EEG Signals Analysis for Epileptic Seizure Detection Using DWT Method with SVM and KNN Classifiers

Asseel Jabbar Almahdi[1*], Atyaf Jarullah Yaseen[2], Ali Fattah Dakhil[3]

[1]Computer Department, Computer science & Mathematics College, University of ThiQar, Thi-Qar, Iraq
[2]Computer Department, Computer science & Mathematics College, University of ThiQar, Thi-Qar, Iraq
[3]Information System Department, University of Sumer, Thi-Qar, Iraq

Abstract

Epilepsy is a critical neurological disorder with critical influences on the way of living of its victims and prominent features such as persistent convulsion periods followed by unconsciousness. Electroencephalogram (EEG) is one of the commonly used devices for seizure recognition and epilepsy detection. Recognition of convulsions using EEG waves takes a relatively long time because it is conducted physically by epileptologists. The EEG signals are analyzed and categorized, after being captured, into two types, which are normal or abnormal (indicating an epileptic seizure). This study relies on EEG signals which are provided by Arrhythmia Database. Thus, this work is a step beyond the traditional database mission of delivering users' inquiries; instead, this work is to extract insight and knowledge of such data. The features are extracted from the signals by applying the Discrete Wavelet transform (DWT) method on the input EEG signals. Two different algorithms Support vector machine (SVM) and k-nearest neighbours (KNN) are applied to the extracted features. After using the above method, two different types of EEG are expected by using classification, either to be normal (refers to the normal activeness of the brain) or abnormal (refers to the non-normal activeness of the brain, which may involve epilepsy). The evaluation is based on three parameters (Precision, Recall, and Accuracy), and also on the implementation time. In this research, two different methods are used, the first is the DWT with SVM, and the second is the DWT with KNN. With regard to the three-parameter values and implementation time, it turned out that the second method was more efficient than the first because of its higher accuracy.

Keywords: EEG Signal, Epilepsy, DWT, SVM, KNN, Medical Signal Processing.

تحليل إشارات EEG للكشف عن نوبة الصرع باستخدام طريقة DWT مع مصنفات SVM وKNN

اصيل جبار المهدي[1]*, أطياف جار الله ياسين[2], علي فتاح دخيل[3]
[1]قسم الحاسبات, كلية علوم الحاسوب والرياضيات, جامعة ذي قار , محافظة ذي قار, العراق
[2]قسم الحاسبات, كلية علوم الحاسوب والرياضيات, جامعة ذي قار , محافظة ذي قار, العراق
[3] قسم نظم المعلومات, جامعة سومر , محافظة ذي قار, العراق

_____
*Email: asseelalmahdi@utq.edu.iq

الخلاصة

الصرع هو اضطراب عصبي خطير له ميزات مثل التشنجات المستمرة، مع تأثير حرج على طريقة عيش ضحايا الصرع ويتبعه بشكل دوري فقدان الوعي. مخطط كهربائية الدماغ (EEG) هو أحد الأجهزة المستخدمة الشائعة الموصى بها للتعرف على النوبات وكشف الصرع. يستغرق التعرف على التشنجات باستخدام موجات EEG بعض الوقت لأنه يتم إجراؤها جسديًا من قبل أطباء الصرع. يتم تحليل إشارات EEG وتصنيفها، بعد التقاطها، إلى نوعين، طبيعي أو غير طبيعي (إشارةً إلى نوبة صرع) .

في هذا البحث، يعتمد الباحث على إشارات EEG التي توفرها مجموعة بيانات MIT BIH. سيتم استخراج الميزات من الإشارات من خلال تطبيق طريقة DWT على إشارات EEG المدخلة، ويتم تطبيق خوارزميتين مختلفتين (SVM & KNN) على الميزات المستخرجة، لتصنيفها إلى فئتين مختلفتين، أي ما إذا كانت إشارة الإدخال تحمل نوبة صرع أم لا. بعد استخدام الطريقة المذكورة أعلاه، يُتوقع نوعان مختلفان من تخطيط كهربائية الدماغ باستخدام التصنيف، إما أن يكون عاديًا (يشير إلى النشاط الطبيعي للدماغ) أو غير طبيعي (يشير إلى أن نشاط الدماغ غير طبيعي، قد يحتوي على الصرع). سيعتمد التقييم على ثلاث معلمات (Precision, Recall, and Accuracy)، وكذلك على وقت التنفيذ. يتم استخدام طريقتين مختلفتين في هذا البحث، الأول هو DWT مع SVM ، والثاني هو DWT مع KNN. فيما يتعلق بقيم المعلمات الثلاث ووقت التنفيذ، اتضح أن الطريقة الثانية كانت أفضل من الأولى بسبب دقة النتائج باستخدام الطريقة الثانية.

Introduction

Based on study of a big database which is tested in this work with over 42GB size available by [1] and MIT BIH Arrhythmia Database [2], many people might be affected by epileptic convulsion, which is a critical neurological brain disorder. With reference to the World Health Organization (WHO), about 50 million people are impacted by epileptic convulsions and therefore it is among the most prevalent non-communicable neurological brain illnesses [3]. Findings from WHO reveal that around 80% of the persons suffering from epileptic convulsions illness are linked to low- and middle-income nations. Advanced chemical reactions in the brain nerve cells as a result of sharp electrical and magnetic fields lead to convulsions. Naturally, a person's brain has equality between the nerve cells that stimulation and those that inhibition. On the contrary, the occurrence of convulsions is due to imbalance between activation and inactivation of brain neurons. The result is unregularly processes during brain cell communication. Medically, the irregular emission of electrical tasks is known as paroxysmal activity, taking place during epileptic convulsions (ictal stage) or in the phase between two epileptic convulsions times (inter-ictal stages) [4]. Generally, two phases of convulsions exist, one is the initial phase and the other is the occurrence phase. The initial phase presents the beginning of convulsions while the occurrence phase represents the exact happening of convulsions. In the medical field, neurophysiologists employ the transcripts of electroencephalogram (EEG) to uncover epileptic convulsions. The first scientist to transcript a person's brain's electric field was Hans Berger, a psychiatrist from Germany, in 1924 [5]. Electrodes are positioned on the human scalp for transcription of neurons passing in the form of messages in a human brain.

The 10–20 system is an internationally acknowledged approach [5] with 21 electrodes, which was followed in the resolution of electrodes on the scalp exterior, as revealed in Figure-1. Nasion is the area on the upper part of the nose situated between with eyes, whereas the area in the mid-back of the head, that is the bony lump situated at the bottom of the skull, is the inion. With reference to these areas, measurement of other boundaries takes place in the axial and longitudinal planes.
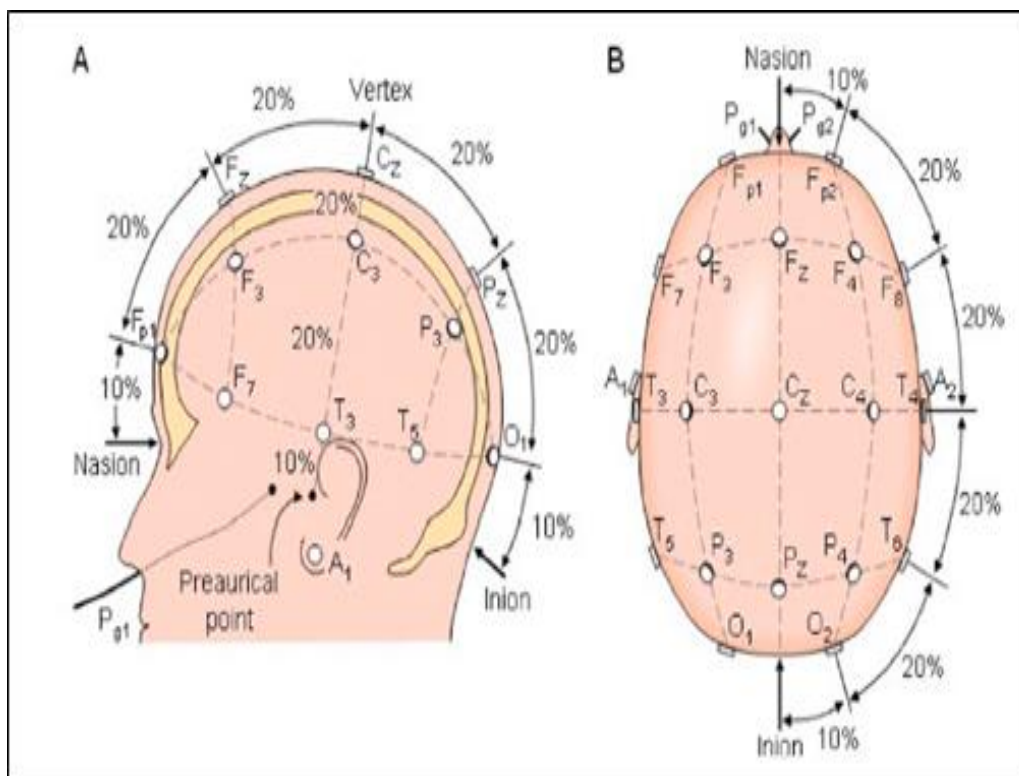
Figure 1- Internationally agreed 10–20 system, earlier presented in 1958.

One of the prominent features characterized by the EEG is the ignition areas (or epochs) connected to epileptic convulsions and, as well as the other signals which are normally grouped as alpha (α), beta (β), delta (δ) and theta (θ). Their characteristics are not as similar as they rely upon a victim under examination. Factors such as magnitude, prevalence, and unpleasant sound vary when an individual is conscious, performing Rapid Eye Movement (REM), sleeping (condition of dreaming with active eye motion) and intensely sleeping [6, 7].

The complexity of EEG waves is a result of intrinsic nonlinear and changeable characteristics (entailing more than two bands of frequency details changing with time) [8]. Hence, conventional detection is tedious and requires long-term monitoring by specialist physicians to decompose EEG signals. With the advance in technology, it is much easier by feeding digital EEG signal data to the computer, that the epileptic seizure is automatically detected by a special software system. The application of computer programming will lower the epilepsy diagnosis period. This will enlarge the outcomes' preciseness and minimize the evaluation period, because automation will confer the ability to nurse several victims within a limited period. With reference to classification algorithms, there is a possibility of an instinctual diagnosis of epileptic convulsions.

In this paper, support vector machine (SVM) and k-nearest neighbors (KNN) classification algorithms proposed a classification algorithm with the feature extraction method, DWT, for the detection of epileptic seizures from signals. The work is performed as follows; First, the EEG signal is taken, noises are reduced and processed by the extraction method (DWT) to extract the features. Second, the extracted features are classified into normal or abnormal.

Literature Survey

The grouping issue deals with the identification of epileptic and non-epileptic EEG waves. It entails the removal, from EEG waves, of the discriminating factors and then executing grouping. Through the succeeding paragraphs, a discussion can be conferred of similar state-of-the-art approaches that employ various factor extraction and grouping approaches for the categorization of epileptic and non-epileptic EEG waves. Jiang *et al*. suggested seizure grouping of EEG waves by employing transductive transfer learning, semi-supervised learning, and TSK fuzzy system [9]. However, the duration of this task is longer for training and testing the dataset. Dash *et al*. proposed an epileptic focus restriction by employing a distinct wavelet modification based on intellectual, intracranial EEG [10]. In that approach, they designed a framework for the usage of DWT and a

support vector machine (SVM) for epileptic focus modification issues based on the EEG. The major drawback was that the final model did not provide good results. Chandan *et al*. performed an analysis of lifting based on DWT and MLPNN for epilepsy seizure detection from EEG [11]. The proposed model classifies the data using a multilayer perceptron neural network, which takes a long time to classify. These drawbacks are overcome by our system which uses a combination of DWT and GLCM for feature extraction and selection. The probabilistic neural network, which has four layers of processes, is implemented for providing better classification.

Authors of another article [12] assessed a set of ensemble techniques, which included bagging technique, boosting technique and random subspace ensembles. Parameters were employed for every classifier. The exactness of this research was low for epileptic capture detection. Another study [13], presented RBFNs as a technique to find out epileptic captures. This approach is called bagging, and it relies on the technique which employs differential evolution (DE). This uses RBFNs algorithm as the base classifier, it has high detection complexity of the technique, higher time complexity, and the need extra work without to enhance the precision of classification.

The Proposed Method
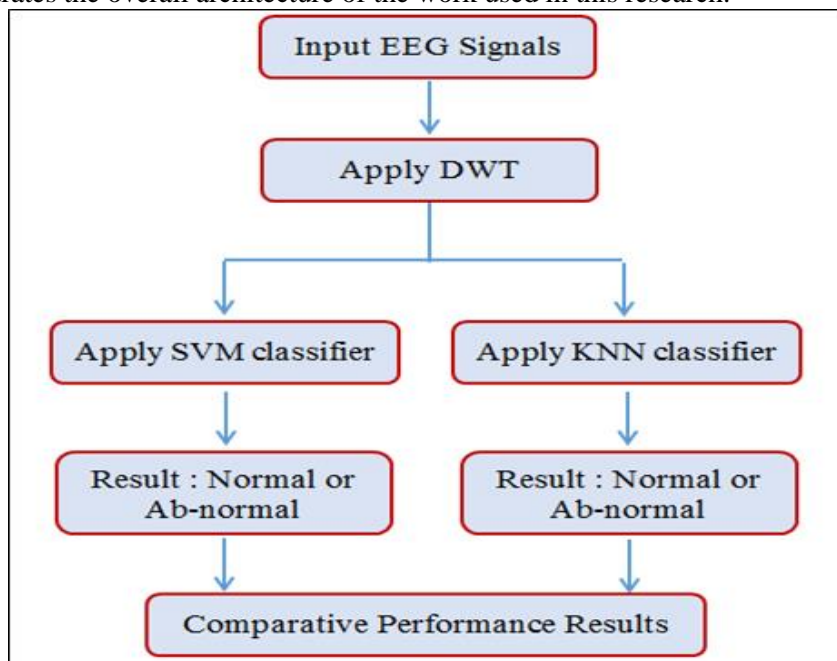Figure-2 illustrates the overall architecture of the work used in this research.

Figure 2- Flow diagram of the proposed method

A.Feature Extraction Using (DWT)
The features removal is one of the most significant processes in braincase diagnosis. Useful features can be obtained depending on the features extraction method. In this work, DWT is engaged to extract the features from EEG [14]. DWT is as considered one of the most important approaches used to extract the features available in electrical signals. DWT is used to analyse EEG signals to a range of different frequencies, as illustrated in Figure-3. Often, the signals of EEG are unstable because they depend a lot on the subject condition. In this research, we will use Daubechies 6 DWT to analyse brain signals for four levels, as shown in Figure-4. At these levels, we will find an adequate amount of information. Here the implementation time will be less while keeping the basic information. When applying DWT to an EEG signal, the noise will be eliminated, and the required features to classify the signals will be obtained. DWT has four filters; each one is obtained on one feature, as follows:
• Low pass filter: used to extract a certain approximation coefficient of the signal.
• Low high pass filter: used for certain detailed coefficient extraction of the received signal.
• High Low pass filter: used to extract a certain vertical coefficient of the signal.
• High pass filter: used to extract a certain diagonal coefficient of the signal.
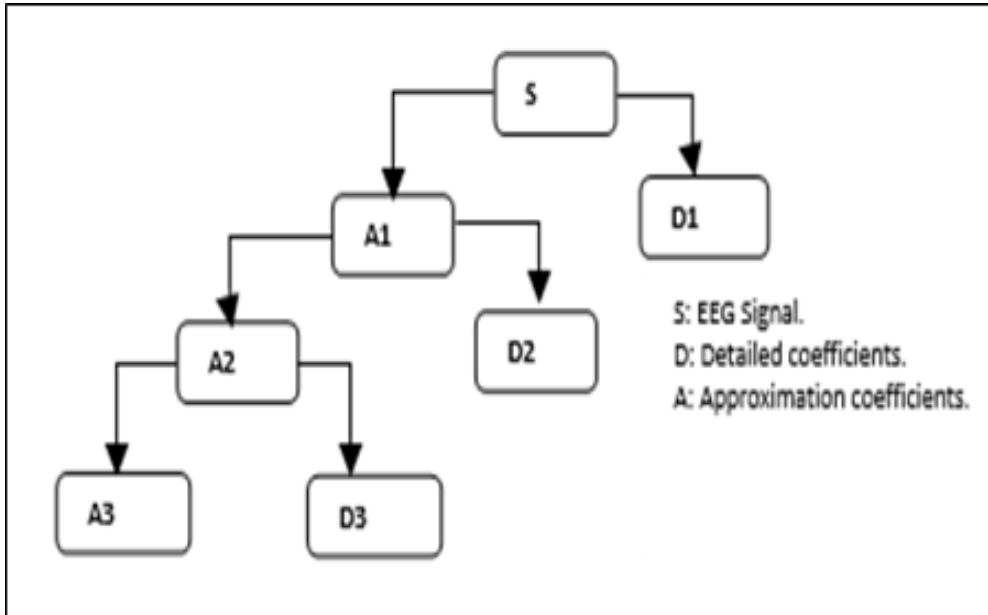
Figure 3- Example of the EEG signals decomposition process by DWT

The expansion of the wavelet in every EEG signal is:

$$f(x) = \sum_{k} c_{j0}(K)\varphi_{j0,k}(x) + \sum_{j=j0}^{\infty} \sum_{k} dj\ (k)\psi_{j,k}(x) \tag{1}$$

where $(x) \in L2\ (R)$, $L2\ (R)$ is relative to the wavelet $\psi(x)$, $\varphi(x)$, and $cj0$ are the estimation coefficients. In the first sum, the estimation coefficient $cj0$ is represented as the result of the interior product process in the original signal (x).

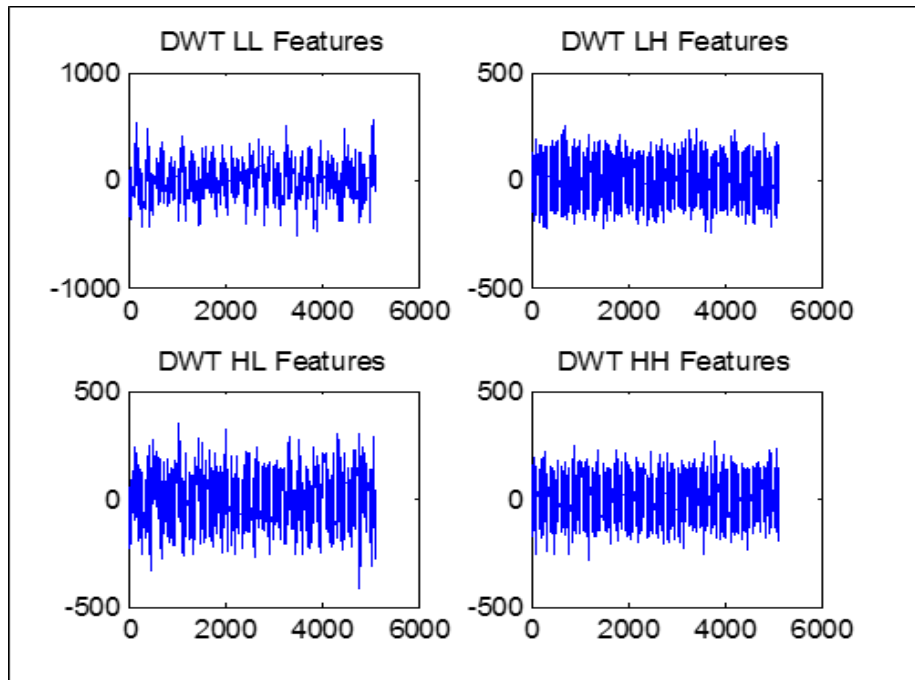It is used to extract features (coefficients). After that, the following functions are applied:

1)   Mean Function:

$$x = \frac{1}{N} \sum_{i=1}^{n} x_i . \tag{2}$$

2)   Standard deviation Function:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \tilde{x})^2 \tag{3}$$

x, k such that $xk > xi$, $\forall i = k, i = 1, ..., n$

Figure 4- The results of DWT filters (LL Filter, LH Filter, HL Filter, HH Filter)

B.**Features Classification:**

After applying the DWT to the EEG brain signals, the characteristics will be ready to apply classification techniques to detect whether the patient is an epileptic or not. The following classification techniques are used to classify EEG signals. Thereafter, cooperation between the results is achieved to find the best one:

1) SVM

To discern the undisclosed validation set of considerations into their proper categories, grouping methods such as SVM are employed based on the training set of some known considerations. The classifier is a mathematical function used by a grouping method to predict the true group of undisclosed considerations. Cortes and Vapnik [15] initiated the SVM classifier in order to diagnose abnormalities of biomedical waves, due to its benefit and robustness to deal with the non-linear and high-dimensional data excellently in the field of biomedical sciences. The major role of SVM is to effectively detect two different sets into their correct groups. SVM is often used as a binary classifier [16]. Its work is to create a surface called hyperplane that separates the elements in two categories, each of which is close to the characteristics of its elements or has the same qualities. SVM depends on the training data to predict the division of input data into two categories. Examples of training set of " M " class are given below.

$T = \{(x_i, y_i), i = 1., M\}$, $x_i \in R_n$, $y_i \in (-1, 1)$, and the test data is calculated by:

$$(x) = \text{sign}\left(\sum_{i=1}^{M} \alpha_i y_i K(x_i, x) + b\right) \qquad (4)$$

2) KNN

The KNN technique is a straightforward statistic-based grouping technique, which is frequently employed in mature grouping algorithms. Its main task is to detect the K samples that are adjacent to the undisclosed sample points and establish the group information of the undisclosed samples from most of the K samples. Its work depends on comparing the features of the input signals with the training signal features, then finding the distance between them using the technique of the measure of similarity (one of the methods used to evaluate the space between cases of data to identify the interchangeable ones). K means a number of interchangeable cases [17]. KNN works to find samples which are similar to the input data to K training samples and closer to their selection during

classification processes. The equation below is used to calculate the distance that lies between two points:

$$\text{Euclidean Dis.} = \sqrt{\sum_{i=1}^{k}(x_{1i} - y_{2i})^2} \qquad (5)$$

where: X1= (x11, x12 …, x1n), Y2= (y21, y22, …, y2n).

3) Ensemble classifiers

Now, after the application of classification algorithms (SVM, KNN) on the input signal features, we can obtain the classification results for the signal to detect the epileptic seizure, which are in two categories, normal or abnormal (Figure-5a, Figure-5ab).
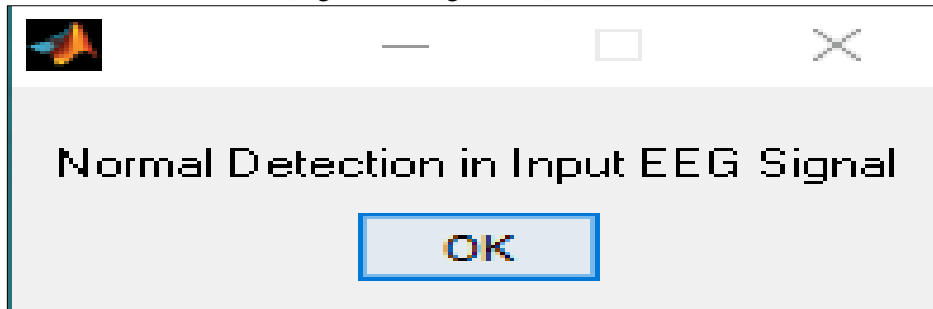
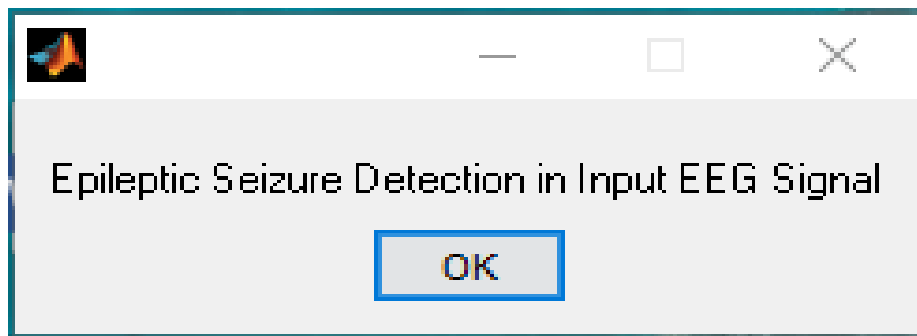

Figure 5- a: normal seizure result

Figure 5- b: epileptic seizure result

Results

In this research, two grouping algorithms (SVM and KNN) are individually applied to compare their efficiency. The performance of the proposed methods was evaluated by three performance metrics (Precision, Recall, Time, and Accuracy) [18]. These metrics are defined as follows:

Precision: means positive prediction value, evaluated by: Precision = tp / (tp+ fp).

Recall (sensitivity): is the true positive rate, which is evaluated by: Recall = tp / (tp + fn).

Accuracy: is the exactness of the identification of representatives, which is evaluated by:

Accuracy = (tp + tn) / (tp + tn+ fp+ fn).

Time: is the time of executing tasks, which is evaluated by Time = end time – start time,

where tp means true positive, fp means false positive, tn means true negative, and fn means false negative. The outcomes of both algorithms are as below:

A. DWT with SVM

After applying the DWT method to the input signal to the system (EEG signal), 4 features were extracted from each signal. Then, the classification algorithm (SVM) was applied to the extracted features so that the epileptic seizure is detected from the inputted EEG. Table-1 shows the results obtained with SVM.

Table 1- Results of the combined application of DWT with SVM

| Parameter	SVM
Precision %	84.6
Recall %	40.9
Time %	0.41
Accuracy %	72.7

B. DWT with KNN

After applying the DWT to the input signal (EEG signals), 4 features were extracted from each signal. Then, the classification algorithm (KNN) was applied to the extracted features to diagnose the epileptic subjection from the inputted EEG. The results of this approach are shown in Table -2.

Table 2- Results of the combined application of DWT with KNN

Parameter	SVM
Precision %	76.9
Recall %	88.8
Time %	0.019
Accuracy %	81.8

Analysis of the results for both classification algorithms is demonstrated in Figure-6, which explains the results for each individual classification algorithm and shows the accuracy of their classification.
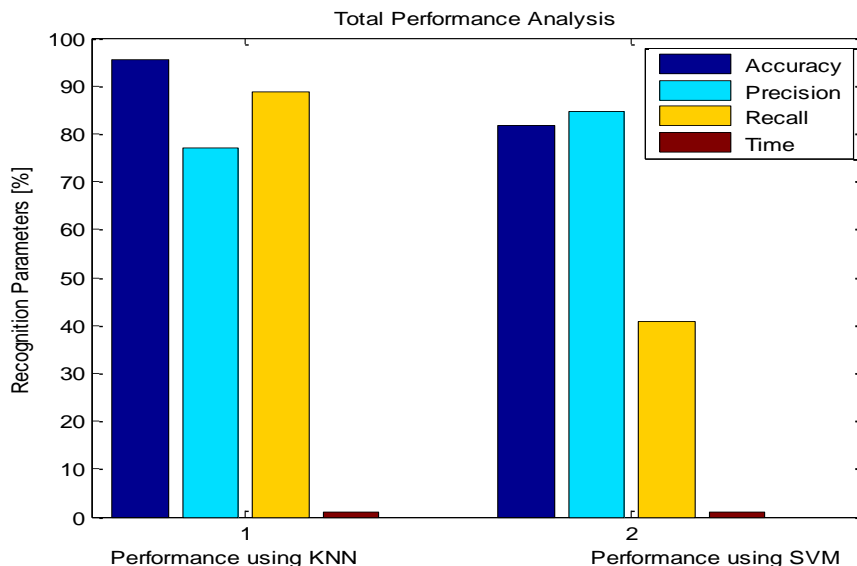
Figure 6- Analysis of results for both Algorithms

Conclusion

In this study, two different classification algorithms (SVM and KNN) were discussed to detect epilepsy, where the DWT method was applied on the input EEG signals to extract the features. Then the above classifiers were individually applied using MATLAB. The KNN algorithm was found to

be more efficient, due to the higher detection accuracy of epilepsy (81.8%). While, detection accuracy with SVM was 72.7%. The increase in accuracy will lead to an increase in classification efficiency.

References

1. https://physionet.org/content/chbmit/1.0.0/ (Accessed 15 October 2019).
2. Moody, G. B. and Mark, R. G. **2001**. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, **20**(3): 45-50.
3. Epilepsy, http://www.who.int/mediacentre/factsheets/fs999/en/. (Accessed 19 October 2016).
4. T.P. Exarchos, A.T. Tzallas, D.I. Fotiadis, S. Konitsiotis, S. Giannopoulos. **2006**. Eeg transient event detection and classification using association rules, *IEEE Trans. Inf. Technol. Biomed.* **10**(3): 451–457.
5. Malmivuo, Plonsey, Jaakko Malmivuo, and Robert Plonsey. **1995**. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.
6. Sazgar, Mona, and Michael G. Young. **2019**. *Normal EEG Awake and Sleep." Absolute Epilepsy and EEG Rotation Review*. Springer, Cham, 2019. 127-139.
7. Proserpio, Paola, Michele Terzaghi, and Lino Nobili. **2019**. "*Local Cortical Activations During REM Sleep and Implications for RBD*." Rapid-Eye-Movement Sleep Behavior Disorder. Springer, Cham, 2019. 389-401.
8. Hussain, Lal, et al. **2018**. "Multiscaled complexity analysis of EEG epileptic seizure using entropy-based techniques." *Archives of Neuroscience* **5.1** (2018).
9. Harshavarthini, S., et al. **2019**. "Automated Epileptic Seizures Detection and Classification." (2019).
10. Dash, Deba Prasad, and Maheshkumar H. Kolekar. **2018**. "A discrete-wavelet-transform-and Hidden-Markov-Model-Based approach for epileptic focus localization." *Biomedical Signal and Image Processing in Patient Care. IGI Global*, 2018. 34-45.
11. Chandani, Manisha, and Arun Kumar. **2018**. "EEG Signal Processing for Epileptic Seizure Prediction by Using MLPNN and SVM Classifiers." *American Journal of Information Science and Technology* **2.2** (2018): 36-41.
12. Chitupe AR, Joshi SA. **2013**. Data Classification Algorithm Using k-Nearest Neighbour Method Applied to ECG Data. *Journal of Computer Engineering*.2013, **14**(4): 13-21.
13. Thomas N, Mathew D. **2016**. "KNN Based ECG Pattern Analysis and Classification", *International Journal of Science, Engineering and Technology Research*. 2016, **5**(5): 1-5.
14. Dehuri, Satchidanada, Alok Kumar Jagadev, and Sung-Bae Cho. **2013**. "Epileptic seizure identification from electroencephalography signal using DE-RBFNs ensemble." *Procedia Computer Science,* **23**(2013): 84-95.
15. Park, Saerom, et al. **2019**. "Security-preserving Support Vector Machine with Fully Homomorphic Encryption." SafeAI@ AAAI.
16. Abualsaud, Khalid, et al. **2015**. "Ensemble classifier for epileptic seizure detection for imperfect EEG data." *The Scientific World Journal* , 2015 (2015).
17. Jaafar, Haryati, Nur Hidayah Ramli, and Aimi Salihah Abdul Nasir. **2018**. "An Improvement To The k-Nearest Neighbor Classifier For ECG Database." IOP Conference Series: Materials Science and Engineering. Vol. 318. No. 1. IOP Publishing, 2018.
18. Thomas, Namita, and Deepthy Mathew. **2016**. "KNN based ECG pattern analysis and classification." *Int. J. Sci. Eng. Technol5*. **5**(2016): 1630-1634.