Asian Journal of Information Technology 15 (14): 2399-2410, 2016
ISSN: 1682-3915

An Innovative Optimization Technique for Drowsiness Detection Based on Feature Extraction Capitalizing Neural Network and Sparse Classifiers

[1]Reena Daphne and [2]A. Albert Raj
[1]Department of Electrical and Electronics Engineering, Stella Mary's College of Engineering, Aruthengavilai, Azhikal Post, Kanyakumari District Tamil Nadu, India
[2]DMI Engineering College, Kumarapuram Road, Aralvaimozhi, Kanayakumari, Kanayakumari District Tamil Nadu, India

Abstract: Drowsiness is considered as a significant risk factor that contributes to large number of accidents. This study, focuses on methodologies developed for counteracting its effects with very accurate classification techniques categorizing the different drowsy states and alerting the person at definite instants. An optimal Bootstrap technique is applied to features extracted by Daubechies Wavelet Transform (DWT) and the drowsy states are classified using Neural Network (NN) classifier. The Receiver Operating Characteristics (ROC) curve shows the classification accuracy and the computation time is also calculated. In order to improve the efficiency of the proposed method, Fractional Fourier Transform (FrFT) based feature extraction is implemented with ABC (Artificial Bee Colony) for optimization and classification done using NN and Sparse classifiers. The three methods exhibit high efficiency in improving the system's performance in terms of accuracy F1 score and computation time. A comparative study of the three methods is also done with the latter showing better results.

Key words: Electroencephalogram (EEG), ROC, drowsiness detection, feature extraction, WPT, bootsrapping, Neural Network (NN) classifier, FrFT, ABC, Sparse classifier

INTRODUCTION

Plausible factor behind road accidents is implied to driver's drowsiness. The National Highway Traffic Safety Administration (NHTSA) reported in 2002 that about 0.7% of drivers had been involved in a crash that they attribute to drowsy driving, amounting to an estimated 800 000-1.88 million drivers in the past 5 years (NHTSA, 2001). The National Sleep Foundation (NSF) also quantified that a significant 51% of adult drivers had driven a vehicle in drowsiness and 17% had felt asleep. The EEG signal is an extra cellular current caused by the cerebral activity of brain and are distinguished as alpha, beta, theta and delta waves (Khalifa *et al*., 2000). EEG signal is primarily used to analyze brain activity. The brain activities are characterized by their frequency, amplitude and ability to react. Delta waves, like any other brain waves are recorded with an Electroencephalogram (EEG) and are usually associated with the deepest stages of sleep also known as Slow-Wave Sleep (SWS) and aid in characterizing the depth of sleep. Since, drowsiness is related with sleep signal, only delta wave is considered which are sleep waves whose frequency is very low ranging between 0.5-2 Hz (Wilson and Bracewell, 2000). Physical and physiological changes of human activities cause drowsiness (Lin *et al*., 2010). Also, several factors that blight the cognitive states are sleepiness, fatigue, monotony, psycho-physiological characteristics and distraction (Desai and Haque, 2006). The main objective of this paper is to develop a novel method of distinguishing different levels of drowsiness based on the EEG signals collected while the subjects were asleep. To improve the classification accuracy by choice of better optimization technique and to ease the computation by ideal selection of classifier.

MATERIALS AND METHODS

The project is focused on maximizing the precision of drowsiness related information collected from the sleep database. The communal block diagram implemented for the three feature extraction methods is given in Fig. 1 which shows the signal processing, feature extraction, optimization and classification steps. The initial step is the acquisition of signals from the database which is used to develop the model. The raw EEG signal is generated from the database. The extracted signal is preprocessed using Low pass median filter which allows the low frequency

Corresponding Author: Reena Daphne, Department of Electrical and Electronics Engineering, Stella Mary's College of Engineering, Aruthengavilai, Azhikal Post, Kanyakumari District Tamil Nadu, India

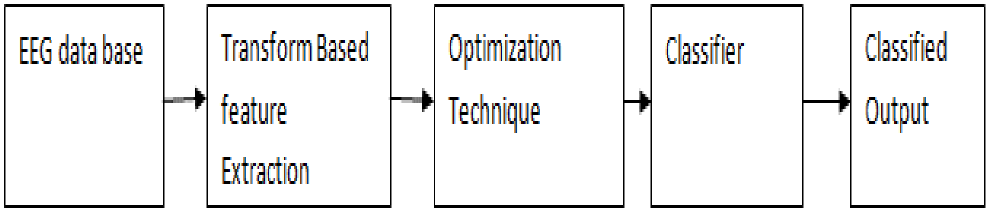

Fig. 1: Basic block diagram

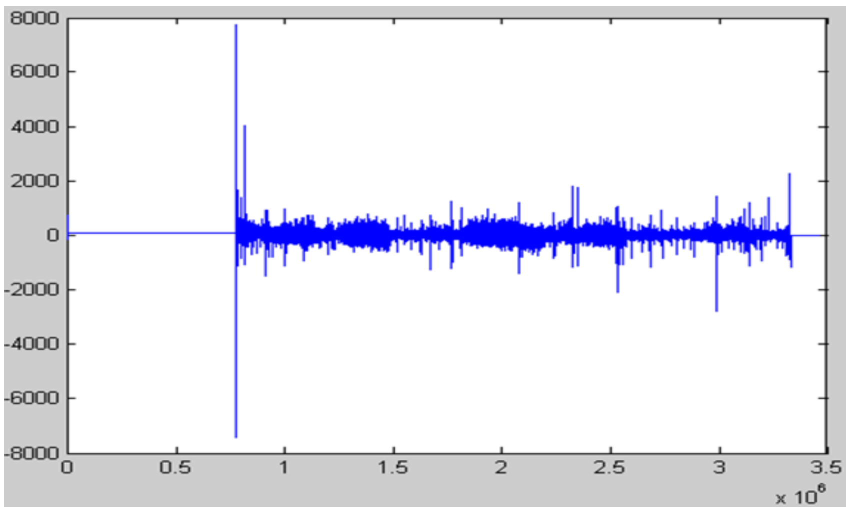

Fig. 2: Generated raw EEG signal

sleep waves and attenuates the high frequency signals. The processed signal is decomposed into different samples of required length. Then, the feature extraction methodologies are implemented and finally the optimality is determined. The resultant output is followed by classification to detect the drowsiness. Further, the efficiency of the system in reducing the computation time is found to be improving.

EEG database: The first and the foremost step is the collection of EEG data which facilitates the classification of results precisely. The training sets are collected from the drowsy subjects and these datasets are acquired from the web.

The EEG data is collected from Sleep-EDF Database in EDF Format [Sleep Recordings and Hypnograms in European Data Format (EDF)]. The drowsy state varies for the subjects based on the somatic condition. A sampling frequency of 100 Hz is chosen in common for all the subjects.

Generation of raw eeg signal: The European Data Format (EDF) is a lucid and flexible format for exchange and storage of multichannel biological and physical signals. EDF is used predominantly in the applications of sleep analysis algorithms. Each database that is being downloaded is an EDF file which cannot be understood as it is a direct recorded file from the radio telemetric system. So in order to convert the file into a readable format, EDF file has to be converted to ASCII file. This is performed by the EDF convertor.

Figure 2 represents the raw EEG signal that has been generated which is ready for feature extraction. Furthermore, it is observed that, the generated EEG signal has artifacts and noise present in it due to the regular human activities. Since the biological signals are low frequency signals a low pass median filter is enough to attenuate the artifacts and noise. Thus, Fig. 3 presents the noise free preprocessed EEG signal which is the analyzed signal that can be subjected to the transform based feature extraction techniques

Feature extraction and classification methods: Three methods have been implemented in this study to classify and detect drowsy states which are tabulated in Table 1.

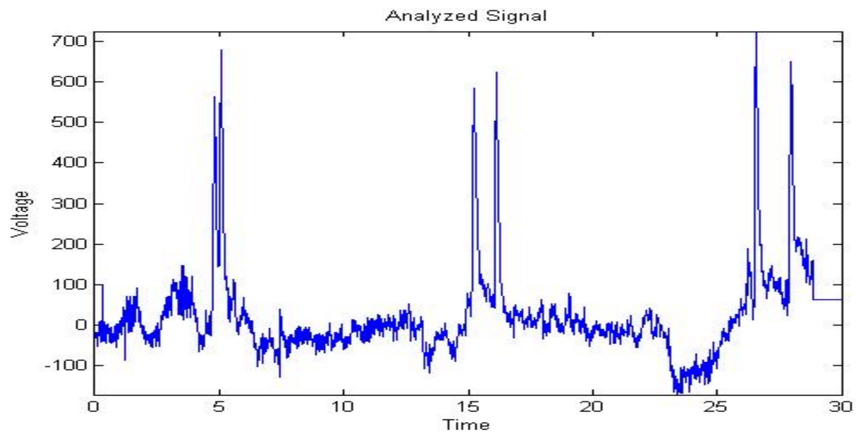

Fig. 3: Analysed signal

Table 1: Methodologies implemented

Methodology	Transform used for feature extraction	Optimization or feature selection	Classification
1	Daubechies wavelet packet transform	Bootstrap	Neural Network (NN) classifier with perceptron learning.
2	Fractional fourier transform	-	NN classifier with Back Propagation (BP) learning
3	Fractional fourier transform	Artificial Bee Colony (ABC)	Sparse classifier

Method-1; Daubechies Wavelet Transform Based feature extraction with bootstrap optimization using Neural Network classifier (DWTBNN)

Wavelet transform: Biomedical signals usually consist of brief high-frequency components closely spaced in time, accompanied by long lasting, low-frequency components closely spaced in frequency. Wavelets are considered appropriate for analysing such signals as they exhibit good frequency resolution along with finite time resolution, i.e., they first localize the low-frequency components and the secondly they resolve the high-frequency components. The wavelet-packet transform, referred to as WPT was introduced generalizing the link between multiresolution approximations and wavelets. WPT is employed to build features that highly draw a parallel with alertness and the different levels of drowsiness. The WPT is chosen due to its capability to deal with stationary, non stationary or transitory characteristics of different signals including unexpected changes, spikes, drifts and trends. Daubechies Wavelet packet Transform (DWT) is based on wavelet packet transform which can uniformly divide the signal frequency range and obtain the frequency time of the signals. Here, the characteristics of WPT and DWT are alike as DWT is only the discrete application of WPT is which then relates more to our study. The WPT is a tree of subspaces starting from root node which represents the original signal space. The root node is represented by $\Omega_{j,k}$ where j represents the scale and k represents the sub band index within the scale. This root node is further decomposed into two orthogonal subspac. Approximation space. Detail space, $l_{j,k} \rightarrow \Omega_{J+1,2k}$.

The approximation space is given $l_{j,k} \rightarrow \Omega_{J+1,2k}$ by and the detail space is given by $l_0 0$ Hence when applied values for (j,k) starting from (0,0), i.e., $l_{1,0}$, the root $_b 1$ node gets decomposed into another two nodes such as $l_2, 0_{-3}$ and $l_{3,0\text{-}7}$. This decomposition process continues until a depth 3 with nodes such as and .Hence the WPT decomposition tree is obtained as shown in the following Fig. 4 and 5.

As, we can see, 15 packets are obtained totally including the root node. In general, the decomposition process is repeated till J times, where $J \leq \log_2 N$. This means that the tree has N packets or coefficients. The scaling and the wavelet functions are given by $\varnothing_{J,k}$ (t) and $_{j,k}$(t) and they are given by:

$$\varnothing_{j,k}(t) = \frac{1}{\sqrt{|2^j|}} \varnothing\left(\frac{t-2^j k}{2^j}\right) \quad (1)$$

$$\varnothing_{j,k}(t) = \frac{1}{\sqrt{|2^j|}} \varnothing\left(\frac{t-2^j k}{2^j}\right) \quad (2)$$

where, the dilation factor 2^j is also known as the scaling parameter and it measures the degree of compression or scaling. On the other hand, the location parameter $2^j k$, also known as the translation parameter determines the time location of the wavelet. The WPT not only decomposes approximation coefficients but also detail coefficients. After the process of decomposition, the energy values has to be calculated for each and every packet.

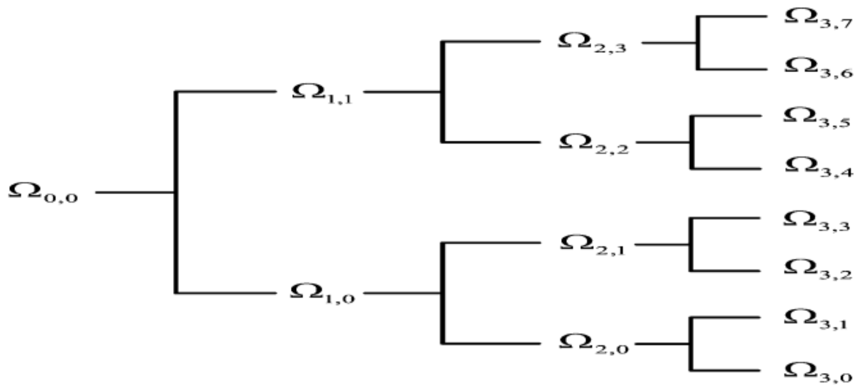

Fig. 4: Wavelet packet decomposition of into a tree

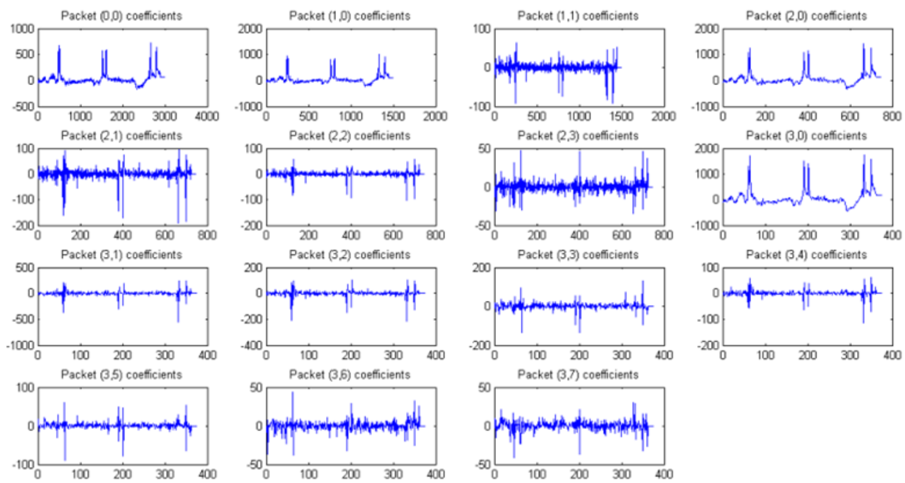

Fig. 5: DWT decomposition

The energy equation is given by:

$$E_{\Omega j,k} = \log\left(\sum\left[\left[\frac{n\left(\left[\left[w^{T}_{j,k,n} X\right)\right.\right.}{\frac{N}{2}}\right]\right]\right) \tag{3}$$

and this gives the normalized logarithmic energy of the wavelet packet coefficients extracted from the subspace $l_{j,k}$ (Amin *et al*., 2010). Wavelet packet transformed signal is given by $w_{j,k}x$ and they are simply the coefficients and they are evaluated at subspace $\Omega_{j,k}$ and $N/2^j$ is the number of coefficients in that particular subspace. The log of energy values is the feature extracted from each packet. The energy values are different for different databases. Thus, for each and every database, the energy values are determined using equation (Khalifa *et al*., 2000).

Bootstrap optimization: Bootstrap is an extremely attractive tool in that it requires very little in the way of modeling, assumptions or analysis and it can be applied in an automatic way. Bradley Efron introduced it in 1979. It is named from the phrase "to pull oneself up by one's bootstraps" which is widely believed to come from "the Adventures of Baron Munchausen". It is popularized in 1980s due to the introduction of computers in statistical practice and it provides a strong mathematical background. It is a method for improving estimators, it is well known as a method for estimating the statistical parameters. It is a computer-based way of estimating statistics (Rieser and Lemon, 2011).

The bootstrap is predominantly a way of finding the sampling distribution, approximately from just one sample The bootstrap technique comprises of two steps which are resampling and bootstrap distribution. We use bootstrap distribution as a way to estimate the variation in a statistic based on the original data. The bootstrap does not replace or add to the original data but just resamples the original samples. Resampling the data means repeatedly sampling the samples with replacement. As a result, the number of samples drawn can be repeated more than once or not at all. Therefore, resampling refers

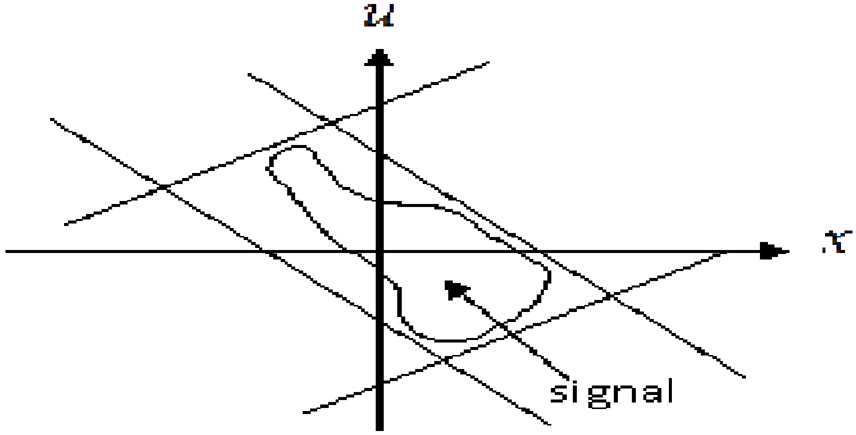

Fig. 6: Random noise removal by filtering in fractional domain

to sampling with replacement. Hence, after randomly drawing an observation from the original sample we put it back before drawing the next observation. Secondly, the bootstrap distribution statistically calculates the mean of the resamples determined. The following algorithm gives an idea regarding the bootstrap resampling and distribution.

- Let the original samples be $X = (x_1, x_2, \dots, x_n)$
- Repeat the samples B times
- End up with bootstrap values

Here, B represents the number of resampling, i.e., no of samples/2 provides the value of B. This is performed by generating a sample x* of size n from X by sampling with replacement. Then, the bootstrap values are computed as θ for x*. This ends up with the bootstrap values $\theta^* = \theta^*_1, \dots, \theta^*_B$

Hence, the determined numerical length from Fig. 6 returns the number of subscripted elements in a specific vector by bootstrapping methodology. The bootstrap mechanism is implemented to all the packets of the five features and the generated features contribute to a matrix of the order 5×15. This matrix is further replicated column wise and transposed to attain a matrix of the order 25×15. This gives the bootstrapped generated features. This acts as the training data for the further classification process. The shows the training data and the testing data that is obtained from the bootstrapped features. Now, the mean of the samples is determined and the values are stored for the process of classifying using NN classifier. Therefore, we obtain 375 samples to be mapped as the training data and testing is done with 120 samples. These samples are further subjected to classification using NN classifier.

Classification using NN classifier with perceptron learning: Classification is one of the most frequently encountered decision making tasks of human activity (Zhang, 2000). In order to classify the samples a neural network classifier is used. Networks that mimic the way the brain works; computer programs that actually learn patterns; forecasting without having to know statistics are some of the claims and attractions of neural networks. A neural network consists of units (neurons), arranged in layers which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally, the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another and it is these weightings which are tuned, in the training phase to adapt a neural network to the particular problem at hand. This is the learning phase. These Neural Networks are non-linear models which make them flexible in modeling real world complex relationships, especially in medical diagnosis. The basic model of a neuron consists of an input with some synaptic weight vector an activation function or transfer function inside the neuron determining output:

$$Y = f\left(w_0 + \sum_{i=1}^{n} w_i x_i\right) \quad (4)$$

where, w_0 is the bias and f is the activation function and y determines the output. The function f(net) is assigned +1 if net is greater than of equal to 1 and -1 for the vice versa condition.A neuron accepts a number of inputs n. Each input is modified by a weight w. Then, all the inputs are summed and get through an activation function. If the output of the function gets over the threshold, the neuron fires. Feed Forward Neural Networks is used where the information is propagated from inputs to the outputs. The number of hidden layer chosen is 2. From this classifier the following can be determined:

- Determination of pertinent inputs
- Collection of data for the learning and testing phase of the neural network
- Finding the optimum number of hidden nodes
- Estimate the parameters (Learning)
- Evaluate the performances of the network
- If performances are not satisfactory then review all the precedent points

A learning vector quantization toolbox serves as the best alternative for adapting the learning rule. The learning rule that is applied is the Perceptron learning rule in which learning signal is the difference between the desired and actual neuron's response. Thus, the learning is supervised and the learning signal is equal to:

$$r = d_i - 0_i \tag{5}$$

Here, d_i is the desired response and O_i is the output response Weight adjustments in this method is given by:

$$\Delta w_i = C\left[d_i - \operatorname{sgn}\left(w_i^t x\right)\right]x \tag{6}$$

where, C is a positive learning constant and it represents the separated classes. Under this rule, weights are adjusted if and only if O_i is increased. The Perceptron Learning rule is of central importance for supervised learning of neural networks. Once when the O_i becomes equal to d_i, the process stops. Therefore, the neural network classifies the various states of drowsiness and detects the exact state the person is in.

Method-2; fractional fourier transform based feature extraction without Feature selection using Neural Network classifier (FrFTNN)

Fractional fourier transform: The Fractional Fourier Transform (FrFT) belongs to the class of time frequency representations that have been extensively used by the signal processing community. In all the time frequency representations, one normally uses a plane with two orthogonal axes corresponding to time and frequency. If we consider a signal x (t) to be represented along the time axis and its ordinary Fourier transform X (f) to be represented along the frequency axis, then the Fourier is consistent with some of the observed properties of the Fourier transform (Narayanan and Prabhu 2003) For example, two successive rotations of the signal through $\pi/2$ will result in an inversion of the time axis. Moreover, four successive rotations will leave the signal unaltered since a rotation through 2π of the signal should leave the signal unaltered. The FrFT is a linear operator that corresponds to the rotation of the signal through an angle which is not a multiple of $\pi/2$, i.e., it is the representation of the signal along the axis u making an angle α with the time axis. A transformation T of a input EEG signal can be made fractional as follows:

$$T^{\alpha}\left\{f(x)\right\} = F_{\alpha}(u) \tag{7}$$

where, f and F are two functions with variables x and u respectively (Almeida, 1994) As seen, we can say that F is a T transform of f. Now, another new transform can be defined as:

$$T^{\alpha}\left\{f(x)\right\} = F_{\alpha}(u) \tag{8}$$

We call T^{α} here the "á-order fractional T transform" and the parameter α is called the "fractional order". Filtering and noise removal in fractional domain is the most imperative feature of FRFT which makes it more efficient than WPT discussed earlier. Here, FRFT acts to be multiresolution filter and also exhibits high frequency resolution in 3D analysis of the EEG signal compared to WPT. For conventional filtering, to remove the noise in frequency domain is impossible. But, we can rotate the Wigner Distribution that is, do the Fractional Fourier Transform, then filtering out the undesired noise and then by choosing proper rotation angle and doing the same process iteratively, we may remove the noise easily as shown in Fig. 6. Here, we initiate the fractional filter design method as in our project. Fractional filter, is defined as:

$$\begin{gathered} x\rrbracket\downarrow^{0}(t) = \Im^{\uparrow}(-\alpha)\left\{\Im^{\uparrow}(-\alpha)\left\{x_{\downarrow}i(t)\right\}\right. \\ H_{\downarrow}\alpha(u) \text{ where } H_{\downarrow}\alpha(u) = \Im^{\uparrow}\alpha\left\{h(t)\right\} \end{gathered} \tag{9}$$

where, x_i (t), x_0 (t) and h (t) correspond to input signal, output signal and the impulse response of the filter. Due to the fact that performing the αth order Fractional Fourier Transform operation corresponds to $\varnothing = \alpha\pi/2$ rotating the Wigner distribution by an angle in the clockwise direction, we can find the fractional domain that signal and noise do not have overlap. Then, we can rotate the Wigner distribution that is do the Fractional Fourier Transform, then filtering out the undesired noise. This is shown in Fig. 6. By the same idea, we can remove random noise by applying fractional filters iteratively as in Fig. 6.

Initially EEG signal is acquired and it is filtered and an analysed signal is obtained. From the 3000 samples we are still more choosing only 1500 samples for our consideration and these 1500 samples are split into five features each having 300 samples. In this way feature assignment is done. Sampling frequency chosen is 100 Hz. Next, each feature that has been selected is subjected to short time fractional fourier transform. The fractional power chosen can take any value of the order of 0.001 between 0 and 1. The significance of using short time fractional fourier transform is to have a high fractional frequency resolution and time resolution and any change of signal within a short time limit can be analysed in depth with ease and precision. Feature extraction is done by calculating the log of energy values (Khushaba *et al*., 2011) of the transform applied feature sets.The energy equation is given by:

$$E\Omega_{j,k} = \log\left(\sum \frac{n\left(W_{x,u}^{T} X\right)^2}{\frac{N}{2^x}}\right) \tag{10}$$

and this gives the normalized logarithmic energy of the fractional fourier transform coefficients extracted from the subspace $\Omega_{x,u}$ Fractional fourier transformed signal is given by $W_{x,u}X$ and they are simply the coefficients and they are evaluated at subspace $\Omega_{x,u}$ and $N/2_x$ is the number of coefficients in that particular subspace. Here, optimization, i.e., selection of the best features is not done. After calculating the energy value of the features we go for classification using NN classifier with back propagation learning.

Classification using nn classifier with back propagation learning: Advantages of neural networks include their high tolerance to noisy data, as well as their ability to classify the samples on which they have not been trained. The most popular neural network algorithm is Back-Propagation (BP) algorithm proposed in the 1980's.

The feed forward, back-propagation architecture was developed in the early 1970's. Currently, this synergistically developed BP architecture is the most popular, effective and easy-to-learn model for complex, multi-layered networks. Its greatest strength is in non-linear solutions to ill-defined problems. The typical back-propagation network has an input layer, an output layer and at least one hidden layer.

During the learning process, a forward flounce is made through the network and the output of each element is computed layer by layer. The difference between the output of the final layer and the desired output is back-propagated to the previous layer(s), usually modified by the derivative of the transfer function and the connection weights are normally adjusted using the Delta Rule (Beale and Jackson, 1990). This process proceeds for the previous layer(s) until the input layer is reached. Here, a supervised learning methodology is implemented.

In this research, two hidden layers are chosen so as to reduce the complexities and the outputs that are obtained are the desired classes which are alert, mild drowsy, moderately drowsy, rather drowsy and very drowsy states. Therefore, based on the comparison between the training and the testing data the desired states are classified. Consider a feed-forward network with n input and m output units. It can consist of any number of hidden units and can exhibit any desired feed-forward connection pattern. We are also given a training set of samples $\{(x_1, t_1), ..., (x_p, t_p)\}$ consisting of p ordered pairs of n and m-dimensional vectors which are called the input and output patterns. When, the input pattern of samples x_i from the training set is presented to this network, it produces an output o_i different in general from the target t_i. The target is the identification of the particular drowsy state of the person. Our aim is to make o_i and t_i identical for i = 1, ..., p by using BP learning algorithm and minimize the error function of the network defined (Rojas, 1996) as:

$$E = \frac{1}{2} \sum_{i=1}^{p} \|o_i - t_i\| \tag{11}$$

After minimizing this function for the training set, new unknown input pattern of samples are presented to the network and we expect it to interpolate. The network must recognize whether a new input vector is similar to learned patterns and produce a similar output. The back propagation algorithm is used to find a local minimum of the error function (Beale and Jackson, 1990; Rojas 1996) The network is initialized with randomly chosen weights. The gradient of the error function is computed and used to correct the initial weights. Our task is to compute this gradient recursively and finally equalize with the target. Thus, the various drowsy states are classified.

Method-3; fractional fourier transform based feature extraction with abc optimization (Feature selection) using Sparse Classifier (FrFTABCS)

Fractional fourier transform: As discussed earlier FrFT is used here with optimization done using Artificial Bee Colony (ABC) algorithm.

ABC algorithm for optimization (feature selection): Here, after the calculation of log of energy values of the features the next step is to select the best features that would match with the desired classes which are alert, mild drowsy, moderately drowsy, rather drowsy and very drowsy states.

The bee colony has got a search experience in searching for its food.Hence, it has the capability to remember every characteristics of its collected food. This bee swarm intelligence makes use of the most important aspect of exchange of information among the bees. They use several mechanisms like waggle dance which help in optimizing the location of food sources and to begin the search for new ones. Through this waggle dance the bee will be able to deliver the information regarding the direction , the distance and the quality of the food to the other bees. By this way, the bees will adjust the search tactic based on shared information to discover the good-quality food, neglecting the poor quality ones. This makes them a good contender for developing a new intelligent search algorithms ,namely the ABC algorithm. It is a very simple, robust and population-based, stochastic optimization algorithm (Pham *et al*., 2005; Karaarslan, 2013). In the ABC algorithm, the bee colony contains two groups of bees which are scout and employed bees. The scout bees take up the job of finding a new food source. The duty of the employed bees is to determine a food source within the quarter of the food source in their memories and share their information with other bees.

The proposed ABC Algorithm requires a number of parameters to be set, namely, the number of scout bees (n); the number of elite bees (e), the number of patches selected out of the n visited points (m), the number of bees recruited for patches visited by "elite bees" (nep), the number of bees recruited for the other (m-e) selected patches (nsp) and the size of patches (ngh). (Karaboga, 2005; Pham *et al*., 2005).

The procedure of the abc algorithm is given as follows:

- Step 1); generate randomly the initial populations of n scout bees, i.e., the n number of solutions available. These initial populations must be feasible candidate solutions that satisfy the constraints which is the range of value set for the fitness function. Set $N_i = 0$
- Step 2); Evaluate the objective function value which is the minimization of the integral square error. The objective function value f (d) at each bee position is evaluated as in (Karaarslan, 2013) using:

$$f(d) = \int_0^n e(k)^2 dt = \int_0^n i_{ref}(k) - i_m(k))^2 dt \quad (12)$$

Where:

$i_{ref}(k)$ = Position of the ith bee with highest value (reference) of the fitness function

$i_m(k)$ = Position of the ith loction

- Step 3); Evaluate the fitness value of the initial populations. In this step, depending on the findings, each worker bee performs their dance. The duration of the dance is a measure of quantum of food and Dd represents the duration of the dance. The duration of the dance of the ith bee in the jth location at iteration k (Karaarslan, 2013) is given in:

$$D_{di(j)}(k) = \frac{1}{f(d)_{i(j)} + 1} \quad (13)$$

The dance duration of the bees is indicated by circles and the size of the circle is used to measure Dd:

- Step 4); Select m best bees for neighborhood search. The bee whose location gives the maximum value for D_d is designated as elite bee. Thus, elite bee E_b in the kth iteration (Karaarslan, 2013) is given by:

$$E_b(k) = \max\left(D_{di(j)}\right)(k)) \quad (14)$$

which for our analysis is the bee with the highest fitness value in the constraint. Our project is concerned with the detection of mild drowsy values. Hence, the elite bee here, is said to the bee with the highest amount of nectar having the fitness function value 1. Our aim is to find values having fitness function values between 0.81 and 0.99.

- Step 5); Separate the m best bees into two groups; the first group has e best sites and another group has nsp best bees. These are the bees which is needed for us which have a nectar content assigned with the fitness function values between 0.81 and 0.99
- Step 6); Determine the size of the neighborhood search of each best size (patch size ngh)
- Step 7); Recruit bees for selected bees (nep)
- Step 8); Select the fittest bees from each patch
- Step 9); Check the stopping criterion.i.e.,the range of bees identified with fitness values between 0.81-0.99. If satisfied, terminate the search; else, $N_i = N_i + 1$ where Ni is the number of iterations
- Step 10); Assign the n - m remaining bees to random search
- Step 11); New population of scout bees. Go to step 2

Table 2: ABC algorithm parameters

Algorithm Parameters	Symbols	Values
Population size	n	120
No. of selected bees	m	60
No. of elite bees	e	21
No. of bees around other selected points	nsp	39
No. of bees around elite	nep	12
Patch size	ngh	25.10^{-6}

The values of the parameters of the ABC algorithm are shown in Table 2. In a similar fashion the other classes features can also be optimized using the above algorithm keeping 0.8 to be the highest fitness function value and calculate values in the range from 0.61-0.8 and so on.

Sparse classifier: There is no guarantee that the classes of features thus optimized have been segregated accurately as there is an uncertainity in the classification of the optimized features. Such class imbalance problems can be handled using a sparse classifier. Sparse classifiers outputs faithful conditional probabilities in the vicinity of the decision boundary.

Cascade classifiers like Neural network classifiers have many drawbacks in both their training and the test phases. The training process requires a lot of hand tuning of control parameters and it is non-trivial how to handle the tradeoff between the performance and the complexity of the cascade. Also, for the network to get trained and to produce the output it takes more computation time which is the important control parameter in our work.

Having generated the optimal feature sets, we go for classification using sparse classifier. It is used to sparsely have a check over the segregated classes whether elements of a particular set is present only in it and not in other sets (Suzuki *et al*., 2013; Hautamaki *et al*., 2013) If so, it classifies it correctly under that particular set. In this way the intra similarity between samples in a particular class is found to be an increasing function and the inter similarity between two classes is found to be a decreasing function.

RESULTS AND DISCUSSION

The computation time may vary for different processors. In order to determine the efficiency of the implemented mechanism, an ROC curve is generated. The ROC curve is plotted between True Positive Rate (TPR) and False Positive Rate (FPR). The curve generated by means of ROC gives the exact efficiency of the proposed methodology. The ROC curve is generated so as to determine the classification accuracy, taking 120 test samples into consideration.

Figure 7 shows the ROC of the three methodologies. The classification accuracy of the three methods is found to be in a rising order with sparse classifier showing a maximum of 83.3% and this can be determined from the confusion plot. The computation time required for executing this methodology is just 0.0056 sec which is very much smaller compared to the two other methods.

The reduction of mean square error determines a better performance for DWTBNN and FrFTNN methodologies which use NN classifiers as shown in Fig. 8 and 9. The confusion matrix of the three methods DWTBNN, FrFTNN and FrFTABCS are shown in Fig. 10-12, respectively.

The parameters, sensitivity, specificity, Precision, Negative Predictive Value (NPV), Accuracy, Fall Out, F1 Score are determined from the confusion matrix and tabulated in Table 3 :

- Sensitivity (or) TPR = TP/P = TP/(TP+FN)
- Specificity (or) TNR = TN/N = TN/(FP+TN)
- Precision (or) Positive Predictive Value (PPV) = TP/(TP+FP)
- Negative Predictive Value (NPV) = TN/(TN+FN)
- Accuracy = Acc = (TP+TN)/(P+N)
- Fall out (or) False Positive Rate = FPR = FP/N
- F1 Score = 2TP/(2TP+FP+FN)

Where:

TP = True Positive = Drowsy people correctly identified as drowsy
FP = False Positive = Alert people incorrectly identified as drowsy
TN = True Negative = Alert people correctly identified as alert.
FN = False Negative = Drowsy people incorrectly identified as alert

From Table 3, it is found that the accuracy and F1 score of the methodologies implemented are found to be in the rising order which shows that the performance of the FrFTABCS surpasses the other two methods. Also, the computation time is found to be in the reducing order which would alarm the person instantly when he deviates from his alert state. Further, studies are being done to improve the classification accuracy and the F1 score of the system and reduce the computation time. Thus, in this study three efficient transform based feature

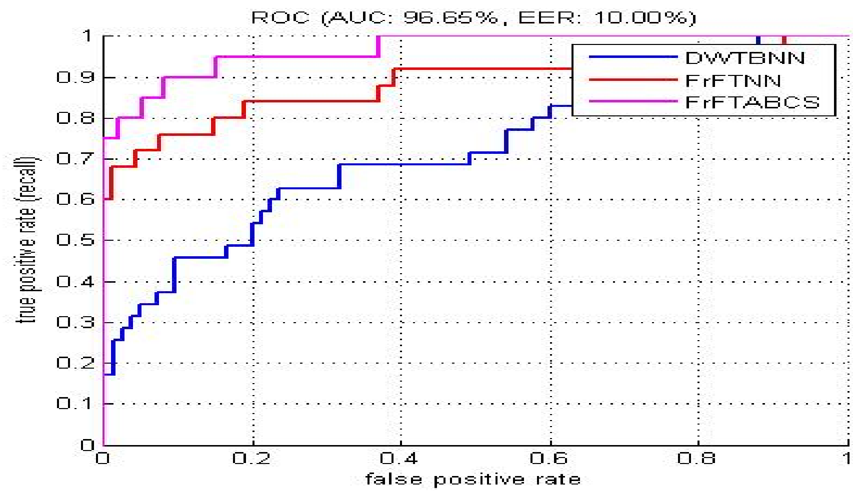

Fig. 7: ROC of implemented methods

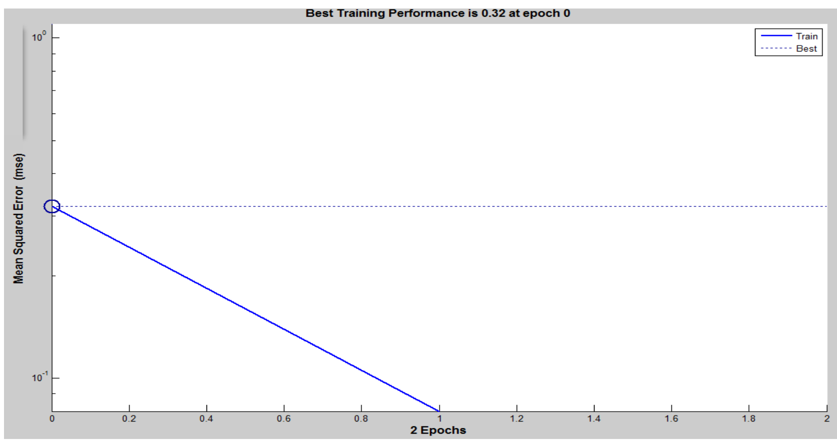

Fig. 8: Minimisation of MSE for DWTBNN

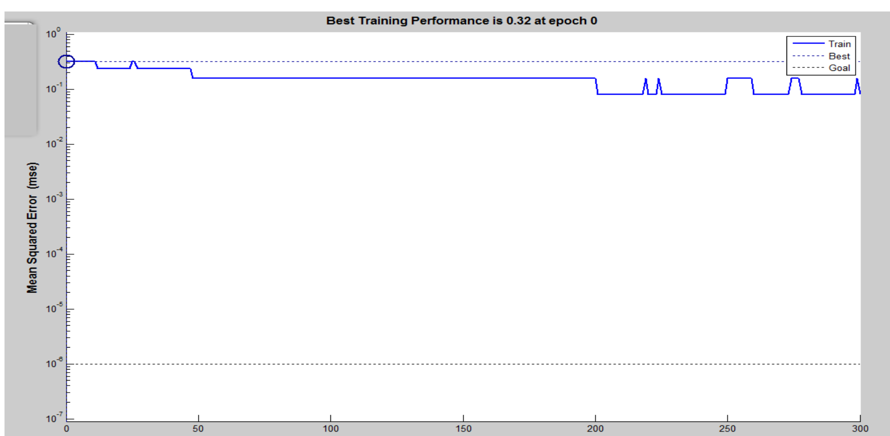

Fig. 9: Minimisation of MSE for FrFTNN

extraction methods are proposed along with a comparative study of their statistical features and it is vivid that the efficiency of the system in reducing the time required for computation of each methodology is found to be increasing in nature.

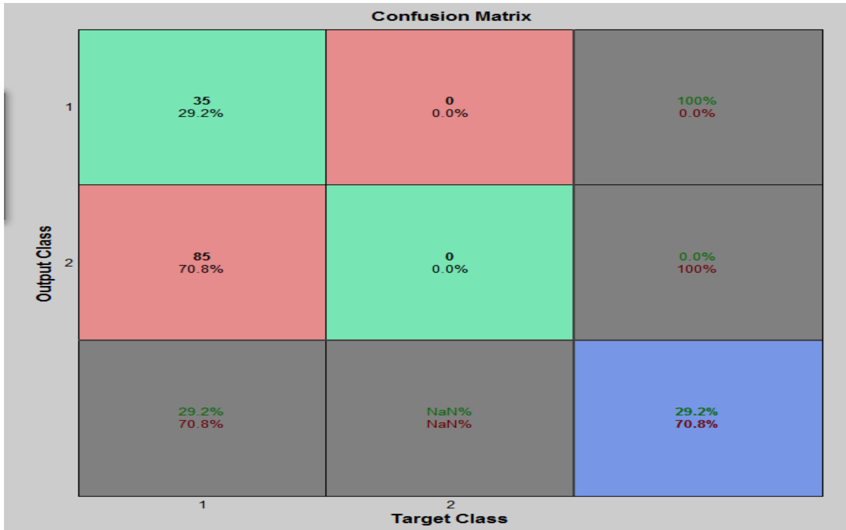

Fig. 10: Confusion matrix for DWTBNN

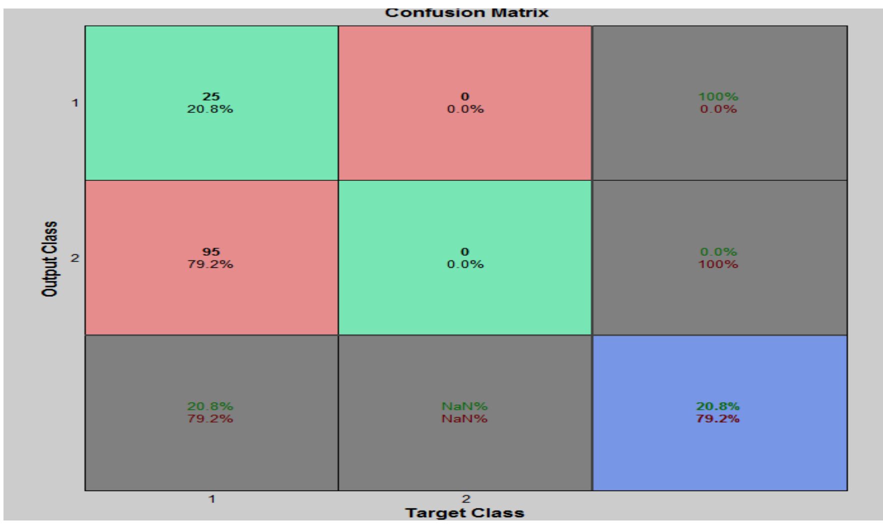

Fig. 11: Confusion matrix for FrFTNN

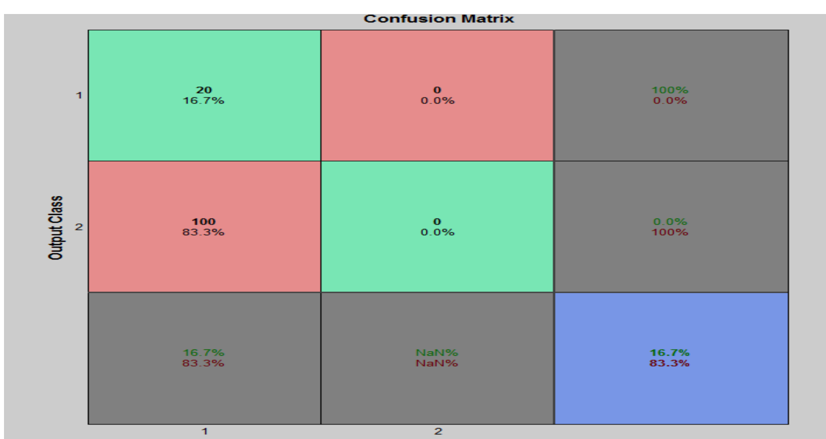

Fig. 12: Confusion matrix for FrFTABCS

Table 3: Performance analysis

Methodology	Observations Accuracy (%)	MR (%)	TPR (%)	TNR (%)	PPV (%)	NPV (%)	FPR (%)	F1 score (%)	Computation time (sec)
DWTBNN	70.8	29.2	70.8	NaN	100	0	NaN	82.92	0.68032
FrFTNN	79.2	20.8	79.2	NaN	100	0	NaN	88.37	0.46085
FrFTABCS	83.3	16.7	83.3	NaN	100	0	NaN	90.9	0.00564

CONCLUSION

Thus, three efficient transform based feature extraction methods are proposed along with a comparative study of their statistical features and found that the efficiency of the system in reducing the time required for computation of each methodology is found to be increasing in nature

ACKNOWLEDGEMENTS

Any task would be incomplete without the thoughtful mentioning of the people who made the success possible. acknowledge their constant guidance and encouragement that had crowned my effort with success. First and foremost I thank the Lord Almighty who had showered his blessings on me by providing me with good health and knowledge and was with me in each and every step of this research work.

I thank all the Teaching staff members of Department of EEE, Stella Mary's College of Engineering for their kind support during the course of my work .At this juncture I gratefully thank my friends for extending their valuable suggestion in bringing out my work successfully.

REFERENCES

Almeida, L.B., 1994. The fractional fourier transform and time-frequency representations. IEEE Trans. Signal Proces., 42: 3084-3091.

Amin, M.S., M.R. Azim, T. Latif, M.A. Hoque and F.M. Hasan, 2010. Spectral analysis of human sleep EEG signal. Proceedings of the 2nd International Conference on Signal Processing Systems, Vol. 3, July 5-7, 2010, Dalian, pp: V103-V6-V3-110.

Beale, R. and T. Jackson, 1990. Neural Computing: An Introduction. Adam Hilger, Bristol, England.

Desai, A.V. and M.A. Haque, 2006. Vigilance monitoring for operator safety: A simulation study on highway driving. J. Safety Res., 37: 139-147.

Hautamaki, V., T. Kinnunen, F. Sedlak, K.A. Lee, B. Ma and H. Li, 2013. Sparse classifier fusion for speaker verification. IEEE Trans. Audio, Speech Language Proces., 21: 1622-1631.

Karaarslan, A., 2013. The implementation of bee colony optimization algorithm to Sheppard-Taylor PFC converter. IEEE Trans. Ind. Elect., 60: 3711-3719.

Karaboga, D., 2005. An idea based on honey bee swarm for numerical optimization. Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, Kayseri, Turkey, October 2005. http://mf.erciyes.edu.tr/abc/pub/tr06_2005.pdf.

Khalifa, K.B., M.H. Bedoui, R. Raytchev and M. Dogui, 2000. A portable device for alertness detection. Proceedings of the 1st Annual International Conference on Microtechnologies in Medicine and Biology, October 12-14, 2000, Lyon, France, pp: 584-586.

Khushaba, R.N., S. Kodagoda, S. Lal and G. Dissanayake, 2011. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. IEEE Trans. Biomed. Eng., 58: 121-131.

Lin, C.T., C.J. Chang, B.S. Lin, S.H. Hung, C.F. Chao and I.J. Wang, 2010. A real-time wireless brain-computer interface system for drowsiness detection. IEEE Trans. Biomed. Circuits Syst., 4: 214-222.

NHTSA., 2001. Traffic Safety Facts 2001: A Compilation of Motor Vehiclecrash Data from the Fatality Analysis Reporting System and the General Estimates System. National Center for Statistics and Analysis, NHTSA., Washington, DC., USA.

Narayanan, V.A. and K.M.M. Prabhu, 2003. The fractional Fourier transform: theory, implementation and error analysis. Microproces. Microsyst., 27: 511-521.

Pham, D.T., A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim and M. Zaidi, 2005. The bees algorithm. Technical note. Manufacturing Engineering Centre, Cardiff University, UK., pp: 1-57.

Rieser, V. and O. Lemon, 2011. Reinforcement Learning for Adaptive Dialogue Systems. Theory and Applications of Natural Language Processing. Springer-Verlag, Berlin Heidelberg.

Rojas, R., 1996. Neural Networks. Springer-Verlag, Berlin.

Suzuki, C.T., J.F. Gomes, A.X. Falcao, J.P. Papa and S. Hoshino-Shimizu, 2013. Automatic segmentation and classification of human intestinal parasites from microscopy images. IEEE Trans. Biomed. Eng., 60: 803-812.

Wilson, B.J. and T.D. Bracewell, 2000. Alertness monitor using neural networks for EEG analysis. Proceedings of the IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing X, Volume 2, December 11-13, 2000, Sydney, Australia, pp: 814-820.

Zhang, G.P., 2000. Neural networks for classification: A survey. IEEE Trans. Syst. Man Cybern. Part C: Applic. Rev., 30: 451-462.