International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2769-2779

A Comparative Analysis on Classification of Sleep Stages Using Different Classifiers

U.S.B.K. Mahalaxmi[1] and M.Ramesh Patnaik[2]

Research Scholar, Andhra University, India.
Sr. Assistant Professor, Andhra University, India.

Abstract

Currently, virtually all new digital polygraphs are equipped with a more or less efficient automatic sleep analyzer. Interest in these systems is growing in sleep study services for significant increase in the demand for recordings and the need for an automatic analysis system freeing the clinician of certain counting tasks. The major difficulty for these systems is the choice of modelling for physiological signals. Different analysis techniques are used: amplitude analysis, period analysis, spectral analysis etc. This paper used Fourier transform for the frequency conversion of EEG signal, which is further process by Butterworth filter .The sleep features are characterise by DWT and further sampled to classifier for 5 sleep stage classification. ABC (ARTIFICIAL BEE COLONY) tuned NN (NEURAL NETWORK) Claimed better accuracy then random forest ,and neural network classifier.

Keywords: ABC, NN, DWT, FT, Random forest, EEG.

INTRODUCTION

In clinical routine, the study of sleep involves the acquisition and recording of a set of physiological signals during a night's sleep, followed by visual analysis to establish the diagnosis. This study is mainly based on three signals: electroencephalogram (EEG), electro-oculogram (EOG) and electromyogram (EMG) [1]. Visual analysis consists of detecting the variations of these signals during the night. These changes define the states of alertness that are awakening and the five stages of sleep: stage 1, 2, 3, 4 and paradoxical sleep (SP) [2]. Each state is characterized by the presence of one or more indicators corresponding to elementary activities and certain grapho-elements. According to these indicators and by applying the standard rules of Rechtschaffen and Kales [3] the clinician associates at a time of 30s a label corresponding to the physiological state.

The difficulty of visual analysis lies in several levels. We can cite for example the rules of interpretation which include a part of subjectivity. In addition, visual detection can be difficult because of noises and artefacts due to poor electrode contact or patient movement. Finally, we must not forget the time consumed by this analysis can last a few hours.

Among the techniques of automatic sleep analysis, Fast Fourier Transform (FFT) spectral analysis is the most used. This choice can be explained by the fact that the visual analysis is essentially based on the detection of some particular frequency waves. Among the processing techniques of this data we will focus on Artificial Neural Networks (ANN). RNAs are widely applied in areas related to neurophysiology: EEG analysis [5], vigilance analysis [6], sleep analysis [7, 8], etc.

A night's sleep is divided into several cycles (four to five), of about 90 minutes, each one respecting the same schema. We distinguish in each cycle several phases that we can essentially recognize from the analysis of electroencephalographic tracings (EEG).

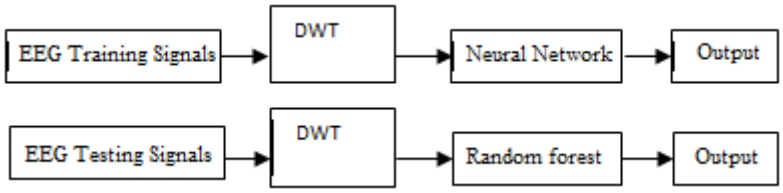
Thus, at the beginning of the night, the subject is awake with his eyes closed. The EEG registers very fast waves (8-12Hz) and very low amplitudes, called alpha. Then the subject passes smoothly in stage 1. We then observe on the EEG a slowdown electric waves up to a frequency of 4 to 6 Hz, these are theta waves. This stadium is pretty close to wakefulness. Moreover, when awake, the patient will say that he is not sleeping. Then the subject switches to stage 2, which can truly be considered as the early sleep. This stage is found on the EEG by the appearance of spindles, puffs fast waves (10-15Hz), and K complexes, large slow waves. Electrical activity is going still slow down, then appear delta waves (3Hz) of great amplitude. It is considered that the subject is passed to stage 3 when at least 20% of the tracing is occupied by delta waves. When this figure exceeds 50%, it is considered to be in stage 4. The stages 3 and 4 are often referred to as deep slow sleep. "Slow", because the waves are very slow, "deep", because a subject awakened during this period will be completely out of place in time and space, as if it were deeply buried in his sleep.

But after a while, the subject's activity changes dramatically. Cardiac activity and respiratory, regular and slow until then, accelerates and becomes irregular. Voltage arterial also undergoes sudden variations. The eyes move very quickly from left to right or from top to bottom, while at the same time the subject reaches a level total muscle relaxation (the muscles of the neck relax), and a depth of increased sleep (the sound stimulation needed for wakefulness needs to be even more important only in deep slow sleep). This contradiction in the symptoms gave his name at this stage, paradoxical sleep. This visual observation is confirmed on the EEG.

The physiological rhythms of which sleep variation has been most studied are the cardiorespiratory rhythm and arterial pressure. This denier decreases as and when from falling asleep to levels of 5 to 15% below sleep [9]. However, it does not vary more during sleep paradoxical, even increase a little. In the same way, the heart rate decreases as the subject sinks

into sleep to reach its minimum value during the deep slow sleep. But above all there is a decrease in the variability of this one [10]. On the other hand, `unlike the blood pressure, the heart rate becomes completely disordered and accelerates on average during REM sleep, which can be to be dangerous for patients who have suffered a myocardial infarction, their heart no longer having the same ability to track these variations [11]. With regard to respiration, there is a periodic variation of the amplitude ventilatory, with a period of seventy to ninety seconds, during the fall asleep, and as long as sleep is not stably reached [12]. When that's the case, sleep becomes extremely regular and the respiratory rate decreases by about 15%. All as for the heart rhythm, a clear change occurs during the paradoxical sleep. Thus, the ventilation becomes irregular, with abrupt variations in its amplitude and its frequency. With regard to blood gases, the partial pressure of CO_2 in the blood is raised during slow sleep but that of O_2 remains constant [13]. As for the temperature body, it decreases during slow sleep [14, 15]. In addition, falling asleep is easier when it occurs on the descending portion of the thermal cycle [16]. Finally, muscle tone varies according to the muscles involved. So the muscles neck and chin are completely relaxed during the transition to REM sleep. Which explains the fall of the head of a sleeper sitting. Conversely, the muscles of the trunk do not do not see their tone evolve during the night. However, in general, the motions less and less of the body as one sinks into sleep.

The organization of research is as follows: further sections discuss the mathematical computation of drowsiness detection followed by testing on CAP database. The results section depicts the effectiveness of algorithm and discussion is made based on it. A conclusion is present at the end.



Figue 1. Process of sleepiness detection

SYSTEM MODEL

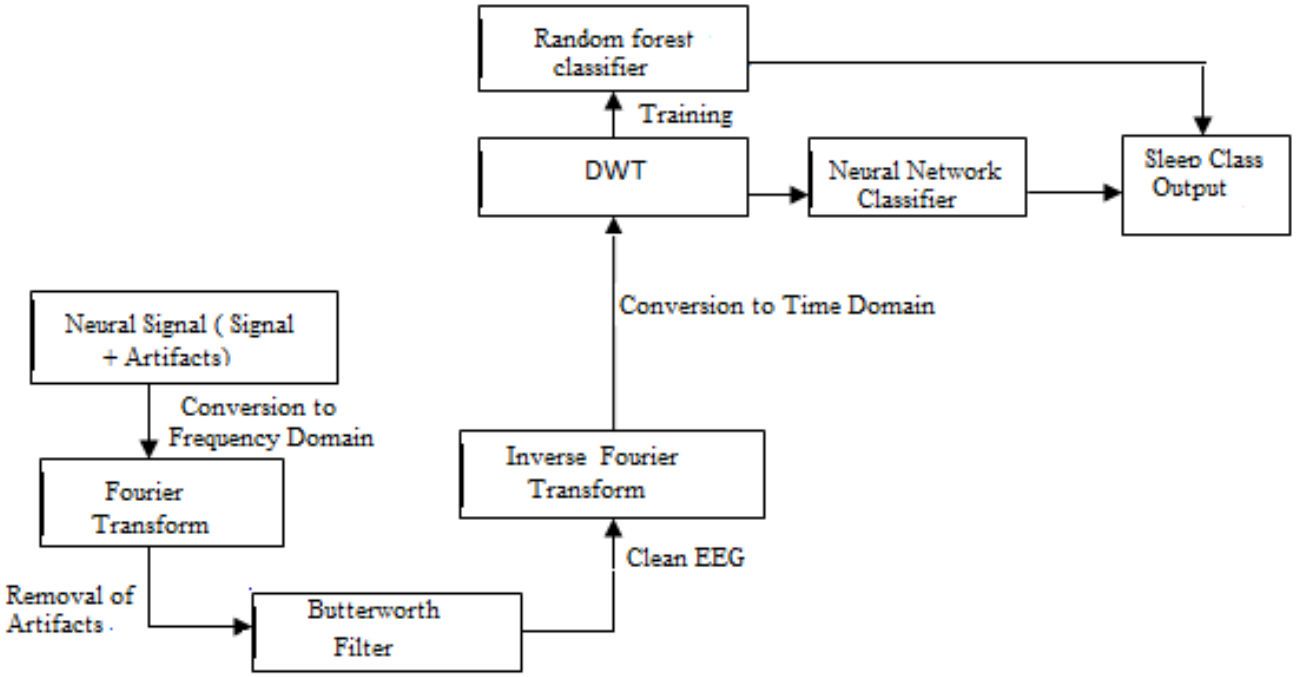
The total system can be divided as

I. Preprocessing of EEG signals

II. Feature Extraction of EEG signals

III. Post Processing of EEG signals

I) **Pre-processing:** EEG frequency domain conversion and Filtering .



Figue 2. Architecture for Detection of Sleepiness

Fast Fourier Transform

Given a finite energy EEG signal s (t), its Fourier transform is a function of the frequency (f), defined by:

$$TF\{s(t)\} = \hat{s}(f) = \int_{-\infty}^{\infty} s(t).e^{-j2\pi tf}dt \qquad (1)$$

For the case of discrete time

$$TFTD\{s(n)\} = \hat{s}(f) = \sum_{n=-\infty}^{\infty} s(n).e^{-j2\pi nf} \qquad (2)$$

$\hat{s}(f)$ is a new representation of s (t) (or s (n)), which facilitates the visualization of the frequency distribution of energy.

Filtering of EEG signal:

For its approximation a function of quadratic magnitude | H (jω) | 2 that satisfies the function criterion maximally plane in ω = 0. To provide the slope downward in the characteristic of the high frequencies, the transmission zeros are located in infinity.

Your frequency response is:

$$|H(j\omega)|^2 = \frac{H_0^2}{1+\left(\frac{\omega}{\omega_p}\right)^{2n}} \qquad (3)$$

where:

|H(jω)| :is the transfer function.

N: is the order of the filter.

Ω_p: is the cutoff frequency (where the response falls -3 dB below the pass band).

Ω: is the complex analog frequency (ω = | jω |).

The low-pass Butterworth functions have the form given in (1) following properties:

1. The frequency range 0 ≤ω≤ 1 rad / s is called the passband.
2. The frequency range ω / ω_p ≥ 1 rad / s is called a blocked band
3. In ω = ω_p rad / s, | H (jω) | = 0.7071H0, independent of value of n.
4. In ω = ω_p rad / s, the slope of $|H(j\omega)|^2$ is proportional to $\alpha -\frac{1}{2} n$.
5. The function | H (jω) | is a monotonic (continuously decreasing) function of ω. The function defined in (1) with ω_p = 1rad / s is known as the normalized function of Butterworth. Since 20 log [| H (j1) | / | H (j0) |] = 20log0.70711 = -3.010dB, at the frequency of ω_p = 1rad / s is known as -3dB frequency.

II) Feature extraction

Frequency Classification

In the multiresolution analysis, or pyramidal algorithm, the idea is the same as in the CWT: to obtain a time-scale representation of a discrete signal, but reducing processing time significantly. The most compact way of describing this process as well as processes to determine the Wavelet coefficients, is the representation of the filters in the form of operator. For a sequence f [n] representing the signal discrete to be decomposed, operators H and G are defined according to the following expressions:

$$(Hf)_k = \sum_n h[n-2k]f[n] \qquad (4)$$

$$(Gf)_k = \sum_n g[n-2k]f[n] \qquad (5)$$

These equations represent the filtering of the signal through the filters digital h [n], g [n]. The factor 2k represents the subsampling and the operators H and G correspond to a step in the Wavelet decomposition [17].The decomposition process begins by applying to the discrete signal a filter low-pass half-band with impulse response h[n]. The filtering of the signal corresponds to the mathematical operation of convolution of this one with h[n]. Each filter eliminates the frequency components above the middle of the bandwidth of the signal.

The highest frequency that exists in the discrete signal is π radians, if the signal is sampled at the Nyquist frequency (which is twice the maximum frequency that exists in signal 2π). After filtering the signal through the low pass filter, half of the samples can be removed following the Nyquist rule, since the signal now has its maximum frequency in π / 2 radians instead of π radians.

By discarding a sample of every two, the signal is sub-sampled by said factor, making the result have half points of the original. So the scale of the DWT is doubled. It must be taken into account that the filtering passes low eliminates the high frequency components but does not modify the scale. The signal is passed in the time domain through various filters, which separate portions of high or low frequencies from the signal. This procedure is repeated several times, and in each of them is removed a fraction of the signal corresponding to a certain band of frequencies, obtaining in each stage a level of coefficients of the DWT.

For example, in the case of applying the DWT to an electrocardiographic signal with sampling frequency 240 Hz and 2048 samples. In a first phase it is divided the signal in two parts passing the signal through a high-pass filter and one passes under which produces two different versions of the same signal: portion of the signal corresponding to the first level of detail. 0-60 Hz (portion low pass), and 60-120 Hz (pass-high pass), each of the parties will have 1024 samples. Subsequently, a portion is taken (usually the low-pass) or both, and does the same procedure again. If you take the low-pass portion, now will have the same signal with half of the samples that had (second level of detail), in the frequencies 0-30 Hz, 30-60 Hz. The low-pass portion is again taken and passed through the filters high pass and low pass; so that we now have two sets of signals corresponding to 0-15 Hz, 15-30 Hz. (third level of detail) each one with half of the samples that had the previous level.

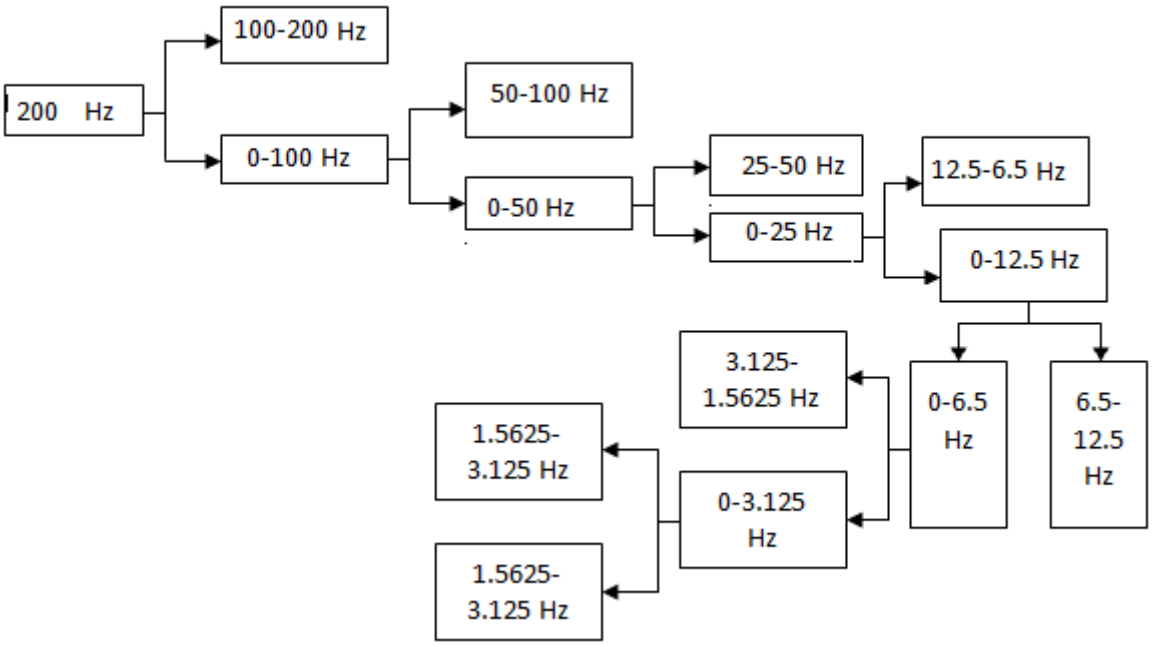


Figure 3: 7- Discrete level of frequency using DWT

III) **Post-Processing**

A) RBF:

Among the different architectures of artificial neural networks we adopted for this study the Multi Layer Perceptrons (MLPs, Multi Layer Perceptron) [18]. MLPs are most commonly used in supervised learning approaches, that is, when an association between two types of data, representing network input and output respectively, must be learned. In an MLP artificial neurons are organized in layers. Neurons belonging to the same layer are not connected to each other. Each neuron receives its inputs from the previous layer and transmits the result of its treatment to the next layer. The two extreme layers correspond to the layer that receives the data (input layer), and the layer that provides the result of the performed processing (output layer). The intermediate layers are called hidden layers, their number is variable. The connectivity between the successive layers is total and each connection is weighted by a weight.

An improved version of genetic algorithms is used in [19] for adjusting the structure and parameters of a neural network. A network of neurons with switches introduced into the links is proposed to predict the sunspots. While [20] has devised a flexible method for solve the problem of the traveling salesman using genetic algorithms to give a maximum approximation of the problem of cost reduction.

A genetic approach has been used for learning neural networks with radial base functions in [21], the algorithm is used to adjust the weights between the hidden layer and the exit to classify four problems known databases with a precision between 91 and 98%. Other evolutionary algorithms have been applied recently for learning neural networks, we can mention: Artificial Bees Colony (ABC) [22] and [23], Particle Swarm Optimization (PSO) [24] and [25], Differential Evolution [26] .

Approximation of RBF functions

Let $y = f(x)$ be a function with $x \in R^D$ and $x \in R^D$ and let $H_i, i = 1, \dots N$ a set or basic functions.

The function f can be written in the form (4.1)

$$y = f(x) = \hat{f}(x) + r(x) \quad (6)$$

Where $r(x)$ is the residue.

The function $f(x)$ can be approximated by $\hat{f}(x)$ given by the form (4.2)

$$y \simeq \hat{f}(x) = \sum_{i=1}^{N} w_i h_i(x) \quad (7)$$

The goal is to minimize the error by adjusting the w_i settings appropriately. A possible choice for the approximation error is the L_2 standard of the residual function $r(x)$ is defined as:

$$(\|r(x)\|L_2)^2 = \int r(x)^2 dx \quad (8)$$

Approximation of functions by RBF networks The output of an RBF network is given by equation (4.4)

$$\hat{y} = \hat{f}(x) = \sum_{i=1}^{N} a_i \emptyset_i(x, \mu_i, \sigma_i) \quad (9)$$

Using the above equation (4.4), the function $\hat{f}(x)$ can be written as:

$$y = \sum_{i=1}^{N} a_i \emptyset_i(x, \mu_i, \sigma_i) + r(x) = \hat{y} + r(x) \quad (10)$$

The error can be minimized by appropriately adjusting the weights a_i, centres μ_i and the widths σ_i.

When the function $f(x)$ is unknown, and having a set of

couples outputs (x_i, y_i) , $i = 1 \ldots . n$,the RBF network can be built and it can be learn it to follow the function $f(x)$.

Learning this network is an optimization problem to choose the values of centres, weights and widths to minimize criterion:

$$J^2(\mu, \sigma, a) = \sum_{i=1}^{n} \|y_i - \hat{y}\|^2{}_i \quad (11)$$

Learning approach of RBF networks by the algorithm ABC-PP Learning neural networks with radial base functions with N neurons can be considered as an optimization problem whose objective is to reduce the error between the desired output of the function to be approximated and the output of the network.

ABC Optimization of Neural Network

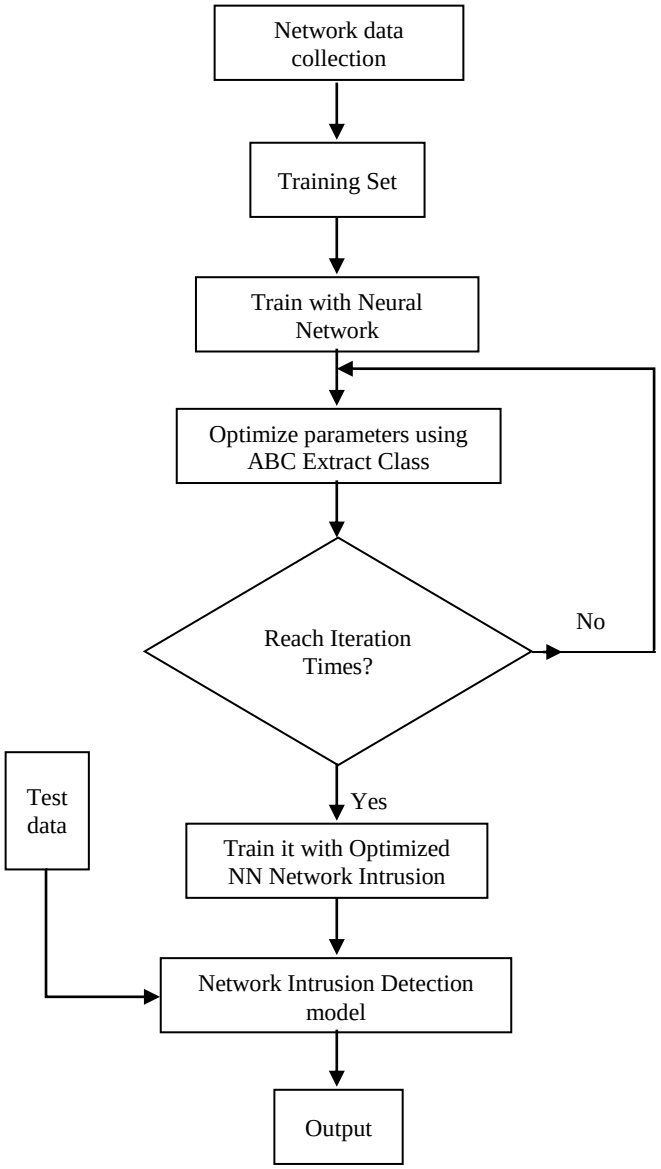


Figure 4: Flow diagram for ABC-BP Neural Network based network intrusion detection model

The ABC (Artificial Bee Colony) algorithm was developed by Karaboga and Basturk in 2005[27], inspecting the behaviours of real bees to find the food source, called nectar, and sharing information from food sources to others bees in the nest.

In this algorithm, artificial bees are defined and classified into three groups, employing bees (bees seeking food), spectators (observation bees) and scouts (scouts) are responsible for finding new foods (nectar from new sources)

If the honey bee from a site fails to find the source of food, it must necessarily become a scout to randomly search for new sources of food. Employer bees share the information with the bees in a beehive so that the honey bees can choose a food source to explore. The process of the ABC algorithm is presented as follows:

Step 1- Initialization:

Start by selecting Fe percentage of population randomly in the search space using the following equation:

$$U_j = U_j^{min} + n_j * \left(U_j^{max} - U_j^{min}\right) \quad n_j \in [0,1] \quad (12)$$

Knowing that each bee carries a vector "U" of "n" solution.

Assess them using the objective function, and calculate their Fitness values, called the amount of liquid by the following equation:

$$Fitness = \frac{1}{F_{objective}+1} \quad (13)$$

Fe represents the ratio of bees in the total population.

Once these populations are placed in the search space, they take the name of the employing bees.

Step 2- Movement of Employer Bees:

Calculate the probability of choosing a food source by the equation:

$$P_i = \frac{0.9 * Fitness_i}{\max(Fitness_i)} + 0.1 \quad (14)$$

Select a food source and determine its amount of nectar. The equation of motion of the observing bees is given as follows:

$$m_{ij}(t+1) = x_{kj} + y\left(x_{ij}(t) - x_{kj}(t)\right) \quad y \in [0,1] \quad (15)$$

x_{ij} is the number of iterations, xij is the honeybee used randomly, j is the size of the solution vector that produces a series of random variables in the range [-1.1]; where k∈ {1,2,3, ... N} and j∈ {1,2, ... D} are chosen randomly with 'k' different from 'i', 'D' is the number of parameters to optimize.

As m_{ij} is the i^{th} position of the bee spectator, t is the number of iterations, x_{ij} is the bee used randomly, $'j'$ represents the size of the solution vector which produces a series of random variables in the range [-1,1]; where $k \in \{1,2,3, \ldots N\}$ and $j \in \{1,2, \ldots D\}$ are chosen randomly with $'k'$ different from $'i'$, $'D'$ is the number of parameters to optimize.

Step 3- Move Scouts:

If the fitness values of the employing bees are not improved by a predetermined number of iterations, these food sources are dropped, and the bee found in this location will randomly

move to explore other new locations. (Employer bees become Scouts). This explanation is mathematically translated by the equation:

$$V_{ij} = V_{ij}^{min} + \varphi_{ij} * \left(V_{ij}^{max} - V_{ij}^{min}\right) \qquad \varphi_{ij} \in [0,1] \qquad (16)$$

Step 4- Update the best food source found so far:

Learn the best fitness value and position, which are found by the bees, and memorize them.

Step 5- Stop Criterion:

Check the calculation process until the number of iterations reaches the predefined maximum value or a solution of the acceptable objective function is found.

B. Random Forest Classifier

A random forest is a classifier comprising of one group of structured tree predictors $[T(x,\ominus_k), k = 1, \ldots.]$ where the $[\ominus_k]$ are random vectors of identical distributions and where each tree delivers a unit vote for the most well-known class for each entry x.

The main advantage of this structure is that it avoids the danger of over-learning for any method of prediction based on induction. BREIMAN, [28] shows that when the number of trees involved in the prediction forest increases, the generalization error rate converges to a limit value, of which an upper bound can be estimated on the basis of the characteristics intrinsic features of the forest.

If the marginal function of a random forest $T(X,\ominus)$

$$mr(X,Y) = P_{\ominus}(T(X,\ominus) = Y) - \max_{j \neq T} P_{\ominus}(T(X,\ominus) = j) \qquad (17)$$

which represents the confidence level of the ranking established by the trees of this forest on the population (X, Y), measured by the difference of probability between the prediction of the correct class Y and the best class erroneous j ≠ Y, one can define the prediction value of a game of trees {T (x, Θ)} by the mathematical expectation of this function

$$s = E_{x,y}[mr(X,Y)] \qquad (18)$$

The dependency between trees in a forest ρ (⊖,⊖ ') is measured by the correlation between their gross marginal function, evaluated for fixed and distinct parameter values ⊖,⊖ '. By means of these definitions, an upper limit to the error in generalization (TEG) of any random forest is given by the relation.

$$TEG \leq \bar{\rho}(1 - s^2)/s^2 \qquad (19)$$

The tree structured classifiers combine to form a random forest classifier (Figure 5(a) and Figure 5(b)). The input objects are sourced to every tree and individual unit vector is voted by every tree. The mode from individual classes with maximum number of votes is given as output class by the forest. More than just prediction the random forest can be trained to achieve information on data proximities for feature generation. The variables closer to responsible variables are predicted by their importance while their relation is the function of partial independencies.

In training of N cases, the trees are trained for sampling. This sampling is a random process that replaces the original data with the bootstrap sample. Out of M number of input variables, m variables are chosen in random manner ($m \ll M$), and best split on these predictors split the node. During the entire operation of training, the value of m is hold for constant. Generally the value of m is selected M times smaller than M inputs. Without pruning, each tree grows to the maximum possible range. This training generates multiple numbers of trees with the maximum value decided by N_{tree}. The length of tree roots (depth of tree) is estimated via parameter node size (i.e. no. of leaf nodes) that is generally fixed to unity. In testing phase the input is fed to forest that runs to all the trees and classification from each individual is recorded as vote. The instance that gets maximum no. of votes is declared as winner or output.

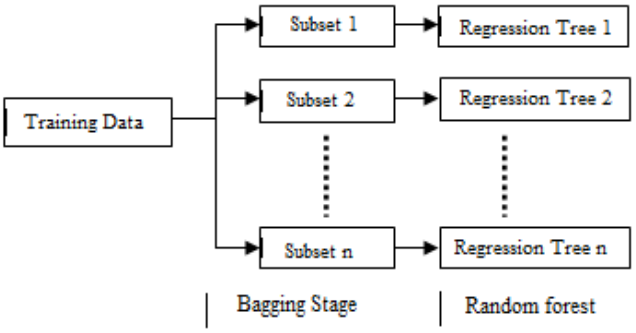


Figure 5(a): Algorithmic Sequence of Random Forest Classifier Training Phase

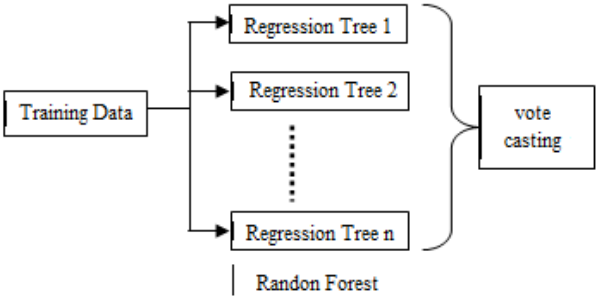
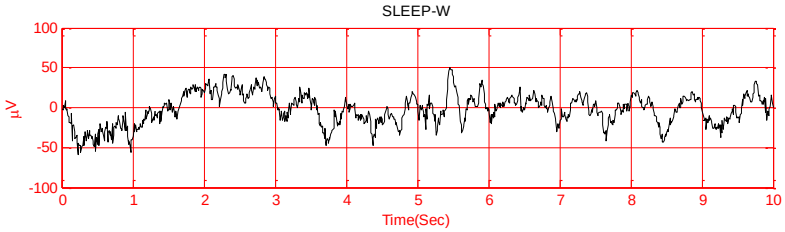


Figure 5(b): Algorithmic Sequence of Random Forest Classifier Testing Phase

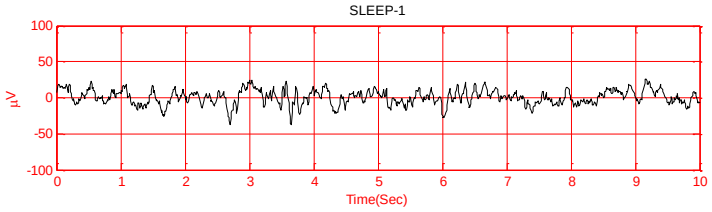
The energy levels from wavelet decomposition are sourced to classifiers. The Neural network and random forest both techniques are well known for this activity in a series of applications. In this paper, these classifiers describe the sleep state in five segments. The parameters of classifiers are as following:

Database Acquisition

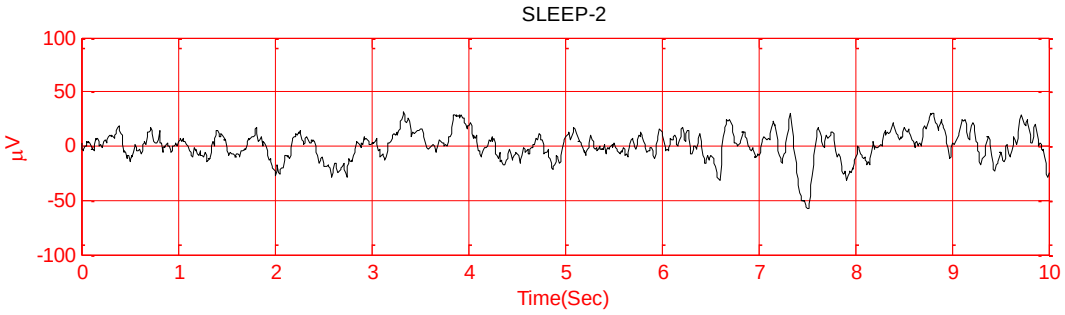
The database is acquired from physionet.org/physiobank/database/sleep-edf. The samples contains the horizontal EOG, P_zOZ EEG with sampling rate of 100 Hz of 25-35 years age


SLEEP-W
μV
Time(Sec)

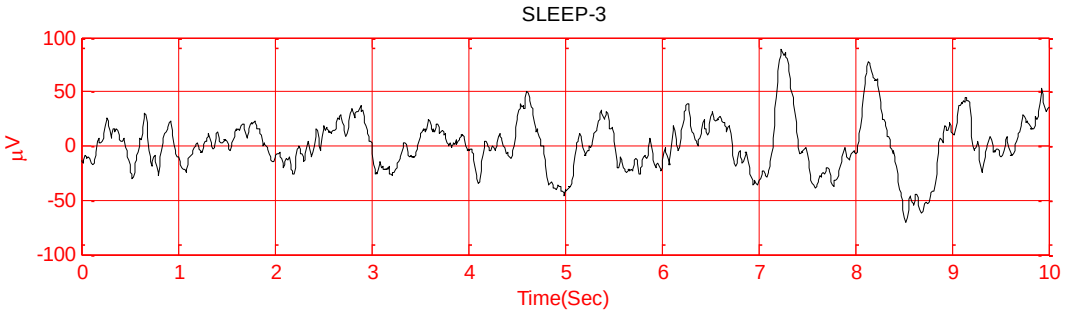
(a)


SLEEP-1
μV
Time(Sec)

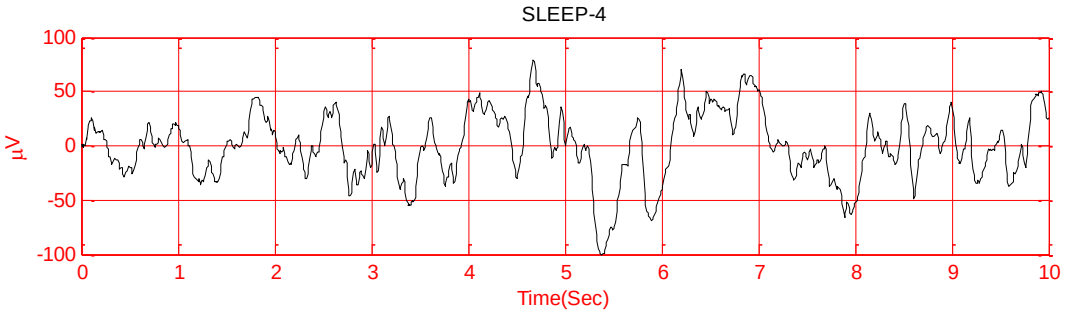
(b)


SLEEP-2
μV
Time(Sec)

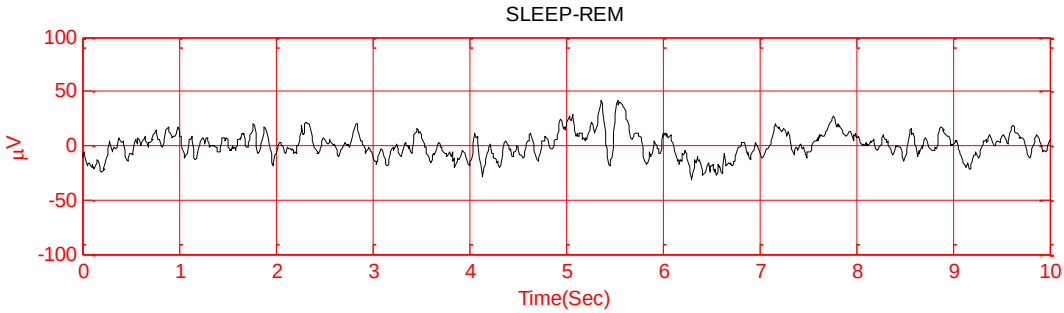
(c)


SLEEP-3
μV
Time(Sec)

(d)


SLEEP-4
μV
Time(Sec)

(e)



(f)

Figure 6: Different sleep stages

RESULTS :

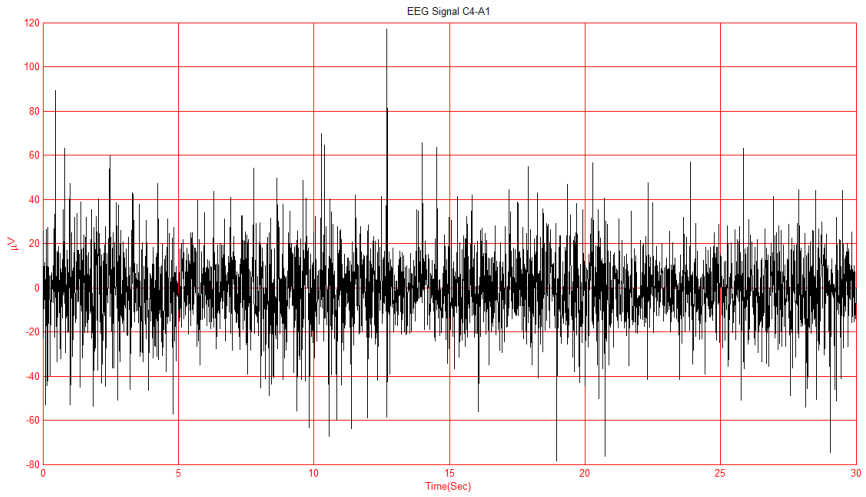


Figure 7: Original EEG signal

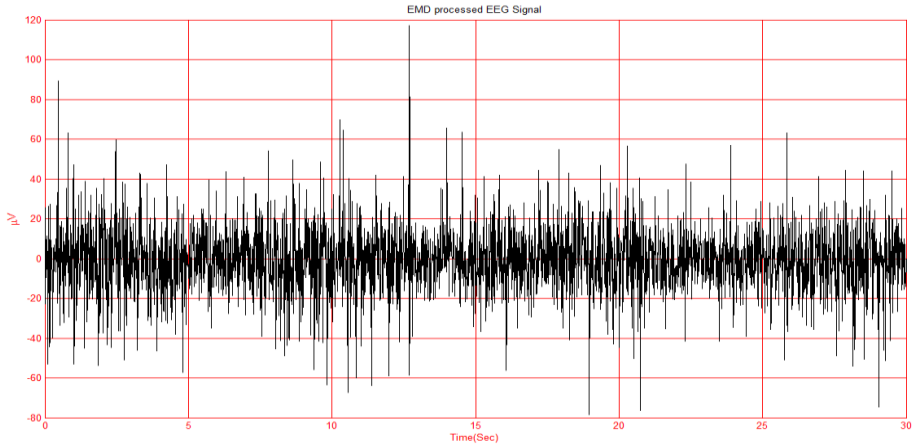


Figure 8: Filtered EEG signal

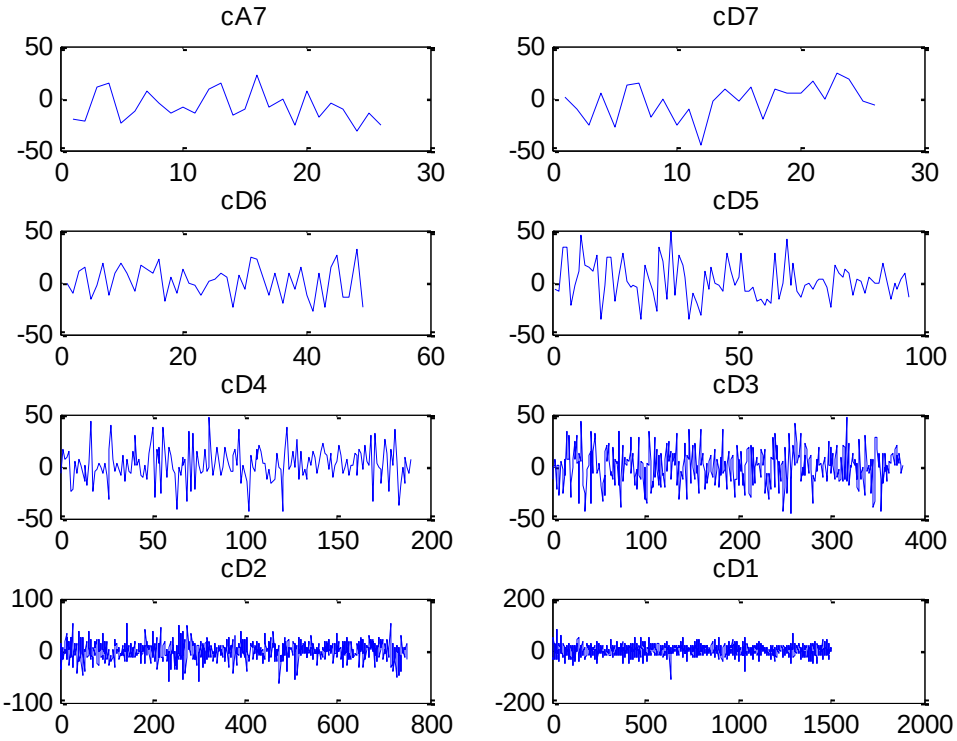


Figure 9: DWT coefficients

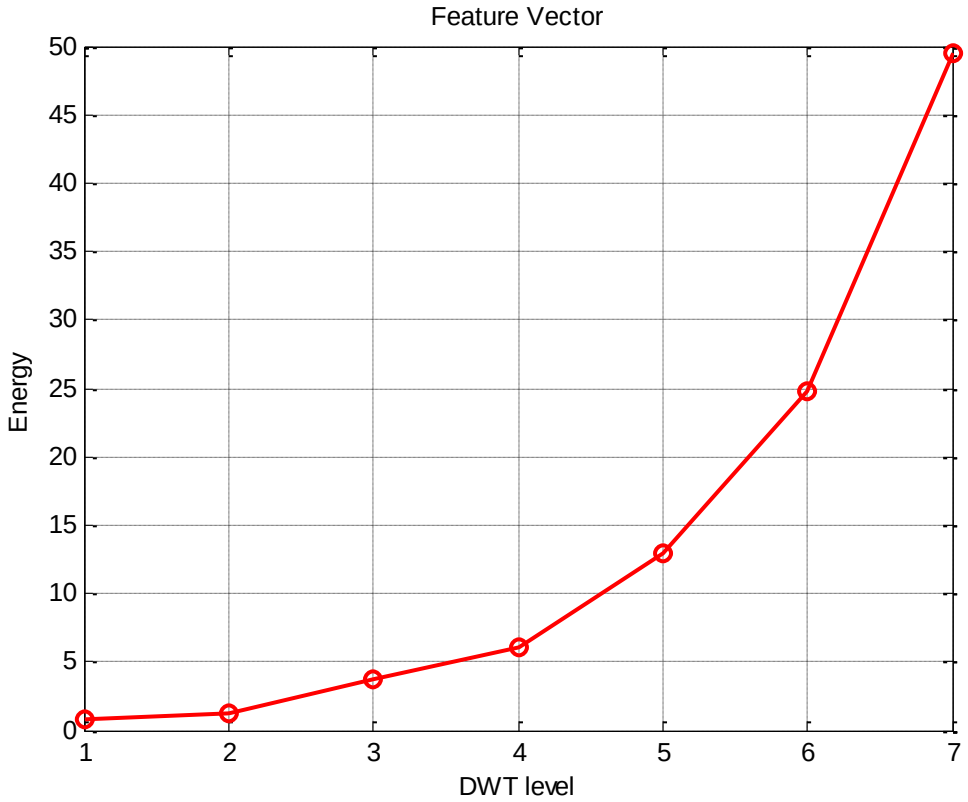


Figure 10: DWT features

Performance evaluation parameter:

Recall:

$$Recall = \frac{TP}{(TP+FN)} \quad (20)$$

Accuracy:

$$ACC = \frac{TP+TN}{P+N} \quad (21)$$

Where, $P + N =$ Total population.

Studying results of [29] and [30] indicates that wavelet transform or neural network both provides the efficient results in their accuracy. Table 1 also depicts that using Random Forest, the accuracy beats all the levels and arguments of previous researches. In our proposed method we also calculated the false and true negative and positive rates and are given in table 2.

Table 1: Sleep stage analysis comparison

Reference	Signal	Method	Classifier	Accuracy
[30]	EEG	WT	Bayesian decision	97.89%
[29]	EEG		RF,SVM	98.27%
Proposed 1	EEG	WT	Neural Network	97.33%
Proposed 2	EEG	WT	Random Forest	98.5677%
Proposed 3	EEG		NN+ABC	99.08%

Table 2: False and True Positive and Negative Rates

Methods	False Negative rate	False positive rate	True positive rate	True negative rate
Proposed 1	.0089	.066	.9024	.9801
Proposed 2	.0091	.054	.9145	.9825
Proposed 3	.0096	.046	.9356	.9963

CONCLUSION

In this paper, first step we presented the study of an automatic sleep classification system. Data has been taken from physionet considering EEG and EOG mixed .To separate the desired EEG data is converted into frequency domain with the help of FFT, afterwards it is filtered with Butterworth filter, once clean EEG is achieved then its energy is calculated with respective bands of alpha, beta, gamma. In a second step, we used the random forest and neural network classifier for classification of sleep stage. Extracted energy value after DWT, further trained for sleep stages in classifier .Keeping objective of higher accuracy neural network classification is further tuned with ABC optimization technique .The evaluated results claims ,the optimized neural network outperform then the traditional classifiers.

REFERENCES

[1] Moretti, D.V., Babiloni, F., Carducci, F., Cincotti, F., Remondini, E., Rossini, P.M., Salinari, S. and Babiloni, C., 2003. Computerized processing of EEG–EOG–EMG artifacts for multi-centric studies in EEG oscillations and event-related potentials. *International Journal of Psychophysiology*, *47*(3), pp.199-216.

[2] Lajnef, T., Chaibi, S., Ruby, P., Aguera, P.E., Eichenlaub, J.B., Samet, M., Kachouri, A. and Jerbi, K., 2015. Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. *Journal of neuroscience methods*, *250*, pp.94-105.

[3] Hori, T., Sugita, Y., Koga, E., Shirakawa, S., Inoue, K., Uchida, S., Kuwahara, H., Kousaka, M., Kobayashi, T., Tsuji, Y. and Terashima, M., 2001. Proposed supplements and amendments to 'a manual of standardized terminology, techniques and scoring system for sleep stages of human subjects', the Rechtschaffen & Kales (1968) standard. *Psychiatry and clinical neurosciences*, *55*(3), pp.305-310.

[4] Kumar, Y., Dewal, M.L. and Anand, R.S., 2014. Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network. *Signal, Image and Video Processing*, *8*(7), pp.1323-1334.

[5] Zhu, G., Li, Y. and Wen, P.P., 2014. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE journal of biomedical and health informatics*, *18*(6), pp.1813-1821.

[6] Shi, L.C., Jiao, Y.Y. and Lu, B.L., 2013, July. Differential entropy feature for EEG-based vigilance estimation. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE* (pp. 6627-6630). IEEE.

[7] Heinrich, A., Geng, D., Znamenskiy, D., Vink, J.P. and de Haan, G., 2014. Robust and sensitive video motion detection for sleep analysis. *IEEE journal of biomedical and health informatics*, *18*(3), pp.790-798.

[8] Frederick, C.A., Kayyali, H., Schmidt, R.N. and Kolkowski, B.M., Cleveland Medical Devices Inc., 2015. *Method and device for sleep analysis*. U.S. Patent 9,202,008.

[9] de Raaff, C.A., Bindt, D.M., de Vries, N. and van Wagensveld, B.A., 2016. Positional obstructive sleep apnea in bariatric surgery patients: risk factor for postoperative cardiopulmonary complications?. *Sleep and Breathing*, *20*(1), pp.113-119.

[10] Halász, P., Bódizs, R., Parrino, L. and Terzano, M., 2014. Two features of sleep slow waves: homeostatic and reactive aspects–from long term to instant sleep homeostasis. *Sleep medicine*, *15*(10), pp.1184-1195.

[11] Nakajima, M. and Al'Absi, M., 2013. Nicotine. In *Encyclopedia of Behavioral Medicine* (pp. 1334-

1335). Springer New York.

[12] Younes, M., Ostrowski, M., Soiferman, M., Younes, H., Younes, M., Raneri, J. and Hanly, P., 2015. Odds ratio product of sleep EEG as a continuous measure of sleep state. *Sleep*, *38*(4), pp.641-654.

[13] Smith, K., Gomez-Rubio, A.M., Harris, T.S., Brooks, L.E. and Mosquera, R.A., 2016. Unusual Ventilatory Response to Exercise in Patient with Arnold-Chiari Type 1 Malformation after Posterior Fossa Decompression. *Case reports in pediatrics*, *2016*.

[14] Shein-Idelson, M., Ondracek, J.M., Liaw, H.P., Reiter, S. and Laurent, G., 2016. Slow waves, sharp waves, ripples, and REM in sleeping dragons. *Science*, *352*(6285), pp.590-595.

[15] Parisky, K.M., Rivera, J.L.A., Donelson, N.C., Kotecha, S. and Griffith, L.C., 2016. Reorganization of sleep by temperature in Drosophila requires light, the homeostat, and the circadian clock. *Current Biology*, *26*(7), pp.882-892.

[16] Becerra-Santacruz, H. and Lawrence, R., 2016. Evaluation of the thermal performance of an industrialised housing construction system in a warm-temperate climate: Morelia, Mexico. *Building and Environment*, *107*, pp.135-153.

[17] Gentilhomme, T., Oliver, D.S., Mannseth, T., Caumon, G., Moyen, R. and Doyen, P., 2015. Ensemble-based multi-scale history-matching using second-generation wavelet transform. *Computational Geosciences*, *19*(5), pp.999-1025.

[18] Tang, J., Deng, C. and Huang, G.B., 2016. Extreme learning machine for multilayer perceptron. *IEEE transactions on neural networks and learning systems*, *27*(4), pp.809-821.

[19] Sastry, K., Goldberg, D.E. and Kendall, G., 2014. Genetic algorithms. In *Search methodologies* (pp. 93-117). Springer US.

[20] Scrucca, L., 2013. GA: a package for genetic algorithms in R. *Journal of Statistical Software*, *53*(4), pp.1-37.

[21] Wu, J., Long, J. and Liu, M., 2015. Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm. *Neurocomputing*, *148*, pp.136-142.

[22] Wang, S., Zhang, Y., Ji, G., Yang, J., Wu, J. and Wei, L., 2015. Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization. *Entropy*, *17*(8), pp.5711-5728.

[23] Karaboga, D., Gorkemli, B., Ozturk, C. and Karaboga, N., 2014. A comprehensive survey: artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review*, *42*(1), pp.21-57.

[24] Dheeba, J., Singh, N.A. and Selvi, S.T., 2014. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of biomedical informatics*, *49*, pp.45-52.

[25] Alexandridis, A., Chondrodima, E. and Sarimveis, H., 2016. Cooperative learning for radial basis function networks using particle swarm optimization. *Applied Soft Computing*, *49*, pp.485-497.

[26] Bas, E., 2016. The training of multiplicative neuron model based artificial neural networks with differential evolution algorithm for forecasting. *Journal of Artificial Intelligence and Soft Computing Research*, *6*(1), pp.5-11.

[27] Karaboga, D. and Basturk, B., 2007. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, *39*(3), pp.459-471.

[28] Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

[29] Şen, B., Peker, M., Çavuşoğlu, A. and Çelebi, F.V., 2014. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *Journal of medical systems*, *38*(3), p.18.

[30] Fu, C., Zhang, P., Jiang, J., Yang, K. and Lv, Z., 2017. A Bayesian approach for sleep and wake classification based on dynamic time warping method. *Multimedia Tools and Applications*, *76*(17), pp.17765-17784.