# Feature Selection Techniques Analysis for Identification of Cognitive and Resting States Based on EEG Data

M. Sh. Murtazina, T. V. Avdeenko
Novosibirsk State Technical University, Karl Marx Avenue 20, Novosibirsk, Russia, 630073

## Introduction

In this paper, we will consider the application of feature selection methods included in the Weka software package to solve the problem of classifying cognitive states and resting states by EEG data.

Feature selection is an important step in classifying mental states from EEG data. In practice, the most significant features for the task of classifying mental states according to EEG data are a priori unknown. For this reason, a large number of features extracted from the EEG signals from each scalp electrode can be included in the feature space. The number of electrodes is determined by the objectives of the study and the equipment used.

## EEG

EEG is now a widely used instrument for recording brain waves. To record the EEG, a number of electrodes are placed on the scalp at the specific points, as well as reference electrodes (usually located on the earlobes) and a ground electrode (usually located on the forehead). The main characteristics of an EEG are frequency and amplitude.

The raw EEG recordings are usually stored in EDF format.

Electrodes are placed on the scalp in accordance with one of the standard international electrode placement schemes, such as 10-20, 10-10 and 10-5 systems. The name of the system indicates the interelectrode distances used in the application of electrodes. The electrode placement scheme according to the 10-20 system is shown in Fig.1.

As features for EEG data, indicators from areas related to time series analysis can be used, such as power spectral density from signal processing, entropy from information theory, and so on.

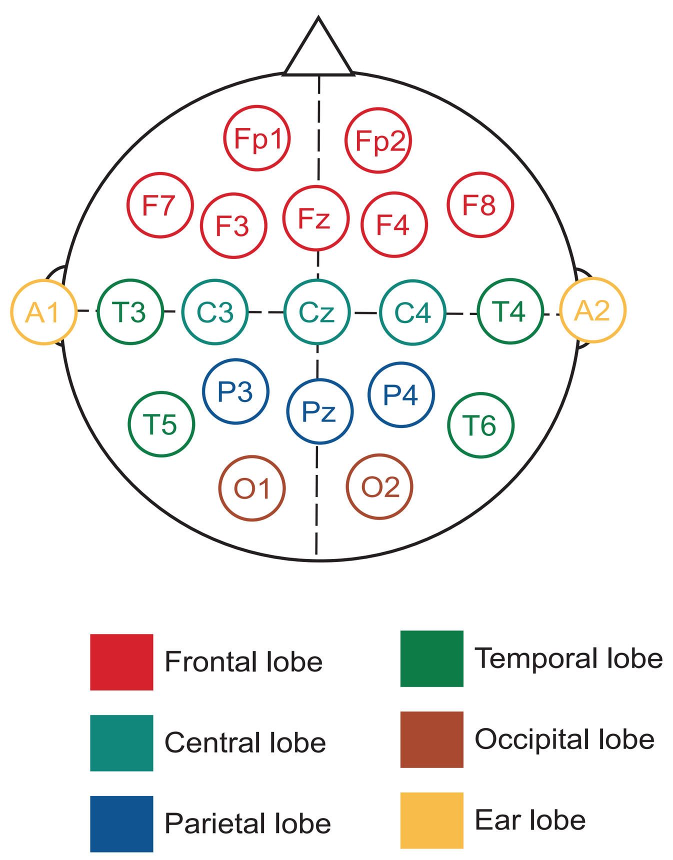


**Fig. 1.** The electrode placement scheme according to the 10-20 system

## Feature Selection in Weka

Free data analysis and machine learning software Weka was used for feature selection and classifier training. This software product can run on any platform because it is a java application.

This study uses the attribute subset evaluators CfsSubsetEval and WrapperSubsetEval. The first attribute evaluator estimates the significance of a subset of features, taking into account the individual predictive power of each feature, as well as the degree of redundancy between them. Second, we evaluate the effectiveness of a feature subset, taking into account the result of the applied classification algorithm to evaluate a set of features. The search strategy may use a different algorithm than what will then be used to train the classifying model.

## Features Formation

For the experiment, the data set "EEG During Mental Arithmetic Tasks" was used, which includes EEG recordings at rest (three minutes) and arithmetic calculations (one minute) for 36 subjects. Files with electroencephalogram records are presented in the data set in EDF format.

To extract features, data from 19 scalp electrodes were used in accordance with the international scheme 10-20. From the EEG records for each electrode, Hjort parameters (activity, mobility and complexity) and power in the delta, theta, alpha and beta frequency ranges were extracted. The power spectral density (PSD) was estimated using the Welch method. The window size for symptom recovery was one minute.

From the original set, EEG recordings for 34 subjects were used because the lengths of EEG recordings at rest in two subjects were less than three minutes. Four instances of data were obtained for each subject. Three of them correspond to the state of rest with closed eyes, one corresponds to the state associated with the cognitive load due to the performance of arithmetic calculations.

To extract features from EEG recordings, the eeglib Python library was used, which provides tools that allow the user to easily create a whole dataset that is compatible with most Python data analysis libraries. This possibility is provided by the creation of a DataFrame object with extractable features. Python has been used to automate the process of creating an ARFF dataset from EEG data. The resulting set includes 136 data instances. For each data instance, 133 features for classification and a target feature are specified.

## Results

As part of the experiment, 25 machine learning algorithms included in the Weka software package were selected, for which the classification accuracy exceeded 75% when trained on the original feature space (133 features for classification). Further, the feature selection was carried out using two methods. The first one uses a combination of the CfsSubsetEval attribute evaluator and the GreedyStepwise search method, the second one uses a combination of the WrapperSubsetEval attribute evaluator with the Logistic classifier and the GreedyStepwise search method.

Using the attribute evaluator CfsSubsetEval, the following subset of features was obtained: Hjort activity for the O1 channel, Hjort complexity for the F8 channel, delta band power for the Fp1 channel, theta band power for the Fp1 channel, delta band power for the F8 channel, beta band power for the C3 channel and alpha range power for the P4 channel.

When using the attribute evaluator WrapperSubsetEval, the following were selected: Hjort complexity for the F8 channel, Hjort mobility for the F8 channel, Hjort mobility for the P4 channel, delta band power for the Fp1 channel, delta band power for the Fp1 channel, delta band power for the Fp2 channel, delta band power for F8 channel, alpha range power for T6 channel.

Classifiers were trained again on the obtained subsets of features. The results are shown in Fig. 2.

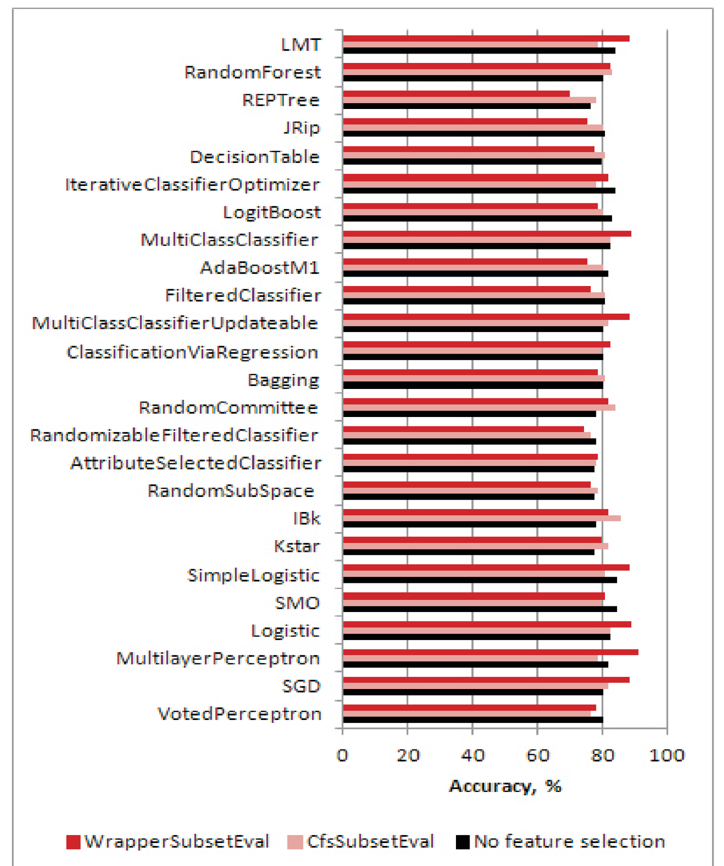


**Fig. 2.** Accuracy of classifiers obtained in WEKA using default settings

Feature selection by the first method was successful and made it possible to improve the accuracy of the classifier for 12 classification algorithms, and by the second method for 13 algorithms

At the same time, the best result by the first method was obtained for the IBk algorithm (the accuracy of the classifier changed from 77.94% to 85.29%). The best result by the second method was obtained for the MultilayerPerceptron algorithm (classifier accuracy changed from 81.61% to 91.17%).

## Conclusion

In this paper, we explored the feature selection methods included in the Weka software package that uses attribute evaluators to reduce the dimension of the feature space. As part of the experimental part, the attribute evaluators CfsSubsetEval and WrapperSubsetEval(Logistic) were explored in combination with the GreedyStepwise search method. Experimental results have shown that WrapperSubsetEval in general enables to get a greater increase in the accuracy of classifiers. Further research should be associated with the search for optimal settings for attribute evaluators that take into account the classification algorithm when assessing the significance of features.

VIII International Conference on Information Technology and Nanotechnology (ITNT-2022), 23–27 May, 2022, Samara