2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Towards Automatic Identification of Epileptic Recordings in Long-term EEG Monitoring

Xuen Hoong Kok, Syed Anas Imtiaz and Esther Rodriguez-Villegas

Abstract— **Electroencephalogram (EEG) is a crucial tool in the diagnosis and management of epilepsy. The process of analyzing EEG is time consuming leading to the development of seizure detection algorithms to aid its analysis. This approach is limited since it requires seizures to occur during monitoring periods and can often lead to misdiagnosis in cases where seizure occurrence is rare. For such cases, it has been shown that the interictal periods in EEG signals, which is the predominant state in long-term monitoring, can be useful for the diagnosis of epilepsy. This paper presents an algorithm, using the information in interictal periods, to discriminate between long-term EEG recordings of epilepsy patients and healthy subjects. It extracts several time and frequency-time domain features from the signals and classifies them using an ensemble classifier, achieving 100% sensitivity and 98.7% specificity in classifying 267 recordings from 105 subjects. The results demonstrate the feasibility of this approach to reliably identify EEG recordings of epilepsy subjects automatically which can be highly useful to facilitate screening and diagnosis of epilepsy, especially in those parts of the world where there is a lack of trained personnel for interpreting EEG signals.**

I. Introduction

Epilepsy is a chronic condition that affects the brain and is characterized by the occurrence of, often debilitating, seizures [1]. It is diagnosed by monitoring the electrical activity of the brain by placing electrodes on the scalp to obtain the electroencephalogram (EEG). This is usually performed in hospitals, where trained staff are able to provide assistance and document seizures as and when they happen. The randomness and unpredictable nature of seizures makes it very difficult to record them during the limited-time EEG monitoring. Hence, in some cases, patients are sent home with ambulatory EEG recording units to increase the likelihood of capturing seizure events.

To help with the diagnosis and management of epilepsy, wearable EEG systems are being developed for long-term monitoring of patients [2], [3]. While providing more context, such systems result in significantly large amount of recording data that need to be analyzed. Manual analysis of this data for the identification of epileptic activities is extremely time consuming as the predominant state is the interictal (between seizure) period. Algorithms for detecting seizures have been developed, to reduce the analysis time and aid epileptologists by preselecting only the regions of interest in the EEG signals for review [4], [5], [6]. Such algorithms are known to have accuracy issues in patients with sporadic seizures where they are likely to miss detection of seizures or end up with large number of false positives, and can consequently lead to misdiagnosis. This can have serious implications on the person, such as increased costs associated with unnecessary prescriptions of antiepileptic drugs (AEDs), restrictions imposed on their normal, work or school lives, and in some cases, being subjected to highly invasive methods for further investigation [7].

While wearable EEG can help to collect significant amount of data, it is also possible for patients not to have any seizures during the period of long-term recording. Despite the absence of seizures, EEG data from epilepsy patients are more likely to contain interictal epileptiform discharges (IEDs) [8], which can be characteristic of epilepsy and thus can be useful for clinicians to analyze. There is therefore a need for an objective method to automatically identify such EEG recordings that can then be used for further investigations of epilepsy [9]. Such a tool would be a useful tool for clinicians to automatically identify recordings that need more attention without spending hours on manual analysis of recordings from healthy subjects. Additionally, it would facilitate screening and diagnosis of epilepsy in parts of the world where there is a lack of trained personnel for interpreting EEG signals. This paper presents an algorithm to discriminate between long-term EEG recordings of epilepsy patients and healthy subjects. It is organized as follow. Section II describes the two datasets used in this paper for the development and validation of the algorithm. Section III presents the algorithm developed, which was based on machine learning, to differentiate between scalp EEG recordings of patients with epilepsy and healthy subjects. In Section IV, the performances of the algorithm and the underlying machine learning classifier are evaluated. Finally, the discussions and conclusion are presented in Section V.

II. Materials

For the development and validation of the algorithm, two datasets were used in this work. The first, *sleep-edfx*, was a publicly available dataset containing scalp EEG recordings from healthy subjects with no epilepsy. The second, *NHNN*, was an anonymized dataset containing EEG recordings from epilepsy patients. A summary of the number of subjects and recording duration in each dataset is shown in Table I.

A. Sleep-edfx

The *sleep-edfx* database [10], [11] on Physionet [12] consists of data from two studies: sleep cassette and sleep

This work was supported by the European Research Council (ERC) grant no 724334 and SUDEP Action.

X. H. Kok, S. A. Imtiaz and E. Rodriguez-Villegas are with the Wearable Technologies Lab, Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, United Kingdom. E-mail: {x.kok17, anas.imtiaz, e.rodriguez}@imperial.ac.uk.

TABLE I

SUMMARY OF THE DATASETS USED IN THIS WORK

	sleep-edfx	NHNN
Number of subjects	78	27
Number recordings	153	116
Total recordings duration (hours)	3470	1362.5

telemetry. Only data from the sleep cassette study were used in this work, since test subjects in the sleep telemetry study took a medication used in treating insomnia. Subjects in the sleep cassette study were monitored using an ambulatory EEG system (four-channel cassette recorder), for a period lasting two days (approx. 48 hours recording). The recordings were obtained using the Fpz-Cz and Pz-Oz EEG montages with a sampling frequency of 100 Hz. In this dataset, the total number of subjects is 78 (M:37, F:41), the median age of subjects is 57 years (range: 25 – 101), and the total recording duration is 3470 hours.

B. NHNN

The *NHNN* dataset is an anonymized research database that was created as part of a study carried out at the National Hospital for Neurology and Neurosurgery (NHNN), London, United Kingdom (approved by the UK Health Research Authority, REC reference number 16/WA/0319). It contains scalp EEG recordings from 27 patients (M:13, F:14) with focal epilepsy. The median age of the patients is 32 years (range: 20 – 53), and the total recording duration is 1362.5h. Of the 27 patients in the database, 14 of them had at least one seizure that occurred during the monitoring period. Although the EEG was obtained with 30 channels, only signals from Fz-Cz montage were used as this corresponded to the closest location available in the *sleep-edfx* dataset.

III. METHODS

A high-level overview of the proposed algorithm is shown in Fig. 1. It consists of the following stages: data preprocessing, feature extraction, and classification, which are described in this section.

A. Data preprocessing

EEG signals from the Fz-Cz montage of both datasets were bandpass filtered in the range of 0.16 Hz and 45 Hz, using a 6^{th} order Butterworth filter, to eliminate any high frequency interference that were captured. Since the recordings in the *NHNN* dataset were recorded at a sampling frequency of 256 Hz, they were resampled to 100 Hz to match those in the *sleep-edfx* dataset. After the filtering and resampling steps, the EEG recordings were split into 2s non-overlapping epochs for further processing.

B. Feature extraction

Several features based on the time domain (TD), Fourier transform (FT) and the discrete wavelet transform (DWT) were extracted from each 2s epoch. These features were explored to determine which ones are relevant using the minimum redundancy maximum relevance (mRMR) feature selection algorithm [13]. A list of the features extracted is shown in Table II. They are based on those used in [14], to characterize EEG signals, and subsequently, to detect the onset of seizure. The equations for the calculation of these features are available in [15].

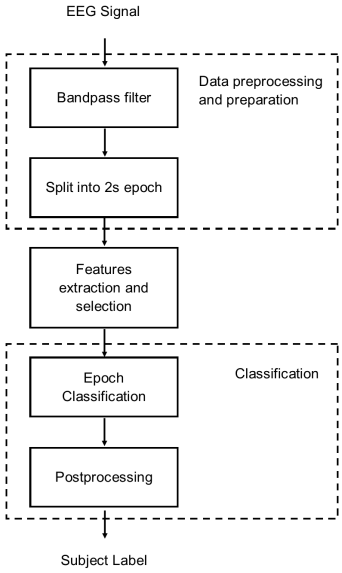

Fig. 1. A high-level overview of the proposed algorithm.

The TD features were calculated directly from the 2s signal epochs. Fourier transform was applied to the signals before FT features were calculated. The DWT features were calculated using the Daubechies 4 ('db4') wavelet, and a four-level decomposition was used. This results in details coefficients (cD) in four frequency bands (cD1, cD2, cD3, and cD4) and approximation coefficients (cA) in one frequency band (cA4), which correspond to frequencies in the range of 25–50 Hz, 12.5–25 Hz, 6.25–12.5 Hz, 3.125–6.25 Hz, and 0–3.125 Hz respectively.

TABLE II

FEATURES EXTRACTED FROM EEG RECORDINGS

Method	Extracted features
Time domain (TD)	Complexity, energy, fractal dimension, minimum, maximum, mean, variance, skewness, kurtosis, line length, mobility, non-linear energy, relative derivative, Shannon entropy, total local maxima and minima, zero crossing, first derivative of zero crossing
Fourier transform (FT)	Median frequency, peak frequency, power, spectral edge frequency, spectral entropy, total spectral power
Discrete wavelet transform (DWT)	Bounded variation, coefficients, energy, entropy, relative bounded variation, relative scale energy, standard deviation

C. Classification

The most relevant features were used with an ensemble classifier based on the gentle adaptive boosting algorithm

(Gentle AdaBoost) [16] to determine if the 2s EEG signal epochs belonged to epilepsy patients or healthy subjects. The models were trained in MATLAB, using a base classifier of a classification and regression tree (CART), and 30 learners trained in each model. Samples from epilepsy patients and healthy subjects were labelled '1' and '0' respectively. The data were partitioned at random into training (70%) and test (30%) sets. As the equipment used in data acquisition were different, the features used were normalized to values in the range of 0 to 1 prior to model training.

The output of the classifier indicated if the epochs were from epilepsy patients or healthy subjects. These were then postprocessed to determine the overall label for the recording (epilepsy or healthy). This was done by first calculating the percentage of epochs for each patient that were classified as positive. If this percentage was greater than a certain threshold, the algorithm considered it to be from the long-term recording of an epilepsy patient.

D. Performance evaluation

The performance of the algorithm was evaluated using the sensitivity $(SENS)$ and specificity $(SPEC)$ metrics that measure the fraction of correctly identified epilepsy and healthy subjects' recordings respectively, and are defined as follows:

$$SENS = \frac{TP}{TP + FN} \tag{1}$$

$$SPEC = \frac{TN}{TN + FP} \tag{2}$$

where,

- TP is the number of recordings from epilepsy patients correctly identified by the algorithm,
- FN is the number of recordings from epilepsy patients incorrectly identified as healthy subjects,
- TN is the number of recordings from healthy subjects correctly identified by the algorithm,
- FP is the number of recordings from healthy subjects incorrectly identified as epilepsy patient.

Additionally, the performance of the epoch classification models can be evaluated using the classification accuracy (ACC_{clf}), classification sensitivity $(SENS_{clf})$ and classification specificity $(SPEC_{clf})$. They are defined as follows:

$$ACC_{clf} = \frac{TP_{ep} + TN_{ep}}{TP_{ep} + TN_{ep} + FP_{ep} + FN_{ep}} \tag{3}$$

$$SENS_{clf} = \frac{TP_{ep}}{TP_{ep} + FN_{ep}} \tag{4}$$

$$SPEC_{clf} = \frac{TN_{ep}}{TN_{ep} + FP_{ep}} \tag{5}$$

where,

- TP_{ep} is the number of epochs belonging to epilepsy patients that were correctly classified,
- TN_{ep} is the number of epochs belonging to healthy subjects that were correctly classified,
- FP_{ep} is the number of epochs belonging to healthy subjects that were misclassified,
- FN_{ep} is the number of epochs belonging to epilepsy patients that were misclassified.

IV. Results

The features extracted in Table II, after being ranked is shown in Table III. Four combinations of the first 5, 10, 15, and 20 features returned by the mRMR algorithm were tested. The results of the epochs classification task on the test sets using the respective number of features are shown in Table IV. With the top 20 ranked mRMR features used in the training of the model, 83.66% of all epochs were correctly classified, and the classification sensitivity and specificity are 63.84% and 91.76% respectively.

TABLE III
FEATURES RANKED BY THE MRMR ALGORITHM

	Features	rank
TD	First deriv. of zero crossing	1
TD	Skewness	2
DWT	cD2 Coefficient	3
DWT	cD3 Coefficient	4
DWT	cD4 Relative energy	5
TD	Energy	6
TD	Total local maxima minima	7
TD	Kurtosis	8
DWT	cA4 Entropy	9
FT	Median frequency	10
TD	Fractal dimension	11
FT	Spectral entropy (12.5–2 5Hz)	12
TD	Minimum	13
DWT	cD4 Coefficient	14
FT	Spectral entropy (0–3.125 Hz)	15
DWT	cD3 Entropy	16
DWT	cD2 Entropy	17
DWT	cD4 Entropy	18
TD	Variance	19
DWT	cD4 Bounded variation	20

TABLE IV
EPOCH-BASED CLASSIFICATION RESULTS USING DIFFERENT NUMBER OF MRMR FEATURES

	Num. mRMR			
	5	10	15	20
Accuracy (%)	78.39	80.75	83.53	83.66
Sensitivity (%)	39.24	51.37	62.63	62.84
Specificity (%)	93.61	92.18	91.66	91.76

The postprocessing steps were applied on to the epoch classification results of model that achieved the highest performance (trained with 20 mRMR features). In epilepsy patients' recordings, the average percentage of positive-class epochs was 59.33±13.3%, but in healthy subjects, this was 8.21±4.4%. A histogram showing the average number of epochs predicted as '1' in a recording sorted by the subject groups are shown in Fig. 2. Applying threshold values from 0.05 to 0.45 (at 0.05 increments), the corresponding sensitivity and specificity values were found, as shown in Table V. At the threshold of 0.2, the sensitivity and specificity of the algorithm are highest, at 100% and 98.7% respectively.

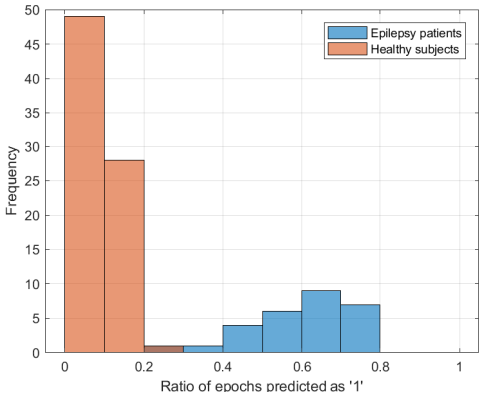

Fig. 2. Histogram showing the ratio of epochs predicted as '1' for each subject. An overlap in the ratios between epilepsy and healthy group at 0.2 and 0.3 can be seen.

TABLE V
ALGORITHM SENSITIVITY AND SPECIFICITY AT VARIOUS THRESHOLDS

Threshold	Sensitivity (%)	Specificity (%)
0.05	100.0	24.4
0.10	100.0	62.8
0.15	100.0	97.4
0.20	100.0	98.7
0.25	96.3	100.0
0.30	96.3	100.0
0.35	92.6	100.0
0.40	92.6	100.0
0.45	85.2	100.0

V. DISCUSSION

This paper presented a novel algorithm to differentiate between long-term EEG recordings acquired from epilepsy patients and healthy subjects. It used an ensemble machine learning classifier to classify individual signals in 2s epochs, which were then postprocessed to label the complete recordings as either epilepsy or healthy. It resulted in high accuracy, with 100% sensitivity and 98.7% specificity when classifying 267 recordings from 27 epilepsy and 78 healthy subjects.

The algorithm presented in this paper can be highly useful for clinicians to filter EEG recordings, focusing on those potentially coming from patients with epilepsy and thus reduce the time taken for diagnosis. It is, however, not intended as a replacement to conventional diagnostic methods, but when used alongside them, it aims to alleviate some of the challenges present in manual analysis of EEG and/or situations where there is incomplete history of the disease or lack of seizures in EEG recordings. This can lead to a reduction the uncertainties, and increase the confidence and accuracy in the diagnosis and management of epilepsy.

The specificity performance of the algorithm can be further improved by identifying and removing artefacts at the preprocessing stage. In the algorithm presented, there were no attempts to remove signal epochs corrupted by artefacts. However, as EEG signals are prone to artefacts and may take on characteristics similar to epileptic activity, the presence of artefacts can compromise the classification performance. Therefore, an artefact rejection stage that automatically cleans the signal or removes corrupted signal epochs, can be added to improve its performance. Additionally, the threshold value can be changed to improve specificity at the expense of detection sensitivity to meet specific clinical requirements. Despite the potential for improvements, the results in this work already demonstrate that characteristic features exist in long-term EEG recordings of epilepsy patients that can be used to reliably discriminate them from the recordings of healthy subjects.

REFERENCES

[1] World Health Organization. (2019, June) Epilepsy. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[2] A. Casson, D. Yates, S. Smith, J. Duncan, and E. Rodriguez-Villegas, "Wearable electroencephalography," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. 3, pp. 44–56, May 2010, DOI: 10.1109/memb.2010.936545.

[3] A. J. Casson, "Wearable EEG and beyond," *Biomedical Engineering Letters*, vol. 9, no. 1, pp. 53–71, Jan. 2019, DOI: 10.1007/s13534-018-00093-6.

[4] L. Logesparan, A. J. Casson, S. A. Imtiaz, and E. Rodriguez-Villegas, "Discriminating between best performing features for seizure detection and data selection," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, July 2013, DOI: 10.1109/embc.2013.6609844.

[5] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, Sept. 2018, DOI: 10.1016/j.compbiomed.2017.09.017.

[6] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 83–94, Jan. 2019, DOI: 10.1109/jbhi.2018.2871678.

[7] C. D. Ferrie, "Preventing misdiagnosis of epilepsy," *Archives of Disease in Childhood*, vol. 91, no. 3, pp. 206–209, Mar. 2006, DOI: 10.1136/adc.2005.088906.

[8] J. Pillai and M. R. Sperling, "Interictal EEG and the diagnosis of epilepsy," *Epilepsia*, vol. 47, no. s1, pp. 14–22, Oct. 2006, DOI: 10.1111/j.1528-1167.2006.00654.x.

[9] M. M. Oto, "The misdiagnosis of epilepsy: Appraising risks and managing uncertainty," *Seizure*, vol. 44, pp. 143–146, Jan. 2017, DOI: 10.1016/j.seizure.2016.11.029.

[10] B. Kemp, A. Zwinderman, B. Tuk, H. Kamphuisen, and J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000, DOI: 10.1109/10.867928.

[11] M. S. Mourtazaev, B. Kemp, A. H. Zwinderman, and H. A. C. Kamphuisen, "Age and gender affect different characteristics of slow waves in the sleep EEG," *Sleep*, vol. 18, no. 7, pp. 557–564, sep 1995, DOI: 10.1093/sleep/18.7.557.

[12] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, jun 2000.

[13] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, aug 2005, DOI: 10.1109/tpami.2005.159.

[14] L. Logesparan, A. J. Casson, and E. Rodriguez-Villegas, "Optimal features for online seizure detection," *Medical & Biological Engineering & Computing*, vol. 50, no. 7, pp. 659–669, apr 2012, DOI: 10.1007/s11517-012-0904-x.

[15] ——, "Erratum to: Optimal features for online seizure detection," *Medical & Biological Engineering & Computing*, vol. 54, no. 8, pp. 1295–1295, jun 2016, DOI: 10.1007/s11517-016-1535-4.

[16] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, apr 2000, DOI: 10.1214/aos/1016218223.