A modified grey wolf optimization based feature selection method from EEG for silent speech classification

Article · November 2019

DOI: 10.1080/02522667.2019.1703262

4 authors:

Rajdeep Ghosh
Gauhati University

Nidul Sinha

Saroj Kr. Biswas
National Institute of Technology, Silchar

Souvik Phadikar
National Institute of Technology, Silchar


Journal of Information & Optimization Sciences
Volume 37 (2016), Number 5, September
Special Issue on
Business Analytics and Intelligence
CONTENTS

Taylor & Francis
Taylor & Francis Group

Journal of Information and Optimization Sciences

ISSN: 0252-2667 (Print) 2169-0103 (Online) Journal homepage: https://www.tandfonline.com/loi/tios20

A modified grey wolf optimization based feature selection method from EEG for silent speech classification

Rajdeep Ghosh, Nidul Sinha, Saroj Kumar Biswas & Souvik Phadikar

Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 40 (2019), No. 8, pp. 1639–1652
DOI : 10.1080/02522667.2019.1703262

A modified grey wolf optimization based feature selection method from EEG for silent speech classification

Rajdeep Ghosh *
Department of Information Technology
Gauhati University
Guwahati 781014
Assam
India

Nidul Sinha †
Department of Electrical Engineering
National Instititue of Technology Silchar
Silchar 788010
Assam
India

Saroj Kumar Biswas §
Department of Computer Science & Engineering
National Institute of Technology Silchar
Silchar 788010
Assam
India

Souvik Phadikar ‡
Department of Electrical Engineering
National Institute of Technology Silchar
Silchar 788010
Assam
India

*E-mail: rajdeep.publication@gmail.com (Corresponding Author)
†E-mail: nidulsinha@hotmail.com
§E-mail: saroj@cse.nits.ac.in
‡E-mail: souvik.phadikar@gmail.com

Abstract

Brain computer interfaces (BCI's) employing electroencephalographic signals are being applied to a wide variety of applications like motor imagery task classification, prosthetics etc. Electroencephalography (EEG) data are inherently non-stationary and noisy, and as such identification of appropriate features for classification is a crucial task. Selection of features based on genetic algorithms (GA) has been applied, but it leads to a redundant set of features. In the present work, grey wolf optimization (GWO) based feature selection method has been applied on EEG data for silent speech classification. The EEG data from the ABISSR (Analysis of Brain Waves and development of intelligent model for silent speech recognition) project was used in the proposed work. An accuracy of 65% was obtained in classifying five imagined vowels /a/, /e/, /i/, /o/ and /u/ from EEG data using support vector machine (SVM). Moreover, it was observed that the GWO outperformed GA in optimization.

Subject Classification: *Computer Science*

Keywords: *Brain computer interface, Electroencephalogram, Feature Selection, Grey wolf optimization*

1. Introduction

Brain computer interface based applications are gaining widespread attention of researchers. A wide variety of BCI applications are being developed such as emotion recognition from EEG, stress monitoring etc. However, such applications require an accurate classifier for identifying the various tasks associated with the concerned BCI. In Silent speech recognition, the subjects imagine individual vowels and the BCI system tries to recognize them from the recorded EEG.

Wester[1] in his thesis recognized different speech modalities like silent, mumbled, unspoken, normal and whispered speech from EEG data. He also established the significance of two regions (Wernicke's area, Broca's area) of the brain. Short time Fourier transform (STFT) and linear discriminant analysis (LDA) were employed for feature extraction. He further used hidden Markov model (HMM) as a classifier and concluded that it was possible to recognize silent speech from EEG, but an appropriate pre-processing method was not employed by him. DaSalla et al. [3] described a BCI system for the imagination of vowel speech. EEG data were collected from three subjects imagining three tasks which were: imagination of the English vowels /a/, /u/, and a state of no action respectively. The authors averaged the trials and observed readiness potentials at 200ms after the presentation of the stimulus and changes in the neural activity related to speech imagination after 350ms.

The authors used spatial filters for discriminating the tasks. The feature vectors thus obtained were classified using a non-linear SVM kernel and obtained a classification accuracy in the range of 68% to 78%. Spatial filters provide an efficient classification of EEG for only two classes, but fail for multiple classes. Moreover, spatial filters are easily affected by noise and artifacts and have an adverse effect on the classification. Brigham et al. [4] considered only two syllables /ku/ and /ba/, the EEG data were pre-processed to reduce the effects of contaminants of EEG like artifacts, noise etc. Autoregressive (AR) coefficients were extracted as features from the EEG data and further classified with k-nearest neighbor (kNN). Their results established that the identification of imagery speech was possible from EEG data, however selecting optimal trials was an important task to be performed during classification [16].

Chi et al. [5] classified five phonemes based on the order of their articulation (tongue, jaw, nasal, fricative and lips) during the production of speech. Naive Bayes and LDA were applied for classification of the recorded EEG data. Results showed that it was possible to discriminate the EEG data during imagination from those where there was no imagination by the subjects. Moreover, the classes were discriminable among the data collected in a single day. The inter-session variability common in BCI applications was not addressed that much specifically in that case. Matsumoto et al. [6] used event related potentials (ERPs) from the subjects for classification of silent speech using brain computer interfaces. An adaptive collection (AC) was used to select suitable output signals from CSP filters. The classification accuracies for the imagination of phonemes /a/ vs. /u/ with 63 channels were in the range of 73-92%. But the selection of the CSP's and the number of channels is challenging and requires supervision.

Kamalakkannan et al. [7] in their work classified five imagined vowels /a/,/e/,/i/,/o/ and /u/ and used variance mean, average power and standard deviation as features and bipolar neural network for classification. They obtained a maximum classification accuracy of 44%. They concluded that the EEG carried some distinctive information.

Recently Ghane et al. [8] adopted principal component Analysis (PCA) and robust PCA (RPCA) to recognize an EEG pattern. The imagined vowels were /a/, /e/, /i/, /o/, /u/. RPCA algorithm achieved a better recognition of patterns. Although the PCA algorithm has been widely used in literature to select the most relevant features, PCA does not give good results due to the inherent noise in EEG data. However, it was observed that the accuracy of such BCI systems depends on the features extracted.

The problem of feature selection is a prime concern in EEG signal processing as the dimension of the feature space is very large (typically 102 – 103 features). Such a large space of potential features excludes the possibility of an exhaustive search of the feature space, which is the only procedure for discovering the optimal subset of features. A variety of feature selection methods based on heuristic algorithms have been used in literature. Cimpanu et al. [9] used GA for feature selection and ranked features based on single objective and multi objective criterion. Bhattacharya et al. [10] used DE to select features for classification of EEG data set and concluded that their methods suffered from stagnation due to the local optima issue. Miao et al. [11] used artificial bee colony (ABC) algorithm to select the features from the EEG datasets. Houssein et al. [16] used whale optimization algorithm (WO) to select the significant features to classify epileptic subjects from normal subjects using SVM. Various heuristic algorithms have been used throughout the literature for selection of features from EEG. However, they suffer from local optimum issue and hence GWO has been adopted in the proposed work for getting a global optimum. Feature selection is important for a twofold purpose, one for leaving out irrelevant features and the other for reducing the curse of dimensionality. The feature selection and classification were done on the data recorded in the ABISSR project. In the ABISSR project, data were recorded for 64 channels from each subject and correspondingly 7 features were evaluated from every channel. Hence 448 features (i.e. 64x7=448) represent each data point. Thus a selection of the most relevant feature is of prime concern as the input space is inherently nonlinear and of higher dimensions. Although various heuristic algorithms available in the literature have been applied for the selection of features from EEG but none has been applied for the selection of features for silent speech classification from EEG data. The paper has been organized in the following order: section 1 presents a brief description of the literary works in the domain of silent speech recognition as well as the methods that have been used for feature selection, section 2 presents the concepts and methods that have been used in the present work followed by section 3, which illustrates about the various steps that have been followed in the proposed work, section 4 furnishes the results that have been obtained during the study followed by section 5 which summarizes the paper as well as describing the scope for future work.

II. Materials

The current section presents the concepts and methods that have been taken up in the present work. In the present work, various features from all the channels have been extracted from the EEG data. GWO has been used for selecting the set of most prominent features from the features extracted and support vector machine has been used for classifying the corresponding phoneme of the silent speech.

A. *Grey Wolf Optimization*

The GWO algorithm was first proposed by Mirjalali et al. [14] inspired by the hunting pattern and social behavior of the grey wolves. Wolves are classified as three types: alpha, beta, delta and omega according to the nature of their behaviour, with alpha representing the fittest individual, followed by beta, delta and omega. The readers may follow the work [14] for a detailed reference. The GWO algorithm can be summarized as:

Initialize *the wolf Population* X_i*, C, a and A.*
Evaluate the fitness value for every wolf.
X_α *represents the alpha wolf (best)*
X_β *represents the beta wolf (second best)*
X_δ *represents the delta wolf (third best wolf)*
While *(t < Maximum_number_of_iterations)*
    For *each wolf*
        Update position in the current iteration
    End for
Update the values of C, a and A
Calculate *the values of the fitness function for all the wolves*
Update *the values* X_α*,* X_β*, and* X_δ
$t \leftarrow t+1$
End While

Parameters a, A and C are related to the encircling behavior of the wolves during hunting. A binary grey wolf optimization (BGWO) algorithm has been used in the proposed work for feature selection based on [13]. The fitness function for the BGWO is described in equation 1 below.

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|C-R|}{C} \tag{1}$$

Where $\gamma_R(D)$ represents the classification accuracy for the selected feature subset R, α and β represent the parameters related to the classification quality, $\alpha \in [0,1]$, $\beta = 1-\alpha$ and C represents the total number of features, $|C\text{-}R|/|C|$ represents the ratio between the unselected features to the total number of features. The above equation is transformed into a minimization problem by replacing classification accuracy with error rate. The minimization problem can be represented as:

$$Fitness = \alpha E_R(D) + \beta\frac{|R|}{C} \tag{2}$$

Where $E_R(D)$ represents the error rate with the selected feature subset. β has been set to 0.01 in the proposed work [13].

B. *Support Vector Machine*

The SVM algorithm was first proposed by Cortes and Vapnik and has been widely used to classify data. The goal of SVM is to split the data into separate classes by creating a hyper-plane. For a data point (x_i, y_i) and $y_i \in \{-1, 1\}$, which denote the class labels, the data can be linearly separated, if there exists a scalar **b** and a vector w such that

$$\mathbf{w}x_i + \mathbf{b} \geq 1, \text{ if } y_i = 1 \tag{3}$$

&

$$\mathbf{w}x_i + \mathbf{b} \leq -1, \text{ if } y_i = -1 \tag{4}$$

Rewriting the above equation as shown below:

$$y_i(\mathbf{w}x_i + \mathbf{b}) \geq 1, i = 1, 2, ...n \tag{5}$$

The maximum margin hyper-plane that separates the training data is:

$$\mathbf{w}_0 x + \mathbf{b}_0 = 0 \tag{6}$$

The hyper-plane can be obtained by solving a constrained optimization equation as described in the paper [15].

C. *Feature Extraction*

From the EEG signal of each channel, the under mentioned features are computed.

(1) Mean: Mean represents the average for a set of values. Given a random variable X such that $X_i = \{X_1, X_2, ...X_n\}$, mean is represented as:

$$\mu_x = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{7}$$

(2) Standard Deviation: Standard deviation measures the maximum distance the data fluctuates from its mean or average.

$$\sigma_x = \sqrt{\sum \frac{(X_n - \mu_n)^2}{N}} \tag{8}$$

(3) Variance: Variance represents the variation of the data from their average or mean value. Given a random variable X such that $X_i = \{X_1, X_2, ...X_n\}$, variance is represented as:

$$Y^2 = \sum \frac{(X-\mu)^2}{N} \tag{9}$$

Where μ and N represents the mean and the number of samples in the dataset.

4) Entropy: Entropy is used to analyze the randomness of a time series data. If $X_i = \{X_1, X_2, ...X_n\}$, represents a single channel of EEG data, then the entropy is given as:

$$e = -\sum\nolimits_1^n (X_i^2 - \log(X_i^2)) \tag{10}$$

(5) Peak to peak amplitude (PPA): PPA measures the difference between the positive and the negative amplitude of a signal. If $X_k = \{X_1, X_2, ...X_n\}$, represents a single channel of EEG data, then peak to peak amplitude can be represented as:

$$p = \text{maximum}\,(X_k) - \text{minimum}\,(X_k) \tag{11}$$

(6) Kurtosis: Kurtosis describes the tailedness of a given distribution with respect to a normal distribution and is numerically represented as:

$$k = m_4 - 3m_2^2 \tag{12}$$

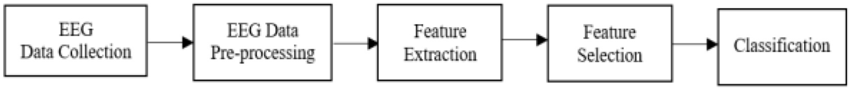

Fig. 1

Overall Methodology.

Where m_4 represents the fourth central moment for the data and m_2 represents the mean of the data.

(7) Band power: In this method the Butterworth infinite impulse response filter is used. So the signal x[t]contains only the required frequency components. The average power for a window size w is:

$$p[n] = \frac{1}{w}\sum_{k=0}^{w} p[n-k] \tag{13}$$

This means the power is averaged over window w. The final feature value is equal to $\ln(\overline{p}[n])$.

III. Proposed methodology

The overall task of silent speech recognition from EEG follows the traditional methodology adopted in a classification task. The procedure starts with collecting data, followed by preprocessing of the data to remove noise from the data. From the pre-processed data features are extracted followed by selection of the most prominent features for classification. In the proposed work, various time domain features from EEG have been adopted, GWO has been adopted for feature selection and SVM has been adopted for classification.

We describe the overall methodology and the various steps that have followed in the present section.

A. *Data Recording*

Data were recorded at National Institute of Technology, Silchar under the ABISSR project. A total of 45 subjects participated in the experiment. The subjects imagined various phonemes in the English language for 15 seconds. The phonemes considered for imagination were the vowels in English language /a/, /e/, /i/, /o/, /u/. The international 10-20 system was adopted to record the EEG data, 64 channels were considered for recording and was sampled at 512Hz. Data were recorded after obtaining the consent from the subjects and adequate permission from the authority.

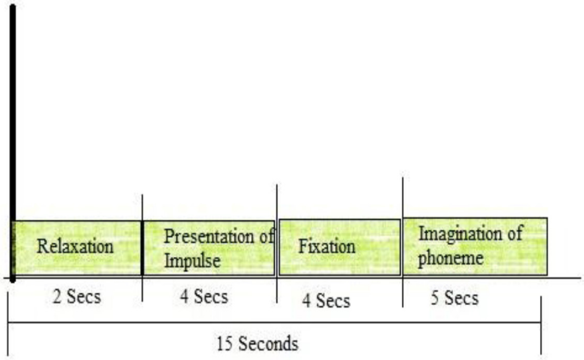

Fig. 2

Trial recording paradigm.

Figure 2 represents the trial recording paradigm where the subjects were initially relaxed for 2 seconds, followed by presentation of the phoneme for 4 seconds, followed by fixation for 4 seconds after which the subjects were asked to imagine the phoneme. Thus a trial consisted of 15 seconds and 5 trials were recorded for each phoneme which resulted in a total of 45x5x5=1125 EEG samples.

Fig. 3

A subject undergoing EEG recording.

B. *Data preprocessing*

After recording the data, the data were pre-processed. Pre-processing involves removal of noise, unwanted components from the EEG. Pre-processing in the proposed work considered the removal of power-line noise, where a notch filter was used at 50 Hz, followed by band pass filtering from 0-40 Hz. After which eyeblink artifacts were removed using autoencoder. The removal of artifact has been described in our previous work [12]. After removal of artifact, the data were trimmed. The trimmed data consisted for only last 5 seconds out of the 15 second data as it contains the EEG corresponding to the imagination of the phoneme.

C. *Feature extraction and selection*

From the pre-processed data, the following features are extracted namely: mean, standard deviation, variance, entropy, peak-to-peak amplitude, kurtosis and band power from each of the 64 channels, therefore a total of 64x7=448 features are calculated for a single trail. In the proposed BGWO algorithm [13] has been adopted for selecting the feature based on the criterion described in equation 1 and 2. The continuous valued GWO algorithm is modified into binary GWO using the equation described below.

$$X_d^{t+1} = \begin{cases} 1, \text{if sigmoid}\left(\dfrac{\sum_1^n x_i}{n}\right) \geq \text{rand} \\ 0, \quad \text{otherwise} \end{cases} \tag{14}$$

Where x_i represents the binary vector of size 7 as because there are 7 features. A 0 or 1 in the corresponding location of the vector specifies as to whether the particular feature is to be included for calculation or not. Thus 7 features described above are mapped to the corresponding locations in the vector.

D. *Classification*

In the proposed work, 5 classes corresponding to the 5 phonemes 'a', 'e', 'i', 'o' and 'u' have been considered. For estimating the classification accuracy SVM was used and accuracy was estimated using one-vs-one classification accuracy. The described method incorporates an ensemble of classifiers where each class is classified from the rest. For n classes

Table I

Classification accuracy for various methods implemented

Classifier	Class	Classification accuracy (in percentage)		
		Using BGWO	Using GA	Without any feature selection
SVM	A	62.2%	60.3%	56.3%
	E	67.2%	61.8%	57.4%
	I	69.5%	62.4%	57.1%
	O	61.2%	60.3%	54.3%
	U	64.9%	59.4%	53.2%
Average Classification Accuracy		65%	60.84%	55.66%

the maximum number of classifiers required for estimating one-vs-one classification accuracy is *(n*(n-1))/2*. In the proposed work, 10 classifiers have been considered as there are 5 classes. Thus, each classifier classifies between two classes and after the classification, the data sample coincident with the maximum number of classes is selected as the particular class for that particular data.

IV. Results

The proposed work was implemented in Matlab 2017b. The configuration of the system was 4GB RAM with Intel i7 processor. The

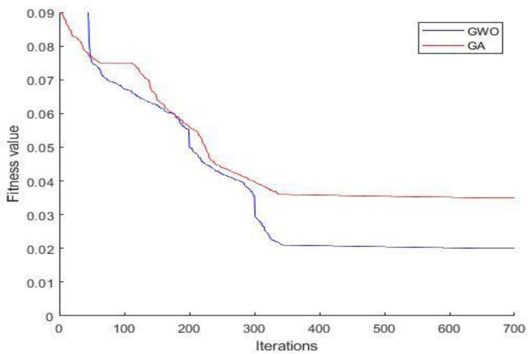

Fig. 4

Fitness values for GWO and GA versus the number of iterations.

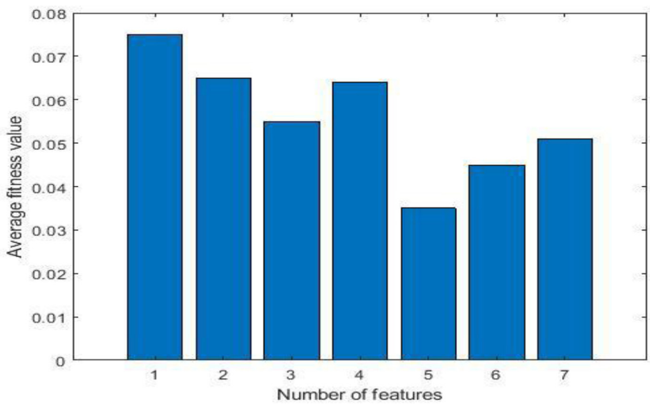

Fig. 5

Number of features versus the average fitness value.

GWO was initialized with 8 search agents, 700 iterations. The wolves are represented as binary values. Figure 4 shows that the overall value of the fitness function reaches an optimum value after 370 iterations. For evaluating the classification accuracy, 10% of the data were used for the purpose of testing. The maximum accuracies obtained for the individual phonemes are presented in table 1. The maximum accuracy achieved was 69.5% for the class representing the phoneme 'i' and the minimum accuracy achieved was 61.2% for the class representing the phoneme 'o'. Also the average classification accuracy obtained using BGWO for all the classes was 65%.

The features selected using BGWO resulted in higher classification accuracy than the method using GA for feature selection and without any feature selection method. The average accuracy obtained after using GA for feature selection is 60.84% and the average accuracy obtained without using any feature selection method is 55.66%.

Figure 4 presents a comparison of GA [9] with GWO and it is evident that GWO outperforms GA in optimizing the fitness function.

Figure 5 depicts the average value of the fitness values corresponding to the number of feature values. From the figure it is evident that the number of features resulting in minimum average fitness value is 5. Although the number of features is less, but due to the multi-channel setup of EEG device, the dimensionality of the problem increases by many folds.

V. Conclusion

EEG recordings relating to the imagination of five individual phonemes /a/, /e/, /i/, /o/ and /u/ have been considered for classification using SVM. From the above study it has been observed that our proposed model classifies EEG data into five classes with reduced number of features selected using BGWO and results in higher classification accuracy than the method using GA for feature selection, but due to the increased dimensions of the data it is difficult to conceptualize about the intrinsic characteristics of the data. Even the run-time for the algorithm is quite high. However, future works might focus on selection of appropriate or the most prominent channels related to the phoneme imagination, ignoring the less prominent ones.

Acknowledgement

The grant for the research was provided by the Ministry of Communications & Information Technology, Govt. of India with the grant number: 13(13)/2014-CC&BT.

References

[1] M. Wester," Unspoken Speech Recognition Based On Electroencephalography", Thesis, Carnegie Mellon University, Pittsburgh, PA, USA,2006.

[2] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan," Toward EEG Sensing of Imagined Speech", Human-Computer Interaction, Part I, HCII 2009, LNCS 5610, pp:40–48, 2009.

[3] C. DaSalla , H. Kambara, M. Sato, Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns", Neural Networks 22 , pp:1334–1339,2009.

[4] K. Brigham, B.V.K. V. Kumar," Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy", 4th International Conference on Bioinformatics and Biomedical Engineering, Chengdu, China, pp: 1–4,2010.

[5] X. Chi, J. B. Hagedorn, D. Schoonover and M. D'Zmura," EEG-Based Discrimination of Imagined Speech Phonemes", *International Journal of Bioelectromagnetism*, Vol. 13, No. 4, pp: 201-206, 2011.

[6] M. Matsumoto," Silent Speech Decoder using Adaptive Collection", Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces, Haifa, Israel , pp-73-76, 2014.

[7] R. Kamalakkannan, R. Rajkumar, M. Madan Raj., and S. Shenbaga Devi," Imagined Speech Classification using EEG", *Advances In Biomedical Science And Engineering,* Vol. 1, Number 2, pp-20-32, 2014.

[8] P. Ghane, G. Hossain and A. Tovar," Robust Understanding of EEG Patterns in Silent Speech", *National Aerospace and Electronics Conference* (NAECON), pp: 345–352, 2015.

[9] C. Cîmpanu, L. Ferariu, F. Ungureanu and T. Dumitriu, "Genetic feature selection for large EEG data with commutation between multiple classifiers", 21st International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, pp:618-623, 2017.

[10] S. Bhattacharyya, A. Sengupta, T. Chakraborti, A. Konar and D. N. Tibarewala, "Automatic feature selection of motor imagery EEG signals using differential evolution and learning automata", *Medical & Biological Engineering & Computing,* Vol. 52, Issue 2, pp:131–139, 2014.

[11] M. Miao, A. Wang and F. Liu, "Application of artificial bee colony algorithm in feature optimization for motor imagery EEG classification", *Neural Computing and Applications,* Vol. 30, Issue 12, pp:3677–3691, 2018.

[12] R. Ghosh, N. Sinha, and S. K. Biswas," Automated eye blink artefact removal from EEG using support vector machine and autoencoder", IET Signal Processing, Vol.13, Issue 2, pp:141-148, 4 2019 .

[13] E.Emarya, H. M. Zawbaa and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection", Neurocomputing, Vol. 172, 8, pp.:371-381, 2016.

[14] S. Mirjalili, S. M. Mirjalili and A. Lewis," Grey Wolf Optimizer", Advances in Engineering Software, Vol. 69, pp:46–61, 2014.

[15] C. Cortes and V. Vapnik, "Support Vector Networks", Machine Learning,Vol.20, pp.273–297,1995.

[16] E. H. Houssein, A. Hamad, A. E. Hassanien and A. A. Fahmy, "Epileptic detection based on whale optimization enhanced support vector machine", J*ournal of Information and Optimization Sciences*, Vol. 40, No. 3, pp.699-723, 2019.