Expert Systems With Applications 163 (2021) 113788

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals

Afshin Shoeibi [a,1], Navid Ghassemi [b,1], Roohallah Alizadehsani [c], Modjtaba Rouhani [b,*], Hossein Hosseini-Nejad [a], Abbas Khosravi [c], Maryam Panahiazar [d], Saeid Nahavandi [c]

[a] Faculty of Electrical and Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran
[b] Computer Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran
[c] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Victoria 3217, Australia
[d] University of California San Francisco, San Francisco, CA, USA

ARTICLE INFO

Keywords:
Epileptic seizures
Electroencephalography (EEG)
Convolutional autoencoder
Feature extraction
Computational complexity

ABSTRACT

Epilepsy, a brain disease generally associated with seizures, has tremendous effects on people's quality of life. Diagnosis of epileptic seizures is commonly performed on electroencephalography (EEG) signals, and by using computer-aided diagnosis systems (CADS), neurologists can diagnose epileptic seizure stages more accurately. In these systems, a mandatory stage is feature extraction, performed by handcrafting features or learning them, ordinarily by a deep neural net. While researches in this field commonly show the value of a group of limited features, yet an accurate comparison between different suggested features is essential. In this article, first, a comparison between the importance of 50 different handcrafted features for seizure detection is presented. Additionally, the computational complexity of features is investigated as well. Then the best features based on Fisher scores are picked to classify signals on a benchmark dataset for evaluation. Additionally, a convolutional autoencoder with five layers is applied to learn features in order to have a complete comparison among feature extraction approaches. Finally, a hybrid method is employed, which combines handcrafted features and encoding of autoencoder to reach high performance in seizure detection in EEG signals.

1. Introduction

The human brain, the center of the neurological system and arguably the most critical body organ, has been the focus of uncountable studies in past centuries (Willis, 1965) to date (Arlotta, 2018), some of which have been concentrating on its disorders (Arlotta, 2018). A recent study (Feigin et al., 2017) has shown that the burden of this type of disease, such as stroke, Alzheimer's, and meningitis has increased substantially in the past 25 years. Among these diseases, epilepsy is a set of neurological disorders associated with seizures (Bajaj & Pachori, 2013; Bhattacharyya & Pachori, 2017; Bhattacharyya et al., 2018; Fisher et al., 2005; Gupta et al., 2017). As reported by the world health organization (WHO), about 65 million people around the world endure epileptic seizures (Fisher et al., 2005).

A seizure is an early and rapid disorder in the brain's electrical activity, usually begins with sudden attacks, affecting some parts or the whole body. This disorder causes loss of consciousness and leads to tremendous physical damage to the patient, such as fractures. Also, these patients may suffer from significant mental pain as a result of embarrassment and discrimination. Epileptic seizures are the result of unusual cortical activity in the brain. They occur in both focal and generalized forms. About 60% of epilepsy patients have focal seizures; in other words, epileptic regions in these patients are a particular part of the brain and are more resistant to drugs (Smelick et al., 2018). Considering all the possible damages, the study of epileptic seizures is fundamental.

Magnetic resonance imaging (MRI) (El Azami et al., 2016), EEG (Li et al., 2017a), magneto-encephalography (MEG) (Khalid et al., 2016), and positron emission tomography (PET) (Tan et al., 2018) are several means used to diagnose epilepsy. Commonly, this diagnosis is performed manually by investigating the EEG signals, a popular method for monitoring brain activity. However, it is very complicated, time-consuming, and may be prone to human error, compelling the need for

* Corresponding author.
E-mail addresses: afshin.shoeibi@gmail.com (A. Shoeibi), navidghassemi@mail.um.ac.ir (N. Ghassemi), ralizadehsani@deakin.edu.au (R. Alizadehsani), rouhani@um.ac.ir (M. Rouhani), hosseini_nezhad@kntu.ac.ir (H. Hosseini-Nejad), abbas.khosravi@deakin.edu.au (A. Khosravi), Maryam.Panahiazar@ucsf.edu (M. Panahiazar), saeid.nahavandi@deakin.edu.au (S. Nahavandi).
[1] Equal contribution.

https://doi.org/10.1016/j.eswa.2020.113788
Received 9 December 2019; Received in revised form 19 July 2020; Accepted 21 July 2020
Available online 30 July 2020

an intelligent diagnostic system with the ability of automatic seizure detection (these systems are also referred to as computer-aided diagnosis system (CADS)).

So far, numerous researches have been conducted for the detection of epileptic seizures from EEG signals. These automated methods usually include two primary stages, feature extraction and classification (Sharaf et al., 2018). Feature extraction is performed to obtain meaningful information from raw EEG signals. Strategies for extracting features can be categorized as automatic feature learning from signals, such as deep neural networks (DNNs), or handcrafting features based on expert knowledge of the specified task. In EEG signals, the second type of features can be divided into the time domain, frequency domain, time–frequency domain, and non-linear ones. In Hassan and Haque (2015), the effectiveness of the statistical features for the diagnosis of epilepsy seizures has been investigated. These features present information on the apparent behavior of EEG signals. Applying frequency features to the same task was done in Faust et al. (2010). Also, Subasi (2007) employed time–frequency features as more appropriate characteristics than the first two groups. Finally, the last group is the non-linear features. Considering the non-stationary and chaotic behavior of EEG signals, these features are expected to improve the overall accuracy (Li et al., 2016, 2017a; Patidar & Panigrahi, 2017; Peker et al., 2015; Reddy & Rao, 2017; Swami et al., 2016).

In this work, we compared non-linear features with features from time, statistical and frequency domains, then selected the best of them using the Fisher algorithm and fed them to two different classifiers; k nearest neighbors (KNN) and support vector machine (SVM) for evaluation on publicly available and widely used Bonn (Andrzejak et al., 2001) benchmark dataset. Then, we used a convolutional autoencoder (CNN-AE) for automatic feature learning and compared the obtained results with the classic methods. This allows us to evaluate how handcrafted features perform overall, considering that CNN-AEs are well-known deep learning based methods used for feature extraction from signals. Finally, a hybrid method using both types of features is presented for achieving the highest accuracy. The block diagram of the proposed CADS is presented in Fig. 1. The study is formed as follows: In Section 2, a brief review of related work is presented. Section 3 discusses the proposed methodology. In Section 4, the evaluation procedure and the results of the experiments are reported. The paper is concluded in Section 5.

2. Related works

Li et al. (2016) presented a double-density discrete wavelet transform (DD-DWT)-based approach with non-linear features. First, the signal was decomposed into 15 sub-bands using the DD-DWT. In the feature extraction stage, the Hurst exponent (HE) and fuzzy entropy (FE) features were extracted from each sub-bands. The analysis of variance (ANOVA) and genetic algorithm based support vector machine (GA-SVM) algorithms were used for feature selection and classification, respectively.

Dual-tree complex wavelet transform (DT-CWT) has been used as an efficient preprocessing for decomposing the EEG signal to different sub-bands (Li et al., 2017a; Peker et al., 2015; Swami et al., 2016). In Li et al. (2017a), fractal dimension (FD) and permutation entropy (PerEn) were extracted as features and SVM was used as the classifier. Peker et al. (2015) gave statistical features of sub-bands to a neural network for detection of seizures. In Swami et al. (2016), the energy, standard division (STD), and Shannon entropy features of the sub-bands were calculated, and the general regression neural network algorithm was applied for classification.

Tunable Q-factor wavelet transform (TQWT) has been applied as another approach to decompose the EEG signal to sub-bands before extracting the features (Al Ghayab et al., 2019; Patidar & Panigrahi, 2017; Reddy & Rao, 2017; Sharaf et al., 2018). Reddy and Rao (2017) calculated the centered correntropy (CCorEn) of sub-bands as features and employed three classifiers, including random forest (RF), multilayer perceptron (MLP) and logistic regression (LR). In Patidar and Panigrahi (2017), the Kraskov entropy (KEn) as a non-linear feature to measure the non-linearity trends of the signal was computed, and the least square support vector machine (LS-SVM) was applied for classification. In Sharaf et al. (2018), statistical, frequency, and non-linear features of sub-bands were calculated, then the firefly and RF algorithms were used for feature selection and classification, respectively. Al Ghayab et al. (2019), after decomposing the EEG signal to sub-bands using TQWT, divided the sub-bands to n-windows and calculated the statistical features on each window. Then the extracted features were given to bagging tree (BT), KNN, and SVM classifiers to detect the seizures.

Zhang et al. (Tao et al., 2016) used frequency slice wavelet transform (FSWT) based on the approximate entropy (ApEn) and fluctuation index (FI) method to diagnose epilepsy. In the classification stage, 50% of the data were selected for training and 50% for testing. The classification was performed using a two-class GA-SVM. A maximum accuracy of 99% in two-class cases and 98.33% accuracy in the three-class case were reported.

Daubechies discrete wavelet transform (DDWT) has been employed as another preprocessing to decompose the EEG signal to sub-bands (Kocadagli & Langari, 2017; Li et al., 2017b). In Li et al. (2017b), after analysis by wavelet, the Hilbert–Huang transform (HHT) of each sub-band and the envelope spectrum of each HHT were calculated. Then, from each envelope spectrum, statistical features including mean, maximum, minimum, and STD were extracted. Finally, a neural network ensemble was used for classification. Kocadagli and Langari (Kocadagli & Langari, 2017) applied Daubechies mother wavelet then extracted the statistical features from each sub-band. Due to the high number of features, they used fuzzy relations (FRs) in the feature selection stage. In the final step, classification was performed by a hybrid artificial neural network (ANN).

Wang et al. (2017) diagnosed epilepsy based on the combination of multi-domain feature extraction and nonlinear analysis of EEG signals. To remove the artifacts, using the wavelet threshold method, they pre-processed EEG signals. Then, features in the time, frequency, time–frequency domain, and nonlinear analysis features were extracted. Features are then extracted in five frequency sub-bands. They reduced feature dimensions using principal component analysis (PCA). Finally, the proposed method achieves 99.25% accuracy on the Bonn dataset.

In Zhang et al. (2018a), the wavelet packet decomposition (WPD) method was applied to decompose the EEG signal to sub-bands, then the fuzzy distribution entropy (FDE) algorithm was used for feature extraction. Finally, Kruskal–Wallis and KNN were employed for feature selection and classification, respectively.

Zhang et al. (2019) introduced the maximal overlap discrete wavelet package transform (MODWPT) to characterize the EEG signal and extracted multi-basis MODWPT-based features. They employed six dimensionality reduction algorithms, namely, PCA, independent component analysis (ICA), kernel PCA, isometric feature mapping (ISOMAP), locally linear embedding (LLEm), and Laplacian eigenmaps (LEM) to map the extracted features into other domains. Finally, the mapped features were applied to an LS-SVM for classification (Zhang et al., 2019).

In Sharma et al. (2017), the analytic time–frequency flexible wavelet transform (ATFFWT) with 16 sub-band decomposition level was used for preprocessing, and fractal dimension (FD) for feature extraction. Then, t-test and LS-SVM algorithms were applied for feature selection and classification, respectively.

Zhang and Chen (2016) used local mean decomposition (LMD) algorithm, which is a time–frequency analytical procedure to decompose the raw EEG signal into a set of product functions (PFs). Then they calculated the time statistical and non-linear features of the first five PFs and fed the features of each PF into five classifiers, containing

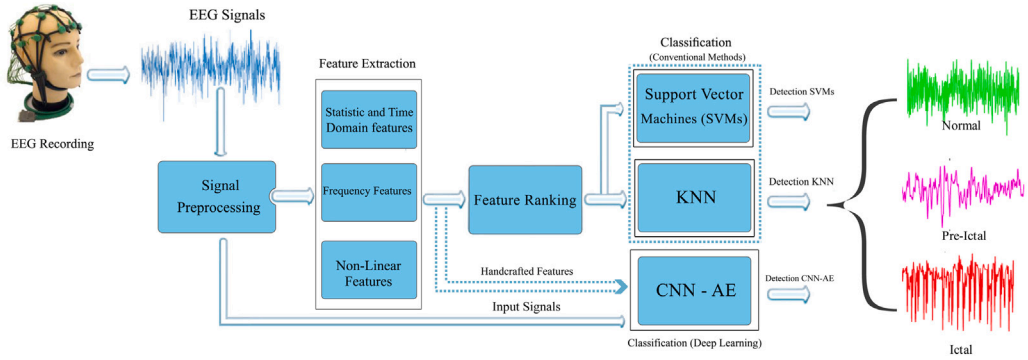


Fig. 1. Proposed method overview.

backpropagation neural network (BPNN), KNN, linear discriminant analysis (LDA), un-optimized SVM, and GA-SVM.

In Jaiswal and Banka (2017), two feature derivation methods based on local pattern alteration containing local neighbor descriptive pattern (LNDP) and one-dimensional local gradient pattern (1D-LGP) were introduced. To classify the epileptic seizures and non-seizure signals, four classifiers including KNN, SVM, ANN, and decision tree (DT) were employed. Zhang et al. (2018b) used a generalized stock-well transform (GST) and singular value decomposition (SVD) for feature extraction. Afterward, they fed the extracted features to random forest (RF) classifier.

To find out whether EEG signals are in inter-ictal or ictal activities, Chen et al. (2019) proposed an automatic classification method using nonlinear dynamics features and nonlinear classifiers. They first applied the discrete wavelet transform (DWT), obtaining de-noised and sub-band signals. Then, based on the entropy theory, they extracted nonlinear dynamic features. After selecting the best features for seizure detection, they were fed to different classifiers for diagnosis of seizures. The experimental results on the Bonn dataset show that the LS-SVM has the accuracy, sensitivity, and specificity of 99.50%, 100%, and 99.40%, respectively.

Moreover, some works have also applied autoencoder based architectures as well. An stacked sparse autoencoder (SSAE) architecture to classify epileptic EEG signals was proposed by Lin et al. (2016). In this study, they proposed a new framework for automatic identification of epilepsy using a 5-layer SSAE with softmax classifier. In this framework, the features are first extracted by SSAE after pre-processing and then sent to the softmax classifier for training and classification. Results with an average accuracy of 96% show the efficiency of their proposed method.

In another research, Sharathappriyaa and colleagues (Sharathappriyaa et al., 2018) proposed the design of an encoding-based automated system for the diagnosis of epilepsy using an EEG signal. Harmonic wavelet packet transform (HWPT) and FDs were used to extract the initial feature vector. The features containing HWPT and FDs are then fed to an autoencoder (AE) for efficient classification. Finally, softmax classification was used for binary classification problems of epilepsy signals and achieved 98.67% accuracy. Lastly, Wen et al. (Wen & Zhang, 2018) used a new CNN-AE architecture to extract features from EEG signals in epileptic patients. After extracting the features, different algorithms were used in the classification step and they achieved 92% accuracy.

Nishad and Pachori (2020) proposed a methodology for classifying normal, non-seizure, and seizure EEG signals involving three steps. The first step contains EEG signals decomposition by the TQWT based filter bank. In the second step, cross-information potential (CIP) is employed to attain the features of each subset, and the RELIEF algorithm is utilized to select the features. The last step is associated with classifying the extracted features using the RF classifier. The TQWT approach has been exploited in other similar investigations to diagnose epileptic seizures (Aydemir et al., 2020; Bhattacharyya et al., 2017; Ghassemi et al., 2019; Sharma & Pachori, 2017a).

In another research, Gupta and Pachori (2019) used the Fourier–Bessel series expansion (FBSE) procedure to decompose EEG signals into numerous sub-bands and subsequently applied weighted multiscale Renyi permutation entropy (WMRPE) to extract the feature. Finally, they employed three RF, LS-SVM, and regression algorithms for classification.

Sharma and Pachori (2017b) have introduced a novel time-frequency representation (TFR) based on the improved eigenvalue decomposition of Hankel matrix and Hilbert transform (IFDHM-HT). Afterward, Hilbert transform (HT) is applied to components to obtain non-stationary TFR signals. Extracted features from acquired TFR EEG signals using the provided technique entails energy concentration measure (ECM), REM, FD, and focus measure operators (FMOs). These features were then ranked using the student's t-test procedure and fed to an LS-SVM classifier with RBF kernel as input.

In another paper, Sharma et al. (2018) decomposed EEG epochs using iterative filtering (IF) to various intrinsic mode functions (IMFs). The amplitude envelope (AmE) function was then computed applying the discrete separation energy algorithm (DESA) algorithm from these modes. The feature set comprises k-nearest neighbor entropy estimator (KNNE), log energy entropy, Shannon entropy, and Poincare plot parameters extracted from AmE functions from IMFs. The discriminating capability of a set of features was evaluated using the Kruskal–Wallis statistical test, and eventually, the efficacious classification features were applied to an RF classifier with ten-fold cross-validation.

Tiwari et al. (2016) recommended a methodology based on the detection of the key points in the EEG signals in order to diagnose epileptic seizures. The feature extraction technique in the proposed procedure involves calculating the local binary patterns (LBP) at any identified key point in the signal. In particular, LBPs are calculated at each key point as well as the corresponding points in the original signal. The LBP histogram is considered as the feature set and is used for SVM classification.

Some research has applied empirical mode decomposition (EMD) techniques to decompose EEG signals, including various IMFs. Sharma and Pachori (2015) first used EMD to decompose signals and subsequently extracted various features from each IMFs. The Kruskal–Wallis statistical test is taken advantage of to assess the discrimination capability of these features. Finally, the LS-SVM was employed as a classifier. Also, in further resembling studies, EMD has been utilized to diagnose epileptic seizures (Bajaj & Pachori, 2011; Pachori, 2009; Pachori & Patidar, 2014).

The proposed Kumar et al. (2015) methodology uses four Gabor filter banks to decompose EEG signals. Then 1D-LBP is computed for each segment. Finally, the 1D-LBPs histograms of the discrete segments of a channel are combined to generate the final exhibition. K-NN classifier with k equal to one is exploited to classify EEG signals. This classifier delineates the correspondence between the two EEG signals

Table 1
Related works.

Work	Preprocessing	Feature Extraction	Feature Selection	Classifier
Swami et al. (2016)	DT-CWT	Energy, STD, Shannon Entropy	CFS, MRMR, T-Test	GRNN
Zhang and Chen (2016)	LMD	Different Features	No	GA-SVM
Tao et al. (2016)	FSWT	APEN, FI	No	GA-SVM
Peker et al. (2015)	DT-CWT	Statistical Features	No	CVANNs
Li et al. (2016)	DD-DWT	HE, FE	ANNOVA	GA-SVM
Li et al. (2017a)	DT-CWT	HE, FD, PEN	No	SVM
Reddy and Rao (2017)	TQWT	CCorEn	No	RF, MLP, LR
Kocadagli and Langari (2017)	DWT	Statistical Features	FRs	GA-ANN
Li et al. (2017b)	DWT, HHT, EnEsp	Mean, Max, Energy, STD, Energy	No	NNE
Patidar and Panigrahi (2017)	TQWT	KEn	Kruskal–Wallis	LS-SVM
Sharma et al. (2017)	Band-Pass Filter, ATFFWT	HFD	T-Test	LS-SVM
Jaiswal and Banka (2017)	No	LNDP, 1D-LGP, 1D-LBP	Histogram	ANN
Wang et al. (2017)	DWT	Different Features	ANNOVA	SVM
Zhang et al. (2018a)	WPD	Fuzzy Distribution Entropy	Kruskal–Wallis	KNN
Sharaf et al. (2018)	TQWT	Chaotic, Statistical, PSD, TCC, Energy, Homogeneity	Firefly	RF
Zhang et al. (2019)	MODWPT	Statistical Features	PCA, ICA, KPCA ISOMAP, LLE, LE	LS-SVM with RBF kernel
Zhang et al. (2018b)	GST	SVD	No	RF
Al Ghayab et al. (2019)	TQWT	Statistical Features	No	KNN
Chen et al. (2019)	DWT	Different Features	ANOVA-FSFS	LS-SVM
Lin et al. (2016)	No	SSAE	SSAE	Softmax
Sharathappriyaa et al. (2018)	HWPT	FDs Features	AE	Softmax
Wen and Zhang (2018)	No	CNN-AE	CNN-AE	Various
Nishad and Pachori (2020)	TQWT	CIP of sub-band Signals	RELIEFF	RF
Sharma and Pachori (2017a)	Filtering, TQWT	FD	Student's T-Test, Kruskal–Wallis, SFS	LS-SVM
Bhattacharyya et al. (2017)	Filtering, TQWT	Multi-Scale KNN Entropy (MKNNEn)	Wrapper	SVM
Ghassemi et al. (2019)	Filtering, TQWT	Different Features	No	Ensemble Learning
Aydemir et al. (2020)	TQWT	Quadruple Symmetric Pattern(QPS)	Neighborhood Component Analysis (NCA)	KNN
Tzimourta et al. (2019)	DWT	Linear & Non-Linear Features	NA	RF
Amin et al. (2020)	Filtering, DWT	Arithmetic Coding (AC)	ANOVA	Different Classifiers
Zhang et al. (2020)	FAWT	CVDistEnt, Logarithmic Energy (LE)	NA	FKNN
Bhati et al. (2017a)	time–frequency Localized Three-Band Synthesis Wavelet Filter Bank (TFLTBSW)	Sub-Bands Norm	NA	MLPNN
Bhati et al. (2017b)	time–frequency Localized Three-Band Analysis Filter Bank (TFLTBAFB)	Different Features	Kruskal–Wallis	MLPNN
Chandel et al. (2019)	Triadic Wavelet Decomposition (TWD)	Statistical Features	NA	KNN
de la O Serna et al. (2020)	5 Band-Pass Filters (BPFs)	Taylor–Fourier EEG-Band Energy (TFEBE)	NA	LS-SVM
Kumar et al. (2015)	EEG Decomposition Using a Bank of Four Parallel Gabor Filters	Histograms of 1D-LBPs	Fusion Strategies for Combination of Information at Feature, Matching Score, and Decision Levels	KNN

(*continued on next page*)

based on the comparison (matching) of the features extracted from the signals. During matching, the histogram features of the query of EEG signal are compared with each training signal.

In the Joshi et al. (2014) study, EEG signals are first passed through a fractional linear prediction (FLP) filter. The filter then calculates the signal FLP coefficients using the least-squares analysis and utilizes them to model the signal. The discrepancy between the actual signal and

Table 1 (*continued*).

Work	Preprocessing	Feature Extraction	Feature Selection	Classifier
Tiwari et al. (2016)	Detection of Keypoints Using Pyramid of Difference of Gaussian (DoG) Filtered Signals	Histograms of LBP	NA	SVM
Rout and Biswal (2020)	Variational Mode Decomposition (VMD), HT	Band-Limited Intrinsic Mode Functions(BLIMFs)		Error-Minimized Random Vector Functional Link Network (EMRVFLN)
Sharma and Pachori (2017b)	TFR Based on IEVDHM–HT	ECM, FD, FMO, REM	Student's T-Test	LS-SVM
Ramos-Aguilar et al. (2020)	STFT	Different Features	No	MLPNN
Sharma et al. (2020)	Filtering Third-Order Cumulant (TOC)	Sparse Autoencoder (SAE)		Softmax
Joshi et al. (2014)	FLP Coefficients Estimation	Prediction Error Energy (PEE), Signal Energy	No	SVM
Gupta and Pachori (2019)	FBSE	WMRPE	Bhattacharyya Space Algorithm (BSA), Entropy, ROC, Student's T-Test, Wilcoxon	RF, LS-SVM, Regression
Sharma et al. (2018)	Decomposition into IMFs Using IF, Extraction of AmE Function from IMFs Using DESA Algorithm	SE, LEE, KNNE, Poincar'e Plot Features	Kruskal–Wallis	RF
Sharma and Pachori (2015)	EMD, 2D and 3D PSRs of IMFs	95% Confidence Area Measure of 2D PSR of IMFs, IQR of Euclidean Distances of 3D PSR of IMFs	Kruskal–Wallis	LS-SVM
Pachori and Patidar (2014)	EMD, SODP	95% Confidence Ellipse Area	Kruskal–Wallis	MLPNN
Bajaj and Pachori (2011)	EMD	Hilbert Transformation of IMF's	Kruskal–Wallis	LS-SVM
Pachori (2009)	EMD, Fourier–Bessel Expansion	Mean Frequency (MF)	Kruskal–Wallis	NA
Raghu et al. (2019)	No	Matrix Determinant	NA	MLP
Jiang et al. (2020)	No	Symplectic Geometry Decomposition (SGD)	No	SVM
Tuncer et al. (2019)	No	Local Senary Pattern (LSP)	NCA	SVM
Akyol (2020)	Normalization	Stacked Ensemble (SEA) based DNN Model		Sigmoid
Gao et al. (2020)	No	ApEn, RQA	Shapiro–Wilk Test, Student's T-Test or The Mann–Whitney U Test	CNN

the modeled signal is identified as the prediction error. After modeling the signal, the energies of predictive error and the actual signal are computed. Finally, these two energies are applied as a feature to the SVM classifier. More details for related work have been demonstrated in Table 1.

3. Material and methods

3.1. Epilepsy dataset

The proposed algorithm was evaluated on publicly available Bonn dataset (Andrzejak et al., 2001), which is commonly used as a benchmark in many epileptic seizure diagnosis studies. This dataset contains five subsets defined as (A–E); each subset contains 100 single-channel EEG segments with the time span of 23.6 s, and each segment contains 4097 signal samples. The sampling rate and resolution are 173.61 Hz and 12 bits, respectively. Subsets A and B are normal signals with open and closed eyes, subsets C and D indicated as intermittent data are related to the preictal phases, and subset E, as ictal data contains epileptic seizures. A detailed description of five subsets of the dataset is presented in Table 2. In this work, five two-class cases and one three-class case, altogether six cases are used. More details about these six cases are shown in Table 3. Samples of EEG signals of the dataset for each class are shown in Fig. 2.

3.2. Feature extraction

As mentioned before, in this work, three types of features are extracted. These features are time domain and statistical, frequency, and non-linear features. This subsection is devoted to the explanation of the details on them.

3.2.1. Time domain and statistical features

For statistical features, mean, variance, standard division, skewness, kurtosis, max, min, mode, and median are calculated (Al Ghayab et al., 2019). Complexity and mobility in Hjorth parameters (Oh et al., 2014), average absolute signal slope (AASS) (Kalatzis et al., 2004), peak-to-peak (PTP) (Kalatzis et al., 2004), time window of peak-to-peak (TWOPP) (Kalatzis et al., 2004), peak-to-peak slope (PPS) (Kalatzis et al., 2004), and slope sign alterations (SSA) (Kalatzis et al., 2004) are also computed as time-domain features. Table 4 presents more details about them.

3.2.2. Frequency features

During the epileptic seizures, the frequency characteristics of the EEG signal are changed; so the frequency features are appropriate candidates for seizure detection. In this work, the power spectrum of frequency sub-bands, including Delta, Theta, Alpha, Beta, and Gamma

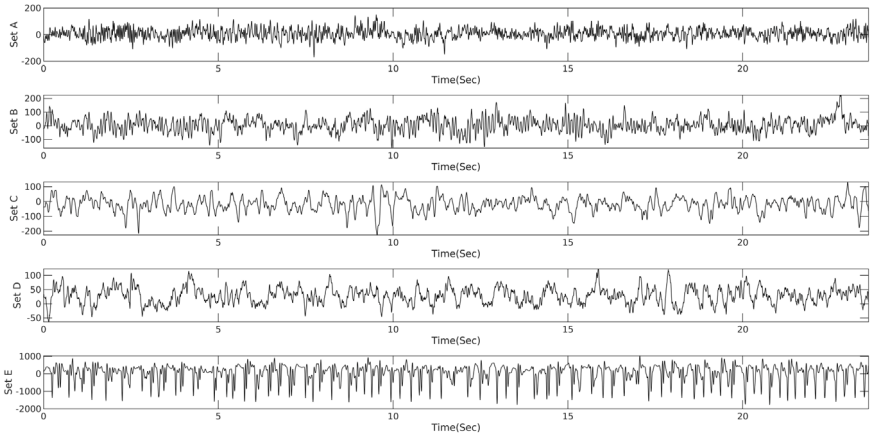


Fig. 2. An example from each set of Bonn dataset.

Table 2
Explanation of five subsets of dataset.

Subjects	Set A	Set B	Set C	Set D	Set E
	Healthy cases	Healthy cases	Epileptic patients	Epileptic patients	Epileptic patients
Patient Stage	Eye open	Eye close	Seizure free	Seizure free	Seizure activity
Electrode Type	Surface	Surface	Intracranial	Intracranial	Intracranial
Electrode Placement	International 10–20	International 10–20	Opposite to epileptogenic zone	Opposite to epileptogenic zone	In epileptogenic zone
Number of cases	5	5	5	5	5
Number of samples	100	100	100	100	100
Length of segment	4097	4097	4097	4097	4097
Sampling Freq. (Hz)	173.61	173.61	173.61	173.61	173.61

Table 3
Details about six cases.

Case	sets	Description	Number of samples
1	A — E	Healthy (Eye open) — Ictal	200
2	B — E	Healthy (Eye close) — Ictal	200
3	C — E	Inter-ictal — Ictal	200
4	D — E	Inter-ictal — Ictal	200
5	ABCD — E	Non-seizure — Seizure	500
6	AB — CD — E	Healthy — Inter-ictal — Ictal	500

(Park et al., 2011) has been used as frequency features. Additionally, three dominant frequencies (Goel et al., 1996) and their domains (six features in total) are calculated as well.

3.2.3. Non-linear features

Because of non-stationarity and non-linearity properties of EEG signals, the non-linear features present a high potential for the detection of epileptic seizures. In this work, non-linear features including large Lyapunov exponent (LLE), FDs and entropy-based features are used.

Largest Lyapunov Exponent

Exponents from Lyapunov (Shayegh et al., 2014) measure the exponential divergence of originally near trajectories of state–space and calculate how much chaos there is in a scheme. To calculate the Lyapunov exponent, each point set of the phase space (s_{n_0}) must be evaluated. This calculates the rate of divergence in the path.

$$\delta(\Delta n) = \frac{1}{N}\sum_{n_0=1}^{N} \ln\left(\frac{1}{|U(S_{n_0})|}\sum_{S_{n_0}\in U(S_{n_0})} |S_{n_0+\Delta n} - S_{n_0}|\right) \quad (1)$$

where

- S_{n_0} is the simulated stage space point at time n_0
- $S_{n_0+\Delta n}$ is the simulated stage space point at time $n_0 + \Delta n$
- $U(S_{n_0})$ is the stage space neighboring point set of S_{n_0}
- and $|U(S_{n_0})|$ is the number of $U(S_{n_0})$ participants.

Table 4
Time domain features.

Type	Name	Theory
Time domain	Mobility	$Mobility = \sqrt{\frac{X_{var}(y(t)\frac{d_y(t)}{dt})}{X_{var}(y(t))}}$
	Complexity	$Complexity = \frac{Mobility(y(t)\frac{d_y(t)}{dt})}{Mobility(y(t))}$
	AASS	$AASS = \frac{1}{N}\sum_{i=0}^{N-\tau}\frac{1}{\tau}\lvert x_{i+\tau} - x_i\rvert$
	PTP	$PTP = X_{Max} - X_{Min}$
	TWOPP	$TWOPP = t_{X_{Max}} - t_{X_{Min}}$
	PPS	$PPS = \frac{PTP}{TWOPP}$
	SSA	$SSA = \sum_{i=\tau}^{N-\tau}\frac{1}{2}\lvert\frac{x_{i-\tau}-x_i}{\lvert x_{i-\tau}-x_i\rvert} + \frac{x_{i+\tau}-x_i}{\lvert x_{i+\tau}-x_i\rvert}\rvert$

After computation of $\delta(\Delta n)$, Lyapunov exponents are calculated as the exponents by which they expend on each dimension. The most

Table 5
Detailed description of fractional dimension methods.

Fractal dimensions	Theory and Description	
	Mathematics	Description
Higuchi Fractal Dimension (HFD)	$X_k^m = \{x(m)x(m+k)x(m+2k)\dots x(m+[\frac{N-m}{k}]k)\}$ $L_m(k) = \frac{1}{k}(\sum_{i=1}^{\frac{(N-m)}{k}} \lvert X(m+ik) - X(m+(i-1)k)\rvert)P$ $P = \frac{N-1}{[\frac{(N-m)}{k}]k}$ $HFD = -\frac{\log(L(k))}{\log(k)}$	In HFD, firstly, X_k^m is sampled from the signal, then by employing the illustrated equations $L_m(k)$ is computed and used for HFD calculation.
Petrosian Fractal Dimension (PFD)	$PFD = -\frac{\log_{10} n}{\log_{10} n + \log_{10}(\frac{n}{n+0.4N\Delta})}$	In PFD, n is the longitude of the data, and $N\Delta$ is the number of dissimilar pairs in the binary data generated.
Katz Fractal	$KFD = -\frac{\log_{10} n}{\log_{10}(\frac{d}{L} + \log_{10} n)}$	Here, L shows the summation of the intervals between successive points, d denotes the distance between the initial point and the furthest point of the sequence.

important of them is the largest one, i.e., the largest Lyapunov exponent, as it determines the overall exponential growth (or decay).

Fractal Dimensions

Fractal Dimensions (FDs) are important tools for complexity measurement in biological data such as EEG signals. In this work, HE (Li et al., 2016), Higuchi (Esteller et al., 2001), Petrosian (Esteller et al., 2001) and Katz (Esteller et al., 2001) methods are utilized to calculate the fractal dimension.

HE (Li et al., 2016) determines the correlation of points in a period. The value of HE of less than 0.5 shows that the timelines are long term anti-correlation and anti-static, while a HE value greater than 0.5 is indicating that the timelines are long term correlated. As a result, HE is widely used to determine the presence or absence of long term correlation. As HE can classify timelines according to the prediction of noise levels, it might be a valuable tool in recognizing the deviation from the normal pattern of brain activity during a seizure. Typically, HE can be estimated using Eq. (2):

$$HE = \frac{\log(\frac{R}{S})}{\log(T)} \tag{2}$$

In this equation, R is the variation between the minimum and maximum deviations associated with the device and S appears for the standard deviation of the timelines. T is the length of the sample data. More details about other fractal algorithms are presented in Table 5.

3.2.4. Entropy based features

In biomedical signal processing, entropies are non-linear parameters that reflect the degree of chaos within the signals. In this work, 19 different entropy-based features are extracted from the EEG signals. Table 6 goes through details about these entropy features.

3.3. Computational complexity of features

Before any further analysis of introduced features, the computational complexity of the calculation of each feature is presented first. Here, the big O notation (O) (Cormen et al., 2009) is used for this matter. All of the computational complexities are demonstrated in Table 7.

A few points should be noted here, first, for counting various patterns of different length we have considered a map data structure (Halim et al., 2013). This requires $O(\log(n) * k)$ for each operation where n is the total number of patterns, and k is the length of patterns. Secondly, the computational complexity for the calculation of base functions such as log and exp depends dramatically on the underlying framework and compiler. For example, calculating based on Taylor series with improvements is done in $O(M * (\log(n))^2)$ as shown in Chudnovsky and Chudnovsky (2004), (M is the complexity of used multiplication algorithm); whereas using Arithmetic–geometric mean iteration this can be reduced to $O(M * \log(n))$ (Brent, 1976). Additionally, as mentioned earlier, M is deeply dependent on the underlying framework; in new CPUs, the floating-point unit usually has separate multiplication and addition units (Eijkhout, 2013), which pushes M closer to a constant. Since nearly all of the variables are double, we consider M as constant in our calculations. To make the equations readable and not dependable on various factors, we use the assumption that the underlying framework is capable of performing a calculation of log and exp functions in optimal computational complexity; thus, we use $O(\log(n))$ for them.

Lastly, it is essential to point out that all these calculations are based on the typical implementation of algorithms. In the case of different implementation, parallelization, or creation of application-specific hardware, these calculations may not be valid due to various limits such as memory or more optimized complicated algorithms; additionally, bad implementations may lead to substantial hidden constants.

3.4. Feature selection

Feature selection is an attempt to overcome the curse of dimensionality (Bishop, 2006), ergo to improve classification accuracy. In this work, we applied the Fisher feature scoring algorithm (Duda et al., 2012); which aims to find features that make data points of the same classes closer and others further simultaneously. Regarding complexities of solving this task, by considering features relevance (Gu et al., 2012), each feature is scored independently. Consider there are D features x^j, $j = 1, \dots, D$ and K classes. Define u^j and σ^j as the mean and standard deviation of the jth feature over all data and u_k^j and σ_k^j as the mean and standard deviation of the jth feature over data in class k. The Fisher score for jth feature is then calculated using following equation:

$$F^j = \frac{\sum_{k=1}^{K} n_k(u_k^j - u^j)^2}{\sum_{k=1}^{K} n_k(\sigma_k^j)^2} \tag{3}$$

where n_k is the size of kth class (Gu et al., 2012).

A higher Fisher score indicates that the feature is more distinguishing (more informative) between different classes. Fig. 3 presents the results of applying this algorithm on extracted features. As explained earlier in Section 3.1, six different classification cases exist, and the Fisher score of each feature is calculated in all of them. In this figure, increment in feature score leads to warmer colors, i.e., red. Also, considering that each four of the classification cases are between binary between two sets of data, the importance of each feature in distinguishing that class from seizures can be deducted.

3.5. Classification

After selecting the best features, the classification process should be performed to make a decision regarding the seizures based on the extracted features. In this work, we applied SVM (Hearst et al., 1998), KNN (Cunningham & Delany, 2007), and CNN-AE (Schmidhuber, 2015) for classification to compare their efficiency and obtain the

Table 6
More details of Entropy features used in this work.

Entropy	Formula	Description
Shannon (Acharya et al., 2018)	$En_{Shan} = -\sum_{n=1}^{x} P_n \log_2 P_n$	P_n is the chance of feature value incidence.
Log Energy (Acharya et al., 2018)	$En_{Log-En} = \sum_{n=1}^{x} \log(S_n^2)$	x and S_n define the longitude of the EEG signal and n_{th} pattern of the EEG signal, respectively.
Threshold (Acharya et al., 2018)	$En_{Thr} = 1$ if $\|S_n\| > r$ and 0 elsewhere so $En_{Thr} = \{ i$ such that $\|S_n\| > r \}$ is the number of time instants when the signal exceeds the r threshold.	
SURE (Acharya et al., 2018)	$En_{su} = X - [nsuchthat\|S_n\| <= m] + \sum_n min(S_n^2, m^2)$	Here, m represents a threshold which always holds a positive value, set to 3 in this work.
Norm (Acharya et al., 2018)	$En_{Norm} = \sum_n \|S_n\|^h = S_h^h, 1 \le h$	h is the power, fixed to 1.1
Approximate (Acharya et al., 2018)	$En_{App} = [\frac{l_{ave}(s)}{l_{ave+1}(s)}]$	$l_{ave}(s)$ defines the mean length of the pattern l and $l_{ave+1}(s)$ represents the mean length of the pattern $m+1$.
Spectral (Tian et al., 2017)	Similar to Shannon, however this entropy works in frequency space and it is normalized	
Permutation (Acharya et al., 2018)	$En_{Per} = -\sum_{n=1}^{x!} S_n \log_2(S_n)$	With $x > 2$, this summation goes through all $x!$ possible arrangements of the selected value x.
Tsallis (Acharya et al., 2018)	$En_{ts} = \frac{1}{\alpha-1}(1 - \sum_{n=1}^{x} P_n^{\alpha}), \alpha \neq 1$	En_{Ts} is a generalized form of En_{Shan} and can be simplified to En_{Shan} when $a = 1$.
Wavelet (Acharya et al., 2018)	$En_{Wa} = \sum_{n<0} S_n \ln(S_n)$	Here, S_n means the chance of distribution of the EEG signal and n symbolizes different resolution levels.
Conditional (Chen et al., 2019)	$En_{Co}(\frac{m}{m-1}) = -\sum_{m-1} p_{m-1} \sum_{m\|(m-1)} p_{m\|(m-1)} \log p_{m\|(m-1)}$	p_{m-1} represents the common probability of the pattern $x_{m-1}(i)$ and $p_{m\|(m-1)}$ represents the conditional probability of the m_{th} sample of the pattern $x_m(i)$ with respect to previous $(m-1)$ cases.
Corrected Conditional (Chen et al., 2019)	$En_{CCo} = En_{Co}(\frac{m}{m-1}) + prec(m)(-\sum_1 p_1 \log p_1)$	$prec(m)$ represents the abundance of singular points in the m-dimensional phase space and p_1 is same as in En_{Co}
Kozachenko (Acharya et al., 2018)	$En_{kl} = \frac{D}{2Ns} \sum_{n=1}^{Ns} \log(\|\|f(n) - f_{NN}(n)\|\|) + K$	The Kozachenko–Leonenko scheme is based on the entropy calculation of the Ns specimens with dispersal of $L2$ range between each sample $(f(n))$ and its closest neighbor $f_{NN}(n)$ among the $NS-1$ available specimens.
Renyi (Acharya et al., 2018)	$En_{rny} = \frac{1}{\alpha-1} \log(\sum_{n=1}^{x} P_n^{\alpha})$, for $\alpha \neq 1$	Another generalized form of En_{Shan}. Here, we choose $\alpha = 2$.
Kraskov (Patidar & Panigrahi, 2017)	$En_{kr} = \phi(n) - \phi(k) + \log(C_d) + \frac{d}{n} \sum_{i=1}^{n} \log(\xi_i^k)$	$\phi(n)$ denotes the digamma function, C_d expresses the volume of the d dimensional unit ball and ξ_i^k presents the distance between x_i and its neighbors in k-NN.
Recurrence (Acharya et al., 2018)	$En_{rec} = -\sum_{n=k_{min}}^{K} RE(n) \ln[RE(n)]$	K is the samples numbers, $k_m in$ denotes the minimum diagonal line length, and $RE(n)$ represents the frequency distribution of the length n of the diagonal lines.
Fuzzy (Acharya et al., 2018)	$\phi^l(F,g) = \frac{1}{S-1} \sum_{x=1}^{S-1} \frac{1}{S-l-1} [\sum_{y=1,y\neq x}^{S-l} (S_{xy}^l)]$ $En_{fuz} = \ln[\phi^l(F,g)] - \ln[\phi^{l+1}(F,g)]$	First, length l sequences are extracted from the EEG signal. Then, the degree of similarity $(S_{x,y}^l)$ between the two sequences (x_{th} and y_{th}) is calculated using the fuzzy function $S_{x,y}^l = u(x,y,F,g)$ whereby u, F, g determine the fuzzy function, gradient, and threshold width of the fuzzy similarity, respectively. After that, En_{fuz} is determined and is finally calculated.
Bubble (Manis et al., 2017)	$En_{Bu} = \frac{(H_{swaps}^{n+1} - H_{swaps}^{n})}{\log(\frac{n+1}{n-1})}$	H_{swaps}^n is calculated with En_{rny} from the distribution of the number of exchanges needed to sort each vector
Multi-Scale Fuzzy & Sample (Azami et al., 2017)	Calculated based on (Azami et al., 2017), for each of these entropies two sets of features based on Mean and Variance are extracted. (En_{MFu}, En_{VMFu} for multi-scale fuzzy and En_{MSa}, En_{VMSa} for multi-scale Sample)	

most appropriate one in order to maintain the best performance for the seizure detection system. The following contains details on selected parameters for each classification algorithm; for the parameters that are not mentioned, the suggested values of the applied library are used.

3.5.1. Support vector machines (SVM)

Support vector machines (Hearst et al., 1998) are used for classification and regression and their power has been demonstrated over the last decades. SVM classifier tends to find the hyperplane which distinguishes data points with the largest margin. To achieve this task, SVM uses pairwise dot product of data points; hence to do classification task in another space, only the dot product operation in that space (also known as the kernel) is needed. Linear (Hearst et al., 1998), RBF (Chung et al., 2003), Polynomial (Boser et al., 1992) and Histogram Intersection (Barla et al., 2003) are kernels we used in this work. Equations for these kernels are shown in Table 8.

3.5.2. K nearest neighbor (KNN)

This classifier does not have a training phase. In test phase, each data point is compared to all others, and voting is performed between k nearest neighbors to label data. Test time grows linearly with training dataset size, causes this method to be inefficient on large datasets. In this work, we applied KNN with k equal to 3 and 5.

3.5.3. Convolutional autoencoder (CNN-AE)

AEs have been around for more than three decades, with written appearances as early as 1987 (Schmidhuber, 2015); however they are still applied as state-of-the-art deep representation learning architectures (Faust et al., 2018). The base idea in AEs is creating a decoder and an encoder to transfer datapoint into a latent space and back. While loss functions of AEs aim to minimize the difference between decoder output and encoder input, yet the main goal here is to learn a representation in the latent space rather than a flawless copy of input

Table 7
Computational complexity of used features.

Features	Computational complexity	Additional details
Statistical & time domain	Mostly $O(n)$, and $O(n\log(n))$ for Mode	Calculation of Median can be performed with randomized partition in $O(n)$ as implemented in $n$th_element function of cpp's algorithm library. (Halim et al., 2013)
Frequency features	The dominant part is usually calculation of FFT which have $O(n\log(n))$ complexity (Cormen et al., 2009).	
LLE	$O(n^2)$	–
HE	$O(n\log(n))$	–
HFD	$O(kn)$	Calculating a single score from the feature vector may require further computations.
PFD	$O(n)$	–
KFD	$O(Q(m)n)$	$Q(m)$ is complexity of the chosen sqrt algorithm
En_{Shan}	$O(n\log(n))$	—
En_{Log-En}	$O(n\log(n))$	–
En_{Thr}	$O(n)$	–
En_{su}	$O(n)$	–
En_{Norm}	$O(P(h)n)$	$P(h)$ is complexity of the chosen power algorithm
En_{App}	$O(n(\log(n))^l)$	Here l is length of patterns.
En_{Spec}	similar to En_{Sha}	
En_{Pre}	$O(nx!x\log(x))$	for each of $x!$ possible arrangement of each of $(n-x)$ pattern of length x, entropy is calculated.
En_{ts}	$O(n\log(n)P(\alpha))$	–
En_{Wa}	$O(n\log(n)l)$	l is maximum level of decomposition.
En_{Co}	$O(n(\log(n))^{2l})$	Here l is length of patterns.
En_{CCo}	$O(n(\log(n))^{2l})$	Here l is length of patterns.
En_{kl}	$O(n^2\log(k))$	k shows chosen number for nearest neighbors.
En_{rny}	$O(n\log(n))$	–
En_{kr}	$O(n^2\log(k))$	k shows chosen number for nearest neighbors.
En_{rec}	$O(n(\log(n))^l)$	Here l is length of patterns.
En_{fuz}	$O(n^2l)$	Here l is length of patterns.
En_{Bu}	$O(n\log(n))$	To count number of swaps for each element efficiently, we loop backward from the end, inserting each element in a balanced binary tree and counting smaller elements which were inserted before, i.e. number of required swaps.
Multi-scale Fuzzy & Sample	For each different scale, entropy is calculated once, so the Computational complexity is scaled of normal entropy by a factor of S where S is maximum scale.	

Table 8
SVM kernels and their equations.

Kernel	Linear	RBF	Polynomial	Histogram intersection
Formula $(K(X_i, X_j) =)$	$X_i^T X_j$	$e^{-\frac{\|X_i - X_j\|^2}{2\sigma^2}}$	$(1 + X_i^T X_j)^p$	$\sum_n min(X_i^{(n)}, X_j^{(n)})$

data to output. After training the AE, the encoder part can be utilized in other tasks, such as classification.

For images and signals, it is common to use CNN-AEs. In this work, a CNN-AE with a structure presented in Table 9 is used. For selecting the most suitable structure and hyperparameters, 10 percent of training data are utilized as validation. An example of a signal and its reconstructed signal by trained AE is illustrated in Fig. 4. As illustrated in this figure, the AE has learned to preserve the overall structure of the signal; therefore it is expected from AE to extract useful information from the signal. After completion of the AE learning process, the decoder is removed and replaced with a flatten layer and a softmax classifier. Here, we finetuned the encoder layers in final training as well. Adadelta (Zeiler, 2012) and stochastic gradient descent (SGD) (Goodfellow et al., 2016) are used as optimizer for autoencoder and final classifier, respectively. The learning rate for these algorithms is set to values mentioned in their base papers.

3.6. Statistical metrics

In biomedical signal processing, statistical metrics are very important for classification performance evaluation. Here, the performance of the classifier algorithms is assessed by seven parameters namely accuracy (Acc), sensitivity (Sens), specificity (Spec), precision (Prec), f1_score (F1) and Matthews correlation coefficient (MCC). Likewise, train–test split would be utilized for preparing and testing the classifiers so that the robust and optimal results would be obtained. The hypothesis of assessment parameters is summarized in Table 10.

In these equations, each two-letter variable shows the number of samples with an assigned label and whether it was classified correctly or not; the first letter referring to classification result, true or false and second letter to the assigned label, positive or negative. So, for example, FN means the number of samples assigned to negative class falsely (they belonged to positive class).

4. Results

In this section, the results of applying the proposed approach on the Bonn dataset are provided. First, a Butterworth filter with a cutoff frequency of 0.5–40 Hz is applied to data to remove low-frequency noises such as eye blinking and high-frequency components. Then, each signal is split into different time-windows. Several choices of window length are examined, and results for 5, 10, and 23 s are presented. Next,

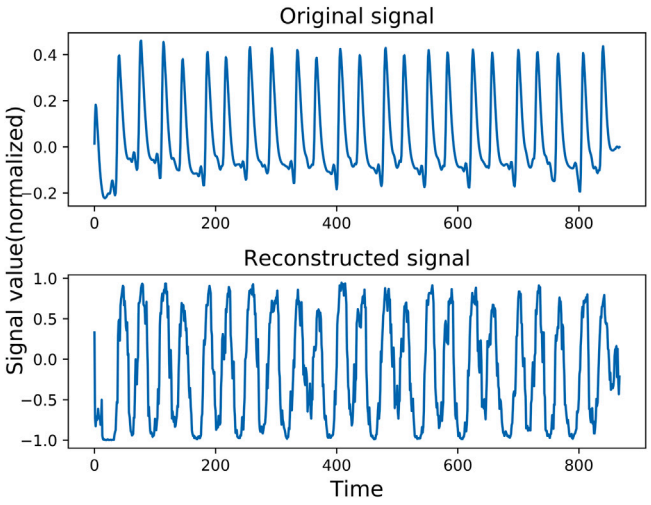
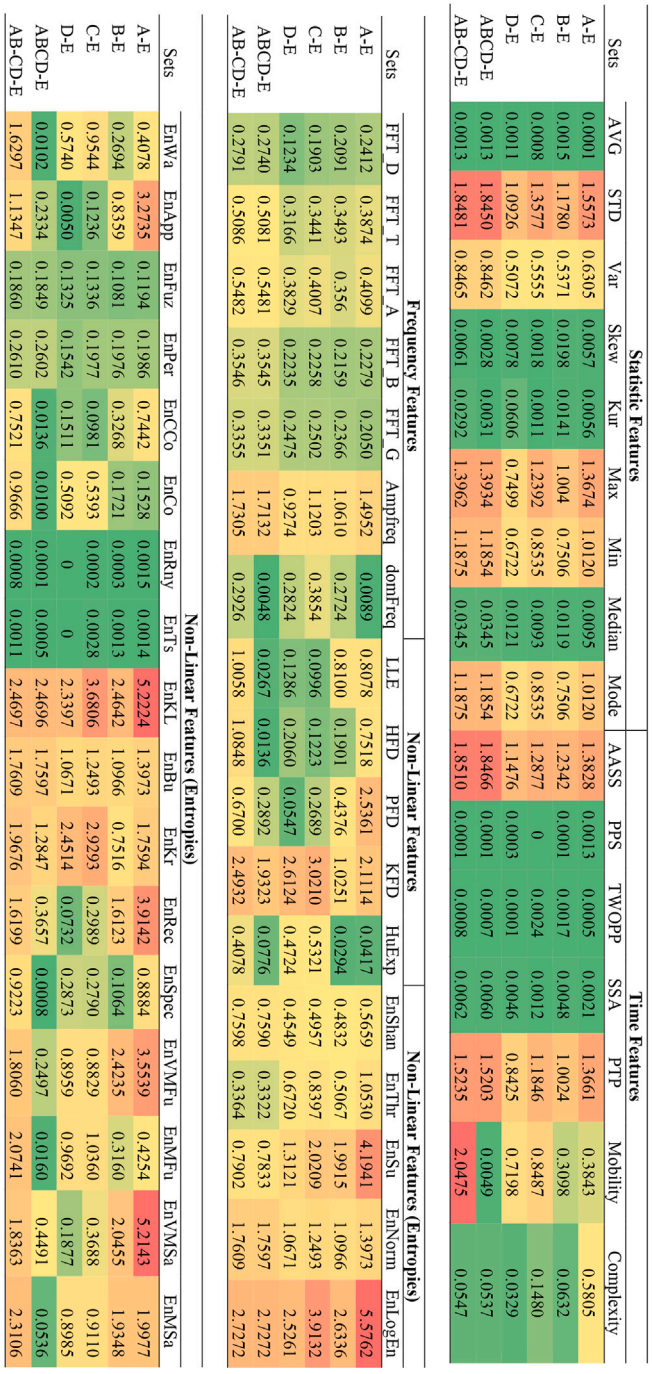


Fig. 4. Convolutional autoencoder reconstructed signal.



Sets	Statistic Features									Time Features				Complexity
	AVG	STD	Var	Skew	Kur	Max	Min	Median	Mode	AASS	PPS	TWOPP	SSA	PTP	Mobility	
A-E	0.0001	1.5573	0.6305	0.0057	0.0056	1.3674	1.0120	0.0095	1.0120	1.3828	0.0013	0.0005	0.0021	1.3661	0.3843	0.5805
B-E	0.0015	1.1780	0.5371	0.0198	0.0141	1.004	0.7506	0.0119	0.7506	1.2342	0.0001	0.0017	0.0048	1.0024	0.3098	0.0632
C-E	0.0008	1.3577	0.5555	0.0018	0.0011	1.2392	0.8535	0.0093	0.8535	1.2877	0	0.0024	0.0012	1.1846	0.8487	0.1480
D-E	0.0011	1.0926	0.5072	0.0078	0.0606	0.7499	0.6722	0.0121	0.6722	1.1476	0.0003	0.0001	0.0046	0.8425	0.7198	0.0329
ABCD-E	0.0013	1.8450	0.8462	0.0028	0.0031	1.3934	1.1854	0.0345	1.1854	1.8466	0.0001	0.0007	0.0060	1.5203	0.0049	0.0537
AB-CD-E	0.0013	1.8481	0.8465	0.0061	0.0292	1.3962	1.1875	0.0345	1.1875	1.8510	0.0001	0.0008	0.0062	1.5235	2.0475	0.0547

Sets	Frequency Features							Non-Linear Features							Non-Linear Features (Entropies)							
	FFT_D	FFT_T	FFT_A	FFT_B	FFT_G	Ampfreq	domFreq	LLE	HFD	PFD	KFD	HuExp	EnRec	EnSpec	EnShan	EnThr	EnSu	EnNorm	EnLogEn	EnWa	EnApp	EnFuz
A-E	0.2412	0.3874	0.4099	0.2279	0.2050	1.4952	0.0089	0.8078	0.7518	2.5361	2.1114	0.0417	0.5659	1.0530	4.1941	1.3973	5.5762	0.4078	3.2735	0.1194		
B-E	0.2091	0.3493	0.356	0.2159	0.2366	1.0610	0.2724	0.8100	0.1901	0.4376	1.0251	0.0294	0.4832	0.5067	1.9915	1.0966	2.6336	0.2694	0.8359	0.1081		
C-E	0.1903	0.3441	0.4007	0.2258	0.2502	1.1203	0.3854	0.0996	0.1223	0.2689	3.0210	0.5321	0.4957	0.8397	2.0209	1.2493	3.9132	0.9544	0.1236	0.1336		
D-E	0.1234	0.3166	0.3829	0.2235	0.2475	0.9274	0.2824	0.1286	0.2060	0.0547	2.6124	0.4724	0.4549	0.6720	1.3121	1.0671	2.5261	0.5740	0.0050	0.1325		
ABCD-E	0.2740	0.5081	0.5481	0.3545	0.3351	1.7132	0.0048	0.0267	0.0136	0.2892	1.9323	0.0776	0.7590	0.3322	0.7833	1.7597	2.7272	0.0102	0.2334	0.1849		
AB-CD-E	0.2791	0.5086	0.5482	0.3546	0.3355	1.7305	0.2926	1.0058	1.0848	0.6700	2.4932	0.4078	0.7598	0.3364	0.7902	1.7609	2.7272	1.6297	1.1347	0.1860		

Sets	Non-Linear Features (Entropies)																					
	EnWa	EnApp	EnFuz	EnPer	EnCCo	EnCo	EnRny	EnTs	EnKL	EnBu	EnKr	EnRec	EnSpec	EnVMFu	EnMFu	EnVMSa	EnMSa					
A-E	0.4078	3.2735	0.1194	0.1986	0.7442	0.1528	0.0015	0.0014	5.2224	1.3973	1.7594	3.9142	0.8884	3.5539	0.4254	5.2143	1.9977					
B-E	0.2694	0.8359	0.1081	0.1976	0.3268	0.1721	0.0003	0.0013	2.4642	1.0966	0.7516	1.6123	0.1064	2.4235	0.3160	2.0455	1.9348					
C-E	0.9544	0.1236	0.1336	0.1977	0.0981	0.5393	0.0002	0.0028	3.6806	1.2493	2.9293	0.2989	0.2790	0.8829	1.0360	0.3688	0.9110					
D-E	0.5740	0.0050	0.1325	0.1542	0.1511	0.5092	0	0	2.3397	1.0671	2.4514	0.0732	0.2873	0.8959	0.9692	0.1877	0.8985					
ABCD-E	0.0102	0.2334	0.1849	0.2602	0.0136	0.0100	0.0001	0.0005	2.4696	1.7597	1.2847	0.3657	0.0008	0.2497	0.0160	0.4491	0.0536					
AB-CD-E	1.6297	1.1347	0.1860	0.2610	0.7521	0.9666	0.0008	0.0011	2.4697	1.7609	1.9676	1.6199	0.9223	1.8060	2.0741	1.8363	2.3106					

Fig. 3. Feature importance in different classification cases.

all features explained in Section 3.2 are extracted from each part of the signal, and 20 most important features among them based on the Fisher score are selected for classification with SVM and KNN algorithms. For choosing this number, we tested various alternatives, such as 15,25 and 30. By testing empirically, it appeared that by choosing 20, algorithms would learn adequately, and overfitting is prevented while acceptable results are reached, showing that essential features are preserved. As shown earlier in Fig. 3, features have ranging importance for different classification cases; however, some features such as log energy entropy, Kozachenko entropy, katz fractal dimension, and average absolute signal slope are presented in picked features in all cases. Also, it is notable to mention that without feature selection, some of the classification algorithms such as SVM with rbf kernel would overfit easily or not learn anything. The results of all these algorithms without feature selection are presented later in this section.

Depending on the length of time windows, each signal may be divided into several time-windows; in those cases, we averaged over the result of the classifier for all segments of the signal to obtain the label for it.

For evaluation, we used train–test splitting, and to make results reliable, different time windows of each signal are not in test and train set simultaneously. Fig. 5 shows the best-obtained accuracy over different cases and time windows among SVMs with different kernels and KNNs. Based on this figure, the best result is reached by the 5-sec window length. Similar findings can be found in a previous work (Ghassemi et al., 2019), indicating that 5-sec is the best time window for these features on this dataset. Next, with picking 5-sec as the window length, a CNN-AE with the mentioned structure is used for the classification of signals without using extracted features. This helps to evaluate the overall quality of the extracted features. Finally, for the final model, a hybrid design of AE features and extracted features is presented; for that, after training the AE and removal of decoder, the extracted features vector is concatenated to encoder output before feeding to a softmax classifier. Results for all these different methods on the chosen window size (5-sec) are presented in Fig. 6. As mentioned earlier, without feature selection, some algorithms suffered from overfitting or did not learn properly; to show this, results of feature-based classification algorithms without feature selection are presented in Fig. 7. As it is observable in this figure, adding feature extraction almost always has helped to improve the results, and in some cases like SVM with rbf and poly kernels, this improvement is significant.

Based on these results, both handcrafted and deep learning based features extract useful information from the signal, and the best results can be obtained by utilizing both types of features, presented in the hybrid model (see Table 11).

5. Discussion, conclusion and future directions

In this work, we compared a total of 50 features and deep representation learning methods for the diagnosis of seizures in EEG signals. Using the Fisher score, an importance score is obtained for each of the extracted features; the same score is used for feature selection and classification on a benchmark dataset. For classification, additional to the group of state of the art algorithms used, a histogram intersection kernel is used for the SVM algorithm; comparing results in Figs. 6

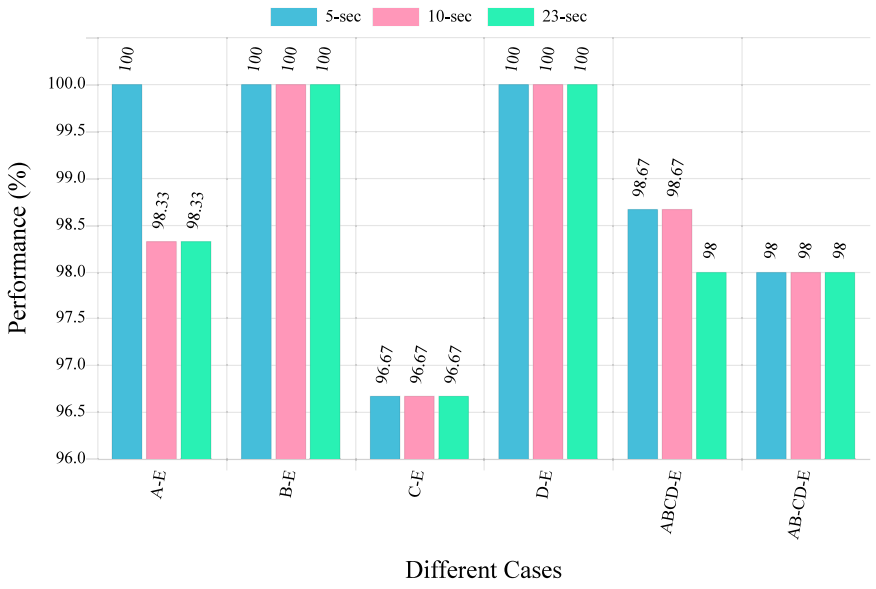


Fig. 5. Best-obtained accuracy for different time windows Using SVM and KNN.

Table 9
Proposed convolutional autoencoder structure.

Layer	Output shape		Layer info	
	Width	Depth	Kernel size	Stride
Input	868	1	–	–
ConVol	868	32	7	1
Relu	868	32	–	–
MaxPool	217	32	4	–
BatchNorm	217	32	Momentum = 0.99	
Dropout	217	32	Rate = 0.5	
ConVol	217	32	7	1
Relu	217	32	–	–
MaxPool	54	32	4	–
ConVol	54	32	7	1
Relu	54	32	–	–
UpSampling	216	32	4	–
BatchNorm	216	32	Momentum = 0.99	
ConVol	216	32	7	1
Relu	216	32	–	–
UpSampling	864	32	4	–
BatchNorm	864	32	Momentum = 0.99	
ZeroPad	868	–	–	–
ConVol	868	1	7	1
Tanh	868	1	–	–

Table 10
Statistical metrics for classification performance evaluation.

Name	Equation
Accuracy	$Acc = \frac{TP+TN}{FP+FN+TP+TN}$
Sensitivity	$Sens = \frac{TP}{FN+TP}$
Specificity	$Spec = \frac{TN}{FP+TN}$
Precision	$Prec = \frac{TP}{FP+TP}$
False positive rate	$FPR = \frac{FP}{FP+TN}$
F1_Score	$F1 = 2\frac{Prec*Sens}{Prec+Sens}$
Matthews correlation coefficient	$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

and 7, this kernel is more robust against large data dimensionalities. Additionally, the computational complexity of each feature is calculated.

To evaluate the overall performance of features, we compared them with a CNN-AE. The structure of this model was tailor-made for this problem. Finally, a hybrid model using both handcrafted features and CNN-AE is presented to reach high accuracy in the mentioned task; which can be used as a core part of a CADS as well. Comparing our method with papers reviewed in Acharya et al. (2019, 2013), Boonyakitanont et al. (2020), Shoeibi et al. (2020), Tzallas et al. (2012), Wang et al. (2017), we have used a more vast group of features, different classifiers, and combinations of handcrafted features with CNN-AEs. Also obtained results and extracted importance score can help to pick the best approach to this task when the resources are limited.

Comparing the advantages and disadvantages of deep learning based representation learning with handcrafting features can be done from various aspects. In terms of ease of use, applying deep learning based models for a new task can be much easier, and it can be done with a minimum level of knowledge on that task. However, to handcraft informative features, detailed knowledge of the task is required. On the other hand, features can be computationally cheap to compute, making them more suitable for hardware or software implementations. Another disadvantage of deep learning models is their need for considerable data in order to learn accurately. Nevertheless, as presented in the

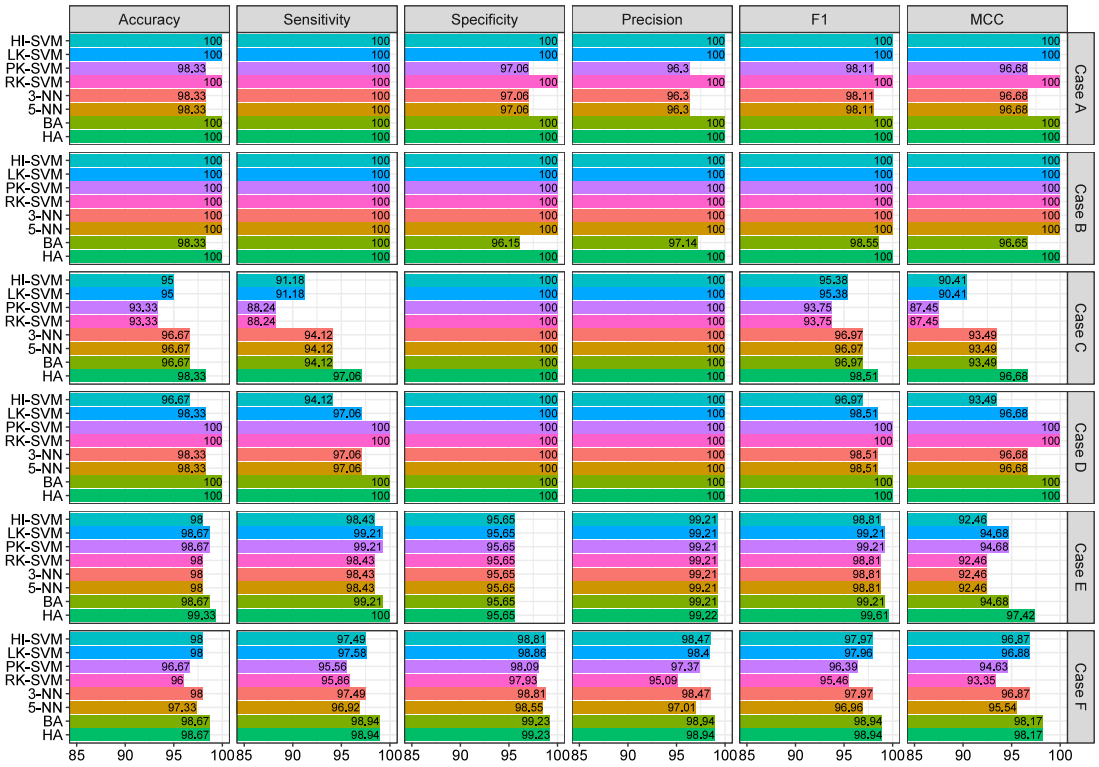


Fig. 6. Results of different classifiers using 5-sec time window for all cases.

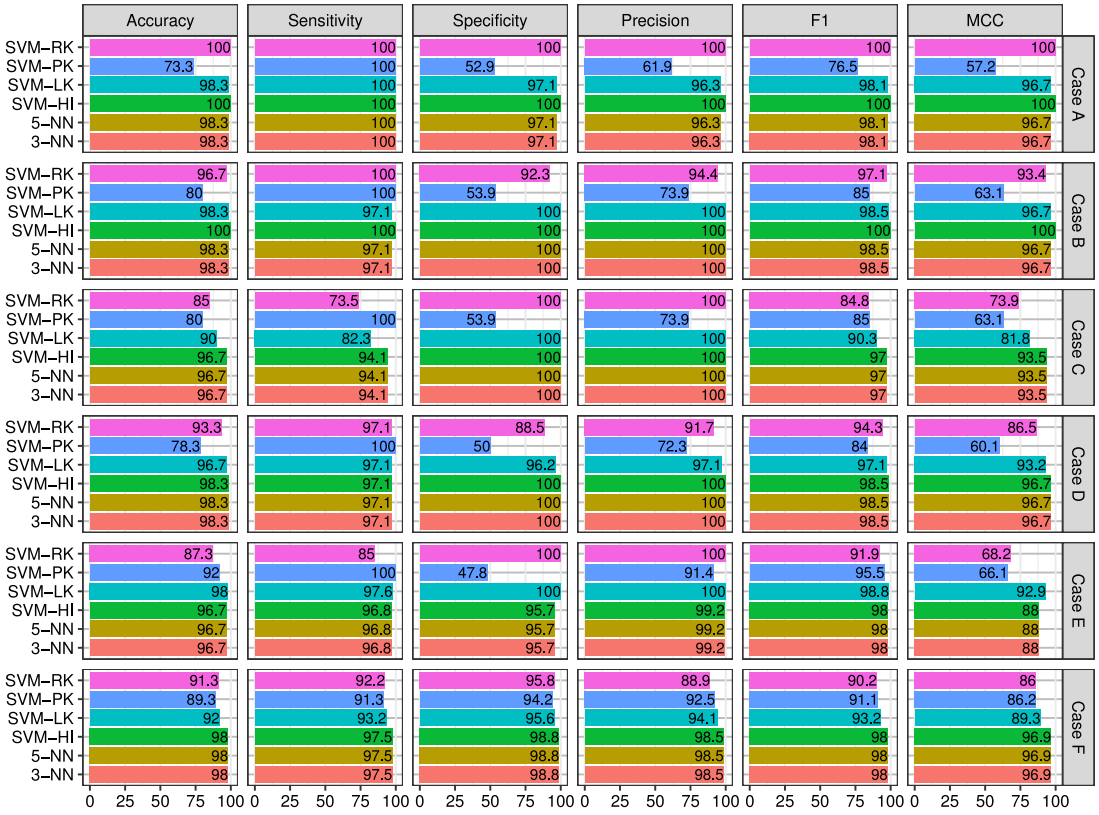


Fig. 7. Results of different classifiers using 5-sec time window for all cases without feature selection.

result section, a combination of both methods can reach a higher accuracy than each of them separately.

In conclusion, according to results, the presented CADS is capable of diagnosing seizures in epilepsy patients with reliable accuracy and can be used to help neurologists and doctors to diagnose with higher

Table 11
Results in comparison to related works.

Work	Samples	Methods	Validation	Sets	Accuracy
Swami et al. (2016)	NA	DT-CWT + Energy, STD, Shannon Entropy + T-Test + GRNN	10-Fold	Two-Classes	97.21%
Zhang and Chen (2016)	NA	LMD + Different Features+ GA - SVM	50% Train - Test	Two-Classes	99.17%
				Multi-Classes	98.27%
Tao et al. (2016)	NA	FSWT + APEN, FI + GA - SVM	50% Train - Test	Two-Classes	93.25%
				Multi-Classes	98.33%
Peker et al. (2015)	NA	DT-CWT + Statistical Features + CVANNs	10-Fold	Multi-Classes	99.30%
Li et al. (2016)	4000	DD-DWT + HE, FE + ANNOVA + GA-SVM	10-Fold	Two-Classes	100%
Li et al. (2017a)	4000	DT-CWT + HE, FD, PEN + SVM	10-Fold	Two-Classes	99.15%
				Multi-Classes	98.22%
Reddy and Rao (2017)	1000	TQWT + CCorEn + RF, MLP, and LR	10-Fold	Two-Classes	98.30%
				Multi-Classes	98.20%
Kocadagli and Langari (2017)	4000	DWT + Max, Min, Range, Mode, Mean, Std + Fuzzy Relations + GA-ANN	NA	Two-Classes	99.90%
Li et al. (2017b)	2000	DWT, HHT, EnEsp + Mean, Max, Energy, STD, + NNE	NA	Multi-Classes	98.78%
Patidar and Panigrahi (2017)	1000	TQWT + KEn + Kruskal–Wallis + LS-SVM	NA	Two-Classes	97.75%
Sharma et al. (2017)	4000	Band-Pass Filter, ATFFWT + HFD + T-Test + LS-SVM	10-Fold	Two-Classes	98.48%
Jaiswal and Banka (2017)	NA	LNDP, 1D-LGP, 1D-LBP + Histogram + Different Classifiers	10-Fold	Two-Classes	99.82%
				Multi-Classes	98.22%
Wang et al. (2017)	NA	DWT + Different Features + ANNOVA + SVM	10-Fold	Two-Classes	99.25%
Zhang et al. (2018a)	NA	WPD + FDE + Kruskal Wallis + KNN	10-Fold	Two-Classes	99.69%
				Multi-Classes	99.07%
Sharaf et al. (2018)	NA	Homogeneity, chaotic, statistical, PSD, TCC, energy + Firefly + RF	10-Fold	Two-Classes	97.00%
Zhang et al. (2019)	NA	MODWPT + Statistic Parameters + Different Feature Selection Methods + LS-SVM	10-Fold	Two-Classes	99.60%
Zhang et al. (2018b)	NA	GST + SVD + RF	10-Fold	Two-Classes	99.12%
Al Ghayab et al. (2019)	Various	TQWT + Statistic Features + KNN	5-Fold	Two-Classes	100%
				Multi-Classes	100%
Chen et al. (2019)	4096	DWT + Different Features + ANOVA-FSFS + LS-SVM	10-Fold	Two-Classes	99.50%
Lin et al. (2016)	NA	SSAE + Softmax	Na	Two-Classes	96.00%
Sharathappriyaa et al. (2018)	347	HWPT + FDs + AE + Softmax	NA	Two-Classes	98.67%
Wen and Zhang (2018)	NA	CNN-AE + Various Classifiers	10	Two-Classes	92.00%
Nishad and Pachori (2020)	4097	TQWT + CIP + RELIEFF + RF	10-Fold	Two-Classes	99.00%
Sharma and Pachori (2017a)	4097	Filtering, TQWT +FD + Different Feature Selection Methods + LS-SVM	10-Fold	Two-Classes	99.66%
				Multi-Classes	99.60%
Bhattacharyya et al. (2017)	NA	Filtering, TQWT + MKNNEn + Wrapper + SVM	10-Fold	Two-Classes	99.30%
				Multi-Classes	98.60%
Ghassemi et al. (2019)	868	Filtering, TQWT + Different Features + Fisher Score + Ensemble Learning	10-Fold	Two-Classes	100%
				Multi-Classes	98.00%
Aydemir et al. (2020)	NA	TQWT + QPS + NCA + KNN	10-Fold	Two-Classes	98.75%
				Multi-Classes	99.67%
Tzimourta et al. (2019)	500	DWT + Linear and Non-Linear Features + RF	10-Fold	Two-Classes	98.75%
				Multi-Classes	91.39%
Amin et al. (2020)	4097	Filtering, DWT + AC + ANOVA + Different Classifiers	10-Fold	Two-Classes	100%
Zhang et al. (2020)	NA	FAWT + CVDistEn1, LE + FKNN	10-Fold	Two-Classes	100%
Bhati et al. (2017a)	NA	TFLTBSW + Sub-Bands Norm + MLPNN	10-Fold	Two-Classes	99.66%
Bhati et al. (2017b)	1728	TFLTBAFB + Different Features + Kruskal–Wallis + MLPNN	10-Fold	Two-Classes	99.33%
Chandel et al. (2019)	NA	TWD + Statistical Features + KNN	NA	Two-Classes	100%
				Multi-Classes	96.00%

(*continued on next page*)

Table 11 (*continued*).

de la O Serna et al. (2020)	4097	5 BPFs + TFEBE + LS-SVM	10	Two-Classes	94.88%
Kumar et al. (2015)	NA	Gabor Filters + Histograms of 1D-LBPs + Different Features + KNN	10-Fold	Two-Classes	99.33%
Tiwari et al. (2016)	4000	DOG + histogram of LBP + SVM	10-Fold	Two-Classes	99.41%
				Multi-Classes	98.80%
Rout and Biswal (2020)	NA	VMD, HT + BLIMFs + EMRVFLN	35% Train - 65% Test	Two-Classes	100%
				Multi-Classes	99.46%
Sharma and Pachori (2017b)	510	IEVDHM–HT + ECM, FD, FMO, REM + Student's T-Test + LS-SVM	10-Fold	Two-Classes	100%
Ramos-Aguilar et al. (2020)	347	STFT + Different Features + MLP	5-Fold	Two-Classes	95.56%
Sharma et al. (2020)	NA	Filtering, TOC + SAE + Softmax	10-Fold	Two-Classes	100%
				Multi-Classes	98.40%
Joshi et al. (2014)	800	FLP + PEE, Signal Energy + SVM	NA	Two-Classes	95.33%
Gupta and Pachori (2019)	4097	FBSE + WMRPE + Different Feature Selection Methods + LS-SVM	10-Fold	Two-Classes	98.93%
Sharma et al. (2018)	4000	IF, DESA + SE, LEE, KNNE, Poincar'e Plot Features + Kruskal–Wallis + RF	10-Fold	Two-Classes	97.97%
				Multi-Classes	98.00%
Sharma and Pachori (2015)	4000	EMD, 2D and 3D PSRs of IMFs + Different Features + Kruskal–Wallis + LSSVM	10-Fold	Two-Classes	98.67%
Pachori and Patidar (2014)	1000	EMD, SODP + 95% Confidence Ellipse Area + Kruskal–Wallis + MLPNN	10-Fold	Two-Classes	97.75%
Bajaj and Pachori (2011)	NA	EMD + HT + Kruskal–Wallis + LS-SVM	NA	Two-Classes	99.50%
Pachori (2009)	NA	EMD, Fourier–Bessel expansion + MF + Kruskal–Wallis	NA	Two-Classes	NA
Raghu et al. (2019)	1388	Matrix Determinant + MLP	10-Fold	Two-Classes	96.94%
				Multi-Classes	95.97%
Jiang et al. (2020)	NA	SGD + SVM	10-Fold	Two-Classes	99.63%
				Multi-Classes	99.51%
Tuncer et al. (2019)	3840	LSP + NCA + SVM	10-Fold	Two-Classes	99.10%
				Multi-Classes	96.50%
Akyol (2020)	NA	Normalization + SEA based DNN Model + Sigmoid	10-Fold	Two-Classes	97.17%
Gao et al. (2020)	1024	ApEn, RQA + Different Feature Selection Methods + CNN	10-Fold	Two-Classes	99.26%
Ours	868	Butterworth Filter + Different Features + Fisher + CNN-AE	70% Train - 30% Test	Two-Classes	99.53%
				Multi-Classes	98.67%

confidence. For future works, evaluation of the effect of wavelet transforms in the preprocessing stage, and comparison of different wavelet transforms can help presenting sound methods for epileptic seizure detection.

CRediT authorship contribution statement

Afshin Shoeibi: Conceptualization, Methodology, Software, Writing. **Navid Ghassemi:** Conceptualization, Methodology, Software, Writing. **Roohallah Alizadehsani:** Writing. **Modjtaba Rouhani:** Conceptualization, Methodology, Validation, Writing, Supervision, Project administration. **Hossein Hosseini-Nejad:** Conceptualization, Methodology, Writing, Supervision, Project administration. **Abbas Khosravi:** Writing, Supervision, Project administration. **Maryam Panahiazar:** Writing. **Saeid Nahavandi:** Writing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Acharya, U. R., Hagiwara, Y., Deshpande, S. N., Suren, S., Koh, J. E. W., Oh, S. L., Arunkumar, N., Ciaccio, E. J., & Lim, C. M. (2019). Characterization of focal EEG signals: a review. *Future Generation Computer Systems*, *91*, 290–299.

Acharya, U. R., Hagiwara, Y., Koh, J. E. W., Oh, S. L., Tan, J. H., Adam, M., & San Tan, R. (2018). Entropies for automated detection of coronary artery disease using ecg signals: A review. *Biocybernetics and Biomedical Engineering*, *38*(2), 373–384.

Acharya, U. R., Sree, S. V., Swapna, G., Martis, R. J., & Suri, J. S. (2013). Automated EEG analysis of epilepsy: a review. *Knowledge-Based Systems*, *45*, 147–165.

Akyol, K. (2020). Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection. *Expert Systems with Applications*, *148*, Article 113239.

Al Ghayab, H. R., Li, Y., Siuly, S., & Abdulla, S. (2019). A feature extraction technique based on tunable Q-factor wavelet transform for brain signal classification. *Journal of Neuroscience Methods*, *312*, 43–52.

Amin, H. U., Yusoff, M. Z., & Ahmad, R. F. (2020). A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques. *Biomedical Signal Processing and Control*, *56*, Article 101707.

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, *64*(6), Article 061907.

Arlotta, P. (2018). Organoids required! A new path to understanding human brain development and disease. *Nature Methods*, *15*(1), 27.

Aydemir, E., Tuncer, T., & Dogan, S. (2020). A tunable-Q wavelet transform and quadruple symmetric pattern based EEG signal classification method. *Medical Hypotheses*, *134*, Article 109519.

Azami, H., Fernández, A., & Escudero, J. (2017). Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis. *Medical & Biological Engineering & Computing*, *55*(11), 2037–2052.

Bajaj, V., & Pachori, R. B. (2011). Classification of seizure and nonseizure EEG signals using empirical mode decomposition. *IEEE Transactions on Information Technology in Biomedicine*, *16*(6), 1135–1142.

Bajaj, V., & Pachori, R. B. (2013). Epileptic seizure detection based on the instantaneous area of analytic intrinsic mode functions of EEG signals. *Biomedical Engineering Letters*, *3*(1), 17–21.

Barla, A., Odone, F., & Verri, A. (2003). Histogram intersection kernel for image classification. In *Proceedings 2003 international conference on image processing (Cat. No. 03CH37429) (Vol. 3)* (pp. III–513). IEEE.

Bhati, D., Pachori, R. B., & Gadre, V. M. (2017). A novel approach for time–frequency localization of scaling functions and design of three-band biorthogonal linear phase wavelet filter banks. *Digital Signal Processing*, *69*, 309–322.

Bhati, D., Sharma, M., Pachori, R. B., & Gadre, V. M. (2017). Time–frequency localized three-band biorthogonal wavelet filter bank using semidefinite relaxation and nonlinear least squares with epileptic seizure EEG signal classification. *Digital Signal Processing*, *62*, 259–273.

Bhattacharyya, A., & Pachori, R. B. (2017). A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform. *IEEE Transactions on Biomedical Engineering*, *64*(9), 2003–2015.

Bhattacharyya, A., Pachori, R. B., Upadhyay, A., & Acharya, U. R. (2017). Tunable-q wavelet transform based multiscale entropy measure for automated classification of epileptic EEG signals. *Applied Sciences*, *7*(4), 385.

Bhattacharyya, A., Singh, L., & Pachori, R. B. (2018). Fourier–Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals. *Digital Signal Processing*, *78*, 185–196.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Boonyakitanont, P., Lek-Uthai, A., Chomtho, K., & Songsiri, J. (2020). A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. *Biomedical Signal Processing and Control*, *57*, Article 101702.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152). ACM.

Brent, R. P. (1976). Multiple-precision zero-finding methods and the complexity of elementary function evaluation. In *Analytic computational complexity* (pp. 151–176). Elsevier.

Chandel, G., Upadhyaya, P., Farooq, O., & Khan, Y. (2019). Detection of seizure event and its onset/offset using orthonormal triadic wavelet based features. *IRBM*, *40*(2), 103–112.

Chen, S., Zhang, X., Chen, L., & Yang, Z. (2019). Automatic diagnosis of epileptic seizure in electroencephalography signals using nonlinear dynamics features. *IEEE Access*, *7*, 61046–61056.

Chudnovsky, D. V., & Chudnovsky, G. V. (2004). Approximations and complex multiplication according to Ramanujan. In *Pi: A source book* (pp. 596–622). Springer.

Chung, K.-M., Kao, W.-C., Sun, C.-L., Wang, L.-L., & Lin, C.-J. (2003). Radius margin bounds for support vector machines with the RBF kernel. *Neural Computation*, *15*(11), 2643–2681.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*. MIT press.

Cunningham, P., & Delany, S. J. (2007). K-nearest neighbour classifiers. *Multiple Classifier Systems*, *34*(8), 1–17.

de la O Serna, J. A., Paternina, M. R. A., Zamora-Méndez, A., Tripathy, R. K., & Pachori, R. B. (2020). EEG-Rhythm specific Taylor-Fourier filter bank implemented with o-splines for the detection of epilepsy using EEG signals. *IEEE Sensors Journal*.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

Eijkhout, V. (2013). *Introduction to high performance scientific computing*. Lulu. com.

El Azami, M., Hammers, A., Jung, J., Costes, N., Bouet, R., & Lartizien, C. (2016). Detection of lesions underlying intractable epilepsy on T1-weighted MRI as an outlier detection problem. *PLoS One*, *11*(9), Article e0161498.

Esteller, R., Vachtsevanos, G., Echauz, J., & Litt, B. (2001). A comparison of waveform fractal dimension algorithms. *IEEE Transactions on Circuits and Systems I*, *48*(2), 177–183.

Faust, O., Acharya, U. R., Min, L. C., & Sputh, B. H. (2010). Automatic identification of epileptic and background EEG signals using frequency domain parameters. *International Journal of Neural Systems*, *20*(02), 159–176.

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, *161*, 1–13.

Feigin, V. L., Abajobir, A. A., Abate, K. H., Abd-Allah, F., Abdulle, A. M., Abera, S. F., Abyu, G. Y., Ahmed, M. B., Aichour, A. N., & Aichour, I. (2017). Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015. *The Lancet Neurology*, *16*(11), 877–897.

Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C., Genton, P., Lee, P., & Engel Jr., J. (2005). Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia*, *46*(4), 470–472.

Gao, X., Yan, X., Gao, P., Gao, X., & Zhang, S. (2020). Automatic detection of epileptic seizure based on approximate entropy, recurrence quantification analysis and convolutional neural networks. *Artificial Intelligence in Medicine*, *102*, Article 101711.

Ghassemi, N., Shoeibi, A., Rouhani, M., & Hosseini-Nejad, H. (2019). Epileptic seizures detection in EEG signals using TQWT and ensemble learning. In *2019 9th international conference on computer and knowledge engineering* (pp. 403–408). IEEE.

Goel, V., Brambrink, A. M., Baykal, A., Koehler, R. C., Hanley, D. F., & Thakor, N. V. (1996). Dominant frequency analysis of EEG reveals brain's response during injury and recovery. *IEEE Transactions on Biomedical Engineering*, *43*(11), 1083–1092.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. ArXiv preprint arXiv:1202.3725.

Gupta, V., Bhattacharyya, A., & Pachori, R. B. (2017). Classification of seizure and non-seizure EEG signals based on EMD-TQWT method. In *2017 22nd international conference on digital signal processing* (pp. 1–5). IEEE.

Gupta, V., & Pachori, R. B. (2019). Epileptic seizure identification using entropy of FBSE based EEG rhythms. *Biomedical Signal Processing and Control*, *53*, Article 101569.

Halim, S., Halim, F., Skiena, S. S., & Revilla, M. A. (2013). *Competitive programming 3*. Lulu Independent Publish.

Hassan, A. R., & Haque, M. A. (2015). Epilepsy and seizure detection using statistical features in the complete ensemble empirical mode decomposition domain. In *TENCON 2015-2015 IEEE region 10 conference* (pp. 1–6). IEEE.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, *13*(4), 18–28.

Jaiswal, A. K., & Banka, H. (2017). Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals. *Biomedical Signal Processing and Control*, *34*, 81–92.

Jiang, Y., Chen, W., & Li, M. (2020). Symplectic geometry decomposition-based features for automatic epileptic seizure detection. *Computers in Biology and Medicine*, *116*, Article 103549.

Joshi, V., Pachori, R. B., & Vijesh, A. (2014). Classification of ictal and seizure-free EEG signals using fractional linear prediction. *Biomedical Signal Processing and Control*, *9*, 1–5.

Kalatzis, I., Piliouras, N., Ventouras, E., Papageorgiou, C. C., Rabavilas, A. D., & Cavouras, D. (2004). Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the p600 component of ERP signals. *Computer Methods and Programs in Biomedicine*, *75*(1), 11–22.

Khalid, M. I., Alotaiby, T., Aldosari, S. A., Alshebeili, S. A., Al-Hameed, M. H., Almohammed, F. S. Y., & Alotaibi, T. S. (2016). Epileptic MEG spikes detection using common spatial patterns and linear discriminant analysis. *IEEE Access*, *4*, 4629–4634.

Kocadagli, O., & Langari, R. (2017). Classification of EEG signals for epileptic seizures using hybrid artificial neural networks based wavelet transforms and fuzzy relations. *Expert Systems with Applications*, *88*, 419–434.

Kumar, T., Kanhangad, V., & Pachori, R. B. (2015). Classification of seizure and seizure-free EEG signals using local binary patterns. *Biomedical Signal Processing and Control*, *15*, 33–40.

Li, M., Chen, W., & Zhang, T. (2016). Automatic epilepsy detection using wavelet-based nonlinear analysis and optimized SVM. *Biocybernetics and Biomedical Engineering*, *36*(4), 708–718.

Li, M., Chen, W., & Zhang, T. (2017a). Automatic epileptic EEG detection using DT-CWT-based non-linear features. *Biomedical Signal Processing and Control*, *34*, 114–125.

Li, M., Chen, W., & Zhang, T. (2017b). Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble. *Biomedical Signal Processing and Control*, *31*, 357–365.

Lin, Q., Ye, S.-q., Huang, X.-m., Li, S.-y., Zhang, M.-z., Xue, Y., & Chen, W.-S. (2016). Classification of epileptic EEG signals with stacked sparse autoencoder based on deep learning. In *International conference on intelligent computing* (pp. 802–810). Springer.

Manis, G., Aktaruzzaman, M., & Sassi, R. (2017). Bubble entropy: an entropy almost free of parameters. *IEEE Transactions on Biomedical Engineering*, *64*(11), 2711–2718.

Nishad, A., & Pachori, R. B. (2020). Classification of epileptic electroencephalogram signals using tunable-Q wavelet transform based filter-bank. *Journal of Ambient Intelligence and Humanized Computing*, 1–15.

Oh, S.-H., Lee, Y.-R., & Kim, H.-N. (2014). A novel EEG feature extraction method using Hjorth parameter. *International Journal of Electronics and Electrical Engineering*, *2*(2), 106–110.

Pachori, R. B. (2009). Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition. *Journal of Electrical and Computer Engineering*, *2008*.

Pachori, R. B., & Patidar, S. (2014). Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions. *Computer Methods and Programs in Biomedicine*, *113*(2), 494–502.

Park, Y., Luo, L., Parhi, K. K., & Netoff, T. (2011). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*, *52*(10), 1761–1770.

Patidar, S., & Panigrahi, T. (2017). Detection of epileptic seizure using kraskov entropy applied on tunable-q wavelet transform of EEG signals. *Biomedical Signal Processing and Control*, *34*, 74–80.

Peker, M., Sen, B., & Delen, D. (2015). A novel method for automated diagnosis of epilepsy using complex-valued classifiers. *IEEE Journal of Biomedical and Health Informatics*, *20*(1), 108–118.

Raghu, S., Sriraam, N., Hegde, A. S., & Kubben, P. L. (2019). A novel approach for classification of epileptic seizures using matrix determinant. *Expert Systems with Applications*, *127*, 323–341.

Ramos-Aguilar, R., Olvera-López, J. A., Olmos-Pineda, I., & Snchez-Urrieta, S. (2020). Feature extraction from EEG spectrograms for epileptic seizure detection. *Pattern Recognition Letters*.

Reddy, G. S., & Rao, R. (2017). Automated identification system for seizure EEG signals using tunable-Q wavelet transform. *Engineering Science and Technology, An International Journal*, *20*(5), 1486–1493.

Rout, S. K., & Biswal, P. K. (2020). An efficient error-minimized random vector functional link network for epileptic seizure classification using VMD. *Biomedical Signal Processing and Control*, *57*, Article 101787.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Sharaf, A. I., El-Soud, M. A., & El-Henawy, I. M. (2018). An automated approach for epilepsy detection based on tunable Q-wavelet and firefly feature selection algorithm. *International Journal of Biomedical Imaging*, *2018*.

Sharathappriyaa, V., Gautham, S., & Lavanya, R. (2018). Auto-encoder based automated epilepsy diagnosis. In *2018 international conference on advances in computing, communications and informatics* (pp. 976–982). IEEE.

Sharma, R., & Pachori, R. B. (2015). Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Systems with Applications*, *42*(3), 1106–1117.

Sharma, M., & Pachori, R. B. (2017a). A novel approach to detect epileptic seizures using a combination of tunable-q wavelet transform and fractal dimension. *Journal of Mechanics in Medicine and Biology*, *17*(07), Article 1740003.

Sharma, R. R., & Pachori, R. B. (2017b). Time–frequency representation using IEVDHM–HT with application to classification of epileptic EEG signals. *IET Science, Measurement & Technology*, *12*(1), 72–82.

Sharma, M., Pachori, R. B., & Acharya, U. R. (2017). A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters*, *94*, 172–179.

Sharma, R., Pachori, R. B., & Sircar, P. (2020). Seizures classification based on higher order statistics and deep neural network. *Biomedical Signal Processing and Control*, *59*, Article 101921.

Sharma, R. R., Varshney, P., Pachori, R. B., & Vishvakarma, S. K. (2018). Automated system for epileptic EEG detection using iterative filtering. *IEEE Sensors Letters*, *2*(4), 1–4.

Shayegh, F., Sadri, S., Amirfattahi, R., & Ansari-Asl, K. (2014). A model-based method for computation of correlation dimension, Lyapunov exponents and synchronization from depth-EEG signals. *Computer Methods and Programs in Biomedicine*, *113*(1), 323–337.

Shoeibi, A., Ghassemi, N., Khodatars, M., Jafari, M., Hussain, S., Alizadehsani, R., Moridian, P., Khosravi, A., Hosseini-Nejad, H., & Rouhani, M. (2020). Epileptic seizure detection using deep learning techniques: A review. ArXiv preprint arXiv: 2007.01276.

Smelick, C., Britton, J. W., Tatum, W. O., & Feyissa, A. M. (2018). Unusual seizure evolution: Focal-general-focal-general. *Epilepsy & Behavior Case Reports*, *10*, 54–56.

Subasi, A. (2007). EEG Signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, *32*(4), 1084–1093.

Swami, P., Gandhi, T. K., Panigrahi, B. K., Tripathi, M., & Anand, S. (2016). A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Systems with Applications*, *56*, 116–130.

Tan, Y.-L., Kim, H., Lee, S., Tihan, T., Ver Hoef, L., Mueller, S. G., Barkovich, A. J., Xu, D., & Knowlton, R. (2018). Quantitative surface analysis of combined MRI and PET enhances detection of focal cortical dysplasias. *Neuroimage*, *166*, 10–18.

Tao, Z., Wan-Zhong, C., & Ming-Yang, L. (2016). Automatic seizure detection of electroencephalogram signals based on frequency slice wavelet transform and support vector machine. *Acta Physica Sinica*, *65*(3), Article 1550040.

Tian, Y., Zhang, H., Xu, W., Zhang, H., Yang, L., Zheng, S., & Shi, Y. (2017). Spectral entropy can predict changes of working memory performance reduced by short-time training in the delayed-match-to-sample task. *Frontiers in Human Neuroscience*, *11*, 437.

Tiwari, A. K., Pachori, R. B., Kanhangad, V., & Panigrahi, B. K. (2016). Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals. *IEEE Journal of Biomedical and Health Informatics*, *21*(4), 888–896.

Tuncer, T., Dogan, S., & Akbal, E. (2019). A novel local senary pattern based epilepsy diagnosis system using eeg signals. *Australasian Physical & Engineering Sciences in Medicine*, *42*(4), 939–948.

Tzallas, A. T., Tsipouras, M. G., Tsalikakis, D. G., Karvounis, E. C., Astrakas, L., Konitsiotis, S., & Tzaphlidou, M. (2012). Automated epileptic seizure detection methods: a review study. *Epilepsy-Histological, Electroencephalographic and Psychological Aspects*, 75–98.

Tzimourta, K. D., Tzallas, A. T., Giannakeas, N., Astrakas, L. G., Tsalikakis, D. G., Angelidis, P., & Tsipouras, M. G. (2019). A robust methodology for classification of epileptic seizures in EEG signals. *Health and Technology*, *9*(2), 135–142.

Wang, L., Xue, W., Li, Y., Luo, M., Huang, J., Cui, W., & Huang, C. (2017). Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis. *Entropy*, *19*(6), 222.

Wen, T., & Zhang, Z. (2018). Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. *IEEE Access*, *6*, 25399–25410.

Willis, T. (1965). *Anatomy of the brain and nerves: Volumes 1 & 2*. McGill-Queen's Press-MQUP.

Zeiler, M. D. (2012). ADADELTA: an adaptive learning rate method. ArXiv preprint arXiv:1212.5701.

Zhang, T., & Chen, W. (2016). LMD Based features for the automatic seizure detection of EEG signals using SVM. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(8), 1100–1108.

Zhang, T., Chen, W., & Li, M. (2018a). Fuzzy distribution entropy and its application in automated seizure detection technique. *Biomedical Signal Processing and Control*, *39*, 360–377.

Zhang, T., Chen, W., & Li, M. (2018b). Generalized stockwell transform and SVD-based epileptic seizure detection in EEG using random forest. *Biocybernetics and Biomedical Engineering*, *38*(3), 519–534.

Zhang, T., Chen, W., & Li, M. (2019). Classification of inter-ictal and ictal EEGs using multi-basis modwpt, dimensionality reduction algorithms and LS-SVM: A comparative study. *Biomedical Signal Processing and Control*, *47*, 240–251.

Zhang, T., Chen, W., & Li, M. (2020). Complex-valued distribution entropy and its application for seizure detection. *Biocybernetics and Biomedical Engineering*, *40*(1), 306–323.