Classification of brain activity using brainwaves recorded by consumer-grade EEG devices

Elizaveta Kosiachenko, Dong Si

Computing and Software Systems, University of Washington Bothell

(elizavet@uw.edu, dongsi@uw.edu)

Abstract. Despite the significant amount of research being conducted in interpreting electroencephalography (EEG) brainwave recordings using machine learning techniques, most of it is done using multi-electrode, medical grade EEG systems. However, these devices are expensive and hard to use by non-professional users. Our research focuses on analyzing the feasibility of using low-cost consumer grade EEG devices for classifying the type of activities performed by the user and for identifying a person based on their EEG measurements. This research is important as consumer grade EEG devices can potentially have a much wider application for the purposes of brain-computer interface than the lab alternatives. We have implemented k-nearest neighbor, decision tree and support vector machine algorithms in our analysis. The results show that consumer-grade EEG devices can be successfully used in some domains to both classify the type of activity performed by a person and to identify people based on their brainwave readings.

Keywords. Brain activity, Consumer-grade non-invasive electroencephalography (EEG), Classification, k-nearest neighbor, decision tree and support vector machine.

1 Introduction

Brain-computer interface is a very popular and fast-developing topic nowadays. One of the most common ways to implement such kind of interface is by using a non-invasive electroencephalography (EEG). There are quite a few research papers exploring the application of modern machine learning techniques to analyze and interpret EEG data: to perform emotional state classification [2], to identify depression patients [3], to control computer cursor [6] and even to use brain electrical activity as a biometric [5]. However, most of the research papers identified perform their experiments and evaluation using multi-electrode, medical grade EEG systems [3], [4], [5], [7], and only very few works focus on low-cost consumer grade EEG devices [8].

The recent availability of such inexpensive EEG devices makes it feasible to take this technology outside of the laboratory and bring it into real-world environment. The benefits of such devices are affordability and ease of use. Considering the current availability of many different commercial consumer-grade EEG devices and their continuous development and update, there is a need in exploring the feasibility of using such low-cost EEG devices for monitoring individuals' EEG signals in their natural environment and making meaningful predictions based on this data.

The goal of this research article is to analyze how the consumer-grade EEG devices could be used for the following two tasks: 1) classifying the type of activity for a particular person based on their EEG reading; and 2) identifying a person based on their EEG measurements while doing a particular activity.

This research is based on the dataset generated by the BioSENSE group in the UC Berkeley using "Mindwave" consumer-grade EEG devices [1], which are produced by the Neurosky (`http://neurosky.com/`). The dataset was published on the UC Berkeley website (`http://biosense.berkeley.edu/indra_mids_5_15_dlpage/`).

2 Method & Approach

The high-level representation of the workflow of the experiments performed is as follows:

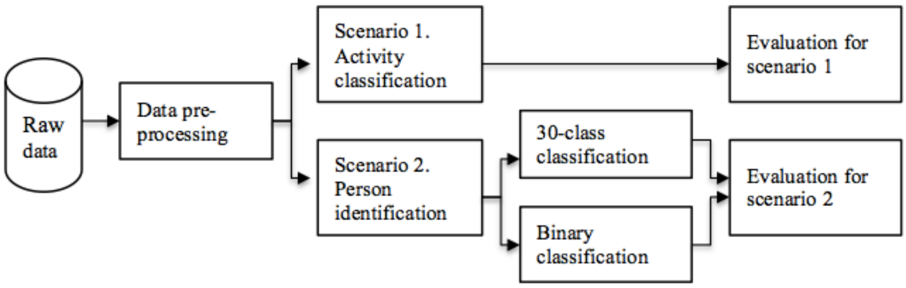

Fig. 1. Experiment workflow

As mentioned above, the dataset analyzed in this paper is provided by the UC Berkeley researchers[1]. According to the description of the experiment, the data was collected using consumer-grade brainwave-sensing headsets "Mindwave" by Neurosky during an in-class group exercise. As part of the exercise participants were presented with various audio-visual stimulus. The dataset includes all subjects' readings during the stimulus presentation, as well as readings from before the start and after the end of the stimulus.

The types of exercises and stimulus presented to the participants are:

- blinking
- relaxing (with eyes closed)
- doing math
- listening to music (with eyes closed)
- watching video (TV commercial)

- thinking of items belonging to a particular category (with eyes closed)
- counting color blocks

The initial dataset contains 29,480 data items and 8 attributes. Each data item represents one reading of the brainwave activity. During the experiment the readings were collected every second from each "Mindwave" device. Data attributes collected during each reading are as follows: id, indra_time, browser_latency, reading_time, attention_esense, meditation_esense, eeg_power, raw_values, signal_quality, createdAt, updatedAt.

Data attributes represent the following information:

- **id**: Integer value in the range of 1 to 30 representing the participant number.
- **indra_time**: The synchronized timing of the reading.
- **browser_latency**: The difference between the time on the subject's computer, and the time on the server. This value is used to calculate the synchronized indra_time, above.
- **reading_time**: The time at which the Neurosky data passed through the bluetooth connection onto the subject's computer.
- **raw_values**: Tuple containing raw sample values acquired by the sensor, at a sampling rate of 512Hz (defined by the Neurosky SDK).
- **attention_esense** and **meditation_esense**: Neurosky's eSense meters for Attention and Meditation levels, in integer values in the range of 0 to 100. The values represent the last values computed in the current period (defined by the Neurosky SDK).
- **eeg_power**: Tuple represents the magnitude of 8 commonly-recognized types of EEG frequency bands -- delta (0.5 - 2.75Hz), theta (3.5 - 6.75Hz), low-alpha (7.5 - 9.25Hz), high-alpha (10 - 11.75Hz), low-beta (13 - 16.75Hz), high-beta (18 - 29.75Hz), low-gamma (31 - 39.75Hz), and mid-gamma (41 - 49.75Hz). These values have no units (defined by the Neurosky SDK).
- **signal_quality**: A zero value indicates good signal quality. A value of 128 greater corresponds to a situation where the headset is not being worn properly (defined by the Neurosky SDK).

Brainwave data is characterized by significant subject-to-subject variability [5], so it was reasonable to perform the classification by the type of activity only in the context of each particular person. We focused on data pre-processing and prediction on two scenarios:

1. Classifying the type of activity for a particular person based on their EEG reading;
2. Identifying a person based on their EEG measurements while doing a particular activity.

Both scenarios outlined above are supervised learning problems, various machine learning algorithms were implemented for this classification task. All pre-processing and classification was performed using Python programming language and a set of libraries for data manipulation and machine learning ("numpy", "pandas", "sklearn", "re", and "pylab").

2.1 Data pre-processing

To prepare the data for training classification models and clear it from noisy readings, the following steps were performed:

- Removed unused columns (data attributes), such as indra_time, browser_latency, reading_time (because activity labels are already assigned to the data readings, time information is irrelevant);
- Removed the "raw_values" column, as, according to the manufacturer, the data from the "raw_values" column is aggregated into 8 EEG bands represented in the "eeg_power" column used in this research;
- Removed the "attention_esense" and "meditation_esense", as these attributes represent the emotional state of the subject and our research aim to classify between tasks performed by the subjects;
- Split "eeg_power" into 8 columns, which represent 8 commonly-recognized types of EEG frequency bands (delta (0.5 – 2.75Hz), theta (3.5 - 6.75Hz), low-alpha (7.5 - 9.25Hz), high-alpha (10 - 11.75Hz), low-beta (13 - 16.75Hz), high-beta (18 - 29.75Hz), low-gamma (31 - 39.75Hz), and mid-gamma (41 - 49.75Hz));
- Filtered out all the entries with "signal quality" > 128, which represent the readings taken when the signal quality was poor or when the device wasn't worn properly;
- Deleted all entries with labels "unlabeled" and "instructionsForX";
- Renamed labels like "blink1", "blink2", etc. into just "blink".

2.2 Methods & Approach for Scenario One – Activity Classification

The aim of the first scenario was classifying the type of activity for a particular person based on their EEG reading.

For the purposes of analyzing this scenario, we selected two distinct activities out of those performed by the subjects of the original research: relaxation with eyes closed and color round (i.e. counting color blocks displayed on the screen). So, the first step in this scenario was to filter out only the readings with the values "relax" and "colorRound" in the column "label".

Then the following steps were performed for each of the participants in the dataset. First, we filtered out other readings and kept the readings related only to this participant. Second, we trained the k-nearest neighbor (kNN)[1], decision tree[2] and support

---

[1] In k-nearest neighbor(KNN) classification, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. Distance-weighted k-nearest neighbor classifier also uses a distance function to weight the evidence of a neighbor depending on their distance from an unclassified observation [9].

[2] Decision trees are a supervised learning method, which maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features [10].

vector machine (SVM)[3] with linear and polynomial kernel models. Finally, the 5-fold cross-validation was performed to assess the quality of model prediction.

For the k-nearest neighbor model we conducted a series of experiments with different numbers of k and selected the one that gave the best accuracy of the prediction (k=25). We also experimented with the different degrees of the polynomial kernel function and the best result for this problem was produced by the degree of 3.

Because the number of data items in the two classes selected ("relax" and "colorRound") is not equal, we used the stratified 5-fold cross-validation as a method of model assessments.

2.3 Methods & Approach for Scenario Two – Person Identification

The aim of the second scenario was to predict the person identification based on their EEG measurements while doing a particular activity. Two different experiments - 30-class classification and binary classification - were designed to achieve that goal.

30-class classification. In this experiment, each class represents readings for a particular person. We selected one of the activities' labels ("math") and filtered out all of the readings with other labels. We then trained the following models to classify the EEG readings between 30 "id" labels: k-nearest neighbor (KNN), decision tree and SVM with linear and polynomial kernels. A 5-fold cross-validation was then performed to assess the accuracy of each model.

Binary classification. Because 30-class classification provided poor prediction accuracy (for details see the section 3.2. Results of Scenario Two), we designed an alternative experiment in hopes of getting better prediction results. For each of the participants and most of the activities in the dataset (excluding blinking activity due to low number of data readings belonging to this class), the following steps were performed:

- Create a new dataset and only copy the readings related to the activity that is currently analyzed from the full pre-processed dataset;
- Substitute all the "id" labels for the readings related to the participant in question with '1's;
- Substitute all other participants "id" labels with '0' and randomly delete rows from this category until the number of rows with '0' in the 'id' column equals the number of rows with '1's;
- Use the resulting data subset to: :
  - Train the k-nearest neighbor (KNN), decision tree and SVM with linear and polynomial kernel models to classify between the two classes = [0, 1];
  - Assess the accuracy using stratified 5-fold cross-validation.

[3] Support Vector Machines (SVMs) are hyperplanes that separate the training data by a maximal margin. All vectors lying on one side of the hyperplane are labeled as one class, and all vectors lying on the other side are labeled as another class. SVMs allow to project the original training data in space to a higher dimensional feature space via a kernel operator [11].

We also performed a grid search for the best k in the k-nearest neighbor algorithm separately for each of the activities, but across subjects. We also tried various degrees of the polynomial kernel in the SVM model and noticed that the best prediction for this scenario was achieved by using the quadratic kernel (degree of 2).

3 Experiment Results

3.1 Results of Scenario One

The results of running the abovementioned models (KNN, Decision tree, SVM – linear kernel and SVM – polynomial kernel) for each of the participants are presented in the table below. The metric used for assessing the quality of prediction is accuracy, which for this scenario is defined as the sum of correctly predicted readings from the "relax" class and correctly predicted readings from the "colorRound" class divided by the total number of observations from both classes.

Table 1. Accuracy of models predictions for Scenario One

	KNN	**Decision Tree**	**SVM (linear)**	**SVM (polyn.)**
Subject 1	73.37%	62.46%	72.46%	72.46%
Subject 2	72.46%	67.79%	72.46%	72.46%
Subject 3	77.10%	74.24%	75.27%	75.27%
Subject 4	73.39%	63.29%	76.20%	77.03%
Subject 5	84.48%	72.39%	82.57%	83.36%
Subject 6	72.46%	55.93%	72.46%	72.46%
Subject 7	76.95%	74.63%	76.20%	70.66%
Subject 8	72.90%	67.09%	72.90%	71.94%
Subject 9	72.46%	60.51%	72.46%	72.46%
Subject 10	73.11%	77.70%	77.87%	78.78%
Subject 11	76.12%	76.99%	81.66%	78.06%
Subject 12	70.38%	66.66%	70.38%	70.30%
Subject 13	91.73%	87.22%	91.68%	92.72%
Subject 14	74.15%	64.02%	75.02%	72.25%
Subject 15	71.14%	58.76%	72.14%	72.14%
Subject 16	73.85%	64.54%	73.76%	70.04%
Subject 17	71.31%	54.63%	71.31%	71.31%
Subject 18	69.22%	60.56%	71.94%	71.94%
Subject 19	78.85%	68.58%	71.31%	71.31%
Subject 20	71.90%	74.80%	74.67%	73.85%
Subject 21	80.43%	72.77%	82.20%	79.43%
Subject 22	71.94%	73.93%	76.58%	73.85%

Subject 23	72.85%	68.22%	72.90%	71.94%
Subject 24	71.90%	66.23%	70.04%	71.94%
Subject 25	71.14%	50.00%	71.14%	71.14%
Subject 26	72.26%	61.34%	72.31%	73.09%
Subject 27	75.75%	86.01%	84.02%	74.80%
Subject 28	71.94%	60.77%	71.94%	71.94%
Subject 29	93.42%	92.53%	78.98%	76.65%
Subject 30	71.57%	64.21%	74.30%	70.66%
Average accuracy of the model among the participants	**75.02%**	**68.30%**	**75.31%**	**74.21%**

As shown in the Table 1, the best average accuracy of prediction of 75.31% is provided by the Support Vector Machine model with the linear kernel. However, there is quite a large variance in the quality of prediction between the participants. For example, for Subject 29 the best model has an 93.42% accuracy of prediction and for Subject 28 the best prediction accuracy is only 71.94%. One of the explanations for such a variance is that some participants might have followed the instructions and focused on the exercise better then the others.

In general, the prediction accuracy of 75.31% is a result, which can provide some useful information (for example, it can be used in a gaming environment where incorrect prediction is not critical), however this accuracy is not good enough for such predictions to be the basis of some important decisions.

3.2 Results of Scenario Two:

As mentioned above, for the purposes of Scenario Two we designed two different experiments: 30-class classification and binary classification.

30-class classification result. Unfortunately, after training the models (KNN, Decision tree, SVM – linear kernel, SVM – polynomial kernel) to classify the EEG readings between 30 'id' labels, the accuracy of prediction provided by these models was pretty low:

Table 2. Accuracy of models predictions for Scenario Two – 30-class classification

Model	Prediction accuracy (using 5-fold cross-validation)
K-nearest neighbor (k = 4)	25.10%
Decision tree	22.45%
SVM – linear kernel	22.55%
SVM – polynomial(quadratic) kernel	15.05%

The accuracy of 25.10%, demonstrated by the k-nearest neighbor model, although is better than the random guess (1/30 = 3.33%), can not realistically be used to solve any real-world problems.

Binary classification. In the hopes of getting a better performance than the 30-class classification we designed another experiment to classify brainwave readings as belonging to a certain individual or not – binary classification (see the details of algorithm steps in section 2.3. Methods & Approach for Scenario Two). The accuracy, precision and recall of different models trained on the data from different activities are presented in the figures below.

For this scenario accuracy is defined as the sum of readings correctly predicted as belonging to the individual and correctly predicted as not belonging to the individual, divided by the total number of readings from both of these classes. We define precision to be the number of readings correctly predicted as belonging to the individual divided by the total number of readings predicted as belonging to the individual. We define recall to be the number of readings correctly predicted as belonging to the individual divided by the total number of readings actually belonging to the individual being analyzed.

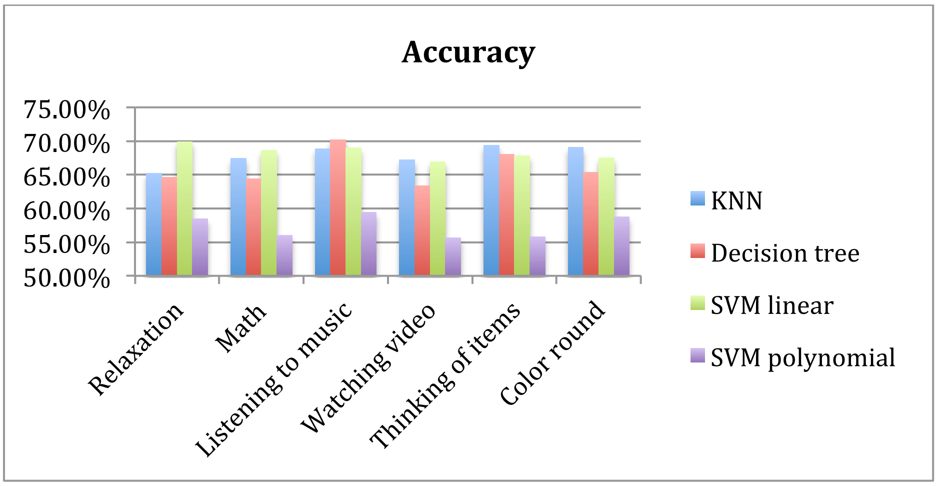

Fig. 2. Average accuracy of models for Scenario Two – binary classification

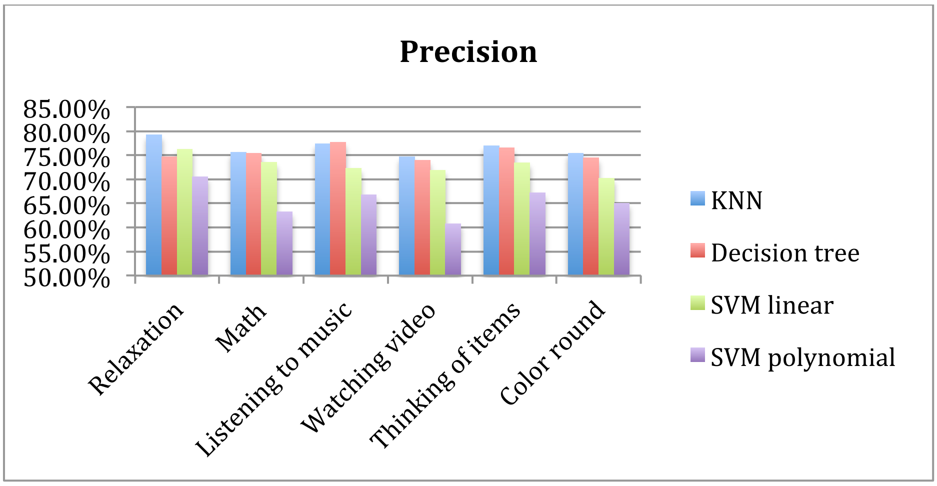

Fig. 3. Average precision of models for Scenario Two – binary classification

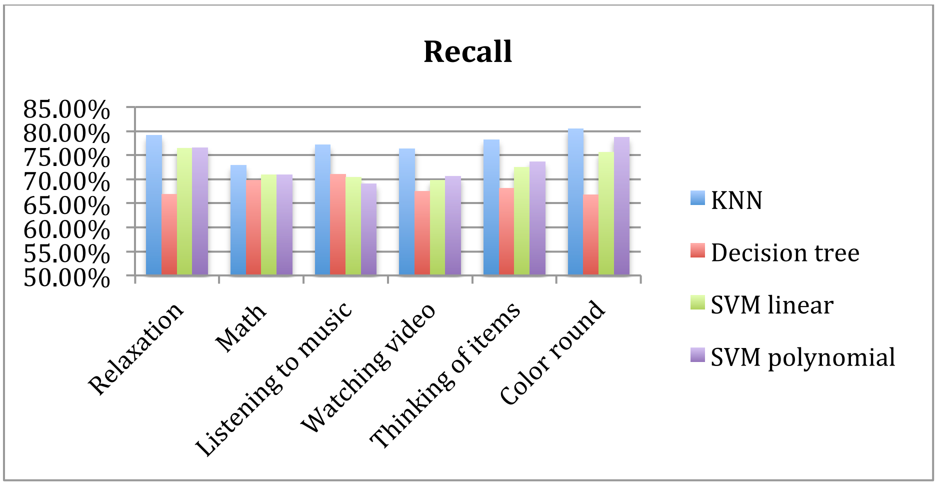

Fig. 4. Average recall of models for Scenario Two – binary classification

As we can see from the Figures 2 to 4, binary classification does provide better results than 30-class classification and proves that consumer-grade EEG devices work. The best accuracy of prediction is provided by applying the decision tree algorithm to the brainwave data collected while participants were listening to music and equals 70.18%. Curiously, the worst activity for identifying people, based on the results of our experiment, is watching video (in the original experiment participants were watching TV commercial). The best accuracy of prediction for the activity of watching video is only 67.23% and is achieved using k-nearest neighbor algorithm with k=8.

However, the nature of the applications that might use person identification from the brainwave recordings would probably require a very high accuracy of prediction, which so far is not achieved by the methods described.

There are two factors that might negatively influence the accuracy of prediction. The first is a relatively small dataset for the purpose of this scenario. We noticed that, despite the fact that the original dataset contained about 30,000 reading, after clearing the data, filtering out irrelevant data and splitting readings by participants and by activities, the amount of data per participant per activity was only about 30 readings, which is not ideal for making highly accurate and precise predictions. And the second factor could be the imperfections of the underlying technology used in the consumer-grade EEG devices used in the research.

4 Conclusion and Discussion

Based on the results from Scenario One, the data from consumer-grade EEG devices produced by the Neurosky, can be successfully used to classify the type of activity for a particular person based on their EEG reading. The results vary greatly from person to person, but the average accuracy of 75.31% allows this prediction to be used in some potential domains, such as gaming or other entertainment.

Scenario Two had a slightly lower quality of prediction with the best average accuracy provided by the 'music' activity at 70.18% and the worst average accuracy provided by the 'video' activity at 67.23%. This result can still be used in some applications and proves that the consumer EEG devices are capable of distinguishing between the brainwaves of different people. However, the nature of applications that use person identification requires a very high accuracy of prediction, which is not delivered by the dataset at hand. So the application of the results of the second experiment is quite limited.

We expect the current fairly good accuracy of prediction could be further improved in the future when the technology used in the consumer-grade EEG devices is improved and more high quality data are collected.

The commercial-grade EEG devices are still a developing technology that needs improvement in the quality of brainwave identification and recording in order to be used to successfully classify brainwaves by activity or to be used to identify people based on their EEG readings. However, these devices are developing really fast, so it is promising that this technology will have a big potential.

For the future work, we would like to continue working on the two scenarios outlined in this paper, in particular to look into improving models' accuracies as new generations of consumer-grade EEG devices are being released. After more and more good quality data has been generated from the devices, some advanced prediction models such as Artificial Neural Networks could be implemented. We think it will be a very interesting and promising concept with wide array of potential applications in brain-computer interface domain.

5 References.

1. Chuang, J., Merrill, N., Maillart, T.: Synchronized Brainwave Recordings from a Group Presented with a Common Audio-Visual Stimulus. UC Berkeley Spring 2015 MIDS Immersion Class. (2015).
2. Wang, X., Nie, D., Lu, B.: Emotional state classification from EEG data using machine learning approach. Neurocomputing. 129, 94-106 (2014).
3. Hosseinifard, B., Moradi, M., Rostami, R.: Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. Computer Methods and Programs in Biomedicine. 109, 339-345 (2013).
4. Blankertz, B., Dornhege, G., Krauledat, M., Muller, K., Kunzmann, V., Losch, F., Curio, G.: The Berlin Brain–Computer Interface: EEG-Based Communication Without Subject Training. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 14, 147-152 (2006).
5. Palaniappan, R.Mandic, D.: Biometrics from Brain Electrical Activity: A Machine Learning Approach. IEEE Transactions on Pattern Analysis and Machine Intelligence. 29, 738-742 (2007).
6. Kostov, A.Polak, M.: Parallel man-machine training in development of EEG-based cursor control. IEEE Transactions on Rehabilitation Engineering. 8, 203-205 (2000).
7. Müller, K., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B.: Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. Journal of Neuroscience Methods. 167, 82-90 (2008).
8. Maskeliunas, R., Damasevicius, R., Martisius, I., Vasiljevas, M.: Consumer grade EEG devices: are they usable for control tasks?. PeerJ. 4, e1746 (2016).
9. Dudani, S.: The Distance-Weighted k-Nearest-Neighbor Rule. IEEE Transactions on Systems, Man, and Cybernetics. SMC-6, 325-327 (1976).
10. Quinlan, J.: Induction of decision trees. New South Wales Institute of Technology, School of Computing Sciences, [Broadway, N.S.W., Australia] (1985).
11. Abe, S.: Support vector machines for pattern classification. Springer, London (2005).