35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Discriminating between best performing features for seizure detection and data selection

Lojini Logesparan, Alexander J. Casson, Syed Anas Imtiaz and Esther Rodriguez-Villegas

Abstract— **Seizure detection algorithms have been developed to solve specific problems, such as seizure onset detection, occurrence detection, termination detection and data selection. It is thus inherent that each type of seizure detection algorithm would detect a different EEG characteristic (feature). However most feature comparison studies do not specify the seizure detection problem for which their respective features have been evaluated. This paper shows that the best features/algorithm bases are not the same for all types of algorithms but depend on the type of seizure detection algorithm wanted. To demonstrate this, 65 features previously evaluated for online seizure data selection are re-evaluated here for seizure occurrence detection, using performance metrics pertinent to each seizure detection type whilst keeping the testing methodology the same. The results show that the best performing features/algorithm bases for data selection and occurrence detection algorithms are different and that it is more challenging to achieve high detection accuracy for the former seizure detection type. This paper also provides a comprehensive evaluation of the performance of 65 features for seizure occurrence detection to aid future researchers in choosing the best performing feature(s) to improve seizure detection accuracy.**

I. INTRODUCTION

Automated seizure detection algorithms to aid diagnosis and treatment of epilepsy have been actively researched for decades [1]–[3] because it is very challenging to obtain high detection accuracy. Recently there has been increased awareness of the methodological factors in the design and evaluation of seizure detection algorithms—such as the choice of performance metrics [4], [5] and EEG amplitude variation over time [6]—that limit the reported detection accuracy. This paper investigates another methodological factor: the precise aim of the seizure detection algorithm and its impact on the design of these algorithms.

Seizure detection algorithms have been developed to solve different *specific* problems such as: seizure occurrence detection [1], onset detection [2], termination detection [7] and seizure recording/data selection [8]. Each type of algorithm discriminates between a specific seizure and non-seizure state at different times within the duration of the seizure. Occurrence detection algorithms detect *any* seizure section; onset detection algorithms detect the start; termination algorithms detect the end; and data selection algorithms detect the *entire duration* of the seizure. Thus the discriminating seizure characteristics suitable for each algorithm are inevitably different. In each case, automated signal processing can be used to extract *features* that describe these characteristics and which can be used in a detection algorithm.

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, UK. Email: {lojini.logesparan04, acasson, anas.imtiaz08, e.rodriguez} @imperial.ac.uk.

The research leading to these results has received funding from the European Research Council under the European Community's 7th Framework Programme (FP7/2007-2013) / ERC grant agreement no. 239749.

Feature comparison studies such as [5], [9]–[11] provide a comprehensive approach for selecting EEG characteristics that best discriminate between the required seizure sections and the irrelevant non-seizure sections. In such studies, the performance metrics selected to evaluate the performance of these features must reflect the seizure detection problem to be solved. For example, the largest feature comparison study on adult scalp EEG [8] evaluates 65 different features using performance metrics pertinent to data selection for low power devices [12]. Another study [9] evaluated 21 features using performance metrics relevant to neonatal data selection [5] in applications where computational complexity/power consumption is not a limiting factor. It is thus understandable that the best performing features presented in these studies would differ due to the different performance metrics selected. In contrast to these studies, it is not always easy to relate the specific seizure detection problem to each feature comparison study as the majority of previous work do not provide sufficient information about the performance metrics used in feature evaluation.

This paper demonstrates that the best features/algorithm bases will not be the same for every type of seizure detection problem, but depend on the testing methodology and type of seizure detection wanted. In particular, [8] evaluated 65 features for the seizure data selection case. This paper re-evaluates the performance of these features when used in the seizure occurrence detection problem case. The performance of the same features for the two kinds of seizure detection problem are not the same and these are contrasted in detail. Section II describes the methodology used to evaluate features for seizure occurrence detection and online data selection. The performance of each feature for seizure occurrence detection is shown in Section III-A and the choice of best features for both seizure detection problems is then discussed in Section III-B.

II. METHODS

A. Feature evaluation

The features evaluated in [8] are listed in Table I where they are split into four groups depending on whether they are calculated in the Time Domain (TD), Fourier Transform (FT) domain, or use coefficients from the Continuous Wavelet Transform (CWT) or Discrete Wavelet Transform (DWT). For a representative comparison of the performance of features for the two seizure detection types, the same

TABLE I

FEATURES INVESTIGATED IN THIS STUDY

TD	Complexity, energy/power, fractal dimension, kurtosis, line length, maximum, mean, minimum, mobility, non-linear energy, relative derivative, Shannon entropy, skewness, total maxima and minima, variance/standard deviation, zero crossing, zero crossing of first derivative.
FT	Median frequency, peak frequency, power*, spectral edge frequency, spectral entropy*, total spectral power.
CWT	Coefficient z-score, energy, entropy, standard deviation of energy.
DWT	Bounded variation*, coefficients*, energy*, entropy*, relative bounded variation*, relative power*, relative scale energy*, variance/standard deviation*.

* calculated across 4 frequency ranges: D3 (12.5–25 Hz), D4 (6.25–12.5 Hz), D5 (3.125–6.25 Hz) and A5 (0–3.125 Hz).

algorithm and database have been used to evaluate the same set of features. However the performance of these features are assessed using metrics pertinent to each seizure detection type as recommended in [5] and described below.

To evaluate the performance of the features in Table I, each feature is placed in turn into the simple seizure detection algorithm proposed in [8] (shown in Fig. 1) and evaluated on the same adult scalp EEG database utilized in [8]. The algorithm and database are described in detail in the Appendix and the calculation procedure for each feature can be found in [8].

B. Performance metrics for seizure occurrence detection

Seizure occurrence detection algorithms, the most common use of seizure detection algorithms, are used to assist neurologists or EEG technicians in the offline review of EEG data by marking sections that contain seizures. Here the algorithm only needs to detect or *mark* any one section of the EEG for the neurologist to view data on either side of the marker. The fraction of correct seizure events detected, event-sensitivity, is an important measure for seizure occurrence detection and is calculated as

$$\text{Event-sensitivity} = \frac{1}{M}\sum_{i=1}^{M}\frac{TP}{TP+FN} \times 100\% \quad (1)$$

where M is the number of EEG records in the test database, each indexed by i, TP is the number of true positives (correctly detected expert marked seizure events) and FN is the number of false negatives (incorrectly rejected seizure events). As a seizure occurrence detection algorithm only needs to detect short explicit sections of seizure EEG for the entire seizure to be considered detected, these algorithms are often not limited by their sensitivity.

To achieve good detection performance, the specificity of rejecting non-seizure EEG is the limiting factor. Specificity is the fraction of non-seizure sections correctly rejected. High specificity is desirable as it corresponds to fewer sections of non-seizure EEG that the neurologist has to review. It is calculated as,

$$\text{Specificity} = \frac{1}{M}\sum_{i=1}^{M}\frac{TN}{TN+FP} \times 100\% \quad (2)$$

where TN is the number of true negatives (correctly rejected non-seizure sections, divided in to 2 s non-overlapping epochs) and FP is the number of false positives (incorrectly detected 2 s epochs).

Event-sensitivity is traded off with the specificity of the algorithm as higher sensitivities can be achieved with lower specificity. The event-sensitivity-specificity trade-off point can be altered by a detection threshold (β) (shown in the seizure detection algorithm in Fig. 1). The threshold can be swept over [0,1] to obtain different sensitivity-specificity pairs which can then be plotted as a sensitivity-specificity trade-off curve. The area under the curve (AUC_{OD}) can then be calculated using trapezoidal estimation and it provides a good overall measure of the *average* sensitivity for all values of specificity or vice versa. Higher values of AUC denote better performance and an ideal feature will achieve AUC=1.

To determine an optimal event-sensitivity-specificity trade-off point, the error between the event-sensitivity (E) and specificity (S) achieved by the feature and the performance of an ideal feature (100% event-sensitivity and 100% specificity) is calculated:

$$\text{Error} = \sqrt{(100\% - E)^2 + (100\% - S)^2}. \quad (3)$$

The event-sensitivity-specificity point that minimizes the error in (3) is then noted for feature comparison.

C. Performance metrics for data selection

In contrast to seizure occurrence detection, data selection algorithms select EEG sections for discontinuous recording and thus only the detected EEG will be available for review by a neurologist. Hence the duration of the seizure correctly detected by the algorithm is more relevant for data selection algorithms instead of the fraction of seizure events correctly detected in (1). An ideal data selection algorithm would detect the entire duration of the seizure and reject all non-seizure EEG. This is intrinsically more challenging than seizure occurrence detection as a data selection algorithm must cope with changes in the EEG as the seizure evolves over time. Thus data selection algorithms are often limited by epoch-sensitivity (percentage of seizure duration detected).

Other metrics of interest for online data selection are [8]: specificity as given in (2); area under the curve (AUC_{DS} based on epoch-sensitivity and specificity); and the relative computational complexity of implementing each feature over another in hardware.

III. RESULTS

The performance of all 65 features is listed in Table II for Time Domain (TD) features, Table III for Fourier Transform (FT) based features, Table IV for Continuous Wavelet Transform (CWT) based features and Table V for Discrete Wavelet Transform (DWT) based features. Each table lists the event-sensitivity and associated specificity at the optimal threshold (β) for seizure occurrence detection, in addition to the area under the sensitivity-specificity curve across all thresholds for occurrence detection (OD) and data selection (DS) (results for DS were obtained from [8]).

TABLE II

RESULTS FOR TIME DOMAIN FEATURES.

Feature	Occurrence detection Sens. (%)	Spec. (%)	β	Comparison AUC OD	DS
Mean	93.62	91.47	0.70	0.95	0.64
Line length	85.11	92.24	0.95	0.93	0.77
Non-linear energy	87.23	85.78	0.75	0.93	0.76
Skewness	89.36	88.15	0.80	0.93	0.58
Maximum	87.23	91.29	0.85	0.93	0.74
Variance	89.36	85.80	0.55	0.92	0.75
Minimum	89.36	92.74	0.90	0.92	0.72
Zero crossing	91.49	84.67	0.98	0.92	0.61
Total maxima and minima	89.36	85.10	0.99	0.91	0.67
Energy/power	89.36	84.12	0.50	0.91	0.74
Kurtosis	85.11	87.50	0.80	0.91	0.54
Complexity	87.23	79.50	0.70	0.90	0.64
Mobility	85.11	82.46	0.95	0.89	0.63
Zero crossing first derivative	82.98	86.41	0.99	0.88	0.60
Relative derivative	76.60	86.27	0.70	0.87	0.66
Shannon entropy*	87.23	69.85	0.50	0.86	0.63
Fractal dimension	74.47	82.69	0.95	0.86	0.53

* feature decreases during seizure.

TABLE III

RESULTS FOR FT-BASED FEATURES.

Feature	Occurrence detection Sens. (%)	Spec. (%)	β	Comparison AUC OD	DS
Spectral entropy (D5)	89.36	92.27	0.65	0.95	0.73
Spectral entropy (D3)	87.23	91.81	0.60	0.94	0.74
Spectral entropy (A5)	85.11	93.34	0.70	0.94	0.70
Power (D5)	89.36	93.05	0.75	0.94	0.73
Power (A5)	85.11	92.91	0.70	0.94	0.70
Spectral entropy (D4)	93.62	82.79	0.45	0.93	0.69
Power (D3)	87.23	92.00	0.75	0.93	0.72
Total spectral power	91.49	90.63	0.65	0.93	0.72
Power (D4)	82.98	91.17	0.75	0.92	0.68
Spectral edge frequency	91.49	81.74	0.95	0.89	0.55
Median frequency	97.87	72.13	0.99	0.85	0.64
Peak frequency	97.87	71.87	0.99	0.85	0.64

A. *Seizure occurrence detection*

The highest area under the curve (AUC=0.97) was achieved by DWT coefficients across all frequency ranges and DWT based relative power in the 3.125 Hz to 6.25 Hz frequency range. All DWT coefficients achieve over 90% event-sensitivity and specificity in Table V, while DWT based relative power achieves higher specificity than the DWT coefficients at lower sensitivities and lower specificity at higher sensitivities. Across all feature categories, DWT based features performed the best with 50% of the features achieving AUC$\geq$ 0.95 whilst only the top feature in the other three categories achieved the same performance.

B. *Comparison with features for online data selection*

There are two significant differences in the performance of the same features for seizure occurrence detection as described above and online data selection as reported in [8]. Firstly, the best performing features are different. For data selection, the highest AUC of 0.83 is achieved by DWT based relative power (12.5 Hz–25 Hz and 3.125 Hz–6.25 Hz) whilst for seizure occurrence detection, DWT coefficients and relative power (3.125 Hz–6.25 Hz) have the best performance. When computational complexity is also considered for *online* data selection, [8] reports that DWT based relative power (12.5 Hz–25 Hz) and line length were the best performers.

TABLE IV

RESULTS FOR CWT-BASED FEATURES.

Feature	Occurrence detection Sens. (%)	Spec. (%)	β	Comparison AUC OD	DS
Coefficient z-score	93.62	92.11	0.75	0.96	0.69
Energy	91.49	90.88	0.65	0.94	0.72
Std. deviation energy	87.23	90.28	0.65	0.93	0.70
Entropy	95.75	83.06	0.95	0.91	0.63

TABLE V

RESULTS FOR DWT-BASED FEATURES.

Feature	Occurrence detection Sens. (%)	Spec. (%)	β	Comparison AUC OD	DS
Coefficients (D3)	97.87	92.29	0.85	0.97	0.69
Coefficients (D4)	95.74	93.87	0.90	0.97	0.66
Coefficients (D5)	95.74	93.19	0.85	0.97	0.65
Coefficients (A5)	93.62	91.73	0.70	0.97	0.68
Rel. power (D5)	89.36	92.86	0.10	0.97	0.83
Rel. power (D3)	93.62	88.63	0.04	0.96	0.83
Rel. power (D4)	85.11	93.02	0.08	0.96	0.81
Rel. scale energy (D5)	95.74	90.46	0.85	0.96	0.65
Bounded var. (D5)	97.87	91.95	0.98	0.96	0.66
Bounded var. (A5)	91.49	91.49	0.96	0.96	0.67
Rel. bounded var. (D4)	95.74	89.31	0.96	0.95	0.63
Rel. bounded var. (D5)	93.62	89.98	0.96	0.95	0.66
Rel. bounded var. (A5)	93.62	89.88	0.94	0.95	0.67
Variance (D5)	91.49	92.15	0.70	0.95	0.75
Energy (D5)	91.49	92.11	0.70	0.95	0.75
Entropy (D5)	93.62	91.49	0.65	0.95	0.75
Rel. power (A5)	89.36	84.17	0.06	0.94	0.73
Bounded var. (D4)	87.23	89.05	0.96	0.94	0.61
Variance (A5)	89.36	91.37	0.60	0.94	0.73
Energy (D3)	89.36	90.05	0.70	0.94	0.71
Energy (A5)	91.49	91.31	0.60	0.94	0.73
Entropy (A5)	91.49	90.76	0.55	0.94	0.73
Rel. scale energy (D3)	85.11	90.53	0.80	0.93	0.62
Rel. scale energy (D4)	87.23	86.95	0.75	0.93	0.61
Rel. scale energy (A5)*	87.23	83.26	0.45	0.93	0.57
Bounded var. (D3)*	89.36	86.58	0.30	0.93	0.53
Rel. bounded var. (D3)	89.36	83.87	0.92	0.93	0.54
Variance (D3)	89.36	90.06	0.70	0.93	0.71
Variance (D4)	91.49	89.50	0.70	0.93	0.70
Energy (D4)	91.49	89.49	0.70	0.93	0.70
Entropy (D3)	87.23	94.86	0.90	0.93	0.71
Entropy (D4)	89.36	88.71	0.65	0.92	0.70

* feature decreases during seizure.

Second, the sensitivity, specificity and area under the curve metrics show worse appearing results for data selection, because epoch-sensitivity appears to be worse than event-sensitivity across all features. The highest area under the curve achieved here AUC_{OD} is 0.97 while the highest AUC_{DS} for data selection in [8] is 0.83.

Based on these results it can be concluded that separate algorithms need to be developed for seizure occurrence detection and (online) data selection, as the same features/algorithm bases would not give the best performance in both cases. Furthermore, data selection algorithm development appears to be more challenging than that of seizure occurrence detection algorithms as the former must consider changes to the EEG as the seizure evolves over time.

IV. CONCLUSION

The performance of 65 features was evaluated for seizure occurrence detection using event-sensitivity, specificity and area under the curve, and the results were compared to previously reported performance of the same features evaluated for online data selection using epoch-sensitivity, specificity, area under the curve and relative computational complexity. For seizure occurrence detection, DWT based features performed the best with 16 out of 32 features achieving an area under the curve $\geq$0.95. The best performing features (AUC=0.97) were DWT coefficients (0 Hz-25 Hz) and DWT based relative power (3.125 Hz-6.25 Hz). As previously reported, line length and DWT based relative power (12.5 Hz-25 Hz) are the best performing features for online data selection. Overall, the performance of the features for seizure occurrence detection appeared to be better than the same features for data selection. This study demonstrates that different feature(s)/algorithm bases should be utilized in the development of data selection and seizure occurrence detection algorithms. It is also the largest systematic comparison of characteristic features for seizure occurrence detection in adult scalp EEG and can aid researchers in choosing the best performing feature(s) to improve the detection accuracy of future seizure occurrence detection algorithms.

APPENDIX

The features presented in this study have been evaluated on over 172 hours of scalp EEG data obtained from 24 adult patients. The database contains 16 EEG channels that are common to all recordings: C3, C4, CZ, F3, F4, FZ, F7, F8, FP1, FP2, O1, O2, T3, T4, T5 and T6. Of the 194 recordings analyzed here, only 47 records contained seizures marked by medical practitioners (total seizure duration of 5396 s). Data was obtained from recordings at the Epilepsy Society (UK), Katholieke Universiteit Leuven (Belgium) and Freiburg University Hospital (Germany).

The features have been evaluated using the simple algorithm shown in Fig. 1. The algorithm consists of a first order high pass filter with cut-off frequency of 0.16 Hz, followed by calculation of the feature $F(e)$ within a 2 s non-overlapping epoch (e). Finally a peak detector is applied to the calculated feature to normalize the feature values by restricting the range of values to [0,1], prior to applying a fixed threshold β. If the normalized feature $N(e)$ exceeds β, the current epoch e is considered a seizure and the same epoch across all 16 channels is selected as a seizure event. Otherwise the epoch is discarded as non-seizure data.

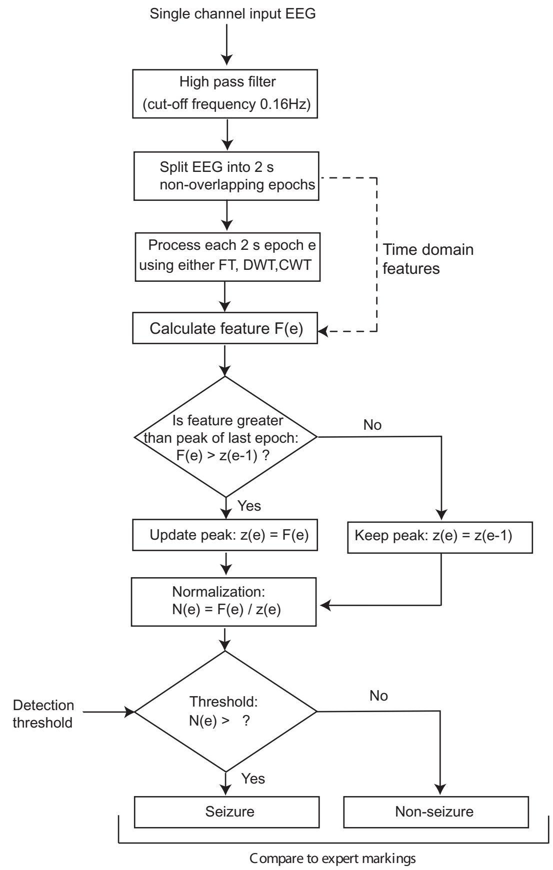

Fig. 1. Flowchart of the seizure detection algorithm.

REFERENCES

[1] J. Gotman, "Automatic detection of seizures and spikes." *J. Clin. Neurophysiol.*, vol. 16, no. 2, pp. 130–140, 1999.

[2] R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless, "Line length: an efficient feature for seizure onset detection," in *IEEE EMBC*, Istanbul, October 2001.

[3] A. J. Casson, D. C. Yates, S. J. Smith, J. S. Duncan, and E. Rodriguez-Villegas, "Wearable electroencephalography," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 3, pp. 44–56, 2010.

[4] A. J. Casson, E. Luna, and E. Rodriguez-Villegas, "Performance metrics for the accurate characterisation of interictal spike detection algorithms," *J. Neurosci. Methods*, vol. 177, no. 2, pp. 479–487, 2009.

[5] L. Logesparan, A. J. Casson, and E. Rodriguez-Villegas, "Improving seizure detection performance reporting: analysing the duration needed for a detection," in *IEEE EMBC*, San Diego, August 2012.

[6] ——, "Assessing the impact of signal normalization: preliminary results on epileptic seizure detection," in *IEEE EMBC*, Boston, September 2011.

[7] A. Shoeb, A. Kharbouch, J. Soegaard, S. Schachter, and J. Guttag, "A machine-learning algorithm for detecting seizure termination in scalp EEG," *Clin. Neurophysiol.*, vol. 22,, no. 1, pp. s36–s43, 2011.

[8] L. Logesparan, A. J. Casson, and E. Rodriguez-Villegas, "Optimal features for online seizure detection," *Med. Biol. Eng. Comput.*, vol. 50, no. 7, pp. 659–669, 2012.

[9] B. R. Greene, S. Faul, W. P. Marnane, G. Lightbody, I. Korotchikova, and G. B. Boylan, "A comparison of quantitative EEG features for neonatal seizure detection," *Clin. Neurophysiol.*, vol. 119, no. 6, pp. 1248–1261, 2008.

[10] L. Kuhlmann, A. Burkitt, M. Cook, K. Fuller, D. Grayden, L. Seiderer, and I. Mareels, "Seizure detection using seizure probability estimation: Comparison of features used to detect seizures," *Ann. Biomed. Eng.*, vol. 37, no. 10, pp. 2129–2145, 2009.

[11] L. Kuhlmann, M. J. Cook, K. Fuller, D. B. Grayden, A. N. Burkitt, and I. M. Y. Mareels, "Correlation analysis of seizure detection features," in *ISSNIP*, Sydney, December 2008.

[12] L. Logesparan, S. A. Imtiaz, A. J. Casson, E. Aguilar-Pelaez, and E. Rodriguez-Villegas, "A 1.8 mW 12 channel wireless seizure detector for miniaturized portable EEG systems," in *uHealthcare*, GyeoungJu, October 2012.