



Multiple robust approaches for EEG-based driving fatigue detection and classification

Sunil Kumar Prabhakar, Dong-Ok Won^{*}

Department of Artificial Intelligence Convergence, Hallym University, Chuncheon, 24252, South Korea

ARTICLE INFO

Keywords:

MDS
SVR
SVM
Hilbert transform
ELM
ANFIS
TQWT

ABSTRACT

Electroencephalography (EEG) signals are used to evaluate the activities of the brain. For the accidents occurring on the road, one of the primary reasons is driver fatigueness and it can be easily identified by the EEG. In this work, five efficient and robust approaches for the EEG-based driving fatigue detection and classification are proposed. In the first proposed strategy, the concept of Multi-Dimensional Scaling (MDS) and Singular Value Decomposition (SVD) are merged and then the Fuzzy C Means based Support Vector Regression (FCM-SVR) classification module is utilized to get the output. In the second proposed strategy, the Marginal Fisher Analysis (MFA) is implemented and the concepts of conditional feature mapping and cross domain transfer learning are implemented and classified with machine learning classifiers. In the third proposed strategy, the concepts of Flexible Analytic Wavelet Transform (FAWT) and Tunable Q Wavelet Transform (TQWT) are implemented and merged and then it is classified with Extreme Learning Machine (ELM), Kernel ELM and Adaptive Neuro Fuzzy Inference System (ANFIS) classifiers. In the fourth proposed strategy, the concepts of Correntropy spectral density and Lyapunov exponent with Rosenstein algorithm is implemented and then the multi distance signal level difference is computed followed by the calculation of the Geodesic minimum distance to the Riemannian means and finally tangent space mapping is implemented to it before feeding it to classification. In the fifth or final proposed strategy, the Hilbert Huang Transform (HHT) is implemented and then the Hilbert marginal spectrum is computed. Then using the Blackhole optimization algorithm, the features are selected and finally it is classified with Cascade Adaboost classifier. The proposed techniques are applied on publicly available EEG datasets and the best result of 99.13% is obtained when the proposed Correntropy spectral density and Lyapunov exponent with Rosenstein algorithm is implemented with the multi distance signal level difference followed by the calculation of the Geodesic minimum distance to the Riemannian means and finally tangent space mapping is implemented with Support Vector Machine (SVM) classifier.

1. Introduction

Staying alert is quite crucial for many jobs such as driving, construction works and security guarding during night shifts etc. [1] The primary reason for all the accidents in the world is due to the drowsiness of the drivers. Drowsy driving can occur in any scenario but predominantly occurs in long-distance transportation. Several factors cause accidents due to fatigueness and drowsiness [2]. For drowsiness related crashes, a lot of factors are responsible such as length of the journey, weather condition, uniformity of the road, driving history of the driver, time of occurrence and the health condition of the driver. Before the onset of drowsiness, the driver must be alerted and it is one of the best methods to prevent the crash caused due to drowsy driving [3]. New

devices have been developed by car industries so that driver drowsiness can be predicted well. Numerous drowsiness detection systems are being offered by car companies. A variety of vehicle-based indices are measured by these built in solutions so that fatigueness and the onset of drowsiness can be easily detected. To detect the drowsiness of the drivers, a variety of methods have been investigated and it includes vehicle-based measures, behavioral measures, or physiological measures [4]. To reflect the performance of the driver during driving, many vehicles based assistive technologies are used such as steering-angle sensors, ultrasonic sensors, GPS etc. Some of the drawbacks of these vehicle based assistive technologies include the interfering factors such as wind and rutted road surfaces, design, usage and assessment of the technology adaptable to each car brand. Behavioral measures include

^{*} Corresponding author.

E-mail address: dongok.won@hallym.ac.kr (D.-O. Won).

<https://doi.org/10.1016/j.array.2023.100320>

Received 12 June 2023; Received in revised form 17 July 2023; Accepted 6 September 2023

Available online 9 September 2023

2590-0056/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

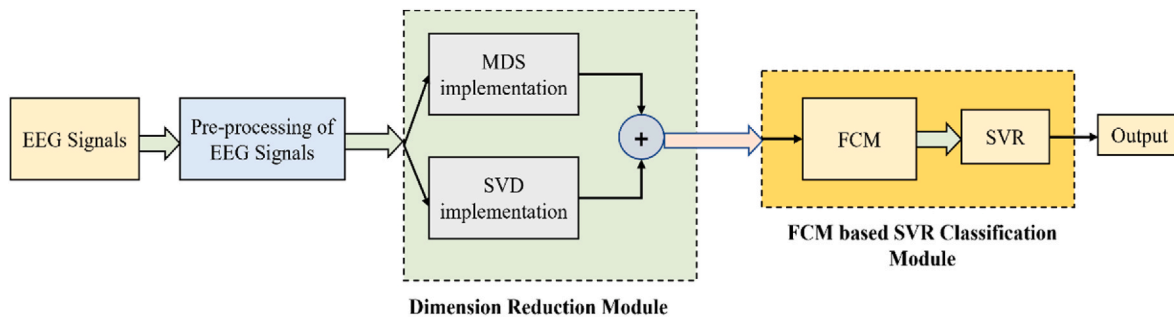


Fig. 1. Implementation of Dimension Reduction with FCM based SVR Classification Module.

the facial behaviour of the driver including eye blinking, yawning, posture of the head, eye closure etc. [5]. Such measures can be observed by a camera and the driver can be alerted immediately if such symptoms are observed. However, being monitored in one's privacy continuously throughout the journey is a huge issue. A lot of factors such as consumption of a heavy meal during daytime, medications, lack of proper sleep, overthinking due to financial and family problems can cause drowsiness [6]. However, a lot of other accompanying factors are also present when drowsiness is referred too. An important impact on the progression of drowsiness is related to age and gender. Other important factors are smoking addiction leading to a poor appetite and sleep pattern, health issues caused due to anxiety, hypertension, high blood pressure etc. If one is feeling drowsy, there is always a state of reduced alertness and so it causes a performance degradation in the work and the person would not be able to think properly. The sudden changes in the behavioral pattern and sudden decline in cognitive functions can be easily measured and quantified using EEG signals [7].

With the advent of EEG and machine learning technologies, a lot of automated techniques have been proposed in the past with respect to the EEG based driving fatigue detection and classification. Using a two-level learning Hierarchy Radial Basis Function, the EEG based driving fatigue detection was analyzed by Ren et al. [8]. Depending on the log-Mel spectrogram and Convolutional Recurrent Neural Network (CRNN), the EEG driving fatigue detection was analyzed by Gao et al. [9]. Using multilevel feature extraction and iterative hybrid feature selection, the EEG based driving fatigue detection was analyzed by Tuncer et al. [10]. The research on fatigue driving detection using forehead EEG depending on adaptive multi-scale entropy was done by Luo et al. [11]. The chaotic entropy analysis of cortical sources obtained from scalp EEG signals was used for the driver fatigue detection by Chaudhuri and Routray [12]. Sparse-deep belief networks were used to improve the EEG based driver fatigue classification by Chai et al. [13]. The analysis of feature fatigue EEG signals based on wavelet entropy was analyzed by Wang et al. [14]. The single channel EEG-based mental fatigue detection based on deep belief networks was analyzed by Li et al. [15]. The EEG based driving fatigue detection using flexible analytic wavelet transform and multi-boosting approaches was done by Subasi et al. [16]. The sample entropy-based method for real driving fatigue detection with multi-channel EEG was done by Zhang et al. [17]. The modified PCANet method was used for driving fatigue detection from EEG by Ma et al. [18]. With the help of multiple entropy fusion analysis in an EEG based system, the driver fatigue detection systems were analyzed by Min et al. [19]. Using wearable EEG based on CNN, the vehicle driver drowsiness detection was analyzed by Zhu et al. [20]. A comprehensive review on the fatigue driving detection based on oculography was done by Tian and Cao [21]. With a dynamical encoder-decoder modelling framework, the driver drowsiness estimation using EEG signals was done by Arefnezhad et al. [22]. A light GBM based on EEG analysis technique for driver mental states classification was done by Zeng et al. [23]. The driving fatigue detection based on EEG signals of forehead area alone was performed by Mu et al. [24]. The EEG based drivers mental fatigue

detection using Event-related synchronization/desynchronization and Hurst exponent was one by Rodriguez et al. [25]. An ensemble classifier for driver's fatigue detection based on a single EEG channel was done by Wang et al. [26]. A new hand-modelled learning framework for driving fatigue detection using EEG signals was done by Dogan et al. [27] and a dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals was done by Tuncer et al. [28].

However, in this paper five efficient strategies are proposed for the detection and classification of driving fatigues from EEG signals.

The main contributions of this paper are as follows.

- The pre-processing of the EEG signals was done using Independent Component Analysis (ICA) [29]. In proposed strategy 1, after pre-processing the MDS and SVD are implemented and then it is merged and sent to the FCM based SVR classification module to get the output.
- In proposed strategy 2, after pre-processing the MFA is implemented and then the conditional feature mapping and cross domain transfer learning is implemented and finally classified with machine learning classifiers.
- In proposed strategy 3, after pre-processing the FAWT and TQWT is implemented and merged, and then it is classified with ELM, Kernel ELM and ANFIS classifiers.
- In the proposed strategy 4, after pre-processing, the Correntropy spectral density and Lyapunov with Rosenstein algorithm is implemented and then the multi distance signal level difference is computed. Then the geometry of the SPD matrices is computed and the Geodesic minimum distance to the Riemannian means is calculated. Finally tangent space mapping is implemented to it before feeding it to classification with machine learning classifiers.
- In the proposed strategy 5, after pre-processing, the HHT is computed and then the Hilbert marginal spectrum is computed. Then using the Blackhole algorithm, the features are selected and finally it is classified with Cascade Adaboost classifier. The reasons for the algorithms chosen in the proposed methods are done after an in-depth study of the individual properties of the algorithms.

The organization of the paper is as follows: In section 2, all the proposed strategies are explained and in section 3, the results and discussion are implemented followed by the conclusion in section 4.

2. Proposed strategies

2.1. Proposed technique 1

The proposed technique 1 uses the idea of MDS and SVD for the purpose of dimensionality reduction. Then an FCM clustering based SVR technique is utilized to assess the performance of driving fatigue detection from EEG signals. The block diagram of the work is shown in Fig. 1.

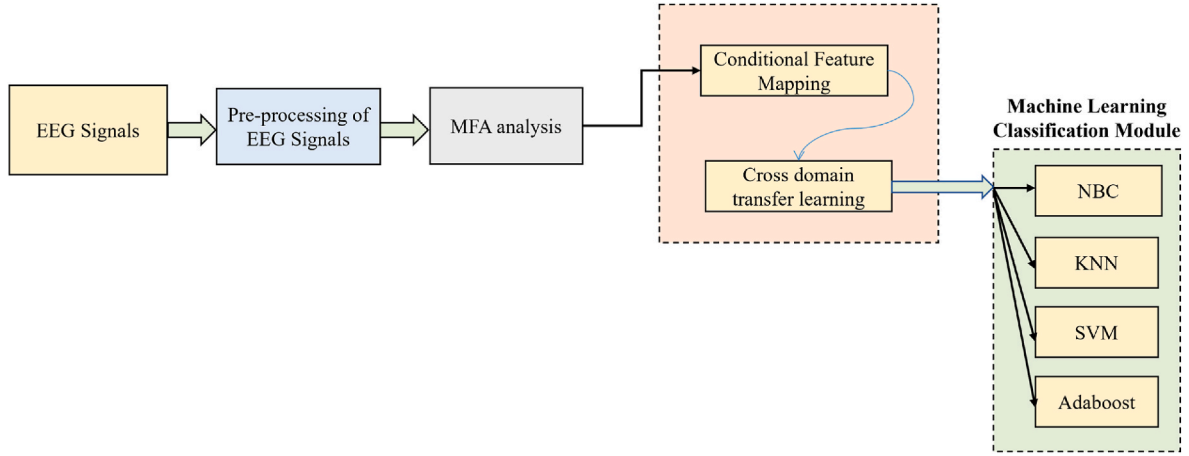


Fig. 2. Implementation of MFA with Conditional Feature Mapping based Cross domain transfer learning with machine learning classification module.

2.1.1. MDS

A lot of irrelevant or useless information is present in the high-dimensional vector which usually leads to a high computational burden to the machine learning classification. MDS is implemented to the pre-processed EEG signals so that the feature dimensions are reduced and therefore the classification accuracy should be improved greatly [30]. If only the similarity matrix can be obtained in between the objects, then reconstructing the Euclidean distance paved the way for developing MDS. Fitting the original data into low dimensional co-ordinate systems is the main idea behind MDS so that the transformation could be minimized which is caused by dimension reduction. In between the paired samples, the similarity is estimated by the MDS. Before and after the dimension reduction, the similarity between the original samples is kept as constant as possible. Assuming a sample vector $p_i (1 \leq i \leq N)$, the detailed computation steps of the classical MDS is as follows:

Step 1. Distance matrix D is constructed. For every pair of vectors p_i and $p_j (1 \leq j \leq N)$, with the help of Euclidean distance d_{ij} , the distance matrix $D = (d_{ij})_{N \times N}$ is assessed easily.

$$d_{ij} = |p_i - p_j| \quad (1)$$

$$d_{ij}^2 = p_i^T p_i + p_j^T p_j - 2p_j^T p_i \quad (2)$$

Step 2. Computation of inner product matrix. From the distance matrix D , the inner product matrix B can be easily obtained and it is expressed as:

$$B = -\frac{1}{2}MD^2M \quad (3)$$

where $M = U - \frac{1}{N}uu^T$, ' U ' denotes a unit matrix and ' u ' denotes a unit column vector with length N .

Step 3. Computation of low dimensional representation. Initially, the first ' k ' maximum eigen values are computed followed by the assessment of the corresponding eigen vectors of matrix B . Then to generate a low dimensional representation Z , the first ' k ' columns of G are employed as follows:

$$G = U_k V_k^{\frac{1}{2}} \quad (4)$$

The diagonal matrix of the k maximum eigen values of B are denoted by V_k , where U_k is a matrix composed of the first ' k ' orthogonal eigen vectors retained by M . The block diagram of the proposed work is shown

in Fig. 2.

2.1.2. SVD

To reduce the high dimensional matrix into a low dimensional matrix, SVD is highly useful [31]. The underlying meaning of the information can be easily retained by SVD and some essential features can be extracted easily. A complex matrix can be represented by multiplying these smaller and simpler matrices. In the SVD theory, it is presumed that a $m \times n$ order matrix is present where all the elements belong to the real field. This can be represented by the SVD matrix M as follows:

$$M = U \Lambda V^T \quad (5)$$

where a $m \times m$ order matrix is represented by U , the diagonal matrix of $m \times n$ is represented by Λ and an $n \times n$ order matrix is represented by V^T .

$$M = U \Lambda V^T = \sum_{i=1}^r \lambda_i u_i v_i^T = \sum_{i=1}^r \lambda_i M_i \quad (6)$$

The i^{th} column vector of the matrices U and V are represented as u_i and v_i . The i^{th} singular value of the matrix M is represented as λ_i . The information contained by the sub matrix is reflected by each singular value λ_i .

2.1.3. FCM- SVR technique

Once the dimensionality is reduced, it is fed to the FCM-SVR classification module. In order to cluster the data in unsupervised learning, a well-known algorithm is FCM [32]. A set of object data $\{p_1, p_2, \dots, p_n\} \subset \mathbb{R}^{s \times n}$ is portioned into ' c ' fuzzy clusters by means of minimizing an objective function $J(U, V)$ as follows:

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|p_k - v_i\|_2^2 \quad (7)$$

where an object datum is represented by $p_k = [p_{1,k}, p_{2,k}, \dots, p_{s,k}]^T$. The i^{th} cluster prototype is indicated by v_i and the j^{th} attribute of p_k is represented by $p_{j,k}$. The matrix of cluster prototype be represented as $V = [v_{ij}] = [v_1, v_2, \dots, v_c]^T \in \mathbb{R}^{c \times s}$. The fuzzification parameter is represented by f and $m \in (1, \infty)$. The Euclidean norm in \mathbb{R}^s is denoted by $\|\cdot\|_2$. The degree which p_k belongs to the i^{th} cluster is represented by the membership u_{ik} and the following condition has to be satisfied as follows:

$$\sum_{i=1}^c u_{ik} = 1 (k = 1, 2, \dots, n; \forall i, k : u_{ik} \in [0, 1]) \quad (8)$$

The partition matrix is represented as $U = [u_{ik}] \in \mathbb{R}^{c \times n}$. With the help of some constraints, the vital conditions for minimization are done and so the following iterative updates formulae are obtained for the prototype and the partition matrix as follows:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m p_k}{\sum_{k=1}^n u_{ik}^m} \quad (i = 1, 2, \dots, c) \quad (9)$$

$$u_{ik} = \left[\sum_{t=1}^c \left(\frac{\|p_k - v_i\|_2^2}{\|p_k - v_t\|_2^2} \right)^{\frac{1}{m-1}} \right]^{-1} \begin{pmatrix} i = 1, 2, \dots, c \\ \text{and} \\ k = 1, 2, \dots, n \end{pmatrix} \quad (10)$$

Unless the changes in the partition matrix values are lower than a particular predetermined threshold, the iterations are proceeded with it.

2.1.4. Implementation of FCM – SVR for EEG based driving fatigue detection

Support Vector Regression is dependent on Support Vector Machine (SVM) whose main intention is to assess the complex relationship between the input and the response of interest by means of mapping the data into a feature space with high dimension [33]. Assuming that the i^{th} input and i^{th} response of interest is indicated by $p_i = (p_{i1}, \dots, p_{id})$ and q_i respectively. The constitution of the regression model of SVR is expressed as:

$$q = w^T \cdot \varphi(p) + b \quad (11)$$

where the feature maps are indicated by φ . The bias term is indicated by b and the weight vector is represented as w . By means of minimizing a cost function (C) which has a penalized regression error, the SVR model is built and is expressed as follows:

$$C = \frac{1}{2} w^T \cdot w + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 \quad (12)$$

In order to penalize large weights and to regularize weight sizes, the first part of cost function C used is a weight decay. For all the training data, the regression error is represented by the second part. The relative weight between them is determined by the parameter γ . Lagrange multiplier technique is utilized to optimize the cost function C as follows:

$$L(w, b, e; \alpha) = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i \times \{w^T \cdot \varphi(p_i) + b + e_i - q_i\} \quad (13)$$

where the Lagrange multipliers are indicated as α_i . The optimal solution is obtained by means of setting the partial first derivatives to zero. Thus,

$$w = \sum_{i=1}^n \alpha_i \varphi(p_i) = \sum_{i=1}^n \gamma e_i \varphi(p_i) \quad (14)$$

where the utilization of a positive definite kernel is done as follows:

$$K(p_i, p_j) = \varphi(p_i)^T \varphi(p_j) \quad (15)$$

The original regression model is thus expressed as follows:

$$q = \sum_{i=1}^n \alpha_i \varphi(p_i)^T \varphi(p) + b = \sum_{i=1}^n \alpha_i \langle \varphi(p_i)^T, \varphi(p) \rangle + b \quad (16)$$

In order to evaluate the point of q_j , it is assessed as:

$$q_j = \sum_{i=1}^n \alpha_i \langle \varphi(p_i)^T, \varphi(p_j) \rangle + b \quad (17)$$

By solving a set of linear equation, α vector can be obtained as follows:

$$\begin{bmatrix} K + \frac{1}{\gamma} & 1_n \\ 1_n^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} q \\ 0 \end{bmatrix} \quad (18)$$

The solution is thus represented as follows:

$$\begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} K + \frac{1}{\gamma} & 1_n \\ 1_n^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} q \\ 0 \end{bmatrix} \quad (19)$$

The Kernel function used is Radial Basis Function (RBF) in our work.

2.2. Proposed technique 2

In this proposed approach 2, once the pre-processing is done, MFA is applied initially and then the concept of conditional feature mapping is implemented. The concept of cross-domain transfer learning is proposed and implemented finally and classified with some machine learning classifiers. The block diagram of the work is shown in Fig. 2.

2.2.1. Marginal Fisher Analysis

To the pre-processed EEG signals, MFA is applied initially [34]. To design the intrinsic graph to explain the intra-class compactness, the famous graph-embedding based framework utilized is MFA. In order to explain the inter-class separation with the penalty graph also, this technique is used widely. In the standard MFA method, the distance measurement technique is the traditional Euclidean distance whereas in this paper the distance measurement technique utilized is the inner product. The solution process of the algorithm can be optimized and simultaneously it is kept consistent with the distance measurement technique. The computation of the proximity degree of same class in the intrinsic graph G_c is done by means of assessing the sum of the distance values between each sample and its nearest K sample points from the similar class, so that the intra-class compactness S_c is achieved and expressed as:

$$S_i = \sum_{i \in N_k(j)} \sum_{or \ j \in N_k(i)} (M^T p_i)^T M^T p_j \quad (20)$$

where the feature transformation matrix is represented as M . An index set of the nearest K sample points from the similar class of sample p_i is represented by $N_k(i)$. The computation of the proximity degree of inter-class marginal points in the penalty graph G_p is assessment by the sum of the distance values between each marginal singular points and its nearest m sample point from various other classes and the expression of inter-class separation S_p is done as follows:

$$S_p = \sum_{i \in p_m(c_i)} \sum_{or \ (i,j) \in p_m(c_j)} (M^T p_i)^T M^T p_j \quad (21)$$

where an index set of the nearest m pairs of marginal samples in $\{(i,j) | i \in \pi_{c_i}, j \notin \pi_{c_i}\}$ is represented by $p(c_i)$. The intra-class compactness is minimized and the inter-class separation is maximized so that with the help of graph embedding, the matrix M can be obtained and the expression of the objective function is done as follows:

$$M = \arg \min_M \frac{S_c}{S_p} \quad (22)$$

The principle domain adaptation implies that a common feature space is built up where a similar distribution is present by both the source domain and target domain. The knowledge learnt in the source domain is transferred to the target domain so that the mapping matrix W is obtained. Assume that the source domain G and target domain H have ' c ' classes. Assuming that ' a ' samples in G and is expressed as $P = [p_1, p_2, \dots, p_a]$ and ' b ' samples in H and is expressed as $Q = [q_1, q_2, \dots, q_b]$. The conditional feature mapping is expressed as follows.

2.2.2. Conditional feature mapping

The transfer learning will be highly influenced when the features of target domain are projected to the source domain. The probability is quite similar for the source domain samples which is being chosen. In the source domain, there are many singular points which may have a weak performance in inter-class separation. To the bounding areas in the source domain, the projection of the target domain will be done if these points are chosen and therefore some negative effects may be produced on the feature mappings [35]. Therefore, choosing the source domain samples in a random manner is unreasonable and so to deal with the negative effects, the feature mappings is restricted from the target

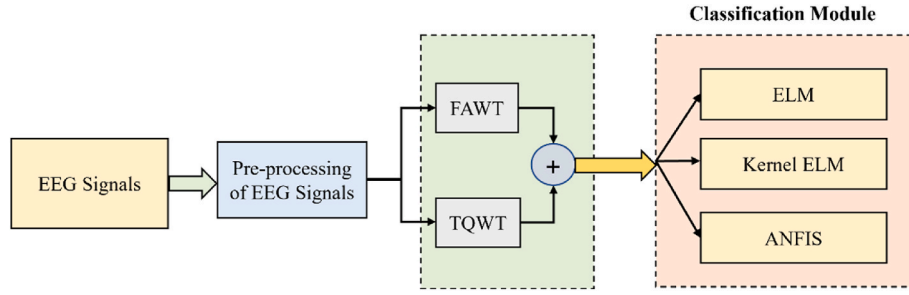


Fig. 3. Analysis of FAWT, TQWT with ELM, ELM kernel and ANFIS

domain to the singular points in source domain. Between each sample point p_i and sample point p_j from the same class, the distance value D_{ij} is calculated. The distance measurement method utilized is inner product and represented as $D_{ij} = p_i^T p_j, i \neq j$. In between each sample point p_i and other sample points from the same class, the sum of the distance value D_i is calculated as $D_i = \sum_{j=1}^{t_k} D_{ij}$, where the number of sample points of class k is represented as t_k . The value of D_i is sorted in ascending order for every point $p_i (1 \leq i \leq t_k)$. The final s_k points are processed as singular points and other points are processed as non-singular ones. The feature mapping from the target domain will be restricted for singular points and the feature mapping will be promoted for non-singular ones and therefore across the domains, the conditional feature mapping is realized.

2.2.3. Cross domain transfer learning

The standard transfer learning technique is implemented based on the conditional feature mapping mechanism [36]. The regularization is applied to W so that the deviation $\psi(w)$ is minimized and therefore the over-fitting can be avoided. The expression of the objective function is as follows:

$$\min_w \psi(w) \quad (23)$$

$$s.t. \quad f_i(P^T W Q) \geq 0, 1 \leq i \leq c$$

where the inter-domain constraints are specified by $f_i(P^T W Q)$. LogDet regularization is used for W and it is specified as $\psi(w)$ where it is projected as the sum of the singular values ($\sigma_1, \sigma_2, \dots, \sigma_x$) of W and it is represented as $\psi(w) = \sum \psi_i(\sigma_j)$, where the scalar function is denoted by ψ_i . In between the samples, the similarity function is expressed with the form of inner product. If the inner product $K_G = P^T P$ and $K_H = Q^T Q$, then the optimal solution of the problem is represented as:

$$W = P K_G^{-1/2} L K_H^{-1/2} Q^T, \quad (24)$$

where L is an $a \times b$ matrix

The respective constraints are built if the intra-class compactness and inter-class separation is considered. To express the labelled samples in source domain G , (p_i, l_i^G) is used, where l_i^G is the label of p_i and to express the labelled sample in target domain H , (q_i, l_i^H) is used, where l_i^H is the label of q_i . The conversion of the constraints is done as follows for the sample pairs composed of the samples in G and H as follows:

$$f_{ij}(p_i W q_j) = p_i^T W q_j - l_i^G = l_j^H \quad (25)$$

$$f_{ij}(p_i W q_j) = u - p_i^T W q_j, l_i^G \neq l_j^H \quad (26)$$

where l indicates the lower limit of the similarity among the sample points and n indicate the upper limit of the similarity among the sample point. The similarity should be as high as possible for the sample pairs with the same classes but for other difficult classes, it is made smaller. A scaling function of $\psi_i(\sigma_j) = \sigma_j^2 / 2$ is adopted so that the mapping learning problem can be converted as follows:

$$\min_w \sum_{j=1}^z \sigma_j^2 / 2 \quad (27)$$

$$s.t. p_i^T W q_j' - l_i^G = l_j^H$$

$$u - p_i^T W q_j', l_i^G \neq l_j^H$$

where $p_i' = M p_i$ and $q_j' = M q_j$. A relaxation coefficient can be added as this is a strict convex optimization problem and the above expression is expressed as follows:

$$\min_w \sum_{j=1}^z \sigma_j^2 / 2 + \lambda \sum_i \sum_j f_{ij} [(M_{p_i})^T W (M_{q_j})] \quad (28)$$

$$s.t. p_i \in P, q_j \in Q$$

Using domain adaption, the mapping matrix W is learnt given the features of both strong class and weak class. The weak class features P are extracted and projected them using W into the latent space so that the corresponding strong class features $Q = P W$ are obtained. The inferred strong class features are used for testing and testing using a machine learning classifier and in our work four classifiers are considered such as NBC, KNN, SVM and Adaboost respectively.

2.3. Proposed technique 3

In this proposed approach 3, once the pre-processing is done, FAWT and TQWT are applied initially and then it is analyzed with ELM, ELM Kernel and ANFIS. The block diagram of the work is shown in Fig. 3.

2.3.1. FAWT

For multi-resolution analysis, processing of non-stationary and nonlinear signals and for the time-frequency localization purposes, wavelet transforms are used widely. Since it has some restrictions, such as dealing with low-resolution at high frequency sub bands and some problems with respect to shift variance, WT is not widely preferred. To overcome the above limitations, FAWT is used [37]. Arbitrary sampling rates are possessed by both low-pass and high-pass filters and so it leads to a flexible time-frequency partition manner. A good shift-invariance is also possessed by FAWT and for the analysis of complex oscillatory signals, FAWT is used widely. The parameters such as redundancy, dilation factor and the quality factor (QF) are controlled by FAWT. The EEG signals can be analyzed easily with pliable parameter such as p, q, r, s and β . The up and down sampling factor are represented as p and q for the low pass channel. The up and down sampling factor are represented as r and s for the low pass channel respectively. The QF is weighed by a position constant β and their relationship is expressed as:

$$QF = 2 - \beta / \beta \quad (29)$$

The decomposition of the raw EEG signal is done into l' level with the help of an iterative filter bank. Every level comprises of 2 high-pass filters and one low pass filter respectively. In our work, the values are as follows: $p = 2, q = 4, r = 0.5, s = 1, \beta = (0.8r) / s$. The level of decomposition is kept at $l = 10$. The sub-band EEG signals reconstructed

by FAWT are specified by sub bands.

2.3.2. Tunable Q-wavelet transform

The analysis of oscillatory EEG signals is analyzed by the TQWT by means of using adjustable input parameters such as Q-factor, no of decomposition level (L) and oversampling rate (r) [38]. The generation of L+1 sub-band signals are done by the L level decomposition of input EEG signal $p(n)$. The low-pass signals are indicated by $a^L(n)$ and the high-pass sub band signals are specified by $b^L(n)$ respectively. The two-channel analysis filter bank is used at every decomposition level in an iterative manner. The sampling rate at every low-pass sub band is represented as αf_s and sampling rate at every high-pass sub-band is represented as βf_s respectively. The scaling factors are represented by α and β respectively, and the sampling frequency of the input EEG signal $p(n)$ is denoted as f_s respectively. The scale factors α and β are used to perform the low-pass scaling and high-pass scaling respectively. The low-pass and high-pass filters are present in the two-channel filter bank and it is denoted as $H_0(w)$ and $H_1(w)$ respectively, where the angular frequency is denoted by w . A restricted redundancy is offered by the filter bank of the TQWT so that a perfect reconstruction is achieved by the following relation: $0 < \alpha < 1, 0 < \beta < 1$ and $\alpha + \beta > 1$. For L-level decomposition of the input EEG signal $p(n)$, the sampling rate for the low-pass sub band signal $a^L(n)$ is $\alpha^L f_s$ while the sampling rate for high pass sub band signal $b^L(n)$ is $\alpha^{L-1} \beta f_s$. At stage L, the frequency response of the low-pass filter $H_0^L(w)$ and the high-pass filter $H_1^L(w)$ is expressed as follows:

$$H_0(w)^L = \begin{cases} \prod_{m=1}^{L-1} H_0(w/\alpha^m), & \text{for } (w) \leq \alpha^L \pi \\ 0, & \text{for } \alpha^L \pi < |w| \leq \pi \end{cases} \quad (30)$$

$$H_1(w)^L = \begin{cases} H_1(w/\alpha^{L-1}) \prod_{m=0}^{L-2} H_1(w/\alpha^m), & \text{for } (1-\beta)\alpha^{L-1}\pi \leq |w| \leq \alpha^{L-1}\pi \\ 0, & \text{for other } w \in [-\pi, \pi] \end{cases} \quad (31)$$

where $L \in N$. In terms of $\theta(w)$, $H_l(w)$ and $H_h(w)$ are defined which is the frequency response of the Daubechies filter and is expressed as follows:

$$H_0(w) = \theta\left(\frac{w + (\beta - 1)\pi}{\alpha + \beta - 1}\right) \quad (32)$$

$$H_1(w) = \theta\left(\frac{\alpha\pi - w}{\alpha + \beta - 1}\right) \quad (33)$$

$$\theta(w) = \frac{1}{2}(1 + \cos w)\sqrt{2 - \cos w} \quad (34)$$

The Q-factor and r are pre-defined in TQWT implementation. The scaling parameters β and α are assessed using these values as follows:

$$\beta = \frac{2}{Q+1} \quad (35)$$

$$\alpha = 1 - \frac{\beta}{r} \quad (36)$$

Using the following relation, the maximum number of decomposition level L_{\max} is assessed as follows:

$$L_{\max} = \left\lceil \frac{\log(\beta N/8)}{\log(1/\alpha)} \right\rceil \quad (37)$$

where the length of the input EEG signal is expressed as $p(n)$. The Q factor helps to control the number of oscillations of the wavelet. The overlapping of the frequency response is governed by the parameter r . If the frequency response is narrowed, then the Q-factor is increased. Therefore, to cover the frequency range of the EEG signal, more and

more number of decomposition levels are required. Overlapping of the adjacent frequency responses can be done by increasing the r value with a fixed Q-factor.

2.3.3. ELM algorithm

The ELM algorithm deals with randomly chosen input weights, hidden nodes and output weights which are analytically determined [39]. It has a pretty good generalization performance. A standard ELM classifier uses J hidden nodes where infinitely differentiable activation functions are used and it could easily approximate arbitrary samples with zero error. It implies that for a given training set $(x_i, T_i), i = 1, 2, \dots, N$ where $X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ and $T_i = [t_{i1}, t_{i2}, \dots, t_{ij}]^T \in \mathbb{R}^J$, there always exist β_q, w_i and b_q that makes the following equation true

$$\sum_{q=1}^J \beta_q f(w_i X_i + b_q) = o_i = T_i, i = 1, 2, \dots, N \quad (38)$$

where the weight vector that helps to connect the q^{th} hidden node with the output node is represented by β_q . The SLFNs output for the i^{th} sample is represented as o_i . The label vector of the i^{th} sample is represented as T_i and the weight vector connecting the i^{th} sample and q^{th} hidden node is denoted as w_i . The bias of the q^{th} hidden node is represented as b_q and the activation function is represented as $f(\cdot)$. The above equation can also be replaced with the following equation as follows:

$$H_{w,b,x,\beta} T = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_J^T \end{bmatrix}_{J \times J} ; T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times J} \quad (39)$$

where the hidden-layer output matrix is termed as $H_{w,b,x}$.

$$H(w_1, \dots, w_J, b_1, \dots, b_J, x_1, \dots, x_N) = \begin{bmatrix} f(w_1.x_1 + b_1) & \dots & f(w_J.x_N + b_J) \\ \vdots & \ddots & \vdots \\ f(w_1.x_N + b_1) & \dots & f(w_J.x_N + b_J) \end{bmatrix}_{N \times J} \quad (40)$$

By means of computing the corresponding least-squares solution $\beta = H_{w,b,x}^\dagger T$, the smallest training error can be achieved, where $H_{w,b,x}^\dagger$ represents the Moore-Penrose generalized inverse of $H_{w,b,x}$. The following three steps are present altogether in the ELM training algorithm.

Step 1: The hidden node parameter w_i and b_q are assigned randomly, $q = 1, 2, \dots, N$.

Step 2: The hidden-layer output matrix $H_{w,b,x}^\dagger$ and its Moore-Penrose generalized inverse $H_{w,b,x}^\dagger$ is computed.

Step 3: The output weight β is calculated.

2.3.4. ELM kernel algorithm

Minimizing the training error $T - H\beta^2$ and the norm of the output weights is aimed by the training process. A constrained optimization issue is used to represent the training process as follows:

Minimize

$$\Psi(\beta, \xi) = \frac{1}{2}\beta^2 + \frac{1}{2}C \sum_{i=1}^N \xi_i^2, \quad (41)$$

Subject to:

$$h(x_i)\beta = t_i - \xi_i, i = 1, 2, \dots, N \quad (42)$$

where the regularization factor C is used to control the tradeoff between the proximity of the training data to the smoothness of the decision function, so that there is a good improvement in the generalization performance. To solve the above optimization problem, Lagrange multiplier technique is used widely. If matrix $((I/C) + H^T H)$ is not singular, then the solution β is obtained as follows:

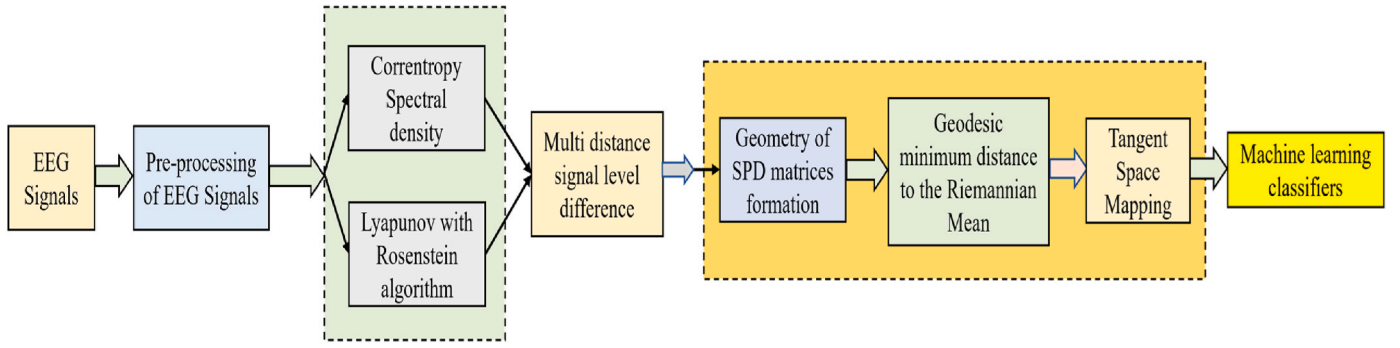


Fig. 4. Block diagram of the proposed technique 4.

$$\beta = H^T (I/C + H^T H)^{-1} T \quad (43)$$

Based on Mercer's condition, Kernel technique can be implemented into ELM [40]. Depending on the above equation, the output vectors $f(x)$ of ELM Kernel is specified as follows:

$$f(x) = h(x)\beta = h(x)H^T \left(HH^T + \frac{I}{C} \right)^{-1} T \quad (44)$$

$$f(x) = \begin{bmatrix} K(x_1, x_1) \\ K(x_1, x_2) \\ \vdots \\ K(x_1, x_N) \end{bmatrix}^T \left(\frac{I}{C} + K \right)^{-1} T \quad (45)$$

where,

$$K = H^T H = \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_N) \\ \vdots & & \vdots \\ K(x_N, x_N) & \dots & K(x_N, x_N) \end{bmatrix} \quad (46)$$

The number of training samples utilized for ELM Kernel is represented by N .

2.3.5. Adaptive neuro-fuzzy inference systems

To evaluate and assess the ability of the features in classifying the EEG signals, ANFIS is also used [41]. The features in the dataset are learnt by the ANFIS so that the system parameters can be adjusted well based on a given error criterion. For analyzing various types of signals, ANFIS is widely used. The training of the ANFIS classifiers is done with the backpropagation gradient descent technique and the least squares method so that the generalization is improved. The training of the ANFIS classifiers is done with these techniques using the input functions. The samples are given the binary target values of (1,0) and (0,1) respectively. By utilizing a generalized bell-shaped membership function, the design of the fuzzy rule architecture of the ANFIS classifiers was done as follows:

$$\mu_{ji}(p_i) = \left(1 + \left[\frac{p_i - c_{ji}}{a_{ji}} \right]^{2b_{ji}} \right)^{-1} \quad (47)$$

where the adaptable parameters are expressed as (a_{ji}, b_{ji}, c_{ji}) . Then first-order Sugeno type ANFIS models with the respective inputs and outputs are implemented. In the following form, the rules of the first order Sugeno fuzzy models are designed as:

$$\begin{aligned} R_i : IF (p_i \text{ is } U_{i1}, \dots, p_m \text{ is } U_{im}) \\ THEN q \text{ is } g_i(p_1, \dots, p_m) = b_0 + b_1 p_1 + \dots + b_m p_m \end{aligned} \quad (48)$$

where the i^{th} rule of the fuzzy system is expressed as $R_i, p_i (i = 1, \dots, m)$ are the inputs to the fuzzy system and q is the output of the fuzzy system. The adaptable parameters are $b_i (i = 0, 1, \dots, m)$. The ANFIS output is expressed as:

$$F = \frac{\sum_j g_j(a_1, a_2, \dots, a_m) \pi_i \mu_{ji}(a_i)}{\sum_j \pi_i \mu_{ji}(a_i)} \quad (49)$$

The degree of membership of $a_i (i = 1, 2, \dots, m)$ is expressed as $\mu_{ji}(a_i)$ to the antecedent linguistic term U_{ji} for the i^{th} rule of the fuzzy system.

2.4. Proposed technique 4

In this proposed approach 4, once the pre-processing is done, CSD and Lyapunov Rosenstein algorithm is implemented and then the Multi distance signal level difference is computed. Then Geometry of SPD matrices are computed following the assessment of geodesic minimum distance to the Riemannian mean is done. Finally, the concept of TSM is implemented and classified with Machine learning classifiers. The block diagram of the work is shown in Fig. 4.

2.4.1. Correntropy spectral density

The corresponding function $F(m)$ for a stationary and discrete signal $p(n), n = 1, 2, \dots, S$ is expressed as:

$$F(m) \triangleq E[K(p(n), p(n-m))] \quad (50)$$

where the number of EEG signal samples are represented by S and the time delta is represented as m . The expected value operated is indicated by $E[\cdot]$ and the symmetric positive-definite kernel is expressed by $K(\cdot, \cdot)$. The centered coentropy function is expressed as follows:

$$F_c(m) \triangleq E[K((p(n), p(n-m)))] - E_{p(n)} E_{p(n-m)} [K((p(n), p(n-m)))] \quad (51)$$

where the marginal expected values of the Kernel $K(\cdot, \cdot)$ with respect to $p(n)$ and $p(n-m)$ are expressed as $E_{p(n)}$ and $E_{p(n-m)}$ respectively. In this work, the kernel function used is a Gaussian Kernel function.

It is expressed as follows [39]:

$$K(p(n), p(n-m)) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(p(n) - p(n-m))^2}{2\sigma^2} \right] \quad (52)$$

The following equation is used to determine the σ bandwidth and it is represented as:

$$\sigma = 0.9AL^{-\frac{1}{3}} \quad (53)$$

where A denote a small value. The inner product of two vectors in feature space is represented by any symmetric, positive-definite Kernel represented as the reproductive kernel Hilbert space (RKHS). The inner product depends on the RKHS operator and $\Phi(p(n))$ is the mapping of signal $p(n)$ into feature space. Correntropy function is similar to an ordinary correlation function and it is also positive-definite and so it is utilized in many signal processing applications. CSD is dependent on the Fourier transform of the centered Correntropy function and it is a positive definite matrix. A favorable performance for signals with Gaussian and non-Gaussian statistics is exhibited by CSD. Spectral estimation

technique can be easily used by CSD as only a few parameters need to be set. The spectral resolution of the CSD is much better than the PSD due to the versatility of the coentropy function. The higher order statistical moments are extracted easily by the coentropy function and so the efficiency of CSD is easily enhanced.

2.4.2. Lyapunov exponents with rosenstein algorithm

The first step lies in the reconstruction aspect of the attractor dynamics from a single time series. Method of delays is used as one goal of the work so that a fast and versatile algorithm is developed [42]. The reconstructed trajectory p is expressed as a matrix where every row is a phase-space vector.

(i) Vector p_i in phase space:

$$p_i = (p(t_i)), p(t_i + \tau), \dots, p(t_i + (e - 1) * \tau) \quad (54)$$

where the reconstruction delay is expressed by τ , embedding dimension is represented as e and $t_i \in [1, T - (e - 1)\tau]$.

(ii) It is defined that λ , in the theory $d(t) = C \exp(\lambda(t))$, the divergence of the j^{th} pair of nearest neighbours diverge at a rate projected by the largest Lyapunov exponent and is represented as:

$$d_j(i) = C_j e^{(\lambda_1 \Delta t)}$$

where the initial separation is expressed as C_j . By computing the logarithm on both sides of the above equation,

$$\ln d_j(i) \approx \ln C_j + \lambda_1 (i \Delta t) \quad (55)$$

A collection of approximately parallel lines for $(j = 1, 2, \dots, m)$ is represented by the above equation where each slope is proportional to λ_1 . The calculation of the largest Lyapunov exponent is done using a least-square fit to the average line and is expressed as follows:

$$q(i) = \frac{1}{\Delta t} \langle \ln d_j(i) \rangle \quad (56)$$

where the average over all values of j is indicated as $\langle \cdot \rangle$. Computing the accurate values of λ_1 is done by this process of averaging. Normalizing the separation of the neighbours is done by C_j .

2.4.3. Multi distance signal level difference

A modification of the gray-level difference (GCD) is the multi distance signal level difference [43]. The absolute value of the difference of two adjacent pixels in the diagonal, horizontal and vertical directions give us the GLD. The calculation of the GLD in the horizontal direction is as follows:

$$q(i, j) = |p(i, j) - p(i, j + D)| \quad (57)$$

where the pixel distance is denoted by D . As the signal utilized is 1D (one dimensional), then the above equation can be modified as follows:

$$q_d(i) = |p(i) - p(i + d)|, \quad (58)$$

where $i = 1, 2, \dots, N - d$ and $d = 1, 2, \dots, k$.

2.4.4. Geometry of SPD matrices

Assuming A_i specifies a short segment of continuous EEG signals, A_i is indicated as follows:

$$A_i = [A_{t+T_s}, \dots, A_{t+T_s+T_s-1}] \in \mathbb{R}^{n \times T_s} \quad (59)$$

where the i^{th} trial of signal movement initiating at time $t = T_i$ is indicated by A_i . The number of sampled points of the selected segment is indicated by T_s . The spatial covariance matrix (SCM) $P_i \in \mathbb{R}^{n \times n}$ for the i^{th} trial is computed as follows:

$$P_i = \frac{1}{T_s - 1} P_i P_i^T \quad (60)$$

There are two ways to classify EEG signals depending on the SCM in the Riemannian manifold and it is through the concept of computation of the Geodesic minimum distance to the Riemannian mean and the calculation of TSM followed by the analysis with machine learning classifiers.

2.4.5. Geodesic minimum distance to the riemannian mean

Between the two SPD matrices M_1 and M_2 in $M(n)$, the Riemannian distance [44] is expressed as follows:

$$\delta_R(M_1, M_2) = \|\log(M_1^{-1} M_2)\|_F = \left[\sum_{i=1}^n \log^2 \lambda_i \right]^{1/2} \quad (61)$$

If there are Z SPD matrices M_1, \dots, M_Z , the geometric mean (GM) in the Riemannian sense is expressed as follows:

$$GM(M_1, \dots, M_Z) = \argmin_{M \in M(n)} \sum_{i=1}^Z \delta_R^2(M, M_i) \quad (62)$$

The Riemannian distance between two unknown class M to the Riemannian mean point of each class is computed so that the unknown class is classified into categories which corresponds to the shortest distance. This is achieved by using Riemannian distance to Riemannian mean (MDRM).

2.4.6. Tangent space mapping (TSM)

With the help of a differentiable Riemannian manifold F , the indication of the SPD matrix of M is done. Every tangent vector S_i is obtained as the derivative at $t = 0$ between M and the exponential mapping $M_i = \text{Exp}_M(S_i)$ and is expressed as follows:

$$\text{Exp}_M(S_i) = M_i = M^{1/2} \exp(M^{-1/2} S_i M^{-1/2}) M^{1/2} \quad (63)$$

The logarithmic mapping is used to express the inverse mapping and is expressed as follows:

$$\log_M(M_i) = S_i = M^{1/2} \log(M^{-1/2} M_i M^{-1/2}) M^{1/2} \quad (64)$$

The Riemannian mean of $I > 1$ SPD matrices is evaluated using the Riemannian geodesic distance as:

$$GD(M_1, \dots, M_I) = \argmin_{M \in M(n)} \sum_{i=1}^I \delta_R^2(M, M_i) \quad (65)$$

With the help of the tangent space which is situated at the geometric mean of the whole set trials

$$M_{GD} = GD(M_i, i = 1, \dots, I) \quad (66)$$

The mapping of each SCM M_i is done into this tangent space so that the set of $z = n(n+1)/2$ dimensional vectors are yielded and it is represented as [45]:

$$S_i = \text{upper} \left(M_{GD}^{-1/2} \log(M_i) M_{GD}^{-1/2} \right) \quad (67)$$

Numerous classification algorithms are implemented in the Riemannian space.

2.5. Proposed technique 5

In this proposed work, once the pre-processing of EEG signals are done, HHT and Hilbert marginal spectrum is computed and the best features are obtained by Black Hole optimization algorithm and finally it is classified by the cascade Adaboost classifier. The block diagram of the

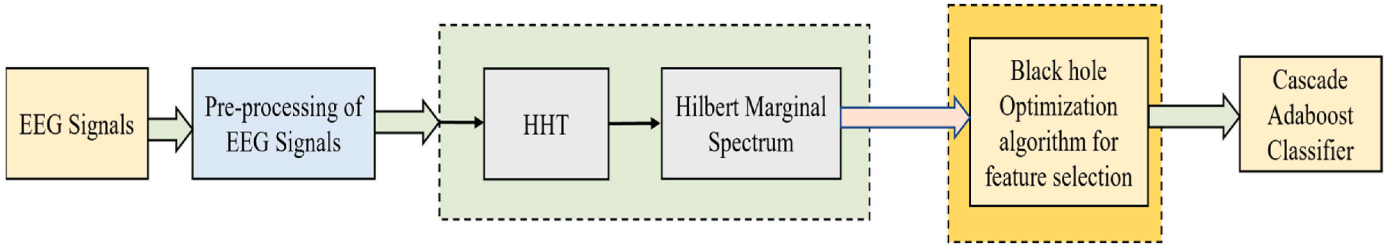


Fig. 5. Block Diagram of the proposed technique 5.

work is shown in shown in Fig. 5.

2.5.1. HHT and hilbert marginal spectrum

To analyze non-linear and non-stationary signals, HHT is highly useful [46]. It is generally not subject to the Heisenberg Uncertainty principle. Empirical Mode Decomposition (EMD) and HT are the two important parts of HHT. A new adaptive EEG signal time-frequency processing technique is EMD. The EEG signal $p(t)$ is decomposed by into a finite intrinsic mode function (IMF) based on the signal scale. The local characteristic EEG signals of various time scales of the original EEG signal are contained by the IMF. Two important criteria are present for understanding the IMF concept. Throughout the data segment, the number of zero crossings and the number of extreme points must be more or less equal. Secondly, the average of the lower envelope found by the local minimum points and the average of the upper envelope formed by the local maximum points is zero. From the upper and lower envelope sequences, the average curve sequences are obtained for a given EEG signal $p(t)$ is as follows:

$$m_1(t) = \frac{p_{\max}(t) + p_{\min}(t)}{2} \quad (68)$$

The residual component of the data is obtained by finding a new time series and is represented as the difference between $p(t)$ and the average curve sequence.

$$h_1(t) = p(t) - m_1(t) \quad (69)$$

Unless the judgement criteria are not satisfied, the above two processes are repeated and the remaining component is obtained as:

$$r_2(t) = p(t) - h_1(t) \quad (70)$$

Until the termination condition is satisfied, the remaining sequence repeats the equations so that multiple IMFs are obtained.

$$p(t) = \sum_{i=1}^n c_i + r_n \quad (71)$$

The IMF component is represented as c_i and the residual function is denoted by r_n . A Hilbert transform is performed on c_i and an analytic signal is constructed and a Hilbert marginal spectrum is obtained [47]. Convolving the signal with $1/\pi t$ is essentially a 90-degree phase shifter and it is performed by HT. The efficiency of the signal is improved by the Hilbert marginal spectrum. A higher frequency resolution is obtained by the Hilbert marginal spectrum and it also represents a variable frequency.

$$H[c_i] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(\tau)}{t - \tau} d\tau \quad (72)$$

$$s_i(t) = c_i(t) + jH[c_i] = a_i(t)e^{j\psi_i(t)} \quad (73)$$

The instantaneous amplitude is represented by $a_i(t)$ and the instantaneous phase is represented by $\psi_i(t)$.

$$H(f, t) = \text{Re} \sum_{i=1}^n a_i(t) e^{j2\pi \int f_i(t) dt} \quad (74)$$

$$h(i) = \int_{-\infty}^{\infty} H(f, t) dt \quad (75)$$

The instantaneous frequency is represented by $f_i(t)$.

2.5.2. Black-hole algorithm

The essential features are then chosen by the Black-hole algorithm [48]. Black-hole algorithm is used here as it is relatively easy to implement and is somewhat free from the tuning parameter issues. This algorithm utilizes a method of exploration/exploitation which is totally free of external components so that the probability of being affected to unexpected changes is reduced drastically. In every evaluation, the black hole algorithm in optimization problem converges to global optimal, whereas some of the other standard algorithms like PSO, ACO and GA might get stuck in local optimal solutions. The black hole algorithm is inspired by the law of attraction/absorption and it is dependent on the phenomenon of the same name which occurs in outer space. Three main fundamentals are present in this algorithm as follows.

- A star present in space is considered as a solution to the problem. The random generation of a certain number of stars are done as it is a population-based algorithm.
- The selection of the black hole is done. A black hole indicates the star which has the best performance of all the solutions.
- By using the absorption formula, the movement and generation of new stars are carried out as follows:

$$x_i^d(t+1) = x_i^d(t) + \alpha [x_{bh}^d - x_i^d(t)], \forall i \in \{1, \dots, n\} \quad (76)$$

where the d^{th} component of the i^{th} star in the iteration t is represented as $x_i^d(t)$. The d^{th} component of the black hole in the search space is represented as x_{bh}^d . The number of solutions is represented by n and a random uniform number of distributions between zero and one is represented as α . The d^{th} component of the location of the i^{th} star in the next iteration is represented by $x_i^d(t+1)$. The event horizon is actually a radius R which is originated by the black hole. The black hole can be easily absorbed and destroyed in case a star crosses the horizon, so that a new star could be created randomly. It is observed as the probability of crossing the even horizon and is computed as follows:

$$R = \frac{f_{bh}}{\sum_{i=1}^n f_i} \quad (77)$$

where the performance value that has the best solution is represented by f_{bh} and the value associated with the quality of the i^{th} star is represented by f_i and the number of stars is represented by n . The star crosses the event horizon when the distance between the black hole and the star is less than the radius. The absorption of this star happens and the random generation of the new one is done. The variability offered by even horizon is highlighted so that the common problem of local optimum stagnation is resolved completely.

Approximation to good solutions is the most vital characteristic of incomplete data processing algorithm and this concept is utilized as stop

criteria. It is quite difficult to assess the quality of found solutions in situations where the optimal solutions are not known apriori. In such a case, the ideal stop criteria depend on the total number of executed iterations in the algorithm. The stop criteria in our work are initially set as 1000 off-line iterations. The optimization procedure is displayed in Algorithm 1. The random generation of the initial n star population for every intrinsic signal starts accordingly.

Algorithm 1: Black hole Algorithm implemented to the concept.

Input: number of stars (solution), stop criteria (maximum of iterations).

Output: the black hole, the best feature.

- (1) Generate the first generation of n stars in the search space.
- (2) Select the ideal solutions as the black hole.
- (3) While a good enough solution has not been reached in a maximum of iterations do
- (4) For all stars s_i , ($\forall i = 1, \dots, n$) do
- (5) Classifier Feature performance
- (6) Location changes of s_i happens according to equation (76)
- (7) If s_i is better than black hole, then
- (8) Select the current solution s_i as black hole.
- (9) End if
- (10) $\{S_{\text{cross}}$ to the event horizon defined by equation (77)}
- (11) If $r > f_{bh} / \sum_{i=1}^n f_i$ then
- (12) Replace s_i with a new star in search space.
- (13) End if
- (14) End for
- (15) End while
- (16) Return results.

The degree of variability in an algorithm is allowed by randomness. The process of absorption of the algorithm in the loop statement is carried out. The calculation of the quality of each solution is done and it is assessed by the performance of the feature. The solution is considered to have a high quality if the rating value is close to 1. The solution is considered to be of low quality if the rating value is close to 0. When the stars are absorbed by the black hole, the solutions are generated. For each intrinsic signal, this process generates a real number of predictabilities. The locations are swapped if a star reaches a value which is better than the black hole. A new value is generated randomly if a star crosses the event horizon of the black hole. It is evaluated based on a random variable with uniform distribution $r \sim [0, 1]$ and this entire procedure is fully iterated. A proportion between the fitness of the star and the combined value of all fitness is used so that the quality of the solutions is measured and this value is termed as event horizon. The absorption of the star will be done of this percentage value is less than r which is generated in a random manner. Thus, a good variability to the solutions is provided by the non-deterministic process. Finally, when an adequate solution is reached, the loop statement terminates. By means of updating the solution in a certain number of iterations, this condition can be easily determined. The best solutions are memorized and evaluated at the end.

2.5.3. Cascade-Adaboost classifier

The features selected by the Black-hole optimization algorithm are then fed to the Cascade-Adaboost classifier [49]. A parallel classifier is Adaboost classifier which combines many linear weak classifiers. The weak classifier can be enhanced by the AdaBoost training algorithm so that self-adaptive goals can be easily achieved. In the given feature set, the focus is on the classification of one dimension by every weak classifier. By adding a lot of weak classifiers, several key features can be easily focused on by the AdaBoost classifier. By adding a lot of weak classifiers, several key features can be easily focused on by the AdaBoost classifier. Once the addition of the weak classifier is done, the employment of the minimum error is calculated so that the weight value of this

weak classifier is computed. Then for all the training samples, the weight values of this weak classifier are computed. Then for all the training samples, the weight values are re-adjusted and ultimately these weight values are chosen as an input to the next training iteration. In every training iteration, a new weak classifier is added and so there is a minimal improvement in the Adaboost classifier. Ultimately, the result of the strong classifier $H(F)$ is expressed as follows:

$$H(F) = \sum_{t=1}^T \beta_t h_t(f_i) \quad (78)$$

where the feature set is expressed as $F = \{f_1, f_2, f_3, f_4, f_5\}$. The weak classifier appended in the t^{th} training iteration is expressed as h_t . The key feature selected by the classifier h_t is expressed as f_i . The corresponding weight value of every weak classifier is represented by β_t and the number of training iteration is represented by T . By training weak classifiers, the discrimination of features is explored and so the Adaboost classifiers are organized in a cascade way. Therefore, simple weak classifiers are considered with which the target of the cascade-Adaboost classifiers is also manipulated. The simple threshold classifiers are chosen as weak classifiers and is expressed as follows.

$$h_t(f_i) = \begin{cases} 1 & T_{\text{lower}} \leq f_i \leq T_{\text{upper}} \\ 0 & \text{otherwise} \end{cases} \quad (79)$$

where T_{lower} and T_{upper} represented the threshold for the weak classifier h_t and is obtained using any search technique. Now the positive training samples and negative training samples are initiated with different weight values. Assuming $\{f_i, y_i\}$ be the training samples, $f_i, i = 1, \dots, n$ is the feature set, $y_i \in \{1, -1\}$ indicate the snoring or non-snoring class. Assuming that there are p positive samples and q negative samples in the training set, their respective weight values can be projected as follows:

$$w_p = q.w_n \quad (80)$$

$$p.w_p + q.w_n = 1 \quad (81)$$

where the number of positive samples is represented by p and the number of negative samples is represented by q . The weight values of positive samples and negative samples are represented by w_p and w_n respectively. The sum of all these T weak classifiers give us the strong classifier $H(F)$. The strong classifier is expressed as follows:

$$H(F) = \max_s \sum_{t=1}^T \beta_t^s h_t^s(f_i^s) \quad (82)$$

where $f_i^s, i = 1, \dots, n$ are the features at scale s . The training of the T weak classifiers are done for each scale s so that the corresponding weight values β_t^s are computed. By adding a lot of new weak classifiers, the overall accuracy of Adaboost classifier can increase but still it could fail if some challenging scenarios are present. Therefore, multiple Adaboost classifiers are organized in a cascade fashion using cascade-Adaboost classifier which can enhance accuracy. If the input feature vector $F = \{f_1, \dots, f_n\}$ which is selected by the Adaboost classifier is assessed as a negative sample, then it would be removed in the training set and as a result it would not progress to the next Adaboost classifier. Consequently, the number of training samples in training set is mitigated if the input feature vector F is assessed as a positive sample, then this particular training sample progresses to the next layer for classification. The training samples on the end layer are quite similar to each other and so there is a good focus on similar training samples by the Adaboost classifiers. As a result, cascade-Adaboost classifiers have the versatility to classify the selected features efficiently giving a high classification accuracy.

Table 1

Results of the MDS + SVD based FCM-SVR technique.

Dimensionality Reduction Technique	Classification Technique	Classification Accuracy
MDS	FCM + SVR	92.46%
SVD	FCM + SVR	91.47%
Proposed MDS + SVD	FCM + SVR	97.99%

Table 2

Results of the proposed MFA-conditional feature mapping – cross domain transfer learning with machine learning classifiers.

Proposed Method	NBC	KNN	SVM	Adaboost
Proposed MFA + Conditional Feature Mapping + Cross domain Transfer learning	98.17%	97.89%	99.12%	98.98%

Table 3

Results of the proposed FAWT-TQWT with ELM, KELM and ANFIS classifiers.

Proposed Method	ELM	KELM	ANFIS
FAWT	89.45%	92.34%	93.33%
TQWT	91.46%	94.56%	91.25%
Proposed FAWT + TQWT	94.54%	97.89%	98.18%

Table 4

Results of the proposed CSD and Lyapunov exponents with multi distance signal level based tangent space mapping with machine learning classifiers.

Proposed Method	NBC	KNN	SVM	Adaboost	DT	RF
Proposed CSD and Lyapunov exponents with Multi distance signal level based Tangent Space Mapping	98.45%	96.54%	99.13%	97.45%	97.89%	95.66%

3. Results and discussion

For this study, a publicly available EEG dataset was used [50]. For the acquisition of EEG signals, a brain cap with 32 electrodes was used. The utilized dataset in this work has EEG signals with 32 channels. The age group of the subjects was between 17 and 25 and the EEG data was collected from 16 healthy subjects (8 males and 8 female). With the help of a driving simulator and a brain cap, the EEG signals were collected. Prior to the experiment, no stimulants were used by subjects such as tea, coffee, alcohol, beer, energy drinks etc. The standard metrics utilized here for the analysis was Sensitivity, Specificity and Accuracy and is represented by the following formulae as follows:

Table 5

Results of the proposed HHT-Hilbert Huang Transform with BH based cascade Adaboost classifier.

Proposed Method	Cascade Adaboost Classifier				
	Black Hole Optimization	Artificial Bee Colony Optimization	Ant Colony Optimization	Particle Swarm Optimization	Glowworm Swarm Optimization
Proposed HHT + Hilbert Marginal Spectral with swarm intelligence-based cascade Adaboost classifier	98.08%	94.34%	91.34%	97.98%	96.99%

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (83)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (84)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (85)$$

where TP means True Positive, TN means True Negative, FP indicates False Positive and FN indicates False Negative respectively. To obtain the results for the proposed works, a 10-fold cross-validation scheme has been utilized in our study. As far as the first proposed strategy is concerned, the kernel function used is RBF in FCM-SVR model and the value of C is set as 50. For the KNN classifier used in this work, the value of K is set as 8 in our experiment. The Gamma value used for SVM is set as 0.04 in our experiment. Table 1 shows the result analysis of the proposed MDS and SVD with FCM-SVR technique where a high classification accuracy of 97.99% is obtained. Table 2 shows the proposed method of using MFA and conditional feature mapping and cross domain transfer learning with machine learning classifiers, where a high classification accuracy of 99.12% is obtained with SVM classifier. Table 3 shows the proposed method of using FAWT and TQWT hybrid classifier with ANFIS reporting a high classification accuracy of 98.18%. Table 4 shows the proposed CSD and Lyapunov exponents with multi distance signal level based tangent space mapping and SVM classifier reporting a classification accuracy of 99.13% for SVM classifier. Table 5 shows the proposed HHT with Hilbert Huang transform with the proposed BH based cascade Adaboost classifier reporting an accuracy of 98.08%.

Fig. 6 shows the performance comparison of classification accuracy among the proposed methods. It is evident a very high classification accuracy is obtained for the proposed CSD and Lyapunov exponents with multi distance signal based tangent space mapping concept with SVM machine learning classifier.

3.1. Comparison with other works

Table 6 shows the current works with the previous works comparison. Our works produced a good classification accuracy when compared to the previous works, however Tuncer et al. [10], produced a slightly higher classification accuracy by more or less 1% increase in a few cases. Our future works aim to improve a very high classification accuracy by incorporating a lot of other versatile methods, but comparatively our methods produce a good classification accuracy.

The proposed MDS + SVD + FCM-SVR produced a classification accuracy of 97.99% and the reason for this could be attributed to the intrinsic properties of MDS and SVD and as it was combined with FCM-SVR, this high classification accuracy was obtained. The Proposed MFA + Conditional Feature Mapping + Cross domain transfer learning with machine learning classifiers produced a high classification accuracy of 99.12% and the reason for obtaining this high classification accuracy could be attributed to the combination factors of MFA and conditional feature mapping accompanied with cross domain transfer learning. The proposed FAWT + TQWT with ANFIS produced a high classification accuracy of 98.18% and this reason is due to the combined performance of the properties of the versions of the wavelet transform with ANFIS

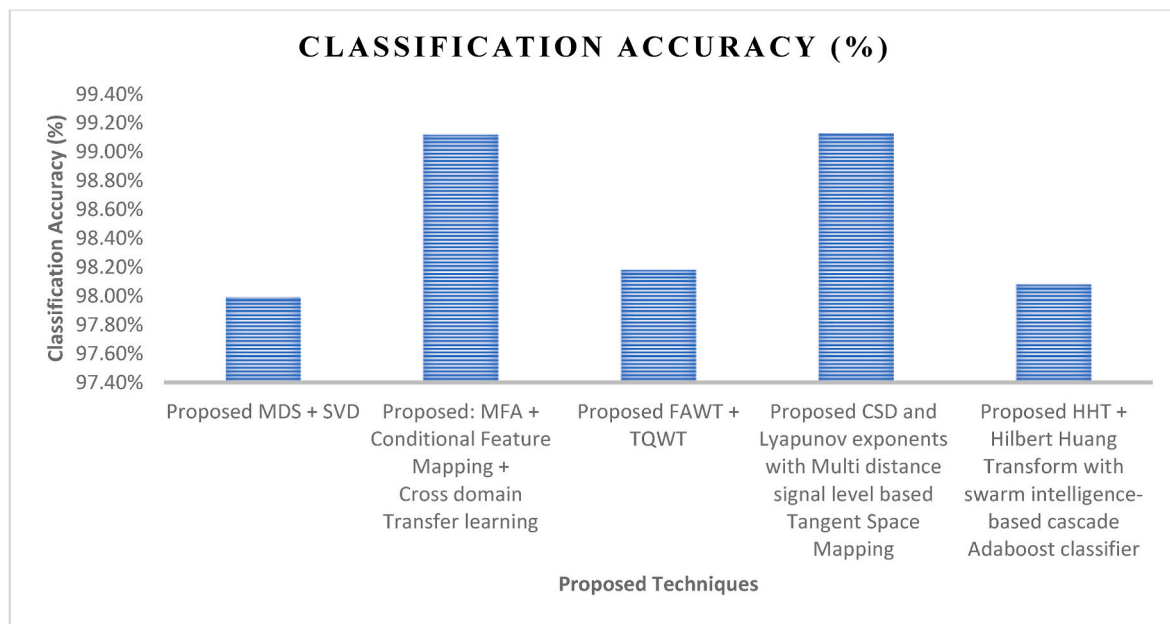


Fig. 6. Performance comparison of Classification Accuracy among the proposed methods.

Table 6

Performance Comparison table with our work and previous works.

Authors	Dataset	Method Used	Classification Accuracy
Luo et al. [11]	Collected Dataset	Adaptive multiscale entropy	95.37%
Wang et al. [14]	Collected Dataset	Wavelet entropy	90.70%
Li et al. [15]	Collected Dataset	Deep Belief Network	98.86%
Tuncer et al. [10]	Qui dataset	1D-DWT + RFINCA + Fine KNN	100%
Proposed Works	Qui dataset	1D-DWT + RFINCA + MEDIUM KNN	99.38%
		1D-DWT + RFINCA + WEIGHTED KNN	99.79%
		1D-DWT + RFINCA + BOOSTED TREE	98.12%
		1D-DWT + RFINCA + BAGGED TREE	98.33%
		1D-DWT + RFINCA + SUBSPACE KNN	99.48%
		Proposed MDS + SVD with FCM-SVR	97.99%
		Proposed MFA + Conditional Feature Mapping + Cross domain Transfer learning	99.12%
		Proposed FAWT + TQWT with ANFIS	98.18%
		Proposed CSD and Lyapunov exponents with Multi distance signal level based Tangent Space Mapping with SVM	99.13%
		Proposed HHT + Hilbert marginal spectrum with BH and cascade Adaboost classifier	98.08%

classifier. The proposed CSD and Lyapunov exponents with multi distance signal level based TSM with SVM produced the highest classification accuracy of 99.13% and the reason can be contributed to the nature of the components used in this experiment, that when it was hybrid together a higher classification accuracy was obtained. Finally, when the proposed HHT with Hilbert marginal spectrum and black hole optimization with cascade Adaboost classifier was obtained, a high

classification accuracy of 98.08% was obtained and this could be due to the contribution of the best feature selection technique chosen here (BH optimization) so that only the best features are considered for classification producing a high classification accuracy.

The merits of the work include the selection of the best components or ingredients for the hybrid models to be developed. Each and every individual model has some beautiful properties embedded within it and when such individual models are hybrid together a good ensemble model is obtained which performs with a higher classification accuracy. The limitations of this work include the non-applicability of the hybrid models to be used in telemedicine and wireless applications for remote health care systems, which the authors intend to address in future.

4. Conclusion and future work

For the main reasons causing traffic accidents, the primary reason can be attributed to the drivers' fatigue. Estimating and preventing driving fatigue is one of the main intentions of automated techniques. In the past decades, various techniques have been proposed to detect the fatigueness of the driver such as video imaging technology, vehicle sensor parameter and physiological signals like EOG, EEG etc. Research has shown that EEG is an effective technique to identify driver's fatigue. As EEG is relatively inexpensive and has a high temporal resolution, EEG is widely preferred. To analyze and detect the fatigueness of the driver, many algorithms based on EEG signals have been proposed. In this work, five efficient strategies were proposed and the high classification accuracy of 99.13% was obtained when the CSD and Lyapunov exponents with multi distance signal level based tangent space mapping and SVM is utilized. The second-best classification accuracy of 99.12% is obtained with the proposed MFA and conditional feature mapping and cross domain transfer learning technique is utilized. The third best classification accuracy of 98.18% is obtained when the proposed FAWT and TQWT technique is utilized with ANFIS classifier. Future works aim to incorporate a plethora of other machine learning techniques, deep learning techniques and transfer learning techniques for developing an efficient fatigue detection system from EEG signals.

Credit author statement

Sunil Kumar Prabhakar: Conceptualization, Methodology, Software,

Data curation, Validation, Writing Original draft preparation, Formal Analysis, Investigation, Dong-Ok Won: Visualization, Supervision, Writing- Reviewing and Editing, Project Administration, Funding Acquisition.

Funding and acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A5A8019303) and partly funded by the Ministry of Education (RS-2023-00250246) and partly supported by the Bio&Medical Technology Development Program of the NRF funded by the Korean government (MSIT) (No. RS-2023-00223501) and partly supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-02068, Artificial Intelligence Innovation Hub).

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.

Data availability

Publicly available dataset is obtained from "T. Qiu, Data for: Research on Fatigue Driving Detection Based on Adaptive Multi-scale Entropy, Mendeley Data, 2019."

References

- [1] Liu JP, Zhang C, Zheng CX. Estimation of the cortical functional connectivity by directed transfer function during mental fatigue. *Appl Ergon* 2011;42(1):114–21.
- [2] Gharagozlou F, Nasl Saraji G, Mazloumi A, et al. Detecting driver mental fatigue based on EEG alpha power changes during simulated driving. *Iran J Public Health* 2015;44(12):1693–700.
- [3] Zhao C, Zheng C, Zhao JM, Tu LiuY. Automatic classification of driving mental fatigue with EEG by wavelet packet energy and KPCA-SVM. *International Journal of Innovative Computing, Information and Control* 2011;7(3):1157–68.
- [4] Yang G, Lin Y, Bhattacharya P. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Inf Sci* 2010;180(10):1942–54.
- [5] Zhang G, Yau KKW, Zhang X, Li Y. Traffic accidents involving fatigue driving and their extent of casualties. *Accid Anal Prev* 2016;87:34–42.
- [6] Fletcher A, McCulloch K, Baulk SD, Dawson D. Countermeasures to driver fatigue: a review of public awareness campaigns and legal approaches. *Aust N Z J Publ Health* 2005;29(5):471–6.
- [7] Hsieh C-S, Tai C-C. An improved and portable eye-blink duration detection system to warn of driver fatigue. *Instrum Sci Technol* 2013;41(5):429–44.
- [8] Ren Z, Li R, Chen B, Zhang H, Ma Y, Wang C, Lin Y, Zhang Y. EEG-based driving fatigue detection using a two-level learning Hierarchy radial Basis function. *Front Neurosci* 2021;15:618408. <https://doi.org/10.3389/fnbot.2021.618408>.
- [9] Gao D, Tang X, Wan M, Huang G, Zhang Y. EEG driving fatigue detection based on log-Mel spectrogram and convolutional recurrent neural networks. *Front Neurosci* 2023;17:1136609. <https://doi.org/10.3389/fnins.2023.1136609>.
- [10] Tuncer T, Dogan S, Subasi A. EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection. *Biomed Signal Process Control* July 2021;68:102591.
- [11] Luo H, Qiu T, Liu C, Huang P. Research on fatigue driving detection using forehead EEG based on adaptive multi-scale entropy. *Biomed. Signal Process. Control* 2019; 51:50–8.
- [12] Chaudhuri A, Routray A. Driver fatigue detection through chaotic entropy analysis of cortical sources obtained from scalp EEG signals. *IEEE Trans Intell Transport Syst* 2019;21.
- [13] Chai R, Ling SH, San PP, Naik GR, Nguyen TN, Tran Y, Craig A, Nguyen HT. Improving EEG-based driver fatigue classification using sparse-deep belief networks. *Front Neurosci* 2017;11:103.
- [14] Wang Q, Li Y, Liu X. Analysis of feature fatigue EEG signals based on wavelet entropy. *Int J Pattern Recogn Artif Intell* 2018;32. Article 1854023.
- [15] Li P, Jiang W, Su F. Single-channel EEG-based mental fatigue detection based on deep belief network 2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC). *IEEE*; 2016. p. 367–70.
- [16] Subasi A, Saikia A, Bagedo K, Singh A, Hazarika A. EEG-based driver fatigue detection using FAWT and multibootstrapping approaches. *IEEE Trans Ind Inf Oct* 2022; 18(10):6602–9. <https://doi.org/10.1109/TII.2022.3167470>.
- [17] Zhang T, Chen J, He E, Wang H. Sample-entropy-based method for real driving fatigue detection with multichannel electroencephalogram. *Appl Sci* 2021;11: 10279. <https://doi.org/10.3390/app112110279>.
- [18] Ma Y, Chen B, Li R, Wang C, Wang J, She Q, Luo Z, Zhang Y. Driving fatigue detection from EEG using a modified PCANet method. *Comput Intell Neurosci* 2019;2019. <https://doi.org/10.1155/2019/4721863>. Article ID 4721863, 9 pages.
- [19] Min J, Wang P, Hu J. Driver fatigue detection through multiple entropy fusion analysis in an EEG-based system. *PLoS One* 2017;12(12):e0188756. <https://doi.org/10.1371/journal.pone.0188756>.
- [20] Zhu M, Chen J, Li H, et al. Vehicle driver drowsiness detection method using wearable EEG based on convolution neural network. *Neural Comput Appl* 2021;33: 13965–80. <https://doi.org/10.1007/s00521-021-06038-y>.
- [21] Tian Y, Cao J. Fatigue driving detection based on electrooculography: a review. *J Image Video Proc* 2021;2021:33. <https://doi.org/10.1186/s13640-021-00575-1>.
- [22] Arefnezhad S, Hamet J, Eichberger A, et al. Driver drowsiness estimation using EEG signals with a dynamical encoder–decoder modeling framework. *Sci Rep* 2022;12:2650. <https://doi.org/10.1038/s41598-022-05810-x>.
- [23] Zeng H, Yang C, Zhang H, Wu Z, Zhang J, Dai G, Babiloni F, Kong W. A LightGBM-Based EEG analysis method for driver mental states classification. *Comput Intell Neurosci* 2019. Article ID 3761203.
- [24] Mu Z, Hu J, Yin J. Driving fatigue detecting based on EEG signals of forehead area. *Int J Pattern Recogn Artif Intell* 2017;31(no. 05):1750011.
- [25] Rodriguez DGG, et al. EEG based drivers mental fatigue detection using ERD/ERS and Hurst exponent, ICMHI'22. *Proceedings of the 6th International Conference on Medical and Health Informatics May* 2022:159–62.
- [26] Wang P, Min J, Hu J. Ensemble classifier for driver's fatigue detection based on a single EEG channel. *IET Intell Transp Syst Dec* 2018;10(issue 10):1322–8.
- [27] Dogan S, Tuncer I, Baygin M, et al. A new hand-modeled learning framework for driving fatigue detection using EEG signals. *Neural Comput Appl* 2023;35: 14837–54. <https://doi.org/10.1007/s00521-023-08491-3>.
- [28] Tuncer T, Dogan S, Ertam F, et al. A dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals. *Cogn Neurodyn* 2021;15:223–37. <https://doi.org/10.1007/s11571-020-09601-w>.
- [29] Xu N, Gao X, Hong B, Miao X, Gao S, Yang F. BCI competition 2003–data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications. *IEEE (Inst Electr Electron Eng) Trans Biomed Eng* 2004;51(6): 1067–72.
- [30] Zhang Y, Zhang X, Liu W, Luo Y, Yu E, Zou K, Liu X. Automatic sleep staging using multi-dimensional feature extraction and multi-kernel fuzzy Support vector machine. *Journal of Healthcare Engineering* 2014;5. Article ID 410705, 16 pages.
- [31] Wu F, Chou C, Lu Y, Chung J. Modeling electromechanical overcurrent relays using singular value decomposition. *J Appl Math* 2012;2012. <https://doi.org/10.1155/2012/104952>. Article ID 104952, 18 pages.
- [32] Bezdek JC. Pattern recognition with fuzzy objective function algorithms. New York, NY, USA: Plenum Press; 1981.
- [33] Peng X. TSVM: an efficient twin Support vector machine for regression. *Neural Network* 2010;23(3):365–72.
- [34] Yan S, Xu D, Zhang B, Zhang H-J, Yang Q, Lin S. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans Pattern Anal Mach Intell* 2007;29(1):40–51.
- [35] Jiang Y, Zheng Y, Hou S, Chang Y, Gee J. Multimodal image alignment via linear mapping between feature modalities. *Journal of Healthcare Engineering* 2017; 2017. <https://doi.org/10.1155/2017/8625951>. Article ID 8625951, 6 pages.
- [36] Li S, Song SJ, Huang G, Wu C. Cross-domain extreme learning machines for domain adaptation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 2019;49 (6):1194–207.
- [37] Sharma M, Pachori RB, Rajendra Acharya U. A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recogn Lett* 2017;94:172–9.
- [38] Baygin M, Barua PD, Chakraborty S, et al. CCPNet136: automated detection of schizophrenia using carbon chain pattern and iterative TQWT technique with EEG signals. *Physiol Meas* 2023 March 14;(3):44.
- [39] Huang GB, Zhu QH, Siew CK. Extreme learning machine: theory and applications. *Neurocomputing* 2006;70(1–3):489–501.
- [40] Huang GB, Wang DH, Lan Y. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics* 2011;2(2):107–22.
- [41] Wu K. Analysis of the scientific meaning of several concepts related to entropy. *J Dialectics Nat* 1996;5:67–74.
- [42] Yang C, Wu Q. On stabilization of bipedal robots during disturbed standing using the concept of Lyapunov exponents. *Robotica* 2006;24(5):621–4.
- [43] Brković M, Simić M. Multidimensional optimization of signal space distance parameters in WLAN positioning. *Sci World J* 2014;2014. <https://doi.org/10.1155/2014/986061>. Article ID 986061, 6 pages.
- [44] O'Neill B. Semi-riemannian geometry. New York, NY, USA: Academic Press; 1983.
- [45] Zhang Z, Zha H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J Sci Comput* 2005;26(1):313–38.
- [46] Gu P, Wen YK. A record-based method for the generation of tridirectional uniform hazard-response spectra and ground motions using the Hilbert-Huang transform. *Bull Seismol Soc Am* 2007;97(5):1539–56.
- [47] Han D, Wei S, Shi P, Zhang Y, Gao K, Tian N. Damage identification of a derrick steel structure based on the HHT marginal spectrum amplitude curvature difference. *Shock Vib* 2017;2017. <https://doi.org/10.1155/2017/1062949>. Article ID 1062949, 9 pages.

- [48] Farahmandian M, Hatamlou A. Solving optimization problems using black hole algorithm. *J Adv Comput Sci Technol* 2015;4:68–74.
- [49] Wang Z, Yoon S, Xie S, Lu Y, Park DS. A high accuracy pedestrian detection system combining a cascade AdaBoost detector and random vector functional-link net. *Sci World J* 2014;2014. <https://doi.org/10.1155/2014/105089>. Article ID 105089, 7 pages.
- [50] Qiu T. Data for: research on fatigue driving detection based on adaptive multi-scale entropy. *Mendeley Data*; 2019.