Journal of King Saud University - Computer and Information Sciences xxx (xxxx) 1–16

Contents lists available at ScienceDirect

Journal of King Saud University - Computer and Information Sciences

journal homepage: www.elsevier.com

Gradient boosting machines fusion for automatic epilepsy detection from EEG signals based on wavelet features

Dwi Sunaryono [a], Riyanarto Sarno [a], Joko Siswantoro [b,*]

[a] Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[b] Department of Informatics Engineering, University of Surabaya Jl. Kali Rungkut, Surabaya 60293, Indonesia

ARTICLE INFO

Article history:
Received 31 May 2021
Received in revised form 17 November 2021
Accepted 19 November 2021

Keywords:
Epilepsy detection
EEG signals
Gradient boosting machines fusion

ABSTRACT

Automatic epilepsy detection from electroencephalogram (EEG) signals is an alternative to manual detection performed by a human expert. High classification performance is needed in automatic epilepsy detection from EEG signals to avoid miss detection. This study aims to propose a classification method for automatic epilepsy detection from EEG signals. The original EEG signals were processed using discrete Fourier transform (DFT) and discrete wavelet transform (DWT) prior to feature extraction. A fusion of 2-class and 3 class gradient boosting machines (GBM), called GBMs fusion, was used to classify EEG signals based on some statistical features and crossing frequency features. In addition, a genetic algorithm was used to select the prominent features before classification. The proposed method has been evaluated using three classes EEG signals (normal-interictal-ictal) included in EEG dataset from University of Bonn. The experimental result shows that the proposed GBMs fusion can improve the performance of a single GBM in classifying EEG signals. Furthermore, the proposed GBMs fusion can perfectly detect epilepsy from EEG signals with an accuracy of 100%. However, the performance of GBMs fusion may not be generalized to the other EEG dataset.

1. Introduction

Epilepsy detection is a growing research topic due to the increasing number of people detected suffering from epilepsy. According to World Health Organization (WHO), globally, an estimated five million people are diagnosed with epilepsy every year, some in high-income countries (World Health Organization, 2019). Epilepsy is a neurological disorder as a result of occurring abnormal electrical activity in the brain. Epilepsy symptoms can be recurring seizures to the unconscious state (Acharya et al., 2018; Peker et al., 2016; Sreekumar et al., 2021). Initially, the detection of epilepsy is manually performed based on an electroencephalogram (EEG) signal. EEG signal is a low-cost tool for evaluating brain activities recorded using some electrodes arranged on the scalp (Parvinnia et al., 2014). The signal is visually examined by expert medical doctors in neurology to determine the occurring of epilepsy. However, manually examining EEG signals is time-consuming, tedious, and often produces a false alarm or missed epilepsy detection (Satyender et al., 2021). To overcome the problem an automatic epilepsy detection from EEG signals can be considered.

Several studies have been conducted to detect epilepsy automatically from EEG signals, especially in classifying EEG signals in one of three classes: EEG signal from a healthy subject (normal), epilepsy subject during seizure-free condition (interictal), and epilepsy subject during seizure activity (ictal). Some previous related studies in EEG signals classification are summarized in Table 1. Automatic epilepsy detection from EEG signals is a classification problem that can be performed in time domain, frequency domain, or time–frequency domain. In time domain classification, the input features are directly extracted from EEG signals. On the other hand, EEG signal is transformed to frequency domain or time–frequency domain prior to feature extraction in frequency domain or time–frequency domain classification. Discrete Fourier transform (DFT) and discrete wavelet transform (DWT) are commonly used to transform EEG signal from time domain to frequency domain and time–frequency domain, respectively. In addition, some studies employed empirical signal decomposition methods, such as empirical mode decomposition (EMD), empirical wavelet transform (EWD), and variational mode decomposition (VMD), to process EEG signals.

Kaya et al. (2014) performed feature extraction in time domain by directly extracting one-dimensional local binary pattern (1D-LBP) features from original EEG signals. The features were inputted to BayesNet, Support vector machine (SVM), artificial neural network

EEG, **electroencephalogram**

Peer review under responsibility of King Saud University.

* Corresponding author.
E-mail addresses: dwis@if.its.ac.id (D. Sunaryono), riyanarto@if.its.ac.id (R. Sarno), joko_siswantoro@staff.ubaya.ac.id (J. Siswantoro).

https://doi.org/10.1016/j.jksuci.2021.11.015
1319-1578/© 2021

Table 1
Some previous related studies in EEG signals classification.

Author	Feature extraction	Feature selection/ reduction	Classifier
Chua et al. (2009)	DFT + HOS features	–	GMM, SVM
Acharya et al. (2011)	DWT + HOS features	ANOVA	SVM, k-NN, RBPNN, GMM, fuzzy classifier
Guo et al. (2011)	DWT + mean, std dev, energy, length, and skewness	GP	k-NN
Du et al. (2012)	DFT + HOS features	PCA	LR
Acharya et al. (2012)	Entropy, HOS features, fractal dimension, and Hurst exponent	ANOVA	Fuzzy classifier
Martis et al. (2013)	ITD + energy, entropy, and FD	ANOVA	SVM, PNN, DT
Kaya et al. (2014)	1-D local binary patterns	–	BayesNet, SVM, ANN, LR, FT
Martis et al. (2015)	DWT + LLE, HFD, HE, SE	–	RBF-SVM, DT, k-NN
Djemili et al. (2016)	EMD + max, min, mean, std dev	–	ANN
Peker et al. (2016)	DT-CWT + max, min, mean, std dev, and median	–	CVANN
Hamad et al. (2017)	DWT	GWO	GWO-SVM
Zhang and Chen (2017)	LMD + temporal statistics, fractal dimension, Renyi entropy, HE	ANOVA	GA-SVM
Li et al. (2017)	DT-CWT + HE, FD and PE		SVM, k-NN, DT, RF
Acharya et al. (2018)	–	–	CNN
Saini and Dutta (2018)	Min, max, zero-crossing rate, skewness, kurtosis, mean, median, mode, energy, entropy, variance, signal to noise ratio, coefficient of variation, std dev, mean power	Kruskal-Wallis Test	PSO-ANN
Houssein et al. (2019)	DWT + max, min, mean, std dev, variance, median, skewness, energy, RWE, entropy	WOA	WOA-SVM
Kumar and Rao (2019)	VMD + DE, PRMS	–	RF
Raghu et al. (2019)	Matrix determinant	–	k-NN, SVM, ANN
Türk and Özerdem (2019)	CWT + Scalogram	–	CNN
Wang et al. (2019)	DWT + Statistical features	PCA	GBM + Grid search optimization
Carvalho et al. (2020)	EMD, EEMD, CEMDAAN, EWT, VMD + spectral features, Hjorth mobility, statistical moments	RFE	SVM, k-NN, GPC, ANN
Goshvarpour and Goshvarpour (2020)	Lagged Poincare plot parameters	–	k-NN, PNN
Zhang et al. (2020)	–	–	MNL-Network
Zhao et al. (2020)	–	–	1D CNN
Sukriti et al. (2021)	MDE, RCMDE	ANOVA	SVM
Li and Chen (2021)	FFT + PCANet	–	SVM

(ANN), logistic regression (LR), and functional tree (FT) for classification. Saini and Dutta (2018) used min amplitude, max amplitude, zero-crossing rate, skewness, kurtosis, mean, median, mode, energy, entropy, variance, signal to noise ratio, coefficient of variation, standard deviation, mean and power directly extracted from EEG signals to classify epilepsy. The features were selected using a nonparametric test, called Kruskal-Wallis test, before classification. An artificial neural network trained with particle swarm optimization (PSO-ANN) was used for classification. Lagged Poincare plot parameters were the other features directly extracted from EEG signals for epilepsy detection, as reported by Goshvarpour and Goshvarpour (2020). The features were used as input to k-NN and probabilistic neural network (PNN) for classification and obtained the best accuracy of 98.33%. Raghu et al. (2019) proposed matrix determinant extracted directly from EEG signal for epileptic seizures classification. The original EEG signal was arranged as a square matrix and the logarithm of absolute value of matrix determinant was calculated as a feature for classification. Three classifiers, including k-NN, SVM, and ANN, were used to classify EEG signals. The best accuracy of 99.45% was achieved using ANN in two classes EEG signals classification. Sukriti et al. (2021) proposed the use of two entropy features, called multiscale dispersion entropy (MDE) and refined composite multiscale dispersion entropy (RCMDE) to detect seizure from EEG signals. One way analysis of variance (ANOVA) was employed select to determine significant features from MDE and RCMDE before inputted to SVM for classification. The best accuracy of 96.67% was achieved using RCMDE features.

Some studies employed a convolutional neural network (CNN) to directly classify EEG signals in time domain. A 13-layers CNN was used by Acharya et al. (2018) to automatically detect seizure from EEG signals. Although CNN has achieved a good performance in image classification, the CNN proposed by Acharya et al. (2018) only produced the classification accuracy of 88.67% in seizure detection from EEG signals. Zhao et al. (2020) proposes a 1D CNN for robust detection of seizure in EEG signals. The proposed CNN model consisted of three convolutional blocks and three fully connected layers and achieved the best accuracy of 98.06% in three classes problem. Zhang et al. (2020) also proposed a 1D CNN called multi-scale non-local (MNL) network for automatic epilepsy detection from EEG signals. To increase CNN performance in epilepsy detection, two specific layers, namely signal pooling layer and multi-scale non-local layer were added to the network. The proposed network achieved the best accuracy of 98. 64% in classifying three classes EEG signals.

Fourier transform was a common technique to analyze EEG signals in frequency domain. Chua et al. (2009) used DFT to transform EEG signals into frequency domain in identifying epileptic EEG signals. The higher-order spectra (HOS) features were extracted from the output of DFT and inputted to Gaussian mixture model (GMM) and SVM for classification. The best accuracy of 93.11% was achieved by GMM. DFT and HOS features were also used by Du et al. (2012). Before classification, the dimension of features was reduced using principal component analysis (PCA). Some machine learning models were used for classification and achieved the best accuracy of 94.5% using simple logistic regression. Li and Chen (2021) employed fast Fourier transform (FFT) and PCA neural network (PCANet) for EEG signals classification. FFT was used to construct a frequency matrix from each EEG signal. A set of features were extracted from the matrix using a deep learning model called PCANet. SVM was used to classify the features and obtained the best accuracy of 99.60% in classifying three classes EEG signals.

Feature extraction in time–frequency domain was most widely used to detect epilepsy from EEG signals in previous studies. Guo et al. (2011) used DWT to decompose EEG signals into five sub-signals. Five features, including mean, standard deviation, energy, length, and skewness, were extracted from each sub-signal to obtain 25-dimensions feature space. Genetic programming (GP) was applied to reduce the dimension of feature space. The reduced features were then classified using k-nearest neighbor (k-NN) into three classes: normal, seizure-free, and seizure with the classification accuracy of 93.5%. DWT was also employed by Acharya et al. (2011), Martis et al. (2015), and Wang et al. (2019) in classifying epileptic EEG signals. Acharya et al. (2011) extracted HOS features from the output of DWT. ANOVA was used to determine the significant features from HOS features. Some classifiers, including SVM, k-NN, radial basis probabilistic neural network (RBPNN), Gaussian mixture model (GMM), and fuzzy Sugeno classifier, were used to detect epileptic EEG signal by Acharya et al. (2011). The best accuracy of 98.5% was obtained using the fuzzy classifier.

Martis et al. (2015) used DWT with Daubechies 6 family to decompose EEG signal up to six decomposition levels. From the output of DWT, some nonlinear features were extracted, including the largest Lyapunov exponent (LLE), Higuchi fractal dimension (HFD), Hurst exponent (HE), and sample entropy (SE). The significant features were selected using ANOVA before inputted to SVM, decision tree (DT), and k-NN. The best performance of 98% was achieved using SVM with radial basis function (RBF) kernel. Wang et al. (2019) proposed the use of DWT with the Symlets 4 family to recognize epileptic EEG signals. The statistical features including mean and standard deviation were extracted from the output of DWT. PCA was employed to reduce the dimension of features before classification using gradient boosting machine (GBM). To obtain the best accuracy of 96.5%, grid search optimization was employed to optimize the hyperparameters of GBM. Hamad et al. (2017) proposed grey wolf optimization (GWO) (Mirjalili et al., 2014) based feature selection to select importance features from the output of DWT in classifying EEG signals. The selected features were then inputted to SVM optimized with GWO (GWO-SVM). Houssein et al. (2019) proposed whale optimization algorithm (WOA) (Mirjalili and Lewis, 2016) based feature selection for EEG signals classification. For levels of DWT with Daubechies 4 family were employed to decompose the EEG signals. Some features, including maximum, minimum, mean, standard deviation, variance, median, skewness, energy, relative wave energy (RWE), and entropy, were extracted from the output of DWT. Some selected features were then inputted to SVM optimized with WOA (WOA-SVM). Although the methods proposed by Hamad et al. (2017) and Houssein et al. (2019) achieved the best classification accuracy of 100%, the methods were only used to classify two classes EEG signals.

Besides DWT, intrinsic time-scale decomposition (ITD), continuous wavelet transformation (CWT), and dual-tree complex wavelet transformation (DTCWT) were also used for feature extraction in time–frequency domain. Martis et al. (2013) proposed a method for seizure prediction from EEG signals using ITD. The EEG signal was decomposed using ITD to obtain low-pass and high-pass signals. Energy, HFD, and SE were extracted from both low-pass and high-pass signals. ANOVA was performed to select the significant features prior to classification using SVM, probabilistic neural network (PNN), and DT. The highest accuracy of 95.67% was obtained using DT. Türk and Özerdem (2019) used CWT to contract a 2D time–frequency scalogram image from EEG signal. The scalogram image was inputted to CNN for classification. The best classification accuracy was 99.00% for three classes EEG signals. Peker et al. (2016) proposed the use of DTCWT for automated epilepsy diagnosis from EEG signals. Five features were extracted from the outputs of DTCWT, that are minimum, maximum, mean, standard deviation. A complex-valued artificial neural network (CVANN) was used to classify EEG signals based on the extracted features and obtained 99.3% accuracy. DTCWT was also used by Li et al. (2017) to automatically detect epilepsy from EEG signals. Three nonlinear features, which are HE, fractal dimension (FD), and permutation entropy (PE), were obtained from the output of DTCWT. The features were then inputted to SVM, k-NN, DT, and random forest (RF) for classification. The best accuracy of 98.87% was obtained using SVM.

Some empirical signal decompositions were also employed in EEG signal classification for epilepsy detection. Djemili et al. (2016) proposed empirical mode decomposition (EMD) for feature extraction in EEG signals. Statistical features, including maximum, minimum, mean, and standard deviation, were extracted from the output of EMD and used for classification using ANN. Although the proposed method achieved the best accuracy of 100%, it was only used in two classes problem. Kumar and Rao (2019) proposed variational mode decomposition (VMD) for epileptic seizure classification in EEG signal. Two semantic features, namely differential entropy (DE) and peak-magnitude of root mean square ratio (PRMS), were extracted from the output of VMD and used to classify EEG signal using random forest (RF). The proposed method achieved the best accuracy of 94.10% in classifying normal and epileptic EEG signal. Carvalho et al.(2020) conducted a study to evaluate five adaptive decomposition methods, including empirical mode decomposition (EMD), extended EMD (EEMD), complete EEMD with adaptive noise (CEEMDAN), empirical wavelet transforms (EWT), and variational mode decomposition (VMD) for EEG signals classification. Spectral features, Hjorth mobility, and statistical moments were extracted from the decomposed EEG signals. Some classifiers were employed in the study to classify EEG signals using the extracted features. The experiment results showed that all decomposition method produced similar performance in classifying EEG signals.

High classification accuracy is the most important factor in automatic epilepsy detection from EEG signals. Missed epilepsy detection can lead to inappropriate treatment for the sufferer. Although several methods have been proposed to automatically detect epilepsy from three classes EEG signals, none of them have achieved 100% accuracy. Therefore, the motivation of this study is to develop a classification method that can achieve 100% accuracy in classifying three classes EEG signals. Furthermore, none of the previous studies used classifiers fusion to detect epilepsy from EEG signals. The use of classifiers fusion to achieve higher classification accuracy is the main advantage of this study compared with others. Furthermore, the use of simple statistical features and crossing frequency features extracted from the output of DFT and DWT are the other advantages of this study. This study proposed a method for automatic epilepsy detection from EEG signals using gradient boosting machines (GBMs) fusion. The proposed method implemented the combination of frequency domain and time–frequency domain feature extraction by applying DFT and DWT. In addition, a genetic algorithm (GA) was used to select the most discriminative features used in classification. The complete flowchart for the proposed method is shown in Fig. 1. The rest of the paper is organized as follows. Section 2 describes the material and method used in

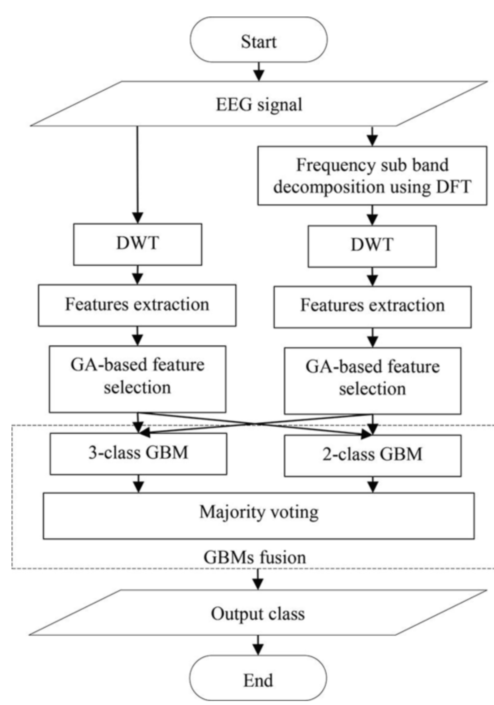

Fig. 1. The complete flowchart for the proposed method.

this study. The experimental result and discussion are provided in Section 3. Finally, the conclusion is drawn in Section 4.

2. Material and method

2.1. EEG dataset

This study used a public EEG signals dataset available from the Department of Epileptology University of Bonn (UoB), Germany, composed by Andrzejak et al. (2001). A 128-channel amplifier system with an average common reference method (Yao, 2001) was used to record all EEG signals. The analog signals were converted to a 12-bit digital and filtered with a bandpass filter setting in the range of 0.53 to 40 Hz before written to computer storage. The dataset initially consisted of five sets denoted by A, B, C, D, and E saved in five different ZIP files. Each ZIP file contained 100 TXT files that represent 100 single-channel EEG signals. Each TXT file had 4096 EEG data in time series with a duration of 23.6 s and a sampling rate of 173.61 Hz.

Set A and B consisted of normal EEG signals from five healthy people recorded in a relaxed state with their eyes open and closed, respectively. On the other hand, set C, D, and E consisted of EEG signals from five patients with epilepsy recorded during the evaluation process before surgery. EEG signals in sets C and D were collected from the hippocampal formation and within the epileptogenic zone, respectively, during seizure-free conditions (interictal period). In contrast, signals in set E were collected during seizure activity (ictal period). This study concentrated on detecting EEG signals to distinguish three types of signals, which are normal, interictal, and ictal. Therefore, this study only used EEG signals from set A (normal), D (inter ictal), and E (ictal) to evaluate the proposed epilepsy detection method. The example of EEG signals from set A, D, and E are depicted in Fig. 2.

2.2. Features extraction

2.2.1. Frequency sub-band decomposition

Discrete Fourier transform has been widely used to analyze a discrete signal, including EEG signal, in the frequency domain (Zhang, 2019). The proposed method employed DFT to decompose the original EEG signal into five frequency sub-bands, which are delta (bellow 4 Hz), theta (between 4 and 8 Hz), alpha (between 8 and 12 Hz), beta (between 12 and 30 Hz) and gamma (above 30 Hz). Suppose $x(n), n = 0, 1, 2, \cdots, N-1$ is a discrete EEG signal in time domain with N sample points. The EEG signal $x(n)$ was transformed into frequency domain using DFT as defined in Eq. (1),

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-\frac{2\pi n i}{N}k}, \quad k = 0, 1, 2, \cdots, N-1 \tag{1}$$

where $i = \sqrt{-1}$.

The transformed EEG signal $X(k)$ was filtered using some bandpass filters with different cutoff frequencies to produce $X_\delta(k), X_\theta(k), X_\alpha(k), X_\beta(k)$, and $X_\gamma(k)$ that correspond to delta, theta, alpha, beta, and gamma frequency subbands, respectively. $X_\delta(k), X_\theta(k), X_\alpha(k), X_\beta(k)$, and $X_\gamma(k)$ were then transformed back into the time domain to obtain decomposed EEG signals in the time domain $x_\delta(n), x_\theta(n), x_\alpha(n), x_\beta(n)$, and $x_\gamma(n)$ using inverse DFT as defined in Eq. (2)

$$x_b(n) = \frac{1}{N}\sum_{k=0}^{N-1} X_b(k)\, e^{\frac{2\pi k i}{N}n}, \quad n = 0, 1, \cdots, N-1, \quad b = \delta, \theta, \alpha, \beta, \gamma \tag{2}$$

This study used the implementation of the fast Fourier transform (FFT) algorithm in NumPy library (Harris et al., 2020) to calculate DFT as well as its inverse. Fig. 3 shows the example of decomposed EEG signal.

2.2.2. Discrete wavelet transform

Wavelet transform is a common technique used to analyze a signal both in time and frequency domains. The fundamental principle used in wavelet transform is to decompose the original signal into a series that is a linear combination of the family of functions. The family of functions is obtained by performing dilations and translations to a function in the different values of dilation and translation parameters. Discrete wavelet transform is a particular case of the wavelet transform that employs discrete values for both dilation and translation parameters (Tan and Jiang, 2019).

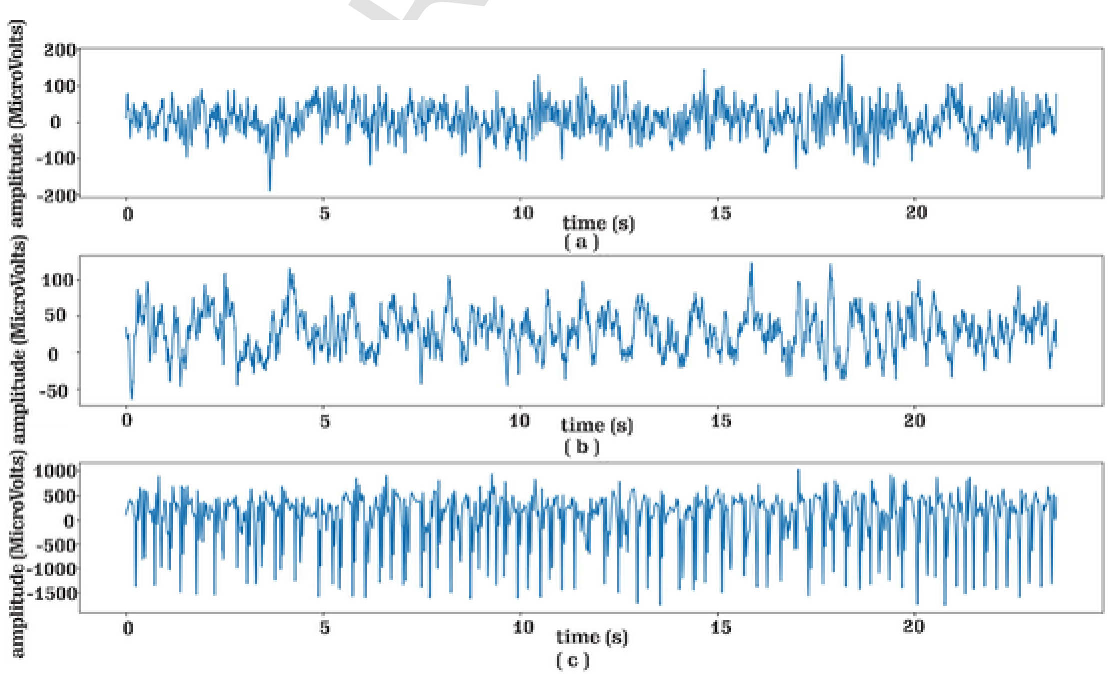

Fig. 2. The example of EEG signal from (a) set A, (b) set D, and (c) set E.

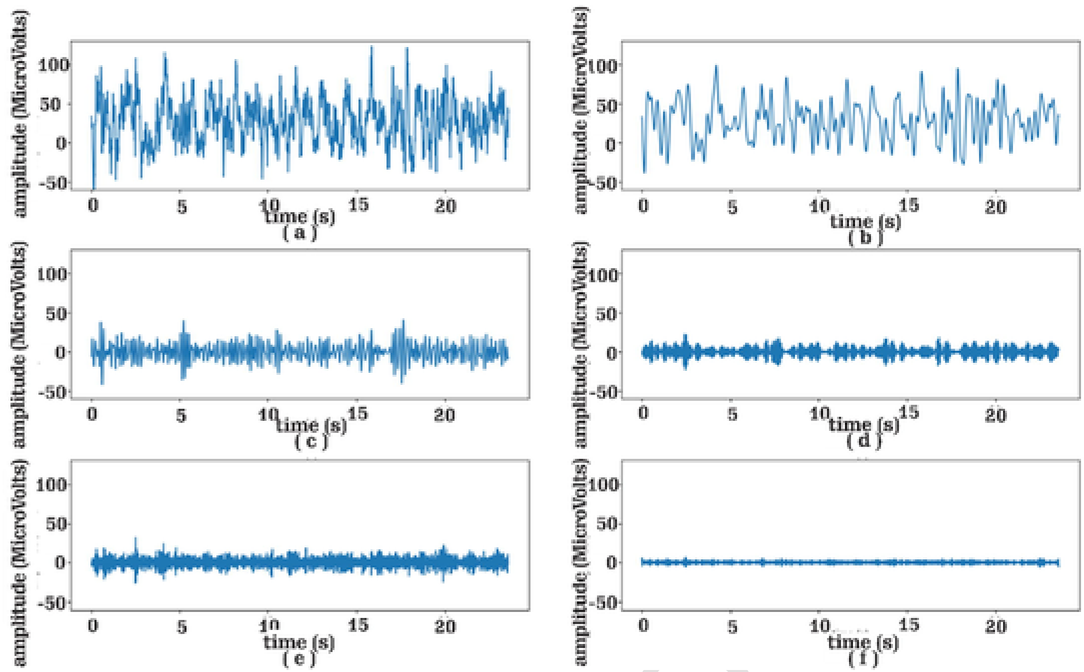

Fig. 3. The example of decomposed EEG signal: (a) original EEG signal, (b) delta subband, (c) theta subband, (d) alpha subband, (e) beta subband, and (f) gamma subband.

Suppose $x(t)$ is a signal in the time domain, $\{\varphi_{jk}(t)\}$ is the family of scalling functions, and $\{\psi_{jk}(t)\}$ is the family of wavelet functions. $\varphi_{jk}(t)$ and $\psi_{jk}(t)$ are obtained from father wavelet $\varphi(t)$ and mother wavelet $\psi(t)$ at scale j and translation k as defined in Eq. (3) and Eq. (4), respectively.

$$\varphi_{jk}(t) = 2^{\frac{j}{2}}\varphi\left(2^{j}t - k\right) \tag{3}$$

$$\psi_{jk}(t) = 2^{\frac{j}{2}}\psi\left(2^{j}t - k\right) \tag{4}$$

DWT with the decomposition level of L transforms the signal $x(t)$ in time domain into a set of coefficients $c_0(k)$ and $d_j(k), j = 0, 1, ..L-1$. The coefficients $c_j(k)$ and $d_j(k)$ are called the approximation coefficients and the detail coefficients, respectively. The DWT of $x(t)$ is defined using the inner product of $x(t), \varphi_{jk}(t)$, and $\psi_{jk}(t)$ as in Eq. (5) and Eq. (6).

$$c_j(k) = \langle x(t), \varphi_{jk}(t)\rangle = \int_{-\infty}^{\infty} x(t)\varphi_{jk}(t)\,dt \tag{5}$$

$$d_j(k) = \langle x(t), \psi_{jk}(t)\rangle = \int_{-\infty}^{\infty} x(t)\psi_{jk}(t)\,dt \tag{6}$$

The invers DWT reconstructs the signal $x(t)$ from the approximation coefficients and the detail coefficients based on wavelet expansion as defined in Eq. (7).

$$x(t) = \sum_{k=-\infty}^{\infty} c_0(k)\varphi_{0k}(t) + \sum_{j=0}^{L-1}\sum_{k=-\infty}^{\infty} d_j(k)\psi_{jk}(t) \tag{7}$$

This study applied DWT to the original EEG signal or the decomposed EEG signals obtained using frequency sub-band decomposition discussed in Section 2.2.1. Some wavelet families were employed in this study to process EEG signals using DWT, including Haar, Daubechies, Symlets, Coiflets, Biorthogonal, Reverse Biorthogonal, and Discrete Meyer. The decomposition level L was heuristically determined in the range of $[1, 12]$ during the experiment such that the best classification performance is achieved. The results of DWT were the coefficient vectors $c_0, d_0, d_1, \cdots, d_{L-1}$. This study used a Python library for wavelet analysis PyWavelets (Lee et al., 2019) to perform DWT. Fig. 4 shows the example of DWT results using Symlet family sym5 with decomposition level 2.

2.2.3. *Statistical features*

From each coefficient vector of DWT results, five statistical features were extracted, including the 5th percentile, the 25th percentile, the 50th percentile, the 75th percentile, and the 95th percentile of the coefficient vector. Therefore, there were $5(L+1)$ statistical features extracted from all coefficient vectors of DWT result for DWT with the decomposition level of L . Percentiles provide information related to the distribution of data by partitioning the data into 100 equal parts. The p^{th} percentile of data is a value such that the portion of data below the value is p%. The elements of the coefficient vector were ordered from smallest to largest to obtain the p^{th} percentile. The index n of the p^{th} percentile in the ordered coefficient vector was determined using Eq. (8),

$$n = \frac{p}{100}(N+1) \tag{8}$$

where N is the length of the coefficient vector. If n is an integer, the p^{th} percentile was the n^{th} element of the ordered coefficient vector. Otherwise, the p^{th} percentile was obtained using linear interpolation based on the fractional part of n , the n^{th} and the $(n+1)^{th}$ elements of the ordered coefficient vector (Filliben and Heckert, 2002). This study used the implementation of percentile in the NumPy library (Harris et al., 2020) to extract the statistical features.

2.2.4. *Crossing frequency features*

Zero-crossing rate (ZCR) is defined as the number of points at which a signal crosses horizontal axis within one second. It is a simple feature and can be used to capture the frequency information of signal. A high ZCR is related to a high frequency signal (Motamedi-Fakhr et al., 2014). This study extracted zero-crossing frequency (ZCF) from the coefficient vector of DWT result to replace ZCR since all EEG signals in the dataset have same duration and the calculation of ZCF is simpler compared to ZCR. ZCF is defined as the frequency of two successive elements of the coefficient vector cross zero or change in sign from negative to positive or vice versa. Therefore, ZCF was be calculated using Eq. (9),

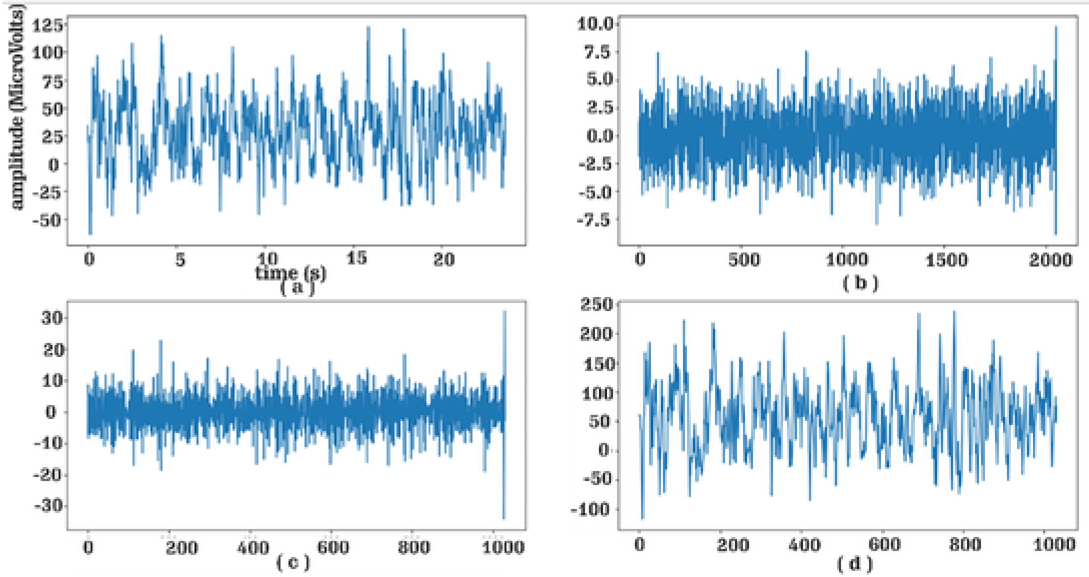

Fig. 4. The example of DWT results using Symlet family sym5 with decomposition level 2: (a) original EEG signal, (b) vector d_1 , (c) vector d_2 , and (d) vector c_0 .

$$ZCF = \frac{1}{2}\sum_{k=1}^{N-1} |\mathrm{sgn}\,(v\,(k+1)) - \mathrm{sgn}\,(v\,(k))| \tag{9}$$

where N is the length of the coefficient vector, $v(k)$ is the k^{th} element of the coefficient vector, and $\mathrm{sgn}\,(x)$ is the sign function defined as in Eq. (10).

$$sgn\,(x) = \begin{cases} \frac{|x|}{x}, x \neq 0 \\ 0, x = 0 \end{cases} \tag{10}$$

Sometimes, a signal is only located in either the above or the bellow of horizontal axis. In this condition ZCR cannot be used to capture the frequency information of the signal. Therefore, this study also extracted mean crossing frequency (MCF) (Myroniv et al., 2017) from the coefficient vector of DWT result as a complement to ZCF in capturing the frequency information of the signal. Suppose m is the mean of the coefficient vector, MCF is defined as the frequency of two successive elements of the vector cross m . MCF was calculated similarly as ZCF from the difference between the coefficient vector and m , as in Eq. (11).

$$MCF = \frac{1}{2}\sum_{k=1}^{N-1} |\mathrm{sgn}\,(v\,(k+1) - m) - \mathrm{sgn}\,(v\,(k) - m)| \tag{11}$$

In total, there were $2\,(L+1)$ crossing frequency features extracted from the coefficient vector of DWT result with the decomposition level of L .This study used some functions from the NumPy library (Harris et al., 2020) to extract ZCF and MCF.

2.3. Feature selection

A large number of features used in classification will increase the complexity of the training process. On the other hand, it does not always produce good classification performance (Siswantoro et al., 2020). One method that can be used to reduce the number of features is feature selection. In the classification problem, feature selection works by selecting the most discriminating subset of features. This study employed a genetic algorithm to perform feature selection. GA is an optimization technique that mimics the principles of natural evolution. GA-based feature selection has been proven to be able to eliminate irrelevant features and improve classification performance. In addition, the time required for feature selection can be reduced by employing GA (Li, 2006).

The GA-based feature selection procedure was started by generating the initial population randomly. The population consisted of chromosomes which are binary mask vectors with the length equal to the number of features. Each chromosome was composed of genes with a value 0 or 1. If the value of i^{th} gene was 1 then the i^{th} feature was selected for classification; otherwise, the feature was not considered. The quality of each chromosome was evaluated using a fitness function. In this study, the fitness function for GA-based feature selection was the classification accuracy of the classifier trained with the selected features that correspond to the chromosome. The population members were iteratively adjusted using three genetic operators, namely crossover, mutation, and selection, until stopping criteria are reached. The parameters of GA used in this study are tabulated in Table 2. This study used the implementation of GA in the DEAP library (Fortin et al., 2012) to perform GA-based feature selection.

2.4. Gradient boosting machine

Gradient boosting machine is a boosting algorithm that uses regression trees as base learners and can be applied in regression and classification problems. For a classification problem, the base learners were used to train a classifier iteratively using the negative gradient of a differentiable loss function. The classifier was expected to minimize the loss function based on training data (Friedman, 2001). Suppose $\{(\mathbf{x}_i, \mathbf{y}_i)\,|i = 1, 2, \cdots, N\}$ is the training data, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \cdots, x_{ip}]$ is the i^{th} input features, $\mathbf{y}_i = [y_{i1}, y_{i2}, \cdots, y_{iK}]$ is the i^{th} desired output, p is the dimension of input features, and K is the number of classes. Furthermore, $y_{ij} = 1$ if $\mathbf{x}_i$ belong to class j and otherwise $y_{ij} = 0$. For K-class

Table 2
The parameters of GA

Parameter	Value
Population size	Equal to the number of features
Number of generations	100
Crossover probability	0.5
Crossover method	Single point
Mutation probability	0.2
Mutation method	Flip bit
Selection	Tournament with size 3

classification problem, GBM used a negative multinomial log likelihood function as loss function as in Eq. (12),

$$L(\mathbf{y}, \mathbf{F}(\mathbf{x})) = -\sum_{j=1}^{K} \log \left(p_j(\mathbf{x})\right)^{y_j} \tag{12}$$

where $\mathbf{F}(\mathbf{x}) = \left[F_1(\mathbf{x}), F_2(\mathbf{x}), \cdots, F_K(\mathbf{x})\right]$ is a vector-valued function that maps the input features $\mathbf{x} = \left[x_1, x_2, \cdots x_p\right]$ to the desired output $\mathbf{y} = \left[y_1, y_2, \cdots, y_K\right]$ and $p_k(\mathbf{x})$ is the probability that $y_k = 1$ given the input features $\mathbf{x}$, which is defined using the softmax function as in Eq. (13).

$$p_k(\mathbf{x}) = \frac{e^{F_k(\mathbf{x})}}{\sum_{i=1}^{K} e^{F_i(\mathbf{x})}} \tag{13}$$

GBM aimed to find an approximation $\widehat{F}_X$ for F(x) that minimize $L(\mathbf{y}, \mathbf{F}(\mathbf{x}))$ iteratively. The iteration was performed M times with the initial value of $\mathbf{F}(\mathbf{x})$ was set to $\mathbf{F}_0(\mathbf{x}) = 0$. At the m^{th} iteration, the approximation $\mathbf{F}_{\mathrm{m}}(\mathbf{x}) = \left[F_{m1}(\mathbf{x}), F_{m2}(\mathbf{x}), \cdots, F_{mK}(\mathbf{x})\right]$ of $\mathbf{F}(\mathbf{x})$ was obtained by fitting K trees with T -terminal nodes, that correspond to regions $R_{mkt}, k = 1, 2, \cdots, K, t = 1, 2, \cdots, T$, with the input features $\mathbf{x}$ and the desired output $\mathbf{g}_m$, where $\mathbf{g}_m = \left[g_{m1}, g_{m2}, \cdots, g_{mK}\right]$ is negative the first partial derivative of $L(\mathbf{y}, \mathbf{F}(\mathbf{x}))$ with respect to $\mathbf{F}$ evaluated at $\mathbf{F}(\mathbf{x}) = \mathbf{F}_{m-1}(\mathbf{x})$, as defined in Eq. (14),

$$g_{mk} = -\left[\frac{\partial L(\mathbf{y}, \mathbf{F}(\mathbf{x}))}{\partial F_k(\mathbf{x})}\right]_{\mathbf{F}(\mathbf{x}) = \mathbf{F}_{m-1}(\mathbf{x})} = y_k - p_{(m-1)k}(\mathbf{x}) \tag{14}$$

and $p_{(m-1)k}(\mathbf{x})$ is $p_k(\mathbf{x})$ as in Eq. (13) evaluated at $\mathbf{F}(\mathbf{x}) = \mathbf{F}_{(m-1)}(\mathbf{x})$.

The regions R_{mkt} were then used to determine the update parameters α_{mkt} for $\mathbf{F}_{\mathrm{m}}(\mathbf{x})$ by solving the optimization problem in Eq. (15),

$$\alpha_{mkt} = \arg\min_{\alpha_{kt}} \sum_{i=1}^{N} \sum_{k=1}^{K} l\left(y_{ik}, F_{(m-1)k}(\mathbf{x}_i) + \sum_{t=1}^{T} \alpha_{kt} f_{R_{tkm}}(\mathbf{x}_i)\right) \tag{15}$$

where $l\left(y_{ik}, F_{mk}(\mathbf{x})\right) = -\log\left(p_{mk}(\mathbf{x})\right)^{y_{ik}}$ and $f_{R_{tkm}}(\mathbf{x})$ is an indicator function defined as in Eq. (16).

$$f_{R_{tkm}}(\mathbf{x}) = \begin{cases} 1, x \in R_{tkm} \\ 0, x \notin R_{tkm} \end{cases} \tag{16}$$

Finally, the approximation $\mathbf{F}_m(\mathbf{x})$ of $\mathbf{F}(\mathbf{x})$ at the m^{th} iteration was be obtained using Eq. (17),

$$F_{mk}(\mathbf{x}) = F_{(m-1)k}(\mathbf{x}) + \eta \sum_{t=1}^{T} \alpha_{mkt} f_{R_{tkm}}(\mathbf{x}) \tag{17}$$

where $\eta \in (0, 1]$ is the learning rate parameter.

After M^{th} iteration the approximation $\mathbf{F}_M(\mathbf{x})$ of $\mathbf{F}(\mathbf{x})$ was obtained. The probability that an unknown input $\mathbf{x}$ belongs to class k was calculated from $p_k(\mathbf{x})$ as in Eq. (13) evaluated at $\mathbf{F}(\mathbf{x}) = \mathbf{F}_M(\mathbf{x})$. The predicted class for $\mathbf{x}$ was then obtained by finding k that maximizes $p_k(\mathbf{x})$. This study used the implementation of GBM in the Scikit-learn library (Pedregosa et al., 2011). The number of iterations M was set to 100, and the learning rate parameter η was set to 0.1.

2.5. Gradient boosting machines fusion

This study proposed a GBMs fusion to improve the classification performance of GBM in classifying EEG signals into three classes, which are ictal (class 0), normal (class 1), and interictal (class 2). Some GBM classifiers were trained as base classifiers to classify the EEG signals into three classes and two classes. The following steps were used to perform training GBMs fusion and predict class label for unknown observation using GBMs fusion.

1. Decompose the original EEG signal using DFT to obtain the decomposed signals $x_\delta, x_\theta, x_\alpha, x_\beta$, and x_γ.
2. Apply DWT with the decomposition level of L to the original EEG signal to obtain the coefficient vectors $c_{o0}, d_{o0}, d_{o1}, \cdots, d_{o(L-1)}$.
3. Apply DWT with the decomposition level of L to the decomposed EEG signal to obtain the coefficient vectors $c_{s0}, d_{s0}, d_{s1}, \cdots, d_{s(L-1)}$, for $s = \delta, \theta, \alpha, \beta, \gamma$.
4. Extract $5(L+1)$ statistical features and $2(L+1)$ crossing frequency features from the coefficient vectors $c_{o0}, d_{o0}, d_{o1}, \cdots, d_{o(L-1)}$ to obtain feature sets $\mathcal{F}_1$.
5. Extract 2 $5(L+1)$ statistical features and $10(L+1)$ crossing frequency features from the coefficient vectors $c_{s0}, d_{s0}, d_{s1}, \cdots, d_{s(L-1)}$, for $s = \delta, \theta, \alpha, \beta, \gamma$ to obtain feature sets $\mathcal{F}_2$.
6. Apply GA-based feature selection to $\mathcal{F}_1$ and $\mathcal{F}_2$ to select some important features.
7. Train two 3-class GBMs, called $\mathcal{G}_{012}^1$ and $\mathcal{G}_{012}^2$, using the selected features from $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, as input features.
8. Train six 2-class GBMs using the selected features from $\mathcal{F}_1$ and $\mathcal{F}_2$ to classify the EEG signal into class 0 and class 1 (called $\mathcal{G}_{01}^1$ and $\mathcal{G}_{01}^2$, respectively), class 0 and class 2 (called $\mathcal{G}_{02}^1$ and $\mathcal{G}_{02}^2$, respectively), or class 1 and class 2 (called $\mathcal{G}_{12}^1$ and $\mathcal{G}_{12}^2$, respectively).
9. The GMBs fusion was performed as follows. Suppose x_0 is an unlabeled EEG signal.
   a. Predict the class label of x_0 using $\mathcal{G}_{012}^1$ and $\mathcal{G}_{012}^2$ to obtain $y_{012}^1 = \mathcal{G}_{012}^1(x_0)$ and $y_{012}^2 = G_{012}^2(x_0)$, respectively.
   b. If $y_{012}^i = 0$ then further predict the class label of x_0 using $\mathcal{G}_{01}^1$ or $\mathcal{G}_{01}^2$ and $\mathcal{G}_{02}^1$ or $\mathcal{G}_{02}^2$ to obtain $y_{01}^i = \mathcal{G}_{01}^1(x_0)$ or $y_{01}^i = \mathcal{G}_{01}^2(x_0)$ and $y_{02}^i = \mathcal{G}_{02}^1(x_0)$ or $y_{02}^i = \mathcal{G}_{02}^2(x_0)$, for $i = 1, 2$.
   c. If $y_{012}^i = 1$ then further predict the class label of x_0 using $\mathcal{G}_{01}^1$ or $\mathcal{G}_{01}^2$ and $\mathcal{G}_{12}^1$ or $\mathcal{G}_{12}^2$ to obtain $y_{01}^i = \mathcal{G}_{01}^1(x_0)$ or $y_{01}^i = \mathcal{G}_{01}^2(x_0)$ and $y_{12}^i = \mathcal{G}_{12}^1(x_0)$ or $y_{12}^i = \mathcal{G}_{12}^2(x_0)$, for $i = 1, 2$.
   d. If $y_{012}^i = 2$ then further predict the class label of x_0 using $\mathcal{G}_{02}^1$ or $\mathcal{G}_{02}^2$ and $\mathcal{G}_{12}^1$ or $\mathcal{G}_{12}^2$ to obtain $y_{02}^i = \mathcal{G}_{02}^1(x_0)$ or $y_{02}^i = \mathcal{G}_{02}^2(x_0)$ and $y_{12}^i = \mathcal{G}_{12}^1(x_0)$ or $y_{12}^i = \mathcal{G}_{12}^2(x_0)$, for $i = 1, 2$.
10. The final predicted class of x_0 was obtained by majority voting from the output of 3-class GBMs and 2-class GBMs, as in Eq. (18),

$$class(x_0) = \arg\max_{k \in \{0,1,2\}} \left(\sum_{i=1}^{2} I_k\left(y_{012}^i\right) + \sum_{\substack{s \in \{01, 02\} \\ \text{or } s \in \{01, 12\} \\ \text{or } s \in \{02, 12\}}} \sum_{i=1}^{2} I_k\left(y_s^i\right) \right) \tag{18}$$

where $I_k(y)$ is an indicator function defined as in Eq. (19).

$$I_k(y) = \begin{cases} 1, y = k \\ 0, y \neq k \end{cases} \tag{19}$$

The using $\mathscr{G}_s^1$ and $\mathscr{G}_s^2$ for $s \in \{01, 02\}$ or $s \in \{01, 12\}$ or $s \in \{02, 12\}$ in step 9b – 9d were heuristically determined such that the best classification performance is achieved. The steps of the proposed GBMs fusion are summarized in Algorithm 1.

Algorithm 1. Algorithm for the proposed GBMs fusion

Input: the training data set $\mathbf{X}$ and $\mathbf{y}$, an unknown EEG signal x_0 .
Output: the predicted class of x_0
Perform frequency sub-band decomposition for each $x \in \mathbf{X}$ to obtain $x_\delta, x_\theta, x_\alpha, x_\beta$, and x_γ
Perform $L-$ level DWT for each $x \in \mathbf{X}$ to obtain $c_{o0}, d_{o0}, d_{o1}, \cdots, d_{o(L-1)}$
Perform $L-$ level DWT for $x_\delta, x_\theta, x_\alpha, x_\beta$, and x_γ to obtain $c_{s0}, d_{s0}, d_{s1}, \cdots, d_{s(L-1)}$, for $s = \delta, \theta, \alpha, \beta, \gamma$
Extract statistical features and crossing frequency features from $c_{o0}, d_{o0}, d_{o1}, \cdots, d_{o(L-1)}$ to obtain $\mathscr{F}_1$
Extract statistical features and crossing frequency features from $c_{s0}, d_{s0}, d_{s1}, \cdots, d_{s(L-1)}$, for $s = \delta, \theta, \alpha, \beta, \gamma$ to obtain $\mathscr{F}_2$
Perform GA-based feature selection on $\mathscr{F}_1$ and $\mathscr{F}_2$
for $i \in \{1, 2\}$
 Train 3 – class GBM using $\mathscr{F}_i$ and $\mathbf{y}$ to obtain $\mathscr{G}_{012}^i$
  for $s \in \{01, 02, 12\}$
   Train 2 – class GBM using $\mathscr{F}_i$ and $\mathbf{y}$ to obtain $\mathscr{G}_s^i$
  end for
end for
Predict $y_{012}^1 = \mathscr{G}_{012}^1(x_0)$
Predict $y_{012}^2 = \mathscr{G}_{012}^2(x_0)$
for $i \in \{1, 2\}$
 if $y_{012}^i = 0$
  Predict $y_{01}^i = \mathscr{G}_{01}^1(x_0)$ or $y_{01}^i = \mathscr{G}_{01}^2(x_0)$
  Predict $y_{02}^i = \mathscr{G}_{02}^1(x_0)$ or $y_{02}^i = \mathscr{G}_{02}^2(x_0)$
 else if $y_{012}^i = 1$
  Predict $y_{01}^i = \mathscr{G}_{01}^1(x_0)$ or $y_{01}^i = \mathscr{G}_{01}^2(x_0)$
  Predict $y_{12}^i = \mathscr{G}_{12}^1(x_0)$ or $y_{02}^i = \mathscr{G}_{12}^2(x_0)$
 else
  Predict $y_{02}^i = \mathscr{G}_{02}^1(x_0)$ or $y_{02}^i = \mathscr{G}_{02}^2(x_0)$
  Predict $y_{12}^i = \mathscr{G}_{12}^1(x_0)$ or $y_{02}^i = \mathscr{G}_{12}^2(x_0)$
 end if
end for
Determine the predicted class of x_0 using Eq. (18)

Fig. 5 shows the toy example for the proposed GBMs fusion. The example contains three classes: class 0, class 1, and class 2. First, the features set $\mathscr{F}_1$ and $\mathscr{F}_2$ are extracted from each EEG signals in the training data. Each features set is then selected using GA-based feature selection to obtain the important features. Two 3-class GBMs are trained using the selected features from $\mathscr{F}_1$ and $\mathscr{F}_2$, called $\mathscr{G}_{012}^1$ and $\mathscr{G}_{012}^2$, respectively to classify the EEG signal into class 0, class 1, and class 2. Furthermore, six 2-class GBMs are trained using the selected features from $\mathscr{F}_1$ and $\mathscr{F}_2$ to classify the EEG signal into class 0 and class 1 (called $\mathscr{G}_{01}^1$ and $\mathscr{G}_{01}^2$, respectively), class 0 and class 2 (called $\mathscr{G}_{02}^1$ and $\mathscr{G}_{02}^2$, respectively), or class 1 and class 2 (called $\mathscr{G}_{12}^1$ and $\mathscr{G}_{12}^2$, respectively). Suppose x_0 is an unknown EEG signal that to be classified. Classify x_0 using $\mathscr{G}_{012}^1$ and $\mathscr{G}_{012}^2$ and suppose the output classes are class 0 and class 2, respectively. Further classify x_0 using $\mathscr{G}_{01}^1$ or $\mathscr{G}_{01}^2$, $\mathscr{G}_{02}^1$ or $\mathscr{G}_{02}^2$, and $\mathscr{G}_{12}^1$ or $\mathscr{G}_{12}^2$ and suppose the output classes are class, 0, class 0, and class 2. The final class of x_0 is obtained by majority voting from all outputs which is class 0.

2.6. Experimental setup

An experiment has been conducted in the laboratory to validate the proposed method in classifying EEG signals to detect epilepsy with three experimental scenarios. In the first scenario, two 3-class GBMs were trained with $\mathscr{F}_1$ and $\mathscr{F}_2$ to classify EEG signals into three classes, which are class 0, class 1, and class 2. In the second scenario, six 2-class GBMs were trained with $\mathscr{F}_1$ and $\mathscr{F}_2$ to classify EEG signals into two classes, which are class 0 and class 1, class 0 and class 2, and class 1 and

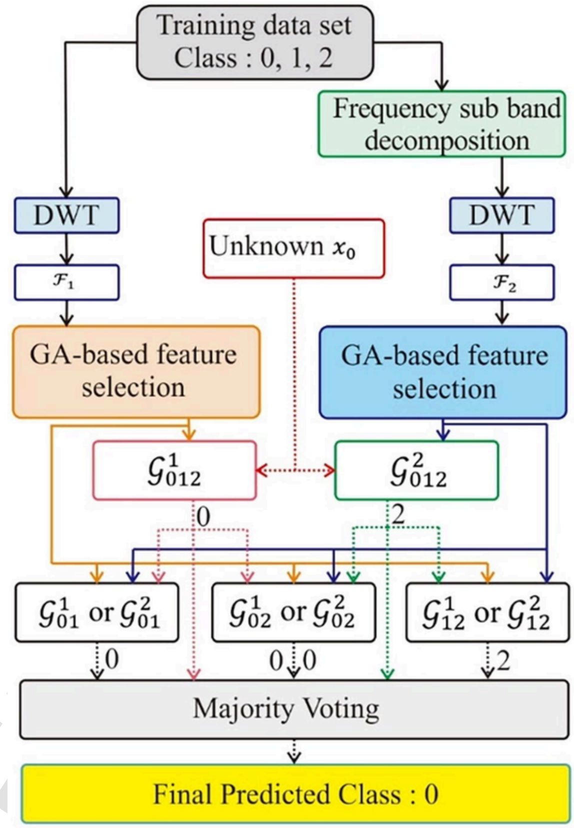

Fig. 5. The toy example for the proposed GBMs fusion.

class 0. In the last scenario, the proposed GBMs fusion was used to classify EEG signals into three classes. In first two scenarios, more than 700 experiments have been carried out to obtain the best classification performance by combining some wavelet families and some decomposition levels and 24,999 experiments in the last scenario. Furthermore, in the first and the second scenarios, all GBMs were trained using the original features and the selected features. On the other hand, in the last scenario, the GBMs fusion was only used the selected features.

The experiment was performed on a high-end computer with specification 3.60 GHz Intel(R) Core (TM) i9-9900 K, 32 GB RAM, NVIDIA GeForce GTX1080Ti GPU, and Ubuntu Linux 16.04.1 operating system. The proposed method was implemented using Python language programming with some Python libraries, including Numpy (Harris et al., 2020), PyWavelets (Lee et al., 2019), DEAP (Fortin et al., 2012), and Scikit-learn (Pedregosa et al., 2011). To ensure that the proposed method can be used in the real clinical applications, the experiment was also performed using a low-end computer with specification 1.40 GHz Intel(R) Core (TM) i3-2367 M, 8 GB RAM, and Ubuntu Linux 16.04.1 operating system.

This study used 10-fold cross-validation to evaluate the performance of the proposed method. The cross-validation worked by splitting the EEG dataset randomly into ten mutually exclusive subsets with the same cardinality, such that all classes have the same proportion in every subset. The training and testing processes were iteratively performed ten times. Each subset was used as the testing data once and the remaining subsets as the training data in each iteration (Alpaydin, 2014). For i^{th} testing data, the accuracy acc_i of the proposed method was measured using Eq. (20),

$$acc_i = \frac{N_{ci}}{N_i} 100\%, \quad i = 1, 2, \cdots, 10 \tag{20}$$

where N_{ci} is the number of EEG signal in i^{th} testing data that correctly classified and N_i is the cardinality of i^{th} testing data. The performance of the proposed method was then obtained from the average of acc_i for $i = 1, 2, \cdots, 10$.

3. Results and discussion

3.1. Classification results of three classes GBM

The classification accuracies of 3-class GBM trained with $\mathscr{F}_1$ and $\mathscr{F}_2$, which are $\mathscr{G}^1_{012}$ and $\mathscr{G}^2_{012}$, are summarized in Table 3. As can be seen in Table 3, the classification accuracy of $\mathscr{G}^1_{012}$ was ranging from 98.67% to 99.67%, with an average of 96.69% when all features in $\mathscr{F}_1$ were used as input features. After applying feature selection, the classification accuracy of $\mathscr{G}^1_{012}$ increased up to 3.67% on the most (719/767) experiments, ranging from 94.67% to 99.67% with an average of 97.84%. The number of experiments in Table 3 came from the number of combinations between the type of wavelet family and the decomposition level applied in 3-class GBM. It can be observed from Table 3 that, on average, the classification accuracy of $\mathscr{G}^2_{012}$ was slightly lower than $\mathscr{G}^1_{012}$. Before feature selection, the accuracy was ranging from 92.67% to 98.00% with an average of 95.39% and improved up to 6.00%, ranging from 96.00% to 99.33% with an average of 97.79%, after applying feature selection on all experiments.

Table 3
The summary of 3-class GBM classification results

Classifier	Features	Number of experiments	Accuracy (%)		
			Min	Max	Mean ± Std. dev.
$\mathscr{G}^1_{012}$	Original	767	93.00	98.67	96.69 ± 0.88
$\mathscr{G}^1_{012}$	Selected	767	94.67	99.67	97.84 ± 0.88
$\mathscr{G}^2_{012}$	Original	768	92.67	98.00	95.39 ± 0.90
$\mathscr{G}^2_{012}$	Selected	768	96.00	99.33	97.79 ± 0.58

Table 4
The best five of 3-class GBM classification results

Wavelet		Original features		Selected features	
Family	Level	Number of features	Accuracy (%)	Number of features	Accuracy (%)
Classifier $\mathscr{G}^1_{012}$					
sym5	6	49	98.33	21	99.67
sym5	8	63	97.00	32	99.67
coif4	4	35	97.33	17	99.33
bior5.5	5	42	97.33	22	99.33
rbio2.6	7	56	97.33	26	99.33
Classifier $\mathscr{G}^2_{012}$					
bior3.3	6	245	96.67	109	99.33
db7	7	280	96.33	128	99.33
bior3.7	8	315	95.33	145	99.33
bior2.4	2	105	96.67	50	99.00
bior2.6	3	140	96.33	56	99.00

The best five of 3-class GBM classification results are tabulated in Table 4. The highest classification accuracy of $\mathscr{G}^1_{012}$ after feature selection, which is 99.67%, was obtained by applying Symlets 5 (sym5) with the decomposition level of 6 and 8 on DWT, as shown in Table 4. However, sym5 with the decomposition level of 6 needed fewer selected features (21 features) compared to sym5 with the decomposition level of 8 (32 features) to achieve the best accuracy. For $\mathscr{G}^2_{012}$, the highest accuracy of 99.33% was achieved after feature selection by applying Biorthogonal 3.3 (bior3.3) with decomposition level of 6, Daubechies 7 (db7) with decomposition level of 7, and Biorthogonal 3.7 (bior3.7) with decomposition level of 8 on DWT. It can be seen from Table 4 that bior3.3, with the decomposition level of 6 needed fewer selected features (109 features) compared to other wavelets to achieve the best accuracy. These results indicate that increasing the decomposition level of DWT does not guarantee to improve the classification accuracy of 3-class GBM. The experimental results also showed that performing frequency sub-band decomposition prior DWT to produce features set $\mathscr{F}_2$ does not always have an impact on improving the classification accuracy of 3-class GBM. Furthermore, more selected features were needed from $\mathscr{F}_2$ to achieve the best accuracy of 3-class GBM compared to $\mathscr{F}_1$.

For further analysis, confusion matrices were used to investigate EEG signals that were misclassified by $\mathscr{G}^1_{012}$ using sym5 wavelet with the decomposition level of 6 and $\mathscr{G}^2_{012}$ using bior3.3 wavelet with the decomposition level of 6 after feature selection, as shown in Fig. 6. As can be seen in Fig. 6(a), all EEG signals from class 0 and class 1 were correctly classified by $\mathscr{G}^1_{012}$. There was only one EEG signal from class 2 misclassified by $\mathscr{G}^1_{012}$ as an EEG signal from class 0. On the other hand, only EEG signals from class 1 were correctly classified by $\mathscr{G}^2_{012}$. There were two EEG signals misclassified by $\mathscr{G}^2_{012}$, one from class 0 classified as class 2 and conversely, as shown in Fig. 6(b). The misclassified signals from class 0 and class 2 came from file S016.txt and file F060.txt, respectively.

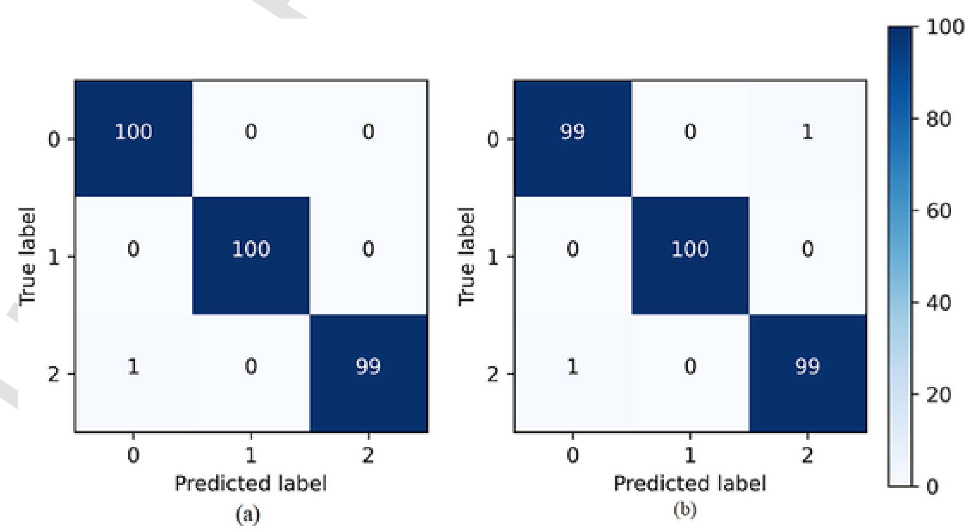

Fig. 6. The confusion matrices for (a) $\mathscr{G}^1_{012}$ using sym5 with the decomposition level of 6 and (b) $\mathscr{G}^2_{012}$ using bior3.3 with the decomposition level of 6 after feature selection.

3.2. Classification result of two classes GBM

The classification accuracies of 2-class GBM trained with $\mathscr{F}_1$ and $\mathscr{F}_2$, which are $\mathscr{G}^1_{01}, \mathscr{G}^2_{01}, \mathscr{G}^1_{02}, \mathscr{G}^2_{02}, \mathscr{G}^1_{12}$ and $\mathscr{G}^2_{12}$, are summarized in Table 5. The number of experiments in Table 5 came from the number of combinations between the type of wavelet family and the decomposition level applied in 2-class GBM. Before applying feature selection, on average the classification accuracies of GMB trained with $\mathscr{F}_1$ ($\mathscr{G}^1_{01}$ and $\mathscr{G}^1_{02}$) were lower than the classification accuracies of GMB trained with $\mathscr{F}_2$ ($\mathscr{G}^2_{01}$ and $\mathscr{G}^2_{02}$) for class 0 vs. class 1 and class 0 vs. class 2. On the other hand, the classification accuracy of GMB trained with $\mathscr{F}_1$ ($\mathscr{G}^1_{12}$) was better than the classification accuracy of GMB trained with $\mathscr{F}_2$ ($\mathscr{G}^2_{12}$) for class 1 vs. class 2. However, before applying feature selection, $\mathscr{G}^1_{01}$ achieved maximum accuracy of 100% for class 0 vs. class 1. After applying feature selection, on average the classification accuracy of 2-class GMB improved for all cases, as shown in Table 5. Furthermore, on average the classification accuracies of GBM trained with $\mathscr{F}_1$ were slightly better than the classification accuracies of GMB trained with $\mathscr{F}_2$ for all cases.

The best five of 2-class GBM classification results with smallest selected features are tabulated in Table 6, Table 7, and Table 8. For class 0 vs. class 1, both $\mathscr{G}^1_{01}$ and $\mathscr{G}^2_{01}$ achieved the highest accuracy of 100% after feature selection. There were 418 experiments produced the accuracy of 100% for $\mathscr{G}^1_{01}$ and 78 experiments for $\mathscr{G}^2_{01}$. As can be seen in Table 6, the best accuracy of $\mathscr{G}^1_{01}$ was achieved with the smallest selected features of four using db24 with the decomposition level of 2 on DWT, followed by five selected features using Coiflets 1 (coif1) with the decomposition level of 1, and seven selected features using bior3.3 and bior3.9 with the decomposition level of 1 and db35 with the decomposition level of 2. For $\mathscr{G}^2_{01}$, the best accuracy was achieved with the smallest selected features of 48 using rbio3.9 with the decomposition level of 2 on DWT, followed by 50 selected features using db13 with the decomposition level of 2, 51 selected features using sym15 with the decomposition level of 2, 64 selected features using coif7 with the decomposition level of 3 and 68 selected features using db15 with the decomposition level of 4.

For class 0 vs. class 2, $\mathscr{G}^1_{02}$ achieved the highest accuracy of 99.00% and $\mathscr{G}^2_{02}$ achieved the highest accuracy of 99.50% after feature selection. There were 49 experiments produced the accuracy of 99.00% for $\mathscr{G}^1_{02}$ and 47 experiments for $\mathscr{G}^2_{02}$. As can be seen in Table 7, the best accuracy of $\mathscr{G}^1_{02}$ was achieved with the smallest selected features of five using rbio1.1 with the decomposition level of 1 on DWT, followed by eight selected features using rbio2.6 with the decomposition level of 3, and nine selected features using db1 with the decomposition level of 2 and sym10 with the decomposition level of 3 and 10 selected features using bior1.1 with the decomposition level of 3. For $\mathscr{G}^2_{02}$, the best accuracy was achieved with 71 selected features using sym20 with the decomposition level of 3 on DWT.

Fig. 7 shows the confusion matrices for $\mathscr{G}^1_{02}$ using rbio1.1 with the decomposition level of 1 and $\mathscr{G}^2_{02}$ using sym20 with the decomposition level of 3 after feature selection. As can be seen in Fig. 7(a), there were one EEG signal from class 0 misclassified as an EEG signal from class 2 and one EEG signal from class 2 misclassified as an EEG signal from

Table 5
The summary of 2-class GBM classification results

Classifier	Features	Number of experiments	Accuracy (%)		
			Min	Max	Mean ± Std. dev
$\mathscr{G}^1_{01}$	Original	783	92.00	100.00	99.28 ± 1.05
$\mathscr{G}^1_{01}$	Selected	783	98.00	100.00	99.67 ± 0.40
$\mathscr{G}^2_{01}$	Original	767	96.50	99.50	99.44 ± 0.23
$\mathscr{G}^2_{01}$	Selected	767	98.00	100.00	99.52 ± 0.21
$\mathscr{G}^1_{02}$	Original	767	93.50	99.00	96.47 ± 0.96
$\mathscr{G}^1_{02}$	Selected	767	95.00	99.00	97.86 ± 0.76
$\mathscr{G}^2_{02}$	Original	767	93.00	98.67	96.69 ± 0.88
$\mathscr{G}^2_{02}$	Selected	767	94.67	99.67	97.84 ± 0.88
$\mathscr{G}^1_{12}$	Original	781	91.50	99.50	96.27 ± 1.50
$\mathscr{G}^1_{12}$	Selected	781	95.00	100.00	98.57 ± 0.89
$\mathscr{G}^2_{12}$	Original	766	88.50	99.50	94.97 ± 1.61
$\mathscr{G}^2_{12}$	Selected	766	94.50	100.00	98.25 ± 0.78

Table 6
The best five of 2-class GBM (class 0 vs. class 1) classification results

Wavelet		Original features		Selected features	
Family	Level	Number of features	Accuracy (%)	Number of features	Accuracy (%)
Classifier $\mathscr{G}^1_{01}$					
db24	2	21	100.00	4	100.00
coif1	1	14	99.00	5	100.00
bior3.3	1	14	99.00	7	100.00
bior3.9	1	14	99.00	7	100.00
db35	2	21	99.50	7	100.00
Classifier $\mathscr{G}^2_{01}$					
rbio3.9	2	105	99.50	48	100.00
db13	2	105	99.50	50	100.00
sym15	2	105	99.50	51	100.00
coif7	3	140	99.50	64	100.00
db15	4	175	99.50	68	100.00

Table 7
The best five of 2-class GBM (class 0 vs. class 2) classification results

Wavelet		Original features		Selected features	
Family	Level	Number of features	Accuracy (%)	Number of features	Accuracy (%)
Classifier $\mathscr{G}^1_{02}$					
rbio1.1	1	14	97.00	5	99.00
rbio2.6	3	28	97.50	8	99.00
db1	2	21	97.00	9	99.00
sym10	3	28	96.50	9	99.00
bior1.1	3	28	96.50	10	99.00
Classifier $\mathscr{G}^2_{02}$					
sym20	3	140	95.50	71	99.50
db27	1	70	96.00	31	99.00
db38	1	70	96.50	31	99.00
sym16	1	70	97.00	33	99.00
db32	1	70	97.00	40	99.00

Table 8
The best five of 2-class GBM (class 1 vs. class 2) classification results

Wavelet		Original features		Selected features	
Family	Level	Number of features	Accuracy (%)	Number of features	Accuracy (%)
Classifier $\mathscr{G}^1_{12}$					
db38	1	14	99.00	7	100.00
coif13	3	28	95.50	11	100.00
coif9	3	28	98.00	12	100.00
sym19	3	28	96.50	12	100.00
db24	4	35	95.00	13	100.00
Classifier $\mathscr{G}^2_{12}$					
db37	1	70	97.00	37	100.00
db21	3	140	99.50	65	100.00
db17	5	210	95.00	102	100.00
db17	6	245	95.00	125	100.00
sym13	6	245	93.00	129	100.00

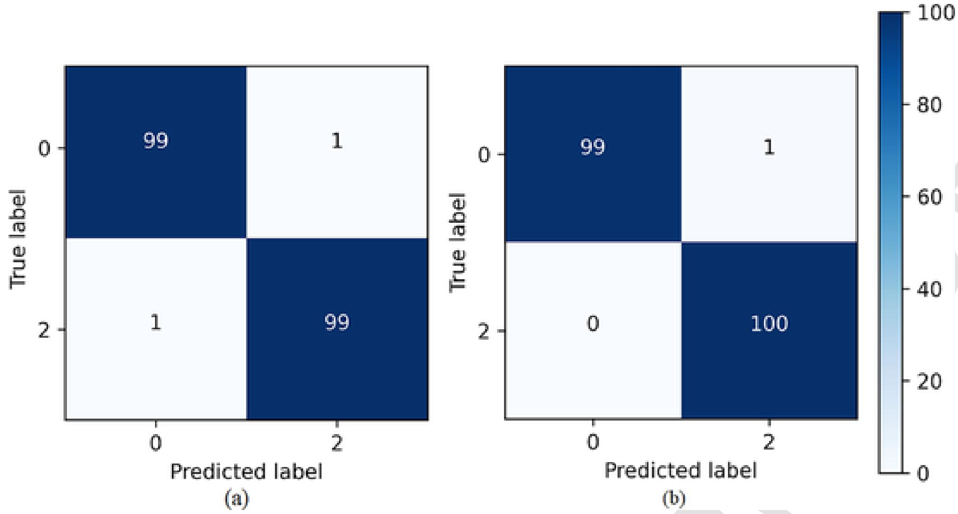

Fig. 7. The confusion matrices for (a) $\mathscr{G}^1_{02}$ using rbio1.1 with the decomposition level of 1 and (b) $\mathscr{G}^2_{02}$ using sym20 with the decomposition level of 3 after feature selection.

class 0 by $\mathscr{G}^1_{02}$. The signals misclassified by $\mathscr{G}^1_{02}$ were the same signals that misclassified by 3-class GMB which are from file S016.txt and file F060.txt. For $\mathscr{G}^2_{02}$, all EEG signals from class 2 were correctly classified, but there was one EEG signals from class 0 misclassified by $\mathscr{G}^2_{02}$ as an EEG signals from class 2, as shown in Fig. 7(b). The signal misclassified by $\mathscr{G}^1_{02}$ was from file S096.txt.

For class 1 vs. class 2, both $\mathscr{G}^1_{12}$ and $\mathscr{G}^2_{12}$ also achieved the highest accuracy of 100% after feature selection. There were 37 experiments produced the accuracy of 100% for $\mathscr{G}^1_{12}$ and five experiments for $\mathscr{G}^2_{12}$. As can be seen in Table 8, the best accuracy of $\mathscr{G}^1_{12}$ was achieved with the smallest selected features of seven using db38 with the decomposition level of 1 on DWT, followed by 11 selected features using coif13 with the decomposition level of 3, and 12 selected features using coif9 and sym19 with the decomposition level of 3 and 13 selected features using db24 with the decomposition level of 4. For $\mathscr{G}^2_{12}$, the best accuracy was achieved with the smallest selected features of 37 using db37 with the decomposition level of 1 on DWT, followed by 65 selected features using db23 with the decomposition level of 3, 102 selected features using db17 with the decomposition level of 5, 125 selected features using db17 with the decomposition level of 6 and 129 selected features using sym13 with the decomposition level of 6. From Table 6, Table 7, and Table 8, it can also be observed that the decomposition level of DWT does not correlate with the classification accuracy of 2-class GBM as in 3-class GBM. Moreover, similar to 3-class GBM, 2-class GBM needed more selected features from $\mathscr{F}_2$ to achieve the best accuracy compared to $\mathscr{F}_1$.

3.3. Statistical analysis for feature selection results

For further analysis, paired *t*-test with level of significant $\alpha = 0.05$ (Walpole et al., 2012) was performed to show that the mean classification accuracies of 3-class and 2 class GBMs using the selected features and the original features are significantly different. The null and alternative hypotheses of the test are formulated as follow,

$$H_0 : \mu_D = 0 vs. H_1 : \mu_D \neq 0 \tag{21}$$

where μ_D is the mean of differences between the classification accuracies of GBM using the selected features and the original features. The results of the paired *t*-test are tabulated in Table 9. The number of experiments in Table 9 came from the number of combinations between the type of wavelet family and the decomposition level applied in 3-class GBM or 2-class GBM.

Table 9
The results of paired *t*-test for the mean of differences between the classification accuracies of GBM using the original features and the selected features

Classifier	Number of experiments	Paired accuracy differences	95% confidence interval for mean difference		p-value
		Mean ± Std. dev. (%)	Lower	Upper	
$\mathscr{G}^1_{012}$	767	1.15 ± 0.71	1.10	1.20	0.00
$\mathscr{G}^2_{012}$	768	2.40 ± 0.83	2.34	2.46	0.00
$\mathscr{G}^1_{01}$	767	0.39 ± 1.01	0.32	0.46	0.00
$\mathscr{G}^2_{01}$	767	0.08 ± 0.30	0.06	0.10	0.00
$\mathscr{G}^1_{02}$	767	1.40 ± 0.87	1.33	1.46	0.00
$\mathscr{G}^2_{02}$	767	2.10 ± 1.00	2.03	2.17	0.00
$\mathscr{G}^1_{12}$	781	2.30 ± 1.68	2.18	2.41	0.00
$\mathscr{G}^2_{12}$	766	3.28 ± 1.53	3.17	3.39	0.00

As can be seen in Table 9, all p-values of the test were less than 0.05. This condition indicates that the hypothesis null, H_0 , should be rejected. Therefore, it can be concluded that the mean classification accuracies of GBMs using the selected features and the original features are significantly different. From Table 9, it can also be seen that the mean of paired accuracy differences for all 3-class and 2-class GBMs were greater than zero, ranging from 0.08% to 3.28%, with standard deviations ranging from 0.30% to 1.68%. In addition, all the lower and upper bound of the 95% confidence intervals for mean difference were greater than zero. These results indicate that all μ_D s are greater than zero for all classifiers. Therefore, it can be inferred that feature selection using genetic algorithm can improve the performance of GBM. Furthermore, the experimental results show that genetic algorithm can reduce the number of features up to 80.95% or 50.03% on average. In term of the type of feature, on average statistical features were more selected compared to crossing frequency features.

3.4. Classification result of GBMs fusion

The classification accuracies of GBMs fusion are summarized in Table 10. The total number of experiments in Table 10 came from the number of combinations between the type of wavelet family and the decomposition level applied in all base classifiers. From 24,999 experiments, GBMs fusion produced higher mean accuracy compared to a sin-

Table 10
The summary of GBMs fusion classification results

Number of experiments	Accuracy (%)
1996	100.00
11,845	99.67
10,863	99.33
287	99.00
8	98.67
Total: 24,999	Mean: 99.45
	Std. dev: 0.22

gle 3-class GBM, which is 99.54% with standard deviation of 0.22%. The maximum accuracy was 100.00%, achieved in 1996 experiments. Furthermore, the remaining experiments produced classification accuracies of 99.67% in 11,845 experiments, 99.33% in 10,863 experiments, 99.00% in 287 experiments, and 98.67% in 8 experiments. These results indicate that GBMs fusion can improve the performance of a single GBM in classifying EEG signals.

Some examples of GBMs fusion classification results with the best accuracy of 100% are shown in Table 11. To achieve the accuracy of 100%, GBMs fusion employed five base classifiers, which are $\mathscr{G}^1_{012}$, $\mathscr{G}^2_{012}$, $\mathscr{G}^1_{01}$ or $\mathscr{G}^2_{01}$, $\mathscr{G}^1_{02}$ or $\mathscr{G}^2_{02}$, and $\mathscr{G}^1_{02}$ or $\mathscr{G}^2_{12}$, as shown in Table 11. The experimental results show that coif4 with the decomposition level of 4, bior5.5 with the decomposition level of 5, sym5 with the decomposition level of 6 and 8, and rbio2.6 with the decomposition level of 7 were used on DWT by $\mathscr{G}^1_{012}$ to obtain the best accuracy. On the other hand, $\mathscr{G}^2_{012}$ needed to use db7 with the decomposition level of 7 on DWT to obtain the best accuracy. For 2-class base classifiers, the wavelet family and the decomposition level used to obtain the best accuracy were varying but generally came from the combination of the wavelet family and the decomposition level that produce the best accuracy in 2-class GBM, as can be seen in Table 6, Table 7, and Table 8.

3.5. Comparison with existing methods

For comparison, the proposed GBM and GBMs fusion were also applied to classify EEG signals for the other various classification cases in UoB dataset, and the results are tabulated in Table 12. The comparison with some existing methods proposed in the previous research con-

Table 11
Some examples of GBMs fusion classification results that produce the best accuracy

Base classifier 1			Base classifier 2			Base classifier 3			Base classifier 4			Base classifier 5			Accuracy (%)
GBM model	Wavelet		GBM model	Wavelet		GBM model	Wavelet		GBM model	Wavelet		GBM model	Wavelet		
	Family	Level		Family	Level		Family	Level		Family	Level		Family	Level	
$\mathscr{G}^1_{012}$	coif4	4	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	bior3.3	1	$\mathscr{G}^2_{02}$	db27	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	coif4	4	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	bior3.9	1	$\mathscr{G}^2_{02}$	db27	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	coif4	4	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	coif1	1	$\mathscr{G}^2_{02}$	db38	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	coif4	4	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	bior3.3	1	$\mathscr{G}^2_{02}$	db38	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	coif4	4	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	coif1	1	$\mathscr{G}^2_{02}$	sym16	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	coif4	4	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^2_{01}$	rbio3.9	2	$\mathscr{G}^1_{02}$	rbio1.1	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	bior5.5	5	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	coif1	1	$\mathscr{G}^2_{02}$	db27	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	sym5	6	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	coif1	1	$\mathscr{G}^2_{02}$	db27	1	$\mathscr{G}^2_{12}$	db37	1	100.00
$\mathscr{G}^1_{012}$	rbio2.6	7	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^1_{01}$	coif1	1	$\mathscr{G}^2_{02}$	db27	1	$\mathscr{G}^1_{12}$	db24	4	100.00
$\mathscr{G}^1_{012}$	sym5	8	$\mathscr{G}^2_{012}$	db7	7	$\mathscr{G}^2_{01}$	db15	4	$\mathscr{G}^1_{02}$	bior1.1	3	$\mathscr{G}^2_{12}$	db17	6	100.00

Table 12
The results of various classification cases for UoB dataset in some existing methods

Author	Accuracy for each case* (%)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Martis et al. (2015)	–	–	–	–	–	–	–	–	–	–	–	–	–	98.00	–	–
Djemili et al. (2016)	–	–	–	**100.00**	–	–	–	–	–	97.70	–	–	–			–
Peker et al. (2016)	–	–	–	**100.00**	–	–	–	–	–		–	–	–	99.30	98.28	–
Zhang and Chen (2017)	–	–	–	**100.00**	–	–	–	–	–	98.10	–	–	–	98.47	98.40	–
Hamad et al. (2017)	–	–	–	**100.00**	–		99.58		99.47	99.23	–	–	–	–	–	–
Li et al. (2017)	–	–	–	–	–	–	–	–			–	–	–	98.87	–	–
Acharya et al. (2018)	–	–	–	–	–	–	–	–			–	–	–	88.67	–	–
Saini and Dutta (2018)	–	–	–	–	–	–	–	–			–	–	–	99.30	–	–
Houssein et al. (2019)	–	–	–	**100.00**	–	–	**100.00**	–	99.80	99.23	–	–	–	–	–	–
Kumar and Rao (2019)	–	–	–	–	–	–	–	–	–	–	–	–	–	94.10	–	–
Raghu et al. (2019)	–	–	–	99.45	–		96.06		97.60	97.60	97.10	96.85	–	–	96.50	–
Türk and Özerdem (2019)	95.50	96.50	**100.00**	99.50	99.00	**100.00**	**100.00**	85.71	98.50	98.50	–	–	–	99.00	–	93.60
Wang et al. (2019)	–	–	–	**100.00**			**100.00**		98.40	98.10	**100.00**	98.10	93.20	–	96.50	–
Carvalho et al.(2020)	–	–	–	–	–	–	–	–	–	–	–	–	–	98.60	–	–
Goshvarpour and Goshvarpour (2020)	–	–	–	–	–	–	–	–	–	–	–	–	–	98.30	–	–
Zhang et al. (2020)	–	–	–	99.52	–	–	99.11	–	98.02	97.63	99.38	98.03	–	97.04	96.97	93.55
Sukriti et al. (2021)	–	–	–	**100.00**	–	–		–		96.50	–	–	–	96.67	–	–
Li and Chen (2021)	–	–	–	**100.00**	–	–	**100.00**	–	**100.00**	99.00	–	–	–	99.33	99.20	–
This study using GBM and $\mathscr{F}_1$	–	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	–	**100.00**		**100.00**	–	**100.00**	99.67	99.40	92.19
This study using GBM and $\mathscr{F}_2$	**100.00**	–	–	–	–	–	–	**94.00**	–	**99.49**	–	**99.33**	–	99.33	**99.80**	93.80
This study using GBMs fusion	–	–	–	–	–	–	–	–	–	–	–	–	–	**100.00**	**99.80**	**97.39**

*) Cases 1: A-B, 2: A-C, 3: A-D, 4: A-E, 5: B-C, 6: B-D, 7: B-E, 8: C-D, 9: C-E, 10: D-E, 11: AB-E, 12: CD-E, 13: AB-CD, 14: A-D-E, 15: AB-CD-E, 16: A-B-C-D-E

ducted during 2015 until 2021 are also provided in Table 12. All methods in Table 12 used EEG signals in UoB dataset. As can be seen in Table 12, the proposed method achieved 100% accuracy in almost all classification cases, except C-D, D-E, CD-E, AB-CD-E, and A-B-C-D-E. However, the proposed method outperformed the other existing methods proposed in the previous research in C-D, D-E, CD-E, AB-CD-E, and A-B-C-D-E cases. From Table 12, it can be also observed that although the method proposed by Wang et al. (2019) used GBM to classify 3-class EEG signals (AB-CD-E), its performance was bellow than GBM trained with features $\mathscr{F}_2$ and $\mathscr{F}_2$. This fact show that the features used in this study are more discriminative compared to the features used by Wang et al. (2019). Finally, all classification accuracies for three classes and five classes from existing methods were lower than the classification accuracy of GBMs fusion proposed in this study. Therefore, the proposed GBMs fusion outperforms existing methods in classifying three and more classes EEG signals.

Fig. 8 shows the receiver operating characteristic (ROC) curves of GBM and GBMs fusion for various classification cases in UoB dataset together with its respective area under the curve. ROC curve has been used to show the trade-off between hit rates and false alarm rates of classifier in signal detecting theory. Currently, it is widely used in medical decision-making to evaluate the performance of a diagnostic test. In machine learning, this curve is used to depict the performance of binary classifier by calculating true positive rate and false positive rate in the several values of threshold. For n-class task, ROC curve can be produced by performing vertical averaging on n ROC curves obtained by splitting n-class classification problem into n binary classifiers one for each class. The $_i{}^{th}$ binary classifier is trained by assuming the $_i{}^{th}$ class as positive class and all remaining classes as negative class (Pedregosa et al., 2011). Another way to compare classifier performance is by calculating the area under the ROC curve (AUC). The higher the AUC value, the better the classifier performance (Fawcett, 2006). As shown in Fig. 8, all ROC curves closed to the upper-left corner for all classification cases. Furthermore, almost all AUCs were equal to 1.0000 except C-D, D-E, CD-E, AB-CD-E, and A-B-C-D-E cases. However, the AUCs of such cases were higher than 0.97. These results indicate that the proposed method produces high performance for all classification cases.

To compare the effectiveness of the feature extraction method and the feature selection method used in this study, the proposed GBMs fusion was also trained using the set of features extracted using matrix determinant (Raghu et al., 2019) and successive decomposition index (Raghu et al., 2020) for classifying 3-class EEG signals in UoB dataset. The results of such comparison are tabulated in Table 13. As can be seen in Table 13, the classification accuracies of GMBs fusion were 76.30% and 75.30 by employing features extracted using matrix determinant and successive decomposition index. This result shows that the feature extraction method and the feature selection method used in this study outperform matrix determinant and successive decomposition index.

The comparison between the proposed GBMs fusion and some existing machine learning classifiers, including k-NN, SVM, ANN, and LR, for classifying 3-class EEG signals in UoB dataset was also performed in this study. In the comparison, all classifiers were trained using the set of features obtained by the same feature extraction method used in the proposed GBMs fusion, which are features set $\mathscr{F}_1$ and $\mathscr{F}_2$. Furthermore, GA-based feature selection was also employed in all features sets before inputted to all classifiers. The classification results of k-NN, SVM, ANN, and LR are tabulated in Table 14. As can be seen from Table 14, almost all classifiers achieved better performance when trained using feature set $\mathscr{F}_1$, except k-NN. In addition, by employing fusion strategy as used in the proposed GBMs fusion on all classifiers, the classification accuracy could be improved up to 99.30%. However, the classification accuracy of GMB outperformed k-NN, SVM, ANN, and LR both for single classifier and fusion strategy.

3.6. Computational time

The computational time needed by the proposed method for signal decomposition, feature extraction, feature selection, training process and testing process were measured in this study. For comparison, an experiment was also conducted on a low-end computer. The computational times for each step both on the high-end computer and the low-end computer are tabulated in Table 15. On high-end computer, the proposed method took 5.38 ms to decompose a signal, 53.95 ms to extract feature from the decomposed signal, 2.81–2535.33 s for feature selection, 1930.89 ms for training process, and 0.03 ms for testing process. On the other hand, the proposed method took 21.97 ms, 285.36 ms, 20.95–22423.30 s, 0,15 ms, respectively, when running on the low-end computer.

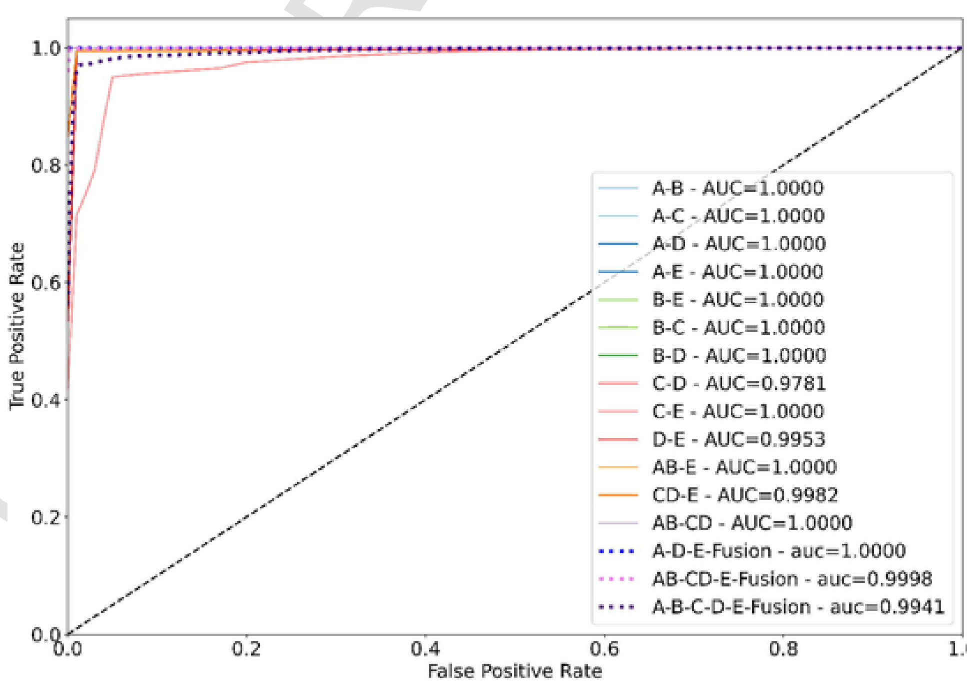

Fig. 8. Receiver operating characteristic curves for two classes GBM and multi class GBMs fusion.

Table 13
The classification accuracy of GBMs fusion trained with different features for classifying 3-class EEG signals in UoB dataset

Feature extraction method	Accuracy (%)
Matrix determinant	76.30
Successive decomposition index	75.30
Frequency sub-band decomposition + DWT + statistical and crossing frequency features + GA-based feature selection	**100.00**

Table 14
Expand
The results of 3-class EEG signals classification using some existing machine learning classifiers

Classifiers	Feature set	Accuracy (%)
k-NN	$\mathscr{F}_1$	98.67
k-NN	$\mathscr{F}_2$	98.67
k-NNs fusion	$\mathscr{F}_1, \mathscr{F}_2$	99.30
SVM	$\mathscr{F}_1$	97.67
SVM	$\mathscr{F}_2$	93.67
SVMs fusion	$\mathscr{F}_1, \mathscr{F}_2$	99.00
ANN	$\mathscr{F}_1$	98.30
ANN	$\mathscr{F}_2$	97.00
ANNs fusion	$\mathscr{F}_1, \mathscr{F}_2$	99.30
LR	$\mathscr{F}_1$	97.00
LR	$\mathscr{F}_2$	95.30
LRs fusion	$\mathscr{F}_1, \mathscr{F}_2$	98.30
GBM	$\mathscr{F}_1$	**99.67**
GBM	$\mathscr{F}_2$	**99.33**
GBMs fusion	$\mathscr{F}_1, \mathscr{F}_2$	**100.00**

Table 15
The computational time needed by the proposed method

Step	Computational time		Percentage increase (%)
	On high-end computer setup	On low-end computer setup	
Signal decomposition	5.38 ms	21.97 ms	75.51
Feature extraction	53.95 ms	285.36 ms	81.09
Feature selection	2.81–2535.33 s	20.95–22423.30 s	645.55–784.43
Training process	1930.89 ms	7135.63 ms	72.94
Testing process	0.03 ms	0.15 ms	80.00

From Table 15, it can be seen that the computational time for all steps were increased between 72.94% and 784.43% when the proposed method running on the low-end computer. However, once important features have been selected and the classifier has been trained, the total times needed to classify an EEG signal using the proposed method were only 59.36 ms and 307.48 ms on the high-end computer and on the low-end computer, respectively. This fact shows that the proposed method can be used to develop an automated system for epilepsy detection in the real clinical application even on the low-end computer. Such system can help neurologist to make a diagnosis and treatment for someone with epilepsy as reported by Raghu et al. (2020).

3.7. Classification results on CHB-MIT dataset

The proposed method was also validated using EEG signals from CHB-MIT dataset (Shoeb, 2009). The dataset was collected from 23 pediatric subjects, ages 1.5–22 years, with seizure at the Children's Hospital Boston. Each EEG signal in the dataset was recorded using the International 10–20 system of EEG electrode positions and sampled at 256 samples per second with 16-bit resolution. Most subjects were recorded in 23 channels and a few subjects were recorded in 24 or 26 channels. Since the channels used by the subjects varied, therefore this study only selected 18 channels that commonly used by all subjects for classification. Such channels were F8-T8, F3-C3, T8-P8, P8-O2, T7-P7, FP1-F7, F7-T7, FP2-F4, CZ-PZ, C3-P3, C4-P4, FZ-CZ, P7-O1, P3-O1, P4-O2, FP2-F8, F4-C4, FP1-F3. Before classification, each EEG signal was partitioned into three classes interictal (Int), preictal (Pre), and ictal (Ict) following the partitioning method used by Ryu and Joe (2021). Frequency sub-band decomposition using FFT and DWT were applied to the EEG signals from all selected channels. The statistical features and crossing frequency features were then extracted from the output of DWT. The important features were selected using GA based feature selection prior to classification.

Some experiments have been performed on CHB-MIT dataset with various wavelet families and decomposition levels. The best classification accuracies of the proposed method in CHB-MIT dataset are tabulated in Table 16. As can be seen in Table 16, GBM achieved the best accuracy of 93.68% for interictal-preictal case after feature selection using features set $\mathscr{F}_1$ and by applying bior1.1 with decomposition level of 5 on DWT. For interictal- ictal case, the best accuracy of 98,82% was achieved after feature selection using features set $\mathscr{F}_1$ and by applying bior3.3 with decomposition level of 6 on DWT. For preictal-ictal case, the best accuracy of 99.26% was achieved after feature selection using features set $\mathscr{F}_1$ and by applying coif3 with decomposition level of 4 on DWT. The classification accuracy of the proposed method in interictal-preictal case was the lowest when compared to other cases. This result can occur since the EEG signals in interictal and preictal classes are very similar. However, the performance of the proposed method in interictal- preictal case is still better than one obtained by Ryu and Joe (2021), which is 93.28%.

For three classes case, interictal-preictal-ictal, best accuracies of 94.06% and 94.56% were achieved by GBM after feature selection using features set $\mathscr{F}_1$ and $\mathscr{F}_2$, and by applying db12 and bior1.5 with decomposition level of 5 and 1 on DWT, respectively. Furthermore, the classification accuracy of three classes case improved to 96.53% by applying GBMs fusion. These results agree with the classification results in UoB dataset.However, the performance of the proposed method in CHB-MIT dataset was lower than its performance in UoB dataset both for 2-class and 3-class problems. Therefore, the performance of the proposed method is highly dependent on the dataset used in the experiment.

3.8. Limitations of the study

Although it has been validated that the proposed method achieves high classification performance in epilepsy detection from EEG signals, there are some limitations that need to be addressed in the future study.

Table 16
The classification accuracies of the proposed method in CHB-MIT dataset

Cases	Classifier and features set	Wavelet		Original features		Selected features	
		Family	Level	Number of features	Accuracy (%)	Number of features	Accuracy (%)
Int-Pre	GBM and $\mathscr{F}_1$	bior1.1	5	756	93.56	385	93.68
Int-Ict	GBM and $\mathscr{F}_1$	bior3.3	6	882	98.73	438	98.82
Pre-Ict	GBM and $\mathscr{F}_1$	coif3	4	630	99.26	320	99.26
Int-Pre-Ict	GBM and $\mathscr{F}_1$	db12	5	756	93.46	394	94.06
Int-Pre-Ict	GBM and $\mathscr{F}_2$	bior1.5	1	1260	94.56	635	94.56
Int-Pre-Ict	GBMs fusion	–	–	–	–	–	96.53

First, the proposed method was only implemented to classify EEG signals in a particular small-sized dataset. These results are only limited in a certain dataset, which are UoB dataset and CHB-MIT dataset, and may not be generalized to the other dataset. Second, the performance of the proposed method is closely related to the wavelet families used in DWT. Therefore, an investigation to determine the appropriate wavelet families need to be performed in a various combination if the proposed method will be applied to the other dataset.

4. Conclusion

In this study, a GBMs fusion was proposed to detect epilepsy from EEG signal automatically. The proposed method achieved the perfect performance with a classification accuracy of 100% in classifying 3-class EEG signals, ictal-normal-interictal, in the dataset from the Department of Epileptology University of Bonn. The GBMs fusion combined the classification result from 3-class and 2-class GBMs. Each EEG signal was processed using DFT and DWT prior to feature extraction. Statistical features and crossing frequency features were extracted from processed EEG signal. Genetic algorithm-based feature selection was employed to find the most discriminative features and to improve the classification performance. Overall, the experimental results show that the proposed GBMs fusion can increase the classification of a single GBM in classifying EEG signals. However, the performance of the proposed method is limited to a particular small-sized dataset and may not be generalized to the other dataset. For future work, the use of the fusion of various wavelet bases in GMBs fusion in classifying EEG signal needs to be investigated. Furthermore, some optimization methods can be considered to optimize the hyperparameters of GBM in order to improve the classification performance in classifying EEG signals.

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.

Acknowledgments

The authors would like to thank Institut Teknologi Sepuluh Nopember (ITS) Surabaya for providing facilities and financial support under scholarship scheme No. T/927/IT2/HK.00.01/2021.

References

Acharya, U.R., Sree, S.V., Suri, J.S., 2011. Automatic detection of epileptic eeg signals using higher order cumulant features. Int. J. Neural Syst. 21 (5), 403–414. https://doi.org/10.1142/S0129065711002912.

Acharya, U.R., Sree, S.V., Ang, P.C.A., Yanti, R., Suri, J.S., 2012. Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals. Int. J. Neural Syst. 22 (2), 1250002. https://doi.org/10.1142/S0129065712500025.

Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adeli, H., 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. 100, 270–278. https://doi.org/10.1016/j.compbiomed.2017.09.017.

Alpaydin, E., 2014. Introduction to Machine Learning, 2nd ed. MIT Press, Cambridge, Massachusetts.

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. Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. 64, 8. https://doi.org/10.1103/PhysRevE.64.061907.

Carvalho, V.R., Moraes, M.F.D., Braga, A.P., Mendes, E.M.A.M., 2020. Evaluating five different adaptive decomposition methods for EEG signal seizure detection and classification. Biomed. Signal Process. Control 62, 102073. https://doi.org/10.1016/j.bspc.2020.102073.

Chua, K.C., Chandran, V., Acharya, U.R., Lim, C.M., 2009. Automatic identification of epileptic electroencephalography signals using higher-order spectra. Proc. Inst. Mech. Eng. Part H J. Eng. Med. 223 (4), 485–495. https://doi.org/10.1243/09544119JEIM484.

Djemili, R., Bourouba, H., Amara Korba, M.C., 2016. Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals. Biocybern. Biomed. Eng. 36 (1), 285–291.

Du, X., Dua, S., Acharya, R.U., Chua, C.K., 2012. Classification of epilepsy using high-order spectra features and principle component analysis. J. Med. Syst. 36 (3), 1731–1743. https://doi.org/10.1007/s10916-010-9633-6.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Filliben, J.J., Heckert, A., 2002. NIST/SEMATECH e-Handbook of Statistical Methods. NIST, Gaithersbg.

Fortin, F.-A., Marc-André Gardner, U., Parizeau, M., Gagné, C., 2012. DEAP: evolutionary algorithms made easy François-Michel De Rainville. J. Mach. Learn. Res. https://doi.org/10.5555/2503308.2503311.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/aos/1013203451.

Goshvarpour, A., Goshvarpour, A., 2020. Diagnosis of epileptic EEG using a lagged Poincare plot in combination with the autocorrelation. Signal, Image Video Process. 14 (7), 1309–1317. https://doi.org/10.1007/s11760-020-01672-w.

Guo, L., Rivero, D., Dorado, J., Munteanu, C.R., Pazos, A., 2011. Automatic feature extraction using genetic programming: an application to epileptic EEG classification. Expert Syst. Appl. 38 (8), 10425–10436. https://doi.org/10.1016/j.eswa.2011.02.118.

Hamad, A., Houssein, E.H., Hassanien, A.E., Fahmy, A.A., 2017. A hybrid EEG signals classification approach based on grey wolf optimizer enhanced SVMs for epileptic detection. In: International Conference on Advanced Intelligent Systems and Informatics. Springer, pp. 108–117.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585 (7825), 357–362.

Houssein, E.H., Hamad, A., Hassanien, A.E., Fahmy, A.A., 2019. Epileptic detection based on whale optimization enhanced support vector machine. J. Inf. Optim. Sci. 40 (3), 699–723. https://doi.org/10.1080/02522667.2018.1453671.

Kaya, Y., Uyar, M., Tekin, R., Yildirim, S., 2014. 1D-local binary pattern based feature extraction for classification of epileptic EEG signals. Appl. Math. Comput. 243, 209–219. https://doi.org/10.1016/j.amc.2014.05.128.

Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K., O'Leary, A., 2019. PyWavelets: A Python package for wavelet analysis. J. Open Source Softw. 4, 1237. https://doi.org/10.21105/joss.01237.

Li, T.-S., 2006. Feature selection for classification by using a GA-based neural network approach. J. Chinese Inst. Ind. Eng. 23 (1), 55–64. https://doi.org/10.1080/10170660609508996.

Li, M., Chen, W., 2021. FFT-based deep feature learning method for EEG classification. Biomed. Signal Process. Control 66, 102492. https://doi.org/10.1016/j.bspc.2021.102492.

Li, M., Chen, W., Zhang, T., 2017. Automatic epileptic EEG detection using DT-CWT-based non-linear features. Biomed. Signal Process. Control 34, 114–125. https://doi.org/10.1016/j.bspc.2017.01.010.

Martis, R.J., Acharya, U.R., Tan, J.H., Petznick, A., Tong, L., Chua, C.K., Kwee Ng, E.Y., 2013. Application of intrinsic Time-scale decomposition (ITD) to EEG signals for automated seizure prediction. Int. J. Neural Syst. 23 (5), 1350023. https://doi.org/10.1142/S0129065713500238.

Martis, R.J., Tan, J.H., Chua, C.K., Loon, T.C., Yeo, S.W.J., Tong, L., 2015. Epileptic EEG classification using nonlinear parameters on different frequency bands. J. Mech. Med. Biol. 15 (3), 1550040. https://doi.org/10.1142/S0219519415500402.

Mirjalili, S., Lewis, A., 2016. The whale optimization algorithm. Adv. Eng. Softw. 95, 51–67.

Mirjalili, S., Mirjalili, S.M., Lewis, A., 2014. Grey wolf optimizer. Adv. Eng. Softw. 69, 46–61.

Motamedi-Fakhr, S., Moshrefi-Torbati, M., Hill, M., Hill, C.M., White, P.R., 2014. Signal processing techniques applied to human sleep EEG signals - a review. Biomed. Signal Process. Control. https://doi.org/10.1016/j.bspc.2013.12.003.

Myroniv, B., Wu, C.-W., Ren, Y., Christian, A., Bajo, E., Tseng, Y.-C., 2017. Analyzing user emotions via physiology signals. Data Sci. Pattern Recognit. 1, 11-25 q.

Parvinnia, E., Sabeti, M., Zolghadri Jahromi, M., Boostani, R., 2014. Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm. J. King Saud Univ. - Comput Inf. Sci. 26, 1–6. https://doi.org/10.1016/j.jksuci.2013.01.001.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peker, M., Sen, B., Delen, D., 2016. A novel method for automated diagnosis of epilepsy using complex-valued classifiers. IEEE J. Biomed. Heal. Informatics 20 (1), 108–118. https://doi.org/10.1109/JBHI.2014.2387795.

Raghu, S., Sriraam, N., Hegde, A.S., Kubben, P.L., 2019. A novel approach for classification of epileptic seizures using matrix determinant. Expert Syst. Appl. 127, 323–341.

Raghu, S., Sriraam, N., Vasudeva Rao, S., Hegde, A.S., Kubben, P.L., 2020. Automated detection of epileptic seizures using successive decomposition index and support vector machine classifier in long-term EEG. Neural Comput. Appl. 32 (13), 8965–8984.

Ravi Kumar, M.R., Srinivasa Rao, Y.S., 2019. Epileptic seizures classification in EEG signal based on semantic features and variational mode decomposition. Cluster Comput. 22 (S6), 13521–13531.

Ryu, S., Joe, I., 2021. A Hybrid DenseNet-LSTM model for epileptic seizure prediction. Appl. Sci. 11 (16), 7661. https://doi.org/10.3390/app11167661.

Saini, J., Dutta, M., 2018. Epilepsy classification using optimized artificial neural

network. Neurol. Res. 40 (11), 982–994.
Satyender, Dhull, S.K., Singh, K.K., 2021. A review on automatic epilepsy detection from EEG signals. In: Hura, G., Singh, A., Siong Hoe, L. (Eds.), Advances in Communication and Computational Technology. Springer, pp. 1441–1454.
Shoeb, A.H., 2009. Application of machine learning to epileptic seizure onset detection and treatment.
Siswantoro, J., Arwoko, H., Siswantoro, M.Z.F.N., 2020. Fruits Classification from Image using MPEG-7 Visual Descriptors and Extreme Learning Machine, in: 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020. doi:10.1109/ISRITI51436.2020.9315523.
Sreekumar, A., Reddy, A.N.S., Ravikanth, D.U., Chowdary, M.C., Nithin, G., Sathidevi, P.S., 2021. Epileptic Seizure Detection Using Machine Learning Techniques. In: Hura, G., Singh, A., Siong Hoe, L. (Eds.), Advances in Communication and Computational Technology. Springer, pp. 919–926.
Sukriti, Chakraborty, M., Mitra, D., 2021. Automated detection of epileptic seizures using multiscale and refined composite multiscale dispersion entropy. Chaos, Solitons Fractals 146, 110939. https://doi.org/10.1016/j.chaos.2021.110939.
Tan, L., Jiang, J., 2019. Digital Signal Processing: Fundamentals and Applications, 3rd ed. Academic Press, London, U.K.
Türk, Ömer, Özerdem, Mehmet Siraç, 2019. Epilepsy detection by using scalogram based convolutional neural network from EEG signals. Brain Sci. 9 (5), 115. https://doi.org/10.3390/brainsci9050115.
Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K., 2012. Probability & Statistics for Engineers & Scientists, ke-9. ed. Prentice Hall, Boston.
Wang, Xiashuang, Gong, Guanghong, Li, Ni, 2019. Automated recognition of epileptic EEG states using a combination of symlet wavelet processing, gradient boosting machine, and grid search optimizer. Sensors (Switzerland) 19 (2), 219. https://doi.org/10.3390/s19020219.
World Health Organization, 2019. Epilepsy [WWW Document]. URL https://www.who.int/news-room/fact-sheets/detail/epilepsy (accessed 8.24.20).
Yao, Dezhong, 2001. A method to standardize a reference of scalp EEG recordings to a point at infinity. Physiol. Meas. 22 (4), 693–711. https://doi.org/10.1088/0967-3334/22/4/305.
Zhang, Zhiguo, 2019. Spectral and Time-Frequency Analysis BT - EEG Signal Processing and Feature Extraction. Springer Singapore, Singapore, pp. 89–116. https://doi.org/10.1007/978-981-13-9113-2_6.
Zhang, Tao, Chen, Wanzhong, 2017. LMD Based Features for the Automatic Seizure Detection of EEG Signals Using SVM. IEEE Trans. Neural Syst. Rehabil. Eng. 25 (8), 1100–1108. https://doi.org/10.1109/TNSRE.2016.2611601.
Zhang, G., Yang, L., Li, B., Lu, Y., Liu, Q., Zhao, W., Ren, T., Zhou, J., Wang, S.-H., Che, W., 2020. MNL-Network: A Multi-Scale Non-local Network for Epilepsy Detection From EEG Signals. Front. Neurosci.
Zhao, Wei, Zhao, Wenbing, Wang, Wenfeng, Jiang, Xiaolu, Zhang, Xiaodong, Peng, Yonghong, Zhang, Baocan, Zhang, Guokai, 2020. A Novel Deep Neural Network for Robust Detection of Seizures Using EEG Signals. Comput. Math. Methods Med. 2020, 1–9. https://doi.org/10.1155/2020/9689821.