Feature Selection Methods Comparison for EEG-based Classifier Constructed Using Discrete Wavelet Transform Features

B Tsybenov, M Svetlakov and I Hodashinsky
Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia

E-mail: svetlakov.m4@gmail.com

Abstract. The paper presents the results of a study in the application of electroencephalography (EEG) for user authentication using discrete wavelet transform. The Leipzig Study for Mind-Body-Emotion Interactions dataset (LEMON) was used. Mean value, standard deviation, and root mean square value are used as features. Feature selection methods based on correlation, on mutual information, and on the χ2 criterion are used for reduce feature space. The SVM model is used for classification. The efficiency of constructed classifier has been tested using cross-validation procedure. Classifier built on feature reduced data via mutual information criteria have improved accuracy (97.4%) with feature space nearly halved (183 features) compared to baseline classifier.

1. Introduction
Nowadays, the most common authentication system is password-based authentication. A user gains access to work by entering a password into the computer, for example, in the form of a predefined word to confirm his identity. The reliability of a password is determined by its length and restrictions on the characters used, so users need to form long sequences to use a strong password, which is not very convenient [1]. The use of password authentication with strong password to confirm the identity of the user in the process of performing any activities is undesirable, due to the inability to implement background verification, without attracting the attention of the user [2].

An alternative to password authentication is biometric authentication [3]. This method uses the unique biological characteristics of a person for recognition. There are some of the most commonly used biometric features such as hand or palm print, the drawing of the eye ring, the timbre and spectral image of the voice, the image of the face, the drawing of the signature [4]. There are some requirements that need to be fulfilled so that biometric signs can be used in real conditions. In particular, the signs should be universal, constant and measurable, and identification systems should have a high performance and recognize the person with sufficient accuracy for practical application [5].

Based on the work [6], where the coefficients of the wavelet transformation of one level were considered and the accuracy of classification on large datasets significantly decreased, it was hypothesized that to increase the accuracy it is necessary to consider the coefficients from all levels.

However, in this case, the number of features increases significantly, which significantly slows down the algorithm. To speed up the algorithm and increase the accuracy of classification, it is necessary to consider methods of feature selection [7].
The Leipzig Study for Mind-Body-Emotion Interactions dataset (LEMON) was taken for the study. The participants of the dataset were thoroughly examined for physical and psychological condition by doctors and represented different age groups [8].

2. Dataset
The LEMON data were collected in 2013-2015 from the Day Clinic for Cognitive Neurology of the University Clinic Leipzig and the Max Planck Institute for Human and Cognitive and Brain Sciences (MPI CBS). The dataset was published in 2019 and consists of 227 participants.

EEG at rest was measured for 216 participants only, where each participant's EEG recordings were kept in sessions of 16 minutes. Each session consisted of 16 blocks of 60 seconds, where 8 blocks were recorded with eyes open and 8 with eyes closed. A 10-20 scheme was used to attach the electrodes.

3. Data preprocessing
Dataset developers excluded data of 13 participants (some information about events was missing) and some channels (poor signal quality). Then the decimation of the signal was performed and a bandpass filter (8-order Butterworth filter) was applied. Also we used Principal Component Analysis (PCA) and Independent Component Analysis (ICA).

We further reduced the number of channels to 17 (AF3, C2, C3, CP3, CP5, F4, Oz, P1, P2, P4, P5, P6, P8, PO3, PO4, PO8, Pz), the data from the channels were divided into epochs of 2 seconds duration. This resulted in records of 203 participants from 17 channels, where each channel has 480 epochs.

4. Feature extraction
A discrete wavelet transform (DWT) was applied for feature extraction. The discrete wavelet transform decomposes the signal into basic wavelet functions (parent wavelets). It applies 2 filters to split the signal into approximating and detailing coefficients, representing low and high frequencies respectively. In the end, the wavelet transform can be described as follows:

\[ WT_i = \int x(t) \psi_{j,k}(t) d(t) \quad (1) \]

where \( x(t) \) is the signal, \( \psi \) is scaled discrete wavelet function and * denotes complex conjugate.

In this paper, a 4th order Daubechies wavelet is used to decompose the signal into 5 levels. Such decomposition allows one to extract certain frequencies of the EEG signal corresponding to "brain rhythms" (Alpha (8-16 Hz), Beta (16-32 Hz), Gamma (32-64), Delta (0.5-4 Hz), Theta (4-8 Hz)). So all 5 levels will be used to extract 3 statistical features from each channel:

\[ STD = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - M)^2 \quad (2) \]

\[ M = \frac{1}{N} \sum_{i=1}^{N} X_i \quad (3) \]

\[ RMS = \left( \frac{1}{N} \sum_{i=1}^{N} |X_i|^2 \right)^{1/2} \quad (4) \]

where \( X_i \) is coefficient obtained from one of the channels using DWT, \( N \) is the amount of such coefficients for one channel.

5. Classifier
In this paper we use Support Vector Machine (SVM) as classifier with following parameters:

- tolerance = 1e-6
- penalty = 12
- max_iter = 2000
- C = 0.15
- kernel = “linear”

6. Feature Selection

Four methods were considered for feature selection: based on correlation, mutual information, $\chi^2$ criterion and analysis of variance (ANOVA).

Feature selection based on correlation can be described as follows: a good subset of features contains such features that correlate well with the target variable (class) while not being correlated with each other.

The $\chi^2$ criterion is used to test the independence of two variables [9]. In this case the independence between a class and different features were tested. The degree of freedom in the chi-square test is calculated by $(n-1)*(m-1)$ where $n$ and $m$ are numbers of data samples (here $n=1325184$) and feature count ($m=306$) respectively. We use $p$-value=0.05 and as we test each feature and get 306 chi-squared-statistic values.

Analysis of variance (ANOVA), which assesses the significance of differences in mean values between groups. The method is easy to implement, fast, and the algorithm works effectively even if the size of groups is different [10]. Null Hypothesis: There is no significant difference between considered for selection feature among all the classes. We use $p$-value=0.05 and as we test each feature and get 306 F-statistic values.

Feature selection based on mutual information is based on the selection of features using mutual information of minimum redundancy and maximum relevance, where the criterion of maximum relevance finds a subset of features such that it maximizes the average value of mutual information of the feature and class on all the features included in the subset. If only the criterion of maximum relevance is used, it may turn out that the selected features have high redundancy, that is, they may be significantly dependent on each other. Therefore, the criterion of minimal redundancy is also considered [11].

7. Experiment

A 5-fold cross-validation was used to verify classifier efficiency. Data from both states (eyes open and eyes closed) were used in a classification task using the SVM classifier.

Based on the work [1] the coefficients of wavelet transform only at one certain level of DWT decomposition were considered and the classification accuracy was low (47%), so it was hypothesized that to increase the accuracy it was necessary to consider the coefficients from all 5 levels of the wavelet transform. The resulting accuracy is 95.3%, even before we apply feature selection. Table 1, table 2, table 3, and table 4 show the results obtained after feature selection. According to the results, the method based on mutual information have the best accuracy of classification (183 features and 96.4% accuracy).

| Percentile of the most informative features based on the $\chi^2$-criterion | Classification Accuracy | Features |
|---|---|---|
| 90 | 0.897 | 275 |
| **80** | **0.900** | **244** |
| 70 | 0.898 | 214 |
| 60 | 0.896 | 183 |
| 50 | 0.881 | 153 |
| 40 | 0.850 | 122 |
| 30 | 0.789 | 91 |
| 20 | 0.670 | 61 |
| 10 | 0.410 | 30 |
Table 2. Feature selection results for ANOVA-criterion

| Percentile of the most informative features based on the ANOVA-criterion | Classification Accuracy | Features |
|------------------------------------------------------------------------|-------------------------|----------|
| 90                                                                     | 0.95600                 | 275      |
| 80                                                                     | 0.96100                 | 244      |
| **70**                                                                 | **0.96300**             | **214**  |
| 60                                                                     | 0.96298                 | 183      |
| 50                                                                     | 0.95800                 | 153      |
| 40                                                                     | 0.94000                 | 122      |
| 30                                                                     | 0.90000                 | 91       |
| 20                                                                     | 0.82000                 | 61       |
| 10                                                                     | 0.69000                 | 30       |

Table 3. Feature selection results for mutual information criterion

| Percentile of the most informative features based on the mutual information criterion | Classification Accuracy | Features |
|--------------------------------------------------------------------------------------|-------------------------|----------|
| 90                                                                                     | 0.955                   | 275      |
| 80                                                                                     | 0.958                   | 244      |
| 70                                                                                     | 0.961                   | 214      |
| **60**                                                                                  | **0.964**               | **183**  |
| 50                                                                                     | 0.958                   | 153      |
| 40                                                                                     | 0.943                   | 122      |
| 30                                                                                     | 0.895                   | 91       |
| 20                                                                                     | 0.820                   | 61       |
| 10                                                                                     | 0.667                   | 30       |

Table 4. Feature selection based on correlation

| Correlation coefficient threshold | Classification Accuracy | Features |
|----------------------------------|-------------------------|----------|
| 0.10                             | 0.9580                  | 200      |
| 0.20                             | 0.9530                  | 170      |
| 0.30                             | 0.9530                  | 170      |
| 0.35                             | 0.9530                  | 170      |
| 0.40                             | 0.9515                  | 166      |
| 0.50                             | 0.9000                  | 136      |
| 0.55                             | 0.8600                  | 120      |
| 0.60                             | 0.6900                  | 102      |
| 0.65                             | 0.4400                  | 40       |

8. Conclusions
The hypothesis of using coefficients from all levels to increase the accuracy of classification on large samples of data was confirmed. According to the results of the work, it was found that in this work, the maximum accuracy of classification after the selection of features 96.4% (1.1542% gain compared to baseline) and the number of features 183.

It is planned to consider other classifiers in the future and test the work on other data sets, or use subject-independent learning methods (right now built classifier depend on fixed list of subjects
presented in training fold) such as Siamese Networks and specific loss functions (Triplet loss, Multi-similarity loss etc.). To overcome EEG data noisiness Huang-Hilbert transform can also be applied.

Acknowledgments
This research was funded by the Ministry of Education and Science of the Russian Federation within the framework of scientific projects carried out by teams of research laboratories of educational institutions of higher education subordinate to the Ministry of Science and Higher Education of the Russian Federation, project number FEWM-2020-0042.

References
[1] Matsumoto T, Matsumoto H, Yamada K and Hoshino S 2002
[2] Nixon K A, Aimale V and Rowe R K pp 403–423
[3] Anokhin A, Steinlein O, Fischer C, Mao Y, Vogt P, Schalt E and Vogel F 1992 Human Genetics 90
[4] Chuang J, Nguyen H, Wang C and Johnson B 2013 pp 1–16
[5] Hu B, Liu Q, Zhao Q, Qi Y and Peng H 2011
[6] Kaur B, Singh D and Roy P P 2018 Neural Computing and Applications 31 pp 5887–5900
[7] Koller D and Sahami M 1996 pp 284–292
[8] Babayan A, Erbey M, Kumral D, Reinelt J D, Reiter A M F, 2019 Sci. Data 6 180308
[9] Mahmood M R 2021 J. Phys. Conf. Ser. 1804 012056
[10] Ding H, Feng P M, Chen W and Lin H 2014 Mol. Biosyst. 10 pp 2229–2235
[11] Bennasar M, Hicks Y and Setchi R 2015 Expert Syst. Appl. 42 pp 8520–8532