A. Ugon et al. (Eds.)

doi:10.3233/978-1-61499-852-5-171

Automatic Recognition of Epileptiform EEG Abnormalities

Alexander BRENNER [a], Ekaterina KUTAFINA [a,b], Stephan M. JONAS [a]
[a] *Department of Medical Informatics, Uniklinik RWTH Aachen, Pauwelsstrasse 30, 52057 Aachen, Germany*
[b] *AGH University of Science and Technology, Faculty of Applied Mathematics, al. Mickiewicza 30, 30-059, Krakow, Poland*

Abstract. Long term EEG examinations, for example during epilepsy diagnosis, can be performed more efficiently with support of automated abnormality detection. Currently, these methods are usually developed based on one specific database, which limits the possibilities of generalizations. Here, we present a machine learning solution for detection of interictal abnormal EEG segments optimized on the publically available TUH Abnormal EEG Corpus. The classifier is further re-trained and tested on several combinations of publicly available data sets. The results achieved internally on the datasets are comparable to the known state of the art, while training and testing on different sources produced accuracy in the range of 67.51% to 99.50%. Lower accuracy is achieved when the training data set is highly preprocessed and relatively small.

Keywords. EEG, epilepsy, spike detection, interictal abnormality

1. Introduction

The electroencephalogram (EEG) is a signal measurement of the brain's electrical activity [1]. It is one of the most commonly used tools for studies on brain functions and neurological disorders [2]. In particular, EEG data collection is conducted in patients with epilepsy [3] as well as to study the effects of antipsychotic medications and abnormal EEG patterns [4]. Current technologies allow long term monitoring, which in turn generates large quantities of data. Manual analysis of this data is time-consuming and inefficient. Hence, automated methods are needed as support for EEG analysis.

Here, we focus on detection of interictal (between seizures) abnormal EEG patterns referred to as epileptiform discharges. These patterns are used by medical professionals for diagnostic and evaluation purposes. To the best of our knowledge, there is no publicly available EEG database containing individually labeled epileptiform events (such as spikes, sharp waves, slow waves and spike-and-slow-wave complexes [5]). Such detailed labelling requires a large amount of work and in the era of well developed machine learning algorithms would be of great interest. In the absence of such data we used the TUH Abnormal EEG Corpus (v1.1.2) [6], where EEG sessions are labelled as normal and abnormal. For this data we developed an algorithm for automatic abnormality detection. On later stages, the data from the so-called Bonn set [7] is added in order to show classifier robustness in dealing with the new data. Six different combinations of training and testing data formed by the subsets of these two databases are considered.

Table 1. Overview of Bonn dataset.

Set	Patients	Setup	Phase
A	healthy	surface EEG	open eyes
B	healthy	surface EEG	closed eyes
C	epilepsy	intracranial EEG	interictal
D	epilepsy	intracranial EEG	interictal
E	epilepsy	intracranial EEG	seizure

2. Methods

2.1. Data

Temple University Hospital (TUH) [6] offers a comprehensive source of clinical EEG data that is continuously updated. The data resources are available in a complete set of 23,218 sessions or in several subsets. In addition to every EEG session, a physician's report is available summarizing the patient's medications, clinical history and the physician's findings. The TUH Abnormal EEG Corpus (v1.1.2), a subset of the TUH EEG Corpus (v0.6.0), provides 3017 EEG sessions that have been categorized into normal and abnormal recordings. Based on this data we develop and test a classification pipeline on automatically categorizing EEG into normal and abnormal recordings.

Another widely used data source for testing automatic spike and seizure detection approaches is provided by the University of Bonn [7]. The database contains five sets with 100 records each (Table 1). Decision process

We categorize the decision process into three stages: pre-processing, feature extraction and machine learning classification (Figure 1). As input to our classification algorithm we use single channel EEG signals. On the multi-channel TUH data we select the channel T5 - O1 derived from the TCP montage as it is proposed by Obeid et al. [8]. Following the same source, we consider the first 60 seconds of each recording in the dataset for classification. Selecting the same parts of the recordings allows a direct comparison of the results.

2.2. Pre-processing

The purpose of the pre-processing stage is to unify recordings from different sources, to reduce noise and to filter out artifacts.

The EEG contained in the TUH dataset were recorded with sampling frequencies of 250 Hz and 256 Hz, while the EEG segments from the Bonn dataset were sampled with 173.61 Hz. We assume to encounter records of various sampling rates in future scenarios. To unify the length of the records we perform a resampling to a fixed sampling frequency. The sampling frequency of 250 Hz was chosen as it corresponds to 87% of all used data. This choice also implies that only 6% of the data loses information through the downsampling (256 Hz files). Epileptiform discharges are known to appear with a minimal duration of 20 milliseconds [5]. Hence, waves with a frequency above 50 Hz are here not of interest for us. In order to remove high frequency noise we perform band-

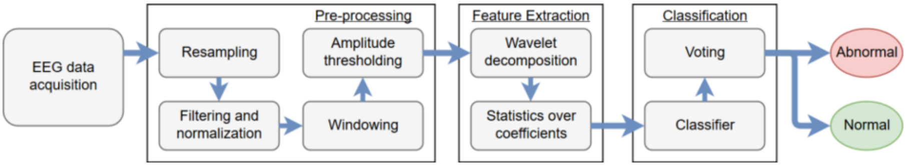

Figure 1. Decision process of EEG data classification.

pass filtering with a second order Butterworth filter, which is known for having a flat frequency response in the pass-band, setting the lower cut-off frequency to 1 Hz and the higher cut-off frequency to 50 Hz.

Intracranial EEG generally have higher amplitude ranges (order of 100 µV) than surface EEG (order of µV) [7]. Since we will test different combinations of training and testing data from both measurement methods, we normalize each recording by the mean absolute amplitude of the corresponding dataset.

Artifacts are generally observed in all EEG, regardless of whether they contain abnormal signal patterns. They can hamper algorithmic analysis and visual inspection. Physiological artifacts can be caused i.a.by saccadic eye movement, eye blinks and activity of the scalp muscles, while technical artifacts result from external influences. To filter out obvious artifacts we perform an amplitude thresholding. Segments that are filled only with zero amplitude values are excluded from further analysis.

2.3. Feature extraction

In the feature extraction phase for each previously pre-processed EEG segment a representative feature vector is computed describing distinct properties of the signal.

Wavelet analysis has been reported to be effective in analyzing non-stationary signals. The wavelet transform decomposes a signal in terms of scaled and translated versions of a mother wavelet and a scaling function [9]. The discrete wavelet transform (DWT) has been frequently used in recent epileptic spike and seizure detection approaches that have shown promising results [10]. The choice on the maximal level of decomposition and an appropriate mother wavelet affect the result of the classification system. After testing various wavelets of the Daubechies, Coiflets and Symlets families, we have selected Symlets (sym) of order 7 as the basis wavelet in combination with a decomposition up to level 6.

The calculated wavelet coefficients are used to form the feature vectors. To reduce the dimensions of the feature vectors statistical features are extracted from the wavelet coefficients in each sub-band. The computed sub-bands consist of the detail coefficients from level 1 to 6 and the approximation coefficient at decomposition level 6. As features we compute the maximum, minimum and the mean of the coefficients and their standard deviation. Additionally, we calculate the relative wavelet energy and the entropy for each sub-band. Consequently, each feature vector consists of 6 statistical features for each set of wavelet coefficients resulting in a feature vector with 42 values.

2.4. Classification

In the classification phase the feature vectors derived from the previous processing are taken as input for machine learning classifiers. Here, feed-forward neural networks are used. After the iterative optimization procedure, the chosen architecture of the networks consists of two hidden layers with 15 nodes and tan-sigmoid activation function. The output of the networks corresponds to a categorical probability distribution that is computed by the softmax function. To reduce the variance of the output we average the results of 40 nets to classify a segment. The final decision of whether an EEG is considered to be abnormal is concluded by the average class probability distribution for all segments in one recording. A recording is classified as abnormal if a threshold of 50% for the corresponding probability is exceeded. For following experiments this probability threshold is further optimized.

Table 2. Overview of the results.

Training data	Testing data	AUC	Accuracy	Literature	Ref.
TUH training	TUH evaluation	.8498	79.78%	78.80%	[11]
Bonn A and D	*Cross-validation*	.9995 (±.0011)	99.50% (±1.12%)	-	-
Bonn A and E	*Cross-validation*	.9920 (±.0179)	99.50% (±1.12%)	97-100%	[12–14]
Bonn A-E	*Cross-validation*	.9961 (±.0046)	98.20% (±2%)	-	-
TUH training	Bonn A-E	.9224	87.20%	-	-
Bonn A-E	TUH evaluation	.7541	67.51%	-	-

3. Results

For the evaluation of our classification system we carried out several tests on both the TUH and the Bonn dataset (Table 2). The probability threshold for classifying an EEG as abnormal was chosen according to the highest Youden index for each training/testing combination. For each of these test cases the AUC and classification accuracy are reported as well as results from prior literature.

4. Discussion

The best results on the TUH dataset slightly outperform the current results of Obeid et al. [8] and Lopez de Diego [11]. We achieve an accuracy of 79.78% which is slightly higher than the reported accuracy of Lopez de Diego of 78.80%. Obeid et al. [8] achieved an accuracy of 83.00% on a smaller subset of the TUH evaluation set, but it is not reported which specific data was chosen. It is noticeable that our method has a high specificity (86.00%) in comparison to the sensitivity (72.44%). The confusion matrix indicates that the most frequent misclassification are abnormal EEG falsely classified as normal. Since abnormal activity may occur temporary, the lower sensitivity may result from the fact that we treat each segment of a longer abnormal recording as abnormal during training phase. The AUC value of the ROC curve of 0.8498 indicates a high discrimination between the classes abnormal and normal.

The tests on the Bonn datasets show competitive results in comparison to other recent methods. The general distinction of abnormal and normal EEG is performed with an accuracy of 98.20% on the full Bonn database.

The test cases combining the Bonn and TUH datasets show reasonable results. Training the classifiers on TUH data achieves an accuracy of 87.20% on the Bonn datasets. This result is close to the accuracy achieved when using cross-validation only on the full Bonn datasets. Training on the Bonn sets and testing on the TUH evaluation data results in an accuracy of 67.51%. A high sensitivity (85.83%) is achieved, while the specificity is lower (52.00%). The classification accuracy is lower in comparison to training and testing on TUH data only, which is to be expected. However, a higher sensitivity is achieved if training is performed on the Bonn data, since the segments for training certainly contain abnormal activity due to their manual selection.

5. Conclusion

We reported the development and evaluation of a machine learning solution for discrimination between normal and abnormal EEG data. While the methodology in principle follows the known ideas, the solution shows a number of advantages. The choice of features allows for fast computations, thanks to relatively small resulting input sizes. At the same time, this choice allows to capture the essential qualities of the data and achieve good classification results. An important characteristic of the proposed solution is its robustness, even when trained and tested on different sources of data. While TUH data set is large and contains raw data, Bonn set is relatively small and pre-processed. Unsurprisingly, the classification results are the best within the same homogeneous dataset, but also training the classifier on the TUH set gives good results on the Bonn set. The combination of the proposed classification algorithm with the TUH Abnormal Corpus as training data showed very promising results not only in terms of accuracy, but also potential usability for the new data sets.

Acknowledgement

This work was funded by the Excellence Initiative of the German federal and state governments (G:(DE-82)ZUK2-SF-OPSF424), and by the Faculty of Applied Mathematics AGH UST statutory tasks within subsidy of Polish Ministry of Science and Higher Education.

References

[1] E. Niedermeyer, and F.L. da Silva, Electroencephalography: basic principles, clinical applications, and related fields, Lippincott Williams & Wilkins, 2005.
[2] F.H. Lopes da Silva, S.I. Gonalves, and J.C. De Munck, Electroencephalography (EEG), in: L.R. Squire (Ed.), Encycl. Neurosci., Academic Press, Oxford, 2009: pp. 849–855.
[3] W.O. Tatum IV, Long-term EEG monitoring: a clinical approach to electrophysiology, *J. Clin. Neurophysiol.* **18** (2001) 442–455.
[4] Y.S. Kikuchi, W. Sato, K. Ataka, K. Yagisawa, Y. Omori, T. Kanbayashi, and T. Shimizu, Clozapine-induced seizures, electroencephalography abnormalities, and clinical responses in Japanese patients with schizophrenia, *Neuropsychiatr. Dis. Treat.* **10** (2014) 1973.
[5] S. Noachtar, C. Binnie, J. Ebersole, F. Mauguiere, A. Sakamoto, and B. Westmoreland, A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings, *Klin. Neurophysiol.* **35** (2004) 5–21.
[6] A. Harati, S. Lopez, I. Obeid, J. Picone, M. Jacobson, and S. Tobochnik, The TUH EEG CORPUS: A big data resource for automated EEG interpretation, in: Signal Process. Med. Biol. Symp., IEEE, 2014: pp. 1–5.
[7] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Phys. Rev. E*. **64** (2001) 061907.
[8] I. Obeid, and J. Picone, Machine Learning Approaches to Automatic Interpretation of EEGs, in: E. Sejdic, and T.H. Falk (Eds.), Biomed. Signal Process. Big Data, CRC Press, 2018.
[9] G. Kaiser, A friendly guide to wavelets, Springer Science & Business Media, 2010.
[10] F.D. De Moraes, and D.A. Callegari, Automated detection of interictal spikes in EEG: A literature review, Pontifícia Universidade Católica do Rio Grande do Sul Av. Ipiranga, 2015.
[11] S.L. de Diego, Automated Interpretation Of Abnormal Adult Electroencephalograms, Master's Thesis, Temple University, 2017.
[12] P. Jahankhani, V. Kodogiannis, and K. Revett, EEG signal classification using wavelet feature extraction and neural networks, in: Mod. Comput. 2006 JVA06 IEEE John Vincent Atanasoff 2006 Int. Symp. On, IEEE, 2006: pp. 120–124.
[13] U. Orhan, M. Hekim, and M. Ozer, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model, *Expert Syst. Appl.* **38** (2011) 13475–13481.
[14] V.P. Nigam, and D. Graupe, A neural-network-based detection of epilepsy, *Neurol. Res.* **26** (2004) 55–60.