VILNIUS UNIVERSITY

Andrius Vytautas
MISIUKAS MISIŪNAS

Investigation of automatic EEG analysis algorithms

SUMMARY OF DOCTORAL DISSERTATION

Natural Sciences,

Informatics N 009

VILNIUS 2020

This dissertation was written between 2015 and 2019 at Vilnius University.

Academic supervisor

prof. dr. Tadas Meškauskas (Vilnius University, Natural Sciences, Informatics – N 009).

This doctoral dissertation will be defended at the public meeting of the Dissertation Defence Panel:

Chairman:

Prof. Dr. Olga Kurasova (Vilnius University, Natural Sciences, Informatics – N 009).

Members:

Prof. Dr. Rūta Mameniškienė (Vilnius University, Medicine and Health Sciences, Medicine – M 001),

Assoc. Prof. Dr. Dalius Matuzevičius (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronical Engineering – T 001),

Prof. Dr. Audris Mockus (The University of Tennessee, USA, Natural Sciences, Informatics – N 009),

Assoc. Prof. Dr. Povilas Treigys (Vilnius University, Natural Sciences, Informatics – N 009).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 1:00 p.m. on the 29$^{\text{th}}$ of June, 2020 in the auditorium 211 of the Institute of Computer Science of Vilnius University.
Address: Didlaukio str. 47, LT-08303 Vilnius, Lithuania.
The summary of the doctoral dissertation was distributed on the 29$^{\text{th}}$ of May, 2020.
The text of this dissertation can be accessed at the Vilnius University Library, as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS

Andrius Vytautas
MISIUKAS MISIŪNAS

Elektroencefalogramų analizės metodų tyrimas

DAKTARO DISERTACIJOS SANTRAUKA

Gamtos mokslai,

Informatika N 009

VILNIUS 2020

Disertacija rengta 2015-2019 metais Vilniaus universitete.

Mokslinis vadovas:

prof. dr. Tadas Meškauskas (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Gynimo taryba:

Pirmininkė:

prof. dr. Olga Kurasova (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Nariai:

prof. dr. Rūta Mameniškienė (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001),

doc. dr. Dalius Matuzevičius (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – T 001),

prof. dr. Audris Mockus (Tenesio universitetas, JAV, gamtos mokslai, informatika – N 009),

doc. dr. Povilas Treigys (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Disertacija ginama viešame Gynimo tarybos posėdyje 2020 m. birželio mėn. 29 d. 13 val. Vilniaus universiteto Matematikos ir informatikos fakulteto Informatikos instituto 211 auditorijoje.
Adresas: Didlaukio g. 47, LT-08303 Vilnius, Lietuva.
Disertacijos santrauka išsiuntinėta 2020 m. gegužės 29 d.
Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir Vilniaus universiteto interneto svetainėje adresu:
https://www.vu.lt/naujienos/ivykiu-kalendorius

CONTENTS

1 INTRODUCTION

Signals and their processing algorithms are an integral part of our every day lives. Photos we view (JPEG compression standard), and music we listen to (MP3 and some other standards) use The Fast Fourier Transform (FFT) algorithm as well as other Digital Signal Processing (DSP) methods [11]. DSP is used in medicine too, for example in analysis of electrocardiograms (ECGs) [4, 16, 21] and electroenecphalograms (EEGs) [3, 13, 14] (which are investigated in this work).

EEG is a form of an instrumental medical examination with its applications, advantages and disadvantages. The main applications of EEGs are diagnosing various forms of epilepsy, sleep disorders and others. The main advantages of EEGs are the high timescale resolution and the non invasive nature of the examination. On the other hand, disadvantages are low spatial resolution and the inability to examine deeper parts of the brain of the test subject.

EEG tests are ubiquitous in both Lithuania and the world. As a result, a multitude of algorithms for EEG analysis have been created: a number of EEG spike detection algorithms [2, 5, 10, 19, 20], ill *vs* healthy classification [1, 6], ictal *vs* inter-ictal EEG classification [17, 18] and many others.

The EEGs of childhood patients (3-17 years old) are investigated in this work. Patients are diagnosed with one of two groups of diagnosis: benign childhood epilepsy (Group I) and structural focal epilepsy (Group II). Although differences between some Group I and Group II EEGs are obvious even to non neurologists, the cases that are difficult to distinguish are investigated in this work. To the author's knowledge, this thesis and publications it is based on is the first attempt to classify Group I and Group II EEGs by diagnosis.

1.1 Goals and tasks of the thesis

The main goal of this thesis is to create an automated algorithm for the classification of Group I and Group II EEGs in complicated (nonobvious or visually identical to neurologists) cases, without knowledge of the patient's case history. Algorithms created are verified with computer modelling based experiments.

To achieve this goal, these tasks were dealt with:

- Choice and optimisation of EEG spike detection algorithm;
- Selection of geometric EEG spike features usable for classification;
- Selection and application of machine learning based methods for EEG classification by diagnosis;
- Combination of chosen methods into algorithm for EEG classification by diagnosis;

- Implementation of proposed algorithms;
- Confirmation by performing necessary experiments of proposed algorithms and other results presented in thesis.

1.2 Means of investigation

Python programming language was employed in the implementation of proposed algorithms and experiments (2.7.10 version in the beginning of preparation of thesis, later moved to 3.5 and 3.6 versions, latest used version – 3.6.8). A number of Python libraries were employed as well: NumPy (reducing time of some calculations), SciPy (implementation of mathematical morphology and other methods), MatPlotLib (graph plotting), Scikit-learn (implementations of various machine learning methods and metrics), Tensorflow-GPU (CNN implementation on GPU), EegTools and PyEdfLib (parsing EDF and EDF+ files), and mpi4py (implementation of MPI in Python).

Most calculations were performed on the author's personal computer with the following parameters: Intel i7-6700K CPU (4.0 GHz, 4 cores, 8 threads), Asus Z170 Deluxe motherboard, 32 GB DDR4 RAM (4 x 8GB Corsair Vengence LPX 2400 MHz), Asus Strix GeForce 980Ti OC GPU (2816 CUDA cores, 6 GB GDDR5 graphical mempory), Noctua NH-D15 CPU cooler, and 5 x Noctua NF-A14-PWM fans. The PC was dual boot with Windows 10 and Linux Ubuntu 14.04 LTS (at the start), Ubuntu 18.04 LTS (upgraded later) operating systems (all OS were 64-bit versions).

Part of the calculations were performed on VU MIF Cluster (PST[1]): 1920 processor cores, 3.6 TB RAM, 620 TB total disk size, about 25 TFLOP/s of computations.

1.3 Scientific novelty of results

1. A three step algorithm has been proposed for classification of EEGs obtained from Group I and Group II patients. This is the first algorithm published in scientific literature to address this task.

2. Parameters of the EEG spike detection algorithm (based on mathematical morphology) were optimised with a genetic algorithm. That is the first optimisation of parameters of the algorithm mentioned using a genetic algorithm.

3. Three strategies of EEG spike data extraction were tested in the third step of the algorithm for EEG classification by diagnosis;

4. Performance of several machine learning-based classifiers was investigated in the final step of proposed algorithm while maximising accuracy and other important metrics.

[1]https://mif.vu.lt/cluster/

1.4 Significance of results in practice

An automatic algorithm able to classify EEGs obtained from Group I and Group II patients by diagnosis has been proposed. Implementation of this algorithm in practice would reduce the number of misdiagnosed cases and would reduce the workload for doctors-neurologists on manual analysis of EEGs.

The EEG spike detection algorithm is already implemented as part of the NKSPS (National clinical decision support information system, No. VP2-3.1-IVPK-10-V-01) project and is already used by doctors. Implementation of the proposed algorithms would reduce the neurologist's load even further.

1.5 Statements defended

1. EEGs obtained from Group I and Group II patients can be classified by diagnosis with the proposed algorithms achieving 75%–82% accuracy.

2. Methodology employing geometric EEG works best with MLP (multilayer perceptron) based classifier.

3. EEG spike signal array classification (when signals of EEG spikes are concatenated) is best performed by an extremely randomized tree algorithm.

4. EEG classification employing signals from all channels in the vicinity of the spike is best performed by CNN combined with majority rule detection classifiers. Additionally, this algorithm has best usability in practise, thus it is recommended to use and investigate further.

1.6 Approbation of the thesis results

The main findings of this thesis are published in peer reviewed periodicals:

1. EEG classification by diagnosis using geometric EEG spike parameters and a MLP based classifier was published in *Biomedical Signal Processing and Control*, and indexed in *Clarivate Analytics Web of Knowledge* database. The author created and implemented the models and significantly contributed to writing the text of the publication.

2. An article has been written and accepted to *Nonlinear Analysis: Modelling and Control*, indexed in *Clarivate Analytics Web of Knowledge* database, detailing EEG classification by diagnosis using CNN and majority rule detection. The author contributed to creating and implementing the models and significantly contributed to writing the text of the publication.

Results were also presented in international and national conferences and their proceedings:

1. DAMSS 2014 (Druskininkai, Lithuania): Data analysis methods for software systems: 6th International Workshop.
2. LMD 56 (Kaunas, Lithuania): 56th conference of Lithuanian mathematical society. June 16-17, 2015.
3. LMD 57 (Vilnius, Lithuania): 57th conference of Lithuanian mathematical society. June 20-21, 2016.
4. NM&A'18 (Borovets, Bulgaria): Ninth International Conference on Numerical Methods and Applications. August 20-24, 2018.
5. DAMSS 2018 (Druskininkai, Lithuania): 10th international workshop on data analysis methods for software systems. November 29 – December 1, 2018.
6. AMiTaNS'19 (Albena, Bulgaria): Eleventh Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences. June 20-25, 2019.

The author of the thesis was the main author and presenter in all conference reports mentioned above. The author was awarded the Young Scientist Award Certificate for successful presentation at the AMiTaNS'19 conference.

2 EEG

EEGs are employed in diagnosing various ailments of the central nervous system: sleep disorders, addiction diseases, and brain tumors [15], however, this work is focused on two groups of patients diagnosed with epilepsy.

EEG recordings of children (3-17 year-old patients) are investigated in this study. The EEGs are from the database of Children's Hospitals, Affiliate of Vilnius University Hospital Santaros Klinikos recorded during the period of 2010—2018. The dataset included only EEGs that a neurologist would identify as visually similar or identical. Exact diagnosis for each EEG recording was known from the clinical record of the patient.

The patients can be assigned into one of the following two groups:

1. Group I: benign childhood epilepsy with centrotemporal spikes;
2. Group II: structural focal epilepsy patients with cerebral palsy, dysplastic brain lesion, gliosis *etc.*

It should be noted that some patients have more than one EEG recording (see Table 1), therefore a strict rule has been imposed: each patient with all their EEG recordings can be assigned to either the training or testing dataset. If EEGs are mixed, pseudo accuracy rises significantly [9].

Table 1: Distribution of EEGs and patients by diagnosis and throughout training and testing data sets. Percentages in parentheses indicate: 1) sample size of EEGs* from whole EEG dataset of the Group, 2) sample size of patients** from whole dataset of patients in a Group.

Number of patients\Group	**Group I**	**Group II**	**Total**
Number of EEGs (Total)	215	48	263
Number of patients (Total)	135	33	168
Number of EEGs (Training set)	43 (20.0%*)	35 (72.9%*)	78
Number of patients (Training set)	37 (27.4%**)	21 (63.6%**)	58

Another important characteristic of the dataset used in this study is that it is imbalanced: there are more EEGs from Group I than from Group II. The main reason for this discrepancy is that Group II EEG recordings that are similar to Group I EEGs are significantly more rare. All trivial cases were omitted in this study.

All EEGs examined in this study are recorded in the 10–20 international EEG system. The main advantage of this system is that all electrodes are always placed over the same regions of the brain for each patient.

3 ALGORITHM FOR EEG CLASSIFICATION BY DIAGNOSIS

3.1 EEG spike detection

The EEG spikes are detected by a morphological filter-based algorithm (for details see [5, 7, 8, 10]). The premise of operation of the morphological filter is that normal brain activity (*e.g.* brain rhythms) is filtered out while abnormal brain activity (*e.g.* EEG spikes) is left out [5]. Any values of filtered signals that are higher than the detection limit are considered to be spike candidates [7].

The spike detection algorithm is implemented employing a combination of morphological filters and operations. The operations used to detect spikes can be expressed through morphological grey erosion and dilation.

These notations are employed: the signal in an EEG channel investigated is signified by $f(t)$, the structuring element is denoted by $g(t)$, while reflection of the structuring element is $g^s(t) = g(-t)$. D denotes the domain of signal $f(t)$. Then erosion is:

$$(f \ominus g^s)(t) = \min_{\tau \in D}\{f(\tau) - g(-(t-\tau))\}. \tag{1}$$

Dilation can be defined as:

$$(f \oplus g^s)(t) = \min_{\tau \in D}\{f(\tau) + g(-(t-\tau))\}. \tag{2}$$

Employing expressions (1) and (2), opening and closing operators can be defined. Opening:

$$(f \circ g)(t) = [(f \ominus g^s) \oplus g](t). \tag{3}$$

The closing operator is defined as:

$$(f \bullet g)(t) = [(f \oplus g^s) \ominus g](t). \tag{4}$$

EEG spikes can exhibit both positive and negative amplitudes, thus both open-closing and close-opening operations are needed to compensate for that. Employing formulas (3) and (4), these operators can be defined. Open-closing:

$$OC(f(t)) = f(t) \circ g_1(t) \bullet g_2(t). \tag{5}$$

Close-opening is defined as:

$$CO(f(t)) = f(t) \bullet g_1(t) \circ g_2(t). \tag{6}$$

Both OC and CO have an impact of the same absolute value, but different signs on the average value of the signal. Thus, to eliminate the change, averaging out the value of (5) and (6) is employed in equations:

$$OCCO(f(t)) = \frac{OC(f(t)) + CO(f(t))}{2}. \tag{7}$$

The expression (7) denotes the value of the morphological filter. In order to apply it, it is still necessary to define the structuring elements employed (see equations (5) and (6)):

$$g_i(t) = a_i\, k_i\, t^2 + b_i, \qquad i = 1, 2. \tag{8}$$

Where k_i is the coefficient used in optimisation (see Subsection 3.2) with a default value of 1, a_i and b_i are defined as:

$$\begin{gathered} a_1 = \frac{2\, Median(|f|)}{Median(W)}, \qquad a_2 = \frac{2\, Median(|f|)}{3\, Median(W)}, \\ b_1 = b_2 = Median(|f|), \end{gathered} \tag{9}$$

Here W is an array of EEG signal arc lengths [5]. Since brain activity of the patient changes with time, coefficients defined in equation (9) need to be recalculated every $t_r = 5\,\mathrm{s}$.

Every part of the EEG that goes over a certain detection limit L is considered to be an EEG spike candidate:

$$L = 2\, k_L\, Median(f_{filtered}), \tag{10}$$

Here k_L is the coefficient used for optimisation (see Subsection 3.2) with a default value of 1, and $f_{filtered}$ is the filtered signal, which can be defined as:

$$f_{filtered}(t) = \left| f(t) - OCCO(f(t)) \right|, \tag{11}$$

See Figure 1 for visualisation of OC, CO and $OCCO$ filter operation.

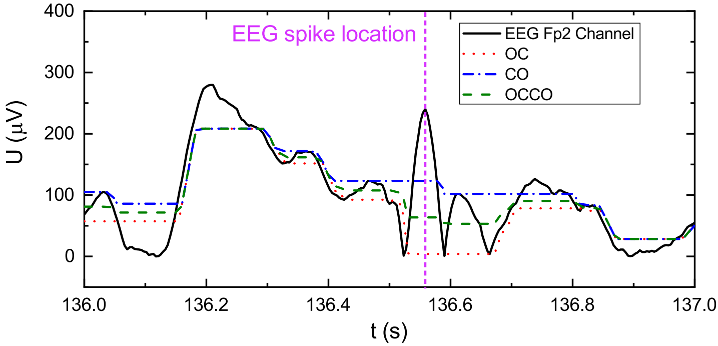
The length of the structuring element is also important, as a structuring element that is too long would result in many false positive spike detections (reduced specificity) and a structuring element that is too short would result in too few spike detections (reduced sensitivity). See Figure 2. It was found that optimal length of a structuring element is:

$$t_e = 4\, k_e\, Median(W) \tag{12}$$

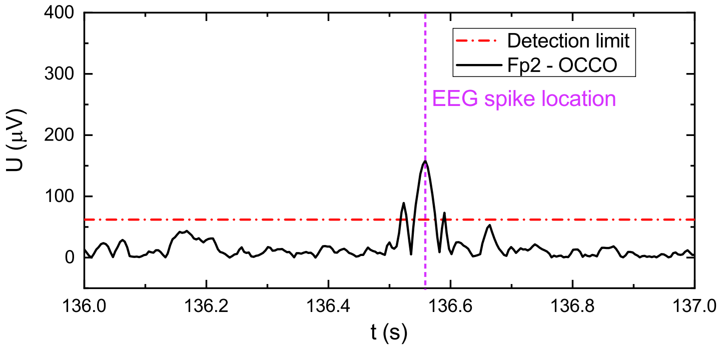
Here k_e is the coefficient for optimisation of length of a structuring element with a default value of 1.

3.2 Optimisation of the parameters of the EEG spike detection algorithm

As noted in Subsection 3.1, the EEG spike detection algorithm has some constants (*e.g.* in equations (9) and 10)) that were introduced in previous studies [5, 10]. However, this study has a different goal compared to these previous studies [5, 7, 10]: instead of just detecting spikes, we tried to classify EEGs by diagnosis. This means that different metrics (*e.g.* accuracy, specificity and sensitivity) of the EEG detection algorithm might be important. Thus the need to optimise the algorithm by these metrics was



(a) EEG signal in Fp2 channel and operation of OC, CO and $OCCO$ filters.



(b) Filtered EEG signal $f_{filtered}(t)$ and detection limit.

Figure 1: Demonstration of raw EEG signal and morphological filter operation. The purple dashed line denotes the position of an EEG spike.

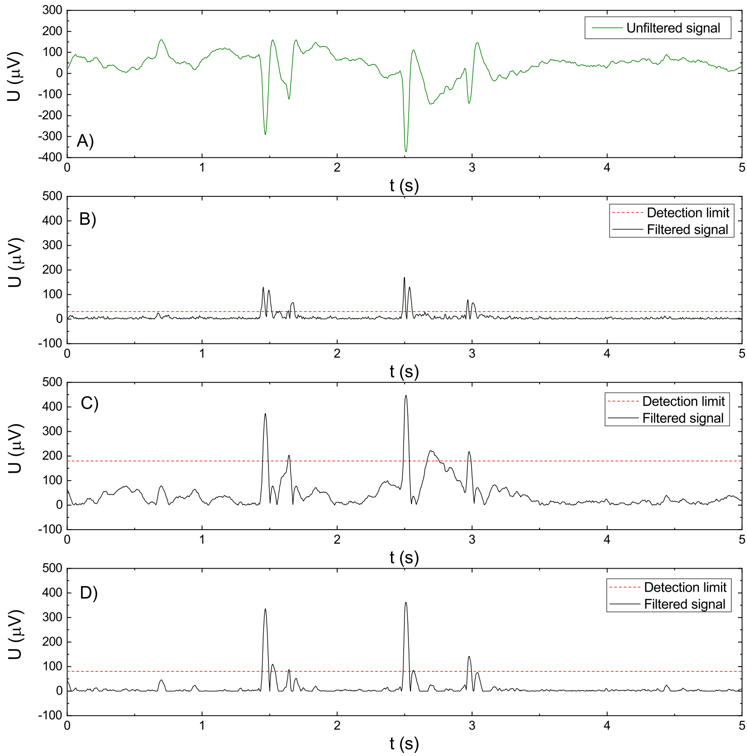


Figure 2: The relation between quality of EEG detection and length of a structuring element. A) shows the original unfiltered signal; B) signal filtered with too short structuring element; C) signal filtered with too long structuring element; D) signal filtered with a structuring element of the right length.

introduced.

For mathematical convenience of optimisation, several coefficients were introduced: k_1 and k_2 in equation (8), k_L in equation (10) and k_e (the value, which is multiplied with t_r). The default starting value of all these coefficients was 1.

Since multiple experiments were done with various fitness functions (accuracy, sensitivity, specificity) and their combinations, any mathematical properties of the fitness function can be guaranteed. It can be presumed that the fitness function is discontinuous, since k_e and k_L values cannot be negative. Furthermore, each evaluation of the fitness function is time and resource consuming. For these reasons the genetic algorithm (GA) was employed in order to optimise the parameters mentioned.

A genetic representation of an individual can be written in the following way: $\left[k_1, k_2, k_L, k_e\right]$. The initial values were generated randomly using normal (Gaussian) distribution with mean $\mu = 1$ and variance $\sigma^2 = 1$. This value generation gave us a selection of new genetic individuals scattered around the known good solution of $\left[1, 1, 1, 1\right]$.

Crossover was implemented by splitting two individuals at a randomly chosen index, swapping the second part and recombining both individuals. Mutation was implemented by modifying a random property of an individual using normal distribution with the mean equal to current value and variance $\sigma^2 = 1$. Elitism

Table 2: Results of the optimisation of parameters of the EEG spike detection algorithm. GA here denotes the genetic algorithm.

Optimisation method	**Sensitivity**	**Specificity**	k_1	k_2	k_L	k_e
Manual optimisation	0.70	0.71	1.00	1.00	1.00	1.00
Sensitivity (GA)	0.92	0.38	0.56	0.61	0.26	0.53
Specificity (GA)	0.11	0.88	1.61	1.63	6.82	1.03
Min(sensitivity, specificity) (GA)	0.73	0.72	1.06	1.08	1.25	1.01

of the selection was applied by carrying over 10% of the best individuals of the current selection to the next one.

Due to the high computational cost of the evaluation of the fitness function of an individual, a population size of 100 individuals was selected. Probability of mutation was 2%. The GA was terminated after 10 populations did not improve the best found solution. For each fitness function, the GA was run five times in order to ensure that it arrived at the same solution within the margin of error. The results are presented in Table 2.

It was determined that the Min(sensitvity, specificity) fitness function displayed optimal classification results for EEG classification by diagnosis. We speculate that the reason this metric works the best is due to both high sensitivity (many EEG spikes are detected) and high specificity (high amount of candidate spikes detected are EEG spikes). High sensitivity fitness function resulted in 52% accuracy of the majority rule voting classifier, and high specificity fitness function resulted in 79% accuracy.

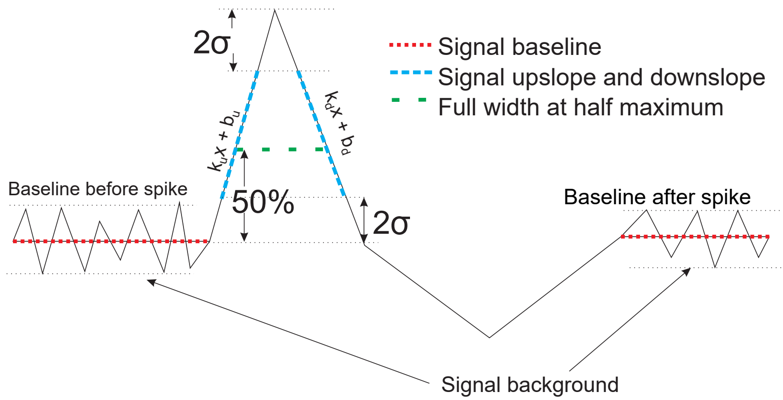


Figure 3: Geometric EEG spike features. Here k_u is upslope, k_d – downslope.

3.3 EEG spike feature extraction

3.3.1 Geometric EEG spike features

After detecting EEG spikes, various features can be extracted. Experimentation has been done with various geometric EEG spike features [9], but upslope and downslope (see Fig. 3) are shown to be the most discriminative ones.

This method has both its advantages and disadvantages. The main advantage is a well-defined feature set that can be used with classical machine learning-based classifiers. However, spike features are not always correctly fitted, thus some additional errors are introduced.

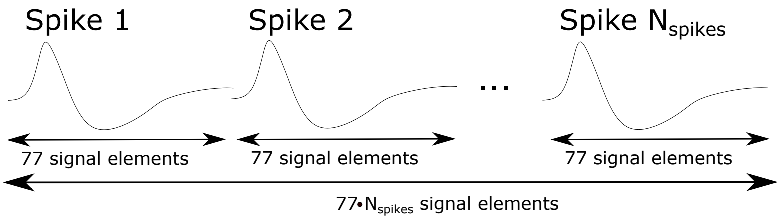


Figure 4: EEG classification strategy where channel in which the spike is detected, is used.

3.3.2 Concatenated EEG spike data

The second possible approach is employing raw EEG spike data. However, since it was determined that more than one spike is necessary to make a diagnosis as accurate as possible, a problem arises: how to pass multiple spike data to the machine learning algorithm-based classifier. The solution proposed is to concatenate EEG spike data from the channel where the spike is detected (see Fig. 4).

3.3.3 EEG spike data in all channels

The third possible strategy is to use EEG data from all channels in the vicinity of the EEG spike detected. This approach works best with classifiers that are tuned to classify image-like (or matrix-shaped) input, like the CNN classifier.

3.4 EEG classification by diagnosis

In this chapter, algorithms for EEG classification and their results are discussed.

3.4.1 EEG classification by diagnosis with geometrical spike parameter data

In this chapter, we try to establish the best classifier for EEG classification by diagnosis using EEG spike geometrical features. In order to achieve this task, some quantifiable parameters of algorithm performance are needed. The most obvious metric for this task is accuracy, which is the sum of true positives and true negatives divided over all detections. This metric is very useful in detecting poorly performing algorithms.

After measuring the accuracy, LDA algorithm was excluded from further analysis due to its poor accuracy: 53%. Multiple supported vector machine (SVM) classifier configurations were tested as well. SVM classifiers with linear and quadratic kernels performed consistently with worse accuracy than SVM with cubic kernels, thus were removed from further analysis.

While accuracy is a great tool for finding some poorly performing algorithms, it does not show all of them. For that reason some true positive rate (TPR) and true negative rate (TNR) analysis was done. Although SVM with both RBF and sigmoid kernels were performing with good accuracy of 75%, they were classifying all the data as Group I. The accuracy was achieved purely due to our data set being biased towards Group I. Due to this reason these algorithms were excluded from further analysis.

Random forest, decision tree, extremely randomized trees, AdaBoost and MLP presented comparable results for both groups and thus were analyzed further. Table 5 presents the commonly used performance metrics [12] for algorithms tested. These tests were performed to evaluate overall quality of the discussed classifiers.

Table 3: Performance metrics [12] for algorithms selected group of algorithms with $N_{spikes} = 100$. Ideal classifier column represents metric values for theoretical ideal classifier. SVM $N_p = 3$ here denotes SVM with a cubic kernel.

Score/ Algorithm	**Random forest**	**Decision tree**	**Extremely randomised tree**	**AdaBoost**	**MLP**	**SVM** $N_p = 3$	**Ideal classi–fier**
Accuracy	0.78	0.76	0.80	0.81	0.75	0.69	**1.00**
TPR	0.79	0.76	0.83	0.90	0.79	0.79	**1.00**
TNR	0.74	0.77	0.71	0.52	0.74	0.48	**1.00**
F1 score	0.76	0.76	0.75	0.64	0.78	0.57	**1.00**
ROC AUC	0.53	0.49	0.56	0.69	0.64	0.49	**1.00**
Cohen kappa	0.06	-0.01	0.12	0.38	0.28	0.26	**1.00**
Matthews correlation coefficient	0.07	-0.01	0.15	0.42	0.38	0.28	**1.00**
Recall score	0.78	0.76	0.81	0.84	0.78	0.69	**1.00**

AdaBoost seems to be the best algorithm by most metrics presented in Table 5, except a couple key ones: TPR and F_1 score. This is due to the fact that AdaBoost classifies Group I (dominant group) correctly 90% of the time and Group II only about

Table 4: EEG classification by diagnosis results using concatenated EEG spike signal data. SD here denotes the standard deviation acquired from k-fold validation.

Algorithm\Metric	**TPR**	**SD**	**TNR**	**SD**
Logistic regression	0.656	0.001	0.6	0.006
Random forest	0.951	0.05	0.768	0.016
Decision tree	0.906	0.008	0.683	0.011
Extremely randomised tree	0.915	0.003	0.805	0.017
AdaBoost	0.765	0.031	0.781	0.053
LDA	0.949	0.001	0.467	0.002
MLP	0.601	0.029	0.58	0.04
SVM $N_p = 3$	0.879	0.02	0.124	0.019
SVM RBF	0.783	0.058	0.264	0.041
SVM sigmoid	0.579	0.063	0.511	0.042

52% of the time. SVM with cubic kernel suffers from the same problem. Despite the good performance of AdaBoost across all other metrics, this algorithm is not suited for the task at hand – detecting rarer Group II cases in the pool of Group I and Group II data. However AdaBoost could be explored further for potential use in the ensemble (voting) type of classifier. This leads to the discussion that some classifier quality metrics can be misleading in this case.

Table 5 shows some more interesting results. Although random forest, decision tree and extremely randomized trees show both high TPR and TNR, their ROC AUC, Cohen kappa and Matthews correlation coefficient are poor. This is probably due to the reason that these metrics are designed to take into account the chance of classifying a record correctly by guessing, therefore these metrics suggest that these algorithms are getting the correct answer by guessing it. Extremely randomized tree suffers less from this problem, yet its Cohen kappa and Matthews correlation coefficient scores are still poor. This means that these algorithms are less suited for EEG classification than MLP and are excluded from further analysis.

This leaves us with MLP, SVM (with cubic kernel) and AdaBoost classifiers. Of these three, the MLP classifier is better considering all metrics, thus it is recommended to be used for automatic classification by diagnosis.

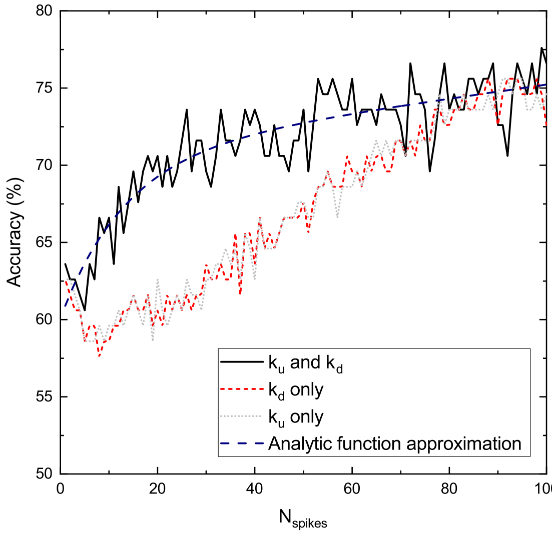


Figure 5: Accuracy of automatic MLP-based classification (between Group I and Group II) vs N_{spikes} (length of lists, containing parameters of spikes, employed in training and testing) for different training strategies.

Table 5: Classifier metrics derived from leave one patient out cross-validation on a single EEG spike classification.

Parameter	**CNN classifier**	**Majority rule classifier**
Accuracy	0.580	0.802
Weighted accuracy	0.572	0.795
F_1 **score**	0.256	0.856
ROC AUC	0.579	0.916
Matthews correlation coefficient	0.144	0.550

The MLP classifier was also tried out with different numbers of spikes. Results show (see Fig. 5) that the accuracy of MLP classifier saturates at about 75%, when 100 spikes are used. However, the algorithm still could be used with a lower amount of spikes, but with lower accuracy.

3.4.2 EEG classification by diagnosis with EEG signal data.

Another approach that was tried in this work is EEG classification by EEG signal data described in Section 3.3.2. Using this approach, the highest TPR and TNR values are displayed by extremely randomised tree-based classifier in both *k-fold* and normal training approaches with accuracy of 82%.

3.4.3 EEG classification by diagnosis using CNN and majority rule vote classifier

The results show that a single EEG spike cannot be decisively classified (58% accuracy) as belonging to either Group I or Group II. Thus, the majority rule voting classifier was proposed. Each detected spike belonging to a patient was classified using CNN. Each classification result of 0.5 or below registered as a vote for assigning a patient to Group I, and each result above 0.5 was a vote assigning a patient to Group II. Figure 6 demonstrates the voting results compared to the real diagnosis of the patient. This did lead to a significant improvement in the average classification accuracy of 80%, which was a 7% increase over previous studies, or 82% (9 % increase) if patients having less than 100 spikes are excluded from analysis as in previous studies [9].

A high accuracy value does not necessarily represent high quality of classification. Therefore, additional investigation is needed to accurately evaluate the quality of the CNN majority rule classifier. This is crucial since our dataset is unbalanced: patients belonging to Group II are much rarer when compared to patients from Group I, resulting in an unbalanced dataset. Figure 6 shows that the majority rule classifier is highly likely to classify both Group I and Group II EEGs correctly (81% and 79% respectively). More metrics are presented in Tables 6 and 7.

The proposed algorithm had a further advantage over the MLP

Table 6: Confusion matrix of leave one patient out of the CNN classifier in a single spike EEG classification.

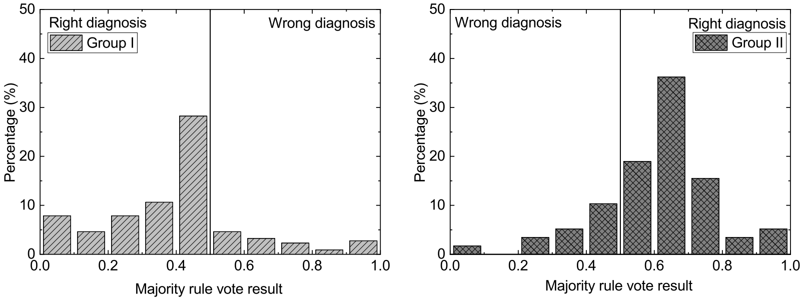
	Group I	Group II
Group I	20872 (TPR = 0.59)	14806 (FPR = 0.41)
Group II	6871 (FNR = 0.43)	9123 (TNR = 0.57)

based classifier proposed in previous studies: a fixed amount of spikes in each EEG was no longer required in order to classify an EEG by diagnosis, since each EEG spike was classified separately by CNN and the final classification result was based on majority rule of all EEG spikes classified. However, a higher number of EEG spikes was still preferred, since rejecting EEGs with less than 100 spikes produced an average accuracy of 82%.

This result was achieved due to the fact that many classification errors of the CNN classifier are spike specific, but not EEG specific. Figure 6 is the majority rule average vote result histogram. It demonstrates that almost all EEGs had spikes classified incorrectly, however, 80% of EEGs on average had the majority of spikes detected correctly leading to correct classification by the majority rule classifier (or 82% if EEGs with less than 100 spikes are not considered like in previous approaches).

Table 7: Confusion matrix of leave one patient out of the majority voting rule classifier in all spike EEG classification.

	Group I	Group II
Group I	128 (TPR = 0.81)	30 (FPR = 0.19)
Group II	13 (FNR = 0.22)	46 (TNR = 0.78)



(a) Majority rule voting results in Group I patients.

(b) Majority rule voting results in Group II patients.

Figure 6: Histograms of majority rule classifier voting results.

4 CONCLUSIONS

- EEG data can be classified by diagnosis (between Group I and Group II) with MLP based classifier and geometric EEG spike features with 75% accuracy (with EEGs containing 100 spikes).

- EEGs can be classified by diagnosis (between Group I and Group II) with extremely randomised tree and concatenated

EEG spike data with 82% accuracy (with EEGs containing 100 spikes).

- EEGs can be classified by diagnosis (between Group I and Group II) with CNN combined with majority rule detection with 80% accuracy (or 82% if EEGs with over 100 spikes are used). This algorithmic version is recommended for practical applications and further studies, since it works with any EEG with any number of spikes without need of retraining the CNN classifier and achieves results that are not worse than other best classifiers.

- All proposed versions of the algorithm are sensitive to the number of EEG spikes available to some extent. Thus EEGs with more spikes are favored in order to make more accurate predictions.

5 PUBLICATIONS ON TOPIC OF THE THESIS

Publications in peer reviewed periodical scientific journals

1. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2019). Algorithm for automatic EEG classification according to the epilepsy type: benign focal childhood epilepsy and structural focal epilepsy. Biomedical Signal Processing and Control 48, p. 118-127. doi: 10.1016/j.bspc.2018.10.006 [Web of Science].

2. A.V. Misiukas Misiūnas, V. Rapševičius, R. Samaitienė, T. Meškauskas (2019). Electroencephalogram spike detection and classification by diagnosis with convolutional neural network. Accepted to Nonlinear Analysis: Modelling and Control [Web of Science].

Publications in peer reviewed continuous scientific journals

1. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2019). Accuracy of Different Machine Learning Type Methodologies for EEG Classification by Diagnosis, Springer Lecture Notes in Computer Science Vol. 11189, p. 441-448.

doi: 10.1007/978-3-030-10692-8_50 [Conference Proceedings Citation Index, Web of Science].

2. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2019). Machine Learning Based EEG Classification by Diagnosis: Approach to EEG Morphological Feature Extraction, accepted to AIP Conference Proceedings [Conference Proceedings Citation Index, Web of Science].

3. A.V. Misiukas Misiūnas, T. Meškauskas, A. Juozapavičius (2015). On the implementation and improvement of automatic EEG spike detection algorithm. Proceedings of the LithuanianMathematical Society, Ser. A (56), p. 60-65. doi: 10.15388/LMR.A.2015.11.

4. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2016). Derivative parameters of electroencephalograms and their measurement methods. Proc. of the Lithuanian Mathematical Society, Ser. A (57), p. 47-52. doi: 10.15388/LMR.A.2016.09.

Publications in books of abstracts and conference programs

1. A. V. Misiukas Misiūnas, T. Meškauskas, A. Juozapavičius (2014). On implementation of automatic EEG spikes detection algorithm. Data analysis methods for software systems:

6th International Workshop. Abstracts book, Druskininkai, Lithuania, December 4-6, 2014. ISBN 9789986680505. p. 41.

2. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2018). Accuracy of different machine learning type methodologies for EEG classification by diagnosis. Numerical Methods and Applications: 9th international conference. Abstracts book, Borovets, Bulgaria, August 20-24, 2018. p. 64.

3. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2018). On implementation of Three-Stage Algorithm for EEG Classification by Diagnosis. Data analysis methods for software systems: 10th International Workshop. Abstracts book, Druskininkai, Lithuania, November 29 - December 1, 2018. ISBN 9786090700433. p. 60.

4. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2019). Machine Learning Based EEG Classification by Diagnosis: Approach to EEG Morphological Feature Extraction. 11th Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences. Abstracts book, Albena, Bulgarija, June 20-25, 2019. p. 63.

6 CURRICULUM VITAE

Education:

- 2015–2019 Vilnius University, PhD studies in informatics.
- 2013–2015 Vilnius University, MSc in computer modelling.
- 2008–2012 Vilnius University, BSc in computer physics.
- 2008 Vilnius „Minties“ Gymnasium.

Work Experience:

- 2017 09 — now, Vilnius University, Institute of Computer Science, junior assistant professor. Teaching theoretical and practical lectures on software system archtitecture for bachelor and PKI students.
- 2015 10 — now, Special Investigation Service of the Republic of Lithuania, administrative department, IT division, chief specialist. Making custom analytical software used in both strategic and tactic analysis, investigations of criminal activities.
- 2016 02 — 2017 06, Vilnius University, Faculty of Mathematics and Informatics, lecturer. Theoretical and practical lectures on software system architecture and practical informatics.

- 2014 05 — 2015 06, EEG spike detection module developer in NKSPS project.
- 2014 07 — 2014 12, UAB iTree Lithuania, Java programmer.
- 2011 10 — 2013 01, Scientific investigation as a student researcher in field of astrophysics.

Additional Information:

- 2006–2008 Winner and participant of National Contest for Young Scientists (I-II places) and EU Contest for Young Scientists (EUCYS).
- 2005–2008 Prize winner in Lithuanian Pupil Astronomy Olympiad (II-III places), participant of International Olympiad on Astronomy and Astrophysics (IOAA).

7 SANTRAUKA LIETUVIŲ KALBA

Signalų analizė ir mašinų mokymosi metodai yra itin plačiai taikomi šiuolaikiniame gyvenime. Ne išimtis ir medicina – joje dažnai atliekami ir analizuojami (mašinų mokymosi ir kitais metodais) tokie signalai kaip elektrokardiogramos (EKG) ir šiame darbe nagrinėjamos elektroencefalogramos (EEG).

Darbe nagrinėjamos dviejų diagnozių grupių EEG: gėrybinę vaikų epilepsija (Rolando epilepsija) (I grupė) ir struktūrinė židininė epilepsija (II grupė). Nagrinėjamos tik tos EEG, kurios yra sunkiai (arba visiškai) neatskiriamos gydytojams neurologams neturint paciento ligos istorijos ar kitų svarbių duomenų.

Disertacijoje aprašytas I grupės ir II grupės EEG klasifikavimo pagal diagnozę algoritmas, turintis tris esminius žingsnius: 1) EEG pikų aptikimas, 2) EEG piko charakteristikų išskyrimas, 3) EEG klasifikavimas pagal diagnozę (I arba II grupė) mašinų mokymosi metodais.

Disertacijoje nagrinėjamas (ir pirmame klasifikavimo pagal diagnozę žingsnyje naudojamas) Nishida ir kt. (1999), Juozapavičiaus ir kt. (2011) pasiūlytas EEG pikų paieškos algoritmas. Minėto algoritmo parametrai optimizuojami genetiniu algoritmu pagal kelias tikslo funkcijas: siekiant kuo didesnio pikų aptikimo tikslumo, jautrumo ir tikslumo bei jautrumo kombinacijos.

Antrame klasifikavimo pagal diagnozę žingsnyje nagrinėjamos trys pagrindiniai EEG piko charakteristikų išskyrimo būdai: 1) geometriniai EEG piko parametrai, 2) EEG signalo atkarpų masyvo naudojimas kanale, kur aptiktas EEG pikas, 3) visų EEG kalanų naudojumas aptiktų EEG pikų aplinkose.

Trečiame siūlomo algoritmo žingsnyje nagrinėjama eilė mašinų mokymusi pagrįstų klasifikavimo metodų: daugiasluoksnis perceptronas (MLP), sprendimų medis, atsitiktinis miškas, labai atsitiktiniai medžiai, logistinė regresija, tiesinė diskriminantinė analizė (LDA), atraminių vektorių mašina (SVM) su įvairiais branduoliais, konvoliuciniai neuroniniai tinklai (CNN), AdaBoost.

Disertacijos tyrimo objektas

Disertacijos tyrimo objektas – vaikų (3–17 m. amžiaus), kuriems nustatyta I arba II grupės diagnozė, EEG.

Disertacijos tikslai ir uždaviniai

Disertacijos tikslas – sukurti algoritmus, kurie automatiškai klasifikuotų I ir II grupių EEG pagal diagnozę, gydytojams (neurologams) tiriant sunkiai atpažįstamus atvejus ir klasifikavimui naudojant tik EEG signalo duomenis, ir verifikuoti šiuos algoritmus kompiuterinio modeliavimo eksperimentais.

Tikslui pasiekti iškelti šie uždaviniai:

- Pasirinkti ir optimizuoti EEG pikų paieškos algoritmą.
- Nustatyti EEG piko geometrines (ir kitas) charakteristikas, tinkamas klasifikuoti pagal diagnozę.
- Pasirinkti mašinų mokymusi pagrįstus klasifikavimo metodus ir pritaikyti juos EEG klasifikuoti pagal diagnozę, atlikti pasirinkimą pagrindžiančius eksperimentus.
- Sujungti pasirinktus algoritmus į EEG klasifikavimo pagal diagnozę algoritmą, eksperimentiškai palyginti įvairių algoritmo versijų veikimą.
- Įgyvendinti (suprogramuoti) pasiūlytus algoritmus.
- Atlikti eksperimentus, reikalingus pasiūlytiems algoritmams ir kitiems gautiems rezultatams patvirtinti.

Mokslinis rezultatų naujumas

- Sukurtas trijų žingsnių algoritmas, skirtas klasifikuoti I ir II grupių pacientų EEG pagal diagnozę. Tai pirmas mokslinėje literatūroje aprašytas šį uždavinį sprendžiantis algoritmas.
- Genetiniu algoritmu optimizuoti EEG pikų paieškos algoritmo, pagrįsto matematinės morfologijos filtru, parametrai.

Tai pirmas mokslinėje literatūroje aprašytas minėto algoritmo parametrų optimizavimas genetiniu algoritmu.

- Ištirti keli mašinų mokymosi algoritmų EEG pikų duomenų charakteringų parametrų išskyrimo būdai antrame EEG klasifikavimo pagal diagnozę algoritmo žingsnyje.
- Ištirta kelių klasifikatorių, pagrįstų mašinų mokymusi, veikla trečiame EEG klasifikavimo pagal diagnozę algoritmo žingsnyje, maksimaliai padidinanti klasifikavimo tikslumą ir kitas svarbias metrikas.

Praktinė rezultatų reikšmė

Sukurtas automatinis algoritmas, leidžiantis klasifikuoti vaikų, kuriems diagnozuota gerybinė epilepsija arba struktūriniai smegenų pažeidimai, EEG. Algoritmo įgyvendinimas praktikoje leistų sumažinti neteisingų diagnozių skaičių, gydytojai neurologai galėtų greičiau įvertinti pacientų EEG.

Disertacijos ginami teiginiai

- Naudojant disertacijoje pristatomus mašinų mokymosi pagrindu veikiančius klasifikavimo algoritmus, I ir II grupių EEG gali būti klasifikuojamos 75–82 proc. tikslumu.

- Naudojant EEG pikų geometrinius parametrus, geriausia klasifikavimo kokybė pasiekiama taikant daugiasluoksnį perceptroną.
- Naudojant EEG pikų signalų atkarpų (kanale, kuriame aptiktas EEG pikas) masyvą, geriausia klasifikavimo kokybė pasiekiama taikant labai atsitiktinio medžio klasifikatorių.
- Konvoliucinio neuroninio tinklo ir daugumos balsavimo pagrindu veikiantis klasifikavimo algoritmas pasižymi geriausiomis klasifikavimo ir panaudojamumo savybėmis, todėl rekomenduojamas tolesniems tyrimams ir taikytinas praktiškai.

Rezultatų patvirtinimas

Disertacijos tema paskelbti du straipsniai periodiniuose recenzuojamuose moksliniuose žurnaluose, indeksuojamuose *Clarivate Analytics Web of Knowledge* duomenų bazėje. Rezultatai pristatyti dvejose tarptautinėse ir keturiose nacionalinėse mokslinėse konferencijose, paskelbtos keturios publikacijos disertacijos tema konferencijų darbuose. Visuose nurodytuose straipsniuose ir konferencijų pranešimuose disertacijos autorius buvo pranešėjas ir pagrindinis straipsnio autorius.

Išvados

1. EEG gali būti klasifikuojamos taikant klasifikavimo pagal diagnozę (I ir II grupių) algoritmą, naudojantį geometrinius pikų parametrus 75 proc. tikslumu (su EEG, turinčiomis 100 pikų). Šiam tikslui pasiekti tinkamiausias MLP klasifikatorius.

2. EEG gali būti klasifikuojamos pagal diagnozę (I ir II grupių) su labai atsitiktinio medžio metodu pagrįstu EEG klasifikatoriumi, klasifikuojančiu pagal EEG signalų atkarpas, naudojančiu kanalo, kuriame aptiktas pikas, duomenis 82 proc. tikslumu (su EEG, turinčiomis 100 pikų).

3. EEG gali būti klasifikuojamos pagal diagnozę (I ir II grupių) su CNN ir daugumos balsavimo klasifikatoriumi, naudojančiu visų EEG kanalų duomenis 80 proc. tikslumu arba 82 proc. tikslumu, jeigu klasifikuojamos EEG, turinčios bent 100 pikų. Šis algoritmas laikytinas geriausiu iš pasiūlytų dėl turimų pranašumų: 1) gali klasifikuoti EEG, turinčias neapibrėžtą pikų kiekį (skirtingai nuo kitų algoritmų, kurie buvo testuojami su EEG, turinčiomis po 100 pikų), nereikia iš naujo mokyti klasifikatorių, 2) nagrinėjant EEG, turinčias 100 ar daugiau pikų, pasiekia ne blogesnį tikslumą negu kiti pasiūlyti algoritmai.

4. Visi rekomenduotini EEG klasifikavimo pagal diagnozę algoritmo variantai yra jautrūs EEG pikų skaičiui, todėl, siekiant

kuo tikslesnio EEG klasifikavimo pagal diagnozę rezultato, esant galimybei, reikėtų naudoti EEG, turinčias kuo daugiau pikų.

Autoriaus publikacijos disertacijos tema

Publikacijos periodiniuose recenzuojamuose moksliniuose žurnaluose:

1. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2019). Algorithm for automatic EEG classification according to the epilepsy type: benign focal childhood epilepsy and structural focal epilepsy. *Biomedical Signal Processing and Control* 48, p. 118–127. doi: 10.1016/j.bspc.2018.10.006 [Web of Science].
2. A.V. Misiukas Misiūnas, V. Rapševičius, R. Samaitienė, T. Meškauskas (2020). Electroencephalogram spike detection and classification by diagnosis with convolutional neural network. *Nonlinear Analysis: Modelling and Control.* [Web of Science] [Priimtas spausdinti].

Publikacijos tęstiniuose recenzuojamuose moksliniuose žurnaluose:

1. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2019). Accuracy of Different Machine Learning Type Meth-

odologies for EEG Classification by Diagnosis. *Springer Lecture Notes in Computer Science* 11189, p. 441–448. doi: 10.1007/978-3-030-10692-8_50 [Conference Proceedings Citation Index, Web of Science].

2. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2019). Machine Learning Based EEG Classification by Diagnosis: Approach to EEG Morphological Feature Extraction. *AIP Conference Proceedings* 2164, p. 080005-1 – 080005-5. doi: 10.1063/1.5130828 [Conference Proceedings Citation Index, Web of Science].

3. A.V. Misiukas Misiūnas, T. Meškauskas, A. Juozapavičius (2015). On the implementation and improvement of automatic EEG spike detection algorithm. *Lietuvos matematikos rinkinys* [*Proc. of the Lithuanian Mathematical Society*], ser. A (56), p. 60–65. doi: 10.15388/LMR.A.2015.11.

4. A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė (2016). Derivative parameters of electroencephalograms and their measurement methods. *Lietuvos matematikos rinkinys* [*Proc. of the Lithuanian Mathematical Society*], ser. A (57), p. 47–52. doi: 10.15388/LMR.A.2016.09.

Trumpos žinios apie autorių

Autorius Vilniaus universitete baigė kompiuterinės fizikos bakalauro (2012 m.) ir kompiuterinio modeliavimo magistro (2015 m.) studijas, 2015–2019 m. studijavo informatikos srities doktorantūrą Vilniaus universiteto Informatikos institute. Nuo 2016 metų dirba Vilniaus universiteto Informatikos instituto jaunesniuoju asistentu, Lietuvos respublikos Specialiųjų tyrimų tarnybos Informacinių technologijų skyriaus vyriausiuoju specialistu.

8 SUMMARY

Automatic algorithm for electroencephalogram (EEG) classification by diagnosis: benign childhood epilepsy with centrotemporal spikes (rolandic epilepsy) (Group I) and structural focal epilepsy (Group II) are presented in this thesis. Manual classification of these groups is sometimes difficult, especially when no clinical record is available, thus presenting the need for an algorithm for automatic classification. A few possible classification by diagnosis algorithm versions are proposed in this thesis: 1) geometric EEG spike parameter and feed-forward multilayer perceptron (MLP) based classifier achieving 75% classification accuracy; 2) extremely randomized tree based algorithm using signal in channel where EEG spikes are classifying 82% accuracy; and 3) convolutional neural network (CNN) and majority rule classifier based algorithm achieving 80% accuracy, or 82% if only EEGs with 100 or more spikes are classified.

REFERENCES

[1] V. Bevilacqua, A. A. Salatino, C. Di Leo, G. Tattoli, D. Buongiorno, D. Signorile, C. Babiloni, C. Del Percio, A. I. Triggiani, and L. Gesualdo. Advanced classification of Azheimer's disease and healthy subjects based on EEG markers. *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–5, July 2015. ISSN 2161-4407. doi: 10.1109/IJCNN.2015.7280463.

[2] J. J. Halford. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. *Clinical Neurophysiology*, 120(11):1909–1915, 2009. ISSN 1388-2457. doi: 10.1016/j.clinph.2009.08.007.

[3] A. R. Hassan and A. Subasi. Automatic identification of epileptic seizures from EEG signals using linear programming boosting. *Computer Methods and Programs in Biomedicine*, 136: 65–77, 2016. ISSN 0169-2607. doi: 10.1016/j.cmpb.2016.08.013. URL `http://www.sciencedirect.com/science/article/pii/S0169260716304928`.

[4] Y. Hsu, J. Wang, W. Chiang, and C. Hung. Automatic ECG-based emotion recognition in music listening. *IEEE Transactions on Affective Computing*, 2017. ISSN 1949-3045. doi: 10.1109/TAFFC.2017.2781732.

[5] A. Juozapavičius, G. Bacevičius, D. Bugelskis, and R. Samaitienė. EEG analysis – automatic spike detection. *Nonlinear Analysis: Modelling and Control*, 16(4): 375–386, 2011. URL `http://www.mii.lt/na/issues/NA_1604/NA16401.pdf`.

[6] H. Komijani, A. Nabaei, and H. Zarrabi. Classification of normal and epileptic EEG signals using adaptive neuro-fuzzy network based on time series prediction. *Neuroscience and Biomedical Engineering*, 4(4): 273–277, 2016.

[7] A. V. Misiukas Misiūnas, T. Meškauskas, and A. Juozapavičius. On the implementation and improvement of automatic EEG spike detection algorithm. *Proc. of the Lithuanian Mathematical Society*, 56(Ser. A): 60–65, 2015.

[8] A. V. Misiukas Misiūnas, T. Meškauskas, and R. Samaitienė. Derivative parameters of electroencephalograms and their measurement methods. *Proc. of the Lithuanian Mathematical Society*, 57(Ser. A): 47–52, 2016.

[9] A. V. Misiukas Misiūnas, T. Meškauskas, and R. Samaitienė. Algorithm for automatic EEG classification according to the epilepsy type: Benign focal childhood epilepsy and structural focal epilepsy. *Biomedical signal processing and control*, 48: 118–127, 2019. ISSN 1746-8094.

[10] S. Nishida, M. Nakamura, A. Ikeda, and H. Shibasaki. Signal separation of background EEG and spike by using mor-

phological filter. *IFAC Proceedings Volumes of 14th World Congress of IFAC*, 32(2): 4301–4306, 1999.

[11] D. Salomon, G. Motta, and D. Bryant. *Data compression: The Complete Reference*. Springer, 2006. ISBN 978-1-84628-602-5.

[12] C. Sammut and G. I. Webb. *Encyclopedia of Machine Learning and Data Mining*. Springer, 2017.

[13] M. Sharma, A. Dhere, R. B. Pachori, and U. Rajendra Acharya. An automatic detection of focal EEG signals using new class of time–frequency localized orthogonal wavelet filter banks. *Knowledge-Based Systems*, 118: 217–227, 2017. ISSN 0950-7051. doi: 10.1016/j.knosys.2016.11.024. URL `http://www.sciencedirect.com/science/article/pii/S0950705116304816`.

[14] A. Supratak, H. Dong, C. Wu, and Y. Guo. Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11): 1998–2008, Nov 2017. ISSN 1534-4320. doi: 10.1109/TNSRE.2017.2721116.

[15] W. O. Tatum, A. M. Husain, S. R. Benbadis, and P. W. Kaplan. *Handbook of EEG Interpretation*. Demos Medical Publishing, 2006. ISBN 978-1-933864-11-2.

[16] M. Thomas, M. Kr Das, and S. Ari. Automatic

ECG arrhythmia classification using dual tree complex wavelet based features. *AEU – International Journal of Electronics and Communications*, 69(4): 715–721, 2015. ISSN 1434-8411. doi: 10.1016/j.aeue.2014.12.013. URL `http://www.sciencedirect.com/science/article/pii/S1434841114003641`.

[17] K. D. Tzimourta, A. T. Tzallas, N. Giannakeas, L. G. Astrakas, D. G. Tsalikakis, and M. G. Tsipouras. Epileptic seizures classification based on long-term EEG signal wavelet analysis. *Precision Medicine Powered by pHealth and Connected Health*, pages 165–169, 2018.

[18] K. D. Tzimourta, A. T. Tzallas, N. Giannakeas, L. G. Astrakas, D. G. Tsalikakis, P. Angelidis, and M. G. Tsipouras. A robust methodology for classification of epileptic seizures in EEG signals. *Health and Technology*, 9(2): 135–142, Mar 2019. ISSN 2190-7196. doi: 10.1007/s12553-018-0265-z.

[19] G. Xu, J. Wang, Q. Zhang, and J. Zhu. An automatic EEG spike detection algorithm using morphological filter. *2006 IEEE International Conference on Automation Science and Engineering*, pages 170–175, Oct 2006. ISSN 2161-8070. doi: 10.1109/COASE.2006.326875.

[20] G. Xu, J. W. Q. Zhang, S. Zhang, and J. Zhu. A spike detection method in EEG based on improved morphological

filter. *Computers in Biology and Medicine*, 37(11): 1647–1652, 2007.

[21] M. Yochum, C. Renaud, and S. Jacquir. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. *Biomedical Signal Processing and Control*, 25: 46–52, 2016. ISSN 1746-8094. doi: 10.1016/j.bspc.2015.10.011. URL http://www.sciencedirect.com/science/article/pii/S1746809415001779.

Vilnius University Pres
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
Print run copies 30