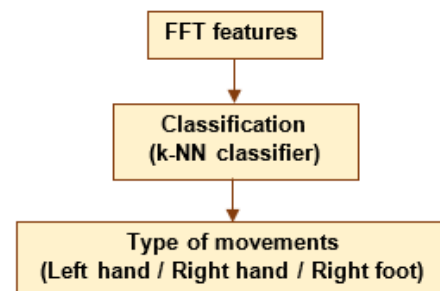


Fig. 3 Block diagram of the classification process

SVM is usually used in BCI research as classifier for this dataset [4-8]. Thus, in this paper, the k-NN is investigated and compared with SVM.

## 4 Result and Discussion

The classification is between two classes which either motor imagery of right hand and left hand or left hand and right foot. In this section, the result of EEG data classification by applying the k-NN classifier with various value of k and distance metrics are tabulated in tables and illustrated in graphs. The results have been analyzed for the performance analysis of k-NN classifier and verified by using 10-fold cross validation technique.

The results have been analyzed for the performance analysis of k-NN classifier. In order to prove the reliability of the classification results obtained, the results is verified using 10-fold cross validation technique. This process randomly divides the data into 10 folds for tested 10 times. The testing process is done on one of the folds while the remaining folds undergoes the training process. The process was repeated until all folds are used for testing and training the classifier. The accuracy of classification was calculated from the average of 10-fold cross validation.

### 4.1. K-NN Accuracy with Varying Parameters

The data classification has been performed with the same classifier which is k-NN with various value of k and

distance measurement. Table 1 shows the result of accuracy for all 4 subjects.

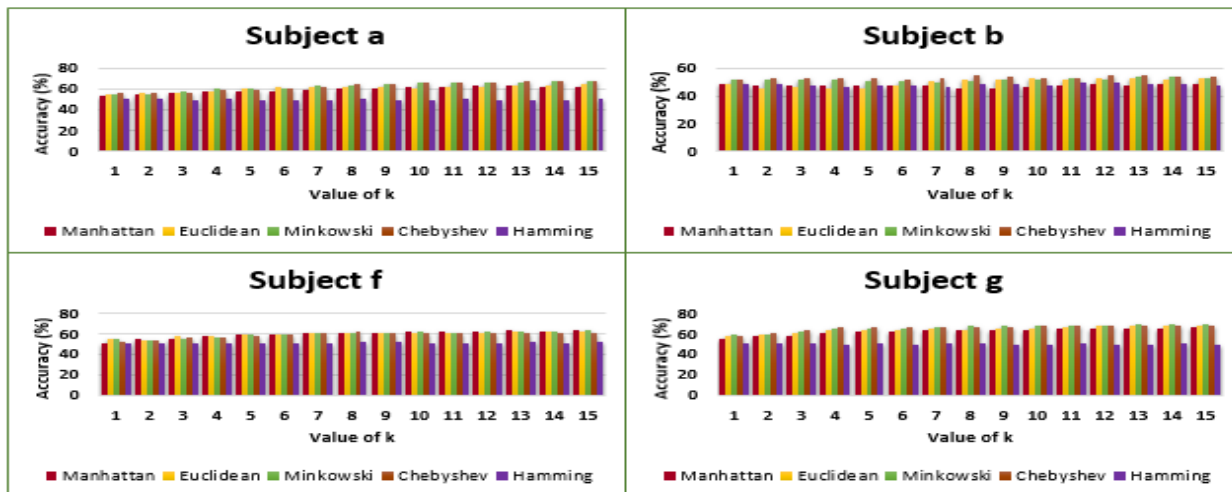
**Table 1.** Result for Subject a, f, b and g

Metric Distance	Value of k	Accuracy (%)			
		Left hand & Right foot		Left hand & Right hand	
		a	f	b	g
Manhattan Distance	1	53.52	51.21	48.15	55.62
	2	54.96	54.28	47.63	57.41
	3	56.98	55.17	48.04	57.85
	4	57.16	57.63	47.54	60.80
	5	57.23	59.02	47.47	61.75
	6	58.34	59.26	47.74	62.19
	7	59.42	60.37	48.01	63.46
	8	60.47	60.92	45.78	63.24
	9	60.96	61.03	45.75	63.86
	10	62.06	61.50	46.33	64.03
	11	62.57	61.90	47.79	64.74
	12	63.08	62.71	48.44	65.14
	13	62.78	63.13	47.88	64.57
	14	62.15	62.87	48.21	65.58
	15	62.46	63.13	48.79	66.37
Euclidean Distance	1	54.73	55.10	48.33	57.52
	2	56.24	53.28	45.76	58.79
	3	56.64	57.17	46.84	61.20
	4	58.24	58.05	45.37	63.77
	5	60.17	58.60	45.74	64.04
	6	61.73	59.98	47.82	63.64
	7	61.37	60.25	50.97	64.40
	8	62.51	60.33	51.74	63.93
	9	61.37	60.31	51.61	65.36
	10	61.10	61.31	53.00	65.55
	11	61.76	60.84	51.87	66.33
	12	62.54	60.99	52.46	67.80
	13	63.10	61.63	52.43	68.40
	14	64.03	61.96	52.05	68.28
	15	64.57	62.26	52.72	68.33
Minkowski Distance (p=11)	1	54.51	54.31	51.37	59.40
	2	55.49	53.71	52.11	59.32
	3	58.39	55.25	51.80	62.73
	4	60.67	56.98	51.59	65.30
	5	61.21	58.67	50.39	65.70
	6	61.22	59.53	50.44	65.70
	7	62.82	61.08	50.14	66.16
	8	63.35	60.80	50.44	67.45
	9	64.65	60.86	52.32	68.03
	10	65.61	61.60	52.09	68.32
	11	65.60	61.24	52.49	68.20
	12	66.15	61.74	51.98	68.54
	13	66.99	61.78	54.13	68.93
	14	67.67	62.07	53.87	69.71
	15	67.66	62.95	52.95	70.08

Chebyshev Distance	1	56.50	51.99	51.52	57.27
	2	56.80	52.90	52.64	60.99
	3	56.85	55.89	52.79	63.92
	4	59.03	56.46	52.99	67.07
	5	59.74	58.21	52.89	67.10
	6	60.79	58.94	51.95	66.69
	7	62.73	61.44	53.33	66.17
	8	64.22	61.91	54.92	65.79
	9	65.39	61.13	54.15	66.81
	10	66.03	61.22	53.39	67.75
	11	66.70	60.62	52.96	67.74
	12	66.73	61.42	54.77	67.78
	13	67.33	61.29	55.26	67.47
	14	67.07	61.09	53.98	67.69
	Hamming Distance	1	50.46	51.05	48.53
2		50.12	51.24	48.88	50.07
3		49.85	50.14	47.81	50.44
4		50.09	50.61	47.01	49.24
5		49.65	50.91	47.87	50.38
6		49.84	50.93	47.29	49.84
7		49.80	51.31	46.34	50.15
8		50.11	51.65	48.31	50.73
9		49.71	51.67	49.01	49.63
10		49.00	50.74	48.06	49.69
11		50.45	50.31	50.01	50.28
12		49.24	50.84	49.67	49.79
13		48.82	52.20	49.00	49.41
14		49.91	50.98	48.43	50.26
15		50.50	51.47	47.21	49.79

The left hand and right foot imaginary movements classification is done by subject a and f. For subject a, the highest accuracy is 67.67% when using Minkowski distance at  $p = 11$  and  $k = 14$ . For subject f, the highest accuracy is 63.13% when using Manhattan distance at  $k = 13, 15$ . Subject b and g is used to classify the right hand and left hand imaginary movements. For subject b, the highest accuracy is 53.87% when using Minkowski distance at  $p = 11$  and  $k = 14$ . For subject g, the best classification accuracy achieved is 70.08% when using Minkowski distance at  $p = 11$  and  $k = 15$ .

The comparison accuracy results between different distance metrics and value of k for motor imaginary task classification of left hand vs right foot imaginary movement and right hand vs left hand imaginary movement are illustrated in Fig. 4. Based on the overall results, the Minkowski distance at  $p = 11$  with the value of k between 13 and 15 give the highest accuracy compared to other distances which should be chosen to get a good performance for BCI motor imagery classification.



**Fig. 4** Comparison of motor imaginary task classification for left hand vs right foot movement (Subject a and Subject f) and right hand vs left hand movement (Subject b and Subject g) by using different distance metrics and value of k

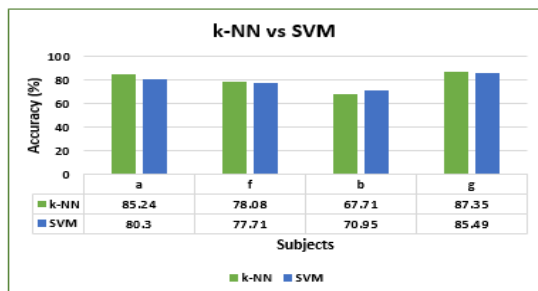
### 4.2 K-NN vs SVM

Although SVM classifier is commonly used for this dataset, the results shows that the k-NN outperformed most of the classification accuracy for subjects when the classification accuracy using SVM and k-NN is compared. The k-NN is very simple algorithm to implement and need least time compared to SVM.

For this comparison, Linear Discriminant Analysis (LDA) is applied before run the classifiers to reduce the data dimensionality in order to get optimum results. As the Minkowski distance at  $p = 11$  give the highest accuracy compared to other distances, this metric distance is used to be compare with SVM. Table 2 shows the result of classification accuracy of k-NN and SVM and Fig. 5 shows the graph of comparison between k-NN and SVM after applying the LDA.

**Table 2.** Results for classification accuracy of k-NN and SVM

Classifier	Accuracy (%)			
	Left hand & Right foot		Left hand & Right hand	
	a	f	b	g
k-NN	85.24	78.08	67.71	87.35
SVM	80.30	77.71	70.95	85.49



**Fig. 5** Comparison classification accuracy between k-NN and SVM

\* Corresponding author: nurulezzati@studentmail.unimap.edu.my

From the results obtained, it can be concluded that the k-NN algorithm is also suitable to use for classification method in BCI system as its result comparable to SVM.

### 5 Conclusion

The motor imagery classification of EEG signal for BCI applications is performed in this paper. Feature extraction of the signals is obtained using FFT, and the k-NN classifier is used to classify the data. The classification of motor imagery task using EEG data is performed to analyze performance analysis of k-NN. This paper demonstrates a detail the performance analysis of k-NN for EEG signal classification. The results of this research show that different accuracy of k-NN is obtained when applied with different k-value and distance metrics. This demonstrates that distance metrics and k-value affect the classifier performance and therefore worth to be considered in designing an EEG signal classifier. The results show that k-NN obtained comparable accuracy to SVM which commonly used for EEG signal classification. Given the expensive computational time of SVM, the lower complexity k-NN is promising for low-cost EEG classification. Since, the Hamming distance metrics perform badly on all datasets as having the lowest accuracy compared to others distance. Thus, Hamming distance not suitable for EEG data classification. In future work, more algorithms will be thoroughly investigated for EEG classification.

The Dataset 1 of BCI Competition IV is a property Berlin BCI group. We like to thank to Berlin BCI group by providing this dataset. In addition, this research is supported by Fundamental Research Grant Scheme (FRGS) from Ministry of Education, Malaysia. [Grant No: 9003-00525]

### References

1. S. A. Sprague, M. T. McBee, E. W. Sellers, *The*

- effects of working memory on brain-computer interface performance*, Clin. Neurophysiol. **127**, 1331–1341 (2016)
2. X. Tang, N. Zhang, J. Zhou, Q. Liu, *Hidden-layer visible deep stacking network optimized by PSO for motor imagery EEG recognition*, Neurocomputing **234**, 1–10 (2016)
  3. Denis R. de Moraes Piazzentin and J. L. G. Rosa, *Motor imagery classification for Brain-Computer Interfaces through a chaotic neural network through a chaotic neural network*, Int. Jt. Conf. Neural Networks (2014)
  4. M. E. A. Abdel-Hadi, R. A. El-Khoribi, M. I. Shoman, M. M. Refaey, *Classification of motor imagery tasks with LS-SVM in EEG-based self-paced BC*, 5th Int. Conf. Digit. Inf. Process. Commun. (CDIPC ), 244–249 (2015)
  5. M. Ming, L. Shaona, M. Haitao, M. Yuliang, G. Yunyuan, *Feature extraction method of motor imagery EEG based on DTCWT sample entropy*, 34th Chinese Control Conference (CCC), 3964–3968 (2015)
  6. J. Hong, X. Qin, , J. Bai, P. Zhang, Y. Cheng, *A combined feature extraction method for left-right hand motor imagery in BCI*, IEEE Int. Conf. on Mechatronics and Automation (ICMA), 2621–2625 (2015)
  7. Z. Qiu, J. Jin, H. K. Lam, Y. Zhang, X. Wang, A. Cichocki, *Improved SFFS method for channel selection in motor imagery based BCI*, Neurocomputing (2015)
  8. B. Yang, H. Li, Q. Wang, Y. Zhang, *Subject-based feature extraction by using fisher WPD-CSP in brain-computer interfaces*, Comput. Methods Programs Biomed. **129**, 21–28 (2016)
  9. J. Birjandtalab, M. B. Pouyan, D. Cogan, M. Nourani, J. Harvey, *Automated seizure detection using limited-channel EEG and non-linear dimension reduction*, Computers in Biology and Medicine **82**, 49-58 (2017)
  10. J. Kevric, A. Subasi, *Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system*, Biomedical Signal Processing and Control **31**, 398-406 (2017)
  11. T. Kayikcioglu, O. Aydemir, *A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data*, Pattern Recognit. Lett. **31**, 1207–1215 (2010)
  12. J. Fan, J. W. Wade, D. Bian, A. P. Key, Z. E. Warren, L. C. Mion, N Sarkar, *A Step towards EEG-based brain computer interface for autism intervention*, Proceedings of the Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS), 3767–3770 (2015)
  13. K. Sabancı, M. Köklü, *The Classification of Eye State by Using kNN and MLP Classification Models According to the EEG Signals*, Int. J. Intell. Syst. Appl. Eng. **3**, 127 (2015)
  14. A. Raheel, S. M. Anwar, M. Majid, B. Khan, Ehatisham-ul-Haq, *Real time text speller based on eye movement classification using wearable EEG sensors*, Proceedings of 2016 SAI Computing Conf., 161–164 (2016)
  15. R. Aldea, M. Fira, A. Lazor, *Classifications of motor imagery tasks using k-nearest neighbors*, 12th Symposium on Neural Network Applications in Electrical Engineering, NEUREL 2014, 115–120 (2015)
  16. F. A. Mousa, R. A. El-Khoribi, M. E. Shoman, *An integrated classification method for brain computer interface system*, 5th Int. Conf. on Digit. Inf. Process. and Commun., ICDIPC 2015, 141–146 (2015)
  17. M. Maleki, T. Kayıkçıoğlu, *Classification of EEG signal during gaze on the different rotating vanes*, 24th Signal Processing and Communication Application Conference, SIU 2016 - Proceedings, 593–596 (2016)
  18. J. T. Oliva, J. L. G. Rosa, *How an epileptic EEG segment, used as reference, can influence a cross-correlation classifier?*, Applied Intelligence, 1-19 (2017)
  19. S. Raschka, *Python Machine Learning*. (Packt Publishing Ltd., (2016)
  20. R. Bose, A. Khasnobish, S. Bhaduri, D. N. Tibarewala, *Performance Analysis of Left and Right Lower Limb Movement Classification from EEG*, 3rd Int. Conf. on Signal Processing and Integrated Networks (SPIN), 174–179 (2016)
  21. M. Mohammadpour, M. Ghorbanian, S. Mozaffari, *Comparison of EEG Signal Features and Ensemble Learning Methods for Motor Imagery Classification*, 8th Int. Conf. on Information and Knowledge Technology (IKT), Hamedan, Iran, 288–292 (2016)
  22. K. Chomboon, P. Chujai, P. Teerarassamee, K. Kerdprasop, N. Kerdprasop, *An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm*, The Proceedings of the 2nd Int. Conf. on Industrial Application Engineering 2015, 280–285 (2015)
  23. B. Blankertz, G. Dornhege, M. Krauledat, K. R. Müller, G. Curio, *The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects*, Neuroimage **37**, 539–550 (2007)
  24. M. Tangermann, K. R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, B. Blankertz, *Review of the BCI competition IV*. Frontiers in Neuroscience (2012)