A Novel Method for Automated Diagnosis of Epilepsy using Complex-Valued Classifiers

Musa Peker, Baha Sen*, and Dursun Delen

Abstract—The study reported herein proposes a new method for the diagnosis of epilepsy from electroencephalography (EEG) signals based on complex classifiers. To carry out the study, first the features of EEG data are extracted using a dual-tree complex wavelet transformation (DTCWT) at different levels of granularity to obtain size reduction. In subsequent phases, five features (based on statistical measurements—maximum value, minimum value, arithmetic mean, standard deviation, median value) are obtained by using the feature vectors, and are presented as the input dimension to the complex-valued neural networks (CVANN). The evaluation of the proposed method is conducted using the k-fold cross-validation methodology, reporting on classification accuracy, sensitivity and specificity. The proposed method is tested using a benchmark EEG dataset and high accuracy rates were obtained. The stated results show that the proposed method can be used to design an accurate classification system for epilepsy diagnosis.

Index Terms—EEG signals, Complex-valued neural networks, Dual-tree complex wavelet transform, Epilepsy.

I. INTRODUCTION

Epilepsy is a neurological disorder that causes uncontrolled and involuntary movements called seizures due to temporary electrical discharges that occur in the brain. Epilepsy can be observed in all regions of the world, in males and females, in all races and in one in every 100 individuals.

EEG is an important diagnostic method used to examine epileptic patients and patients with suspicious attack problems. Brief information about history of the EEG, EEG measurement and used areas is presented below.

The presence of electrical activity in the brain was detected for the first time in 1875 by Caton [1]. In 1929, Hans Berger [2] did a study to prove this claim. Berger, revealed the presence of electrical activity in the human brain. He realized that with the help of electrodes placed on the head and a galvanometer connected to these electrodes. Then, Berger observed that these signals called EEG are changed with the opening and closing of the eyes. With these developments, the presence of EEG signals was scientifically proved. In 1934, Adrian and Matthews [3] ensured of strengthening and recording of EEG signals obtained with the electrodes. The number of research dealing with this issue has increased day by day with the impact of the big developments in the field of electronics and computer. Today, EEG signals are used in many areas such as diagnosis of epilepsy, controlling of anesthesia stage in surgical operations and determining the depth of anesthesia, sleep disorders, investigation of sleep psychology and diagnosis of migraine.

Brain-computer interface is used to measure the EEG signals. The first part of this interface is metal electrodes which are used for measuring the electrical activity of the brain from the head surface. Generally, measurements are made according to an electrode placement scheme called the international 10-20 system. EEG signals have very low amplitude (in μV levels) values. To be able to interpret these signs, they should be amplified by EEG device. This amplified EEG signal is drawn on a paper in continuous form. The recorded EEG signals are similar to the multi-channel seismograph record. Evaluation of EEG recordings is performed by neurologists who specialized in this subject.

The five main wave types in EEG signals are: alpha (8–13 Hz); beta (13–30 Hz); delta (0–4 Hz), gamma (30~100+ Hz) and; theta (4–7 Hz) [4], [5]. EEG signals are not periodic and their amplitude, phase and frequencies change continuously. Therefore, these signals are very difficult to interpret. EEG data are visually analysed by neurologists [5]. This process is rather stressful and tiring. Also, EEG signals analysed by neurologists who have received different training may generate inconsistent information. Therefore, it is imperative to analyse EEG signals with a consistent and suitable method in order to ensure correct epilepsy diagnosis and treatment.

The purpose of this study is to identify an algorithm with better classification accuracy than existing algorithms, and to ensure the provision of gradable information through the use of EEGs.

The rest of the paper is organized in the following manner: Section II presents a brief literature review about the diagnosis of epilepsy including information regarding the method suggested in the current study. Section III briefly describes the data set of the EEG signals employed in our research. Section IV details the methods utilized in the present study. Section V provides the experiments undertaken in the framework of the study, assessment procedures that are used and experiential results that are obtained. Finally, Section VI summarizes the
conclusions derived from the study and the future work.

II. RELATED WORK

An investigation of recent studies shows that various different models have been suggested to assist neurologists in identifying epileptic activities. Some of the studies with high classification accuracies are provided below. A direct comparison of our study to those of others reported in the recent literature in terms of classification accuracy is also given in the form of a table in the result section of this paper.

Pradhan et al. [6] used two layered learning vector quantization networks in order to analyze EEG signals. Petrosian et al. [7] applied EEG signals to recurrent neural networks instead of applying them to statistical properties in the identification of epileptic seizures. In extracting features from EEG signals, Srinivasan et al. [8] used five features based on time and frequency. High accuracy rates were obtained in the system where Elman recurrent neural network was used as a classifier. Nigam and Graupe [9] used two feature values expressed as relative spike amplitude and spike occurrence frequency. A multistage non-linear preprocessing filter was used to extract these features. Later, these features were used as input data, especially for artificial neural networks (ANNs). Subasi and Ercelebi [10] used both MLPNN (multilayer perceptron neural network) and logistic regression models in the identification of epileptic seizures in the analysis of EEG signals. In another study, Subasi [11] compared the two outputs (normal and epileptic) MLPNN and used a mixture of expert model in the classification of EEG data. Kannathal et al. [12] arrived at 92% classification accuracy by using different entropy values and an adaptive neuro-fuzzy inference system (ANFIS) algorithm. In another study, Kannathal et al. [13] used chaotic measures in the identification of features from EEG signals. These measures included correlation dimension, the largest Lyapunov exponent, Hurst exponent and entropy values. The authors arrived at a classification accuracy of higher than 90%. Tzallas [14] employed Wigner-Ville coefficients that have time-frequency conversion characteristics along with ANNs for epilepsy diagnosis. Polat et al. [15] proposed a hybrid model containing decision tree (DT) algorithms and a fast Fourier transformation base. Chua et al. [16] used a variety of higher order spectral (HOS) attributes to distinguish normal, ictal and preictal EEG signals. And they also indicated unique ranges for these features for various classes with high confidence level (p-value of less than 0.05). Faust et al. [17] used Burg autoregressive coefficients in modeling of epileptic seizures. In the study in which support vector machines was preferred as classification algorithm, 93.33% classification accuracy was obtained. Acharya et al. [18] used chaotic features such as Hurst exponent, largest Lyapunov exponent, fractal dimension and approximate entropy (ApEn) to classify normal, ictal and preictal states using EEG signals. In the study in which SVM and GMM algorithms were used as classification algorithm, better results were obtained with the GMM. The average 95% classification accuracy was obtained with the GMM algorithm. Guo et al. [19] obtained features from EEG signals by relative wavelet energy (RWE) algorithm in different frequency bands. A classification accuracy of 95.2% was obtained in the system where ANN was used as a classifier. Acharya et al. [20], used HOS based attributes (specifically cumulants) for automatic detection of normal, ictal and preictal states using EEG signals. And as a result, they achieved 95.5% classification accuracy. Acharya et al. [21] used ten recurrence quantification analysis (RQA) attribute parameters for the classification of the same problem. These attributes were classified with different classification algorithms. SVM algorithm gave the best performance among these algorithms. As a result, 95.6% classification accuracy was obtained. Acharya et al. [22] employed wavelet coefficients and eigenvalues to extract features from EEG signals. Wavelet packet decomposition (WPD) was utilised to obtain wavelet coefficients and eigenvalues were determined with the help of principal component analysis (PCA) algorithm and, later, an analysis of variance (ANOVA) test was implemented to select the significant eigenvalues. The experiments were undertaken with the help of 10-fold cross validation. 99% classification accuracy was obtained in the system where the Gaussian mixture model (GMM) algorithm was used as a classifier. Acharya et al. [23] proposed a method for automatic detection of normal, ictal and pre-ictal states using EEG signals. They used four algorithms (ApEn, sample entropy and two HOS based entropy) in feature selection stage. Obtained attributes were classified with 7 different classification algorithms. These algorithms are as follows: Fuzzy classifier (FC), SVM, k-nearest neighbor (KNN), probabilistic neural network (PNN), DT, GMM and naive Bayes classifier (NBC). FC gave the best result in experiments performed. As a result, three-class problem have been distinguished with 98.1% accuracy rate. Niknazar et al. [24] proposed a RQA based method for automatic detection of normal, ictal and preictal states using EEG signals. The proposed method is based on combination of RQA-based measures of the original signal and its subbands. As a result, 98.67 % accuracy rate was obtained. The use of complex-valued methods for the classification of EEG data was found for the first time in Peker and Sen [25]. In their studies, they obtained 8 real-valued attribute values from EEG data. They turned these attribute values into 4 attribute values in the form of complex numbers. The resulting 4 complex-valued attribute values are presented as an introduction to a complex-valued neural network. The parameter values of the complex-valued neural network were obtained by trial-and-error. The results indicate that high accuracy values were obtained.

This study proposes a new method in the diagnosis of epilepsy that is different from the approaches presented in the previous studies. In the current study, complex-valued classifiers are used to diagnose epilepsy. The proposed hybrid method produced accuracy results that are better than any reported in literature (a comparative table is provided in the results section of the paper). The following is a short summary of our proposed method:

The study investigates the effects of complex-valued classifiers on the diagnosis of epilepsy. Complex numbers are
known to be important in the basis of signal processing. It is possible to hold this view through the Fourier transform method which is one of the most basic and important topics of signal processing. The Fourier transform method works with complex-valued numbers and generates complex values as an output. If we want to classify the outputs of this transform with a real-valued classifier, we need to present the real and virtual values of the complex number to the input separately. This will cause an extra load and a decrease in capacity. When complex numbers are preferred, two-dimensional information will be presented to the network as one dimensional information because complex numbers have both real and imaginary parts. This is presented in more detail in Chapter IV.B.1. Sections IV.A and IV.B.1 describe the advantages of these algorithms while presenting complex-valued algorithms. In line with this information, a model that works with complex numbers was targeted in order to classify the EEG values in high accuracy ratios. DTCWT has been applied in various levels for feature extraction and statistical features are obtained from the obtained complex-valued feature vector. The same number of features for each set of EEG data is presented as entry to the CVANN. The reason for using the DTCWT-CVANN hybrid model is related to solving the problem with an algorithm that works completely in the complex domain. This way, it is possible to observe the effects of the complex-valued methods and compare them with the real-valued methods.

III. DATA

The EEG signals used as the data set in the study were obtained from the database of the Epileptology Department of the University of Bonn [26]. This section provides some information regarding the data. More detailed information can be obtained from the sources provided [26]. All data are composed of five clusters (A-E) and each cluster consists of 23.6 second single channel 100 EEG segments. Clusters A and B consist of surface EEG recordings received from five healthy volunteers whilst eyes were open and closed respectively. Cluster C consists of EEG recordings obtained from the hippocampal opposing hemisphere of sick patients prior to seizures. Cluster D consists of EEG recordings obtained from the epileptogenic region in the sick patients prior to seizures. Cluster E consists of the recordings of seizures from sick volunteers. The signals were transferred to a digital environment with a 173.61 Hz sampling frequency (a total of 4097 samples) after being converted with a 12-bit analog digital converter. A 0.53-40 Hz band pass filter was applied to the EEG signals obtained from the participants.

IV. METHOD

A. The Dual-Tree Complex Wavelet Transformation

Discrete wavelet transformation (DWT) has a wide range of applications and is frequently used in many fields such as signal and image compression, feature extraction, noise reduction, channel coding, image processing and the numerical solution of partial differential equations [27], [28]. Although it offers a proficient computational algorithm and limited representation, the DWT displays four fundamental disadvantages compared to DTCWT: (a) Shift sensitivity: If there is an unpredictable change in transformation coefficients when there is a shift (change) in the time for the input signal, this transformation is defined based on shift sensitivity. Shift sensitivity is an unwanted characteristic. DWT coefficients are unsuccessful in differentiating the input-signal changes [29]. DTCWT is insensitive to change. When compared with standard DWT, it has a developed sensitivity towards shifts in time. (b) Poor directionality: Standard DWT is unsuccessful in selecting diagonal characteristics. DTCWT has 12 wavelets (six for the real tree and six for the imaginary tree) in two dimensions oriented in ±15, ±45, ±75 degrees. (c) Absence of phase information: One of the important shortcomings of the DWT algorithm is the absence of phase information. All natural signals are mainly real-valued and therefore complex-valued filtering is necessary in order to benefit from local phase information [30]. As opposed to normal wavelet transform, the DT-CWT algorithm uses complex functions instead of real-valued main wavelet functions. This way, amplitude and phase information can be examined separately. Due to the existence of all these problems, the use of a DTCWT was proposed by Kingsbury [31].

The DTCWT uses analytic filters to perform the wavelet analysis. DTCWT has a more complex structure compared to standard DWT and is composed of two DWTs that work parallel to each other, as shown in Fig. 1. One of these trees is called the real tree whereas the other is called the imaginary tree. DTCWT uses a pair of filters for mother wavelet function $\psi(t)$ and scaling function $\phi(t)$ and these filters are $h_{0}(n)$, $h_{1}(n)$, which is the low-pass/high-pass filter pair and $(g_{0}(n)$, $g_{1}(n)$, which is the low-pass/high-pass filter pair [32]. Wavelet function and scaling functions are calculated by using Equations (1) and (2).
\[ \psi_h(t) = \sqrt{2} \sum_{n} h_1(n) \phi_h(2t - n) \]  
\[ \psi_g(t) = \sqrt{2} \sum_{n} h_0(n) \phi_h(2t - n) \]

Together with the real and imaginary components, the complex wavelet transformation is expressed as \( \psi(t) = \psi_h(t) + j\psi_g(t) \). Here, \( \psi_h(t) \) and \( \psi_g(t) \) are real-valued. \( \psi_h(t) \) is calculated, as seen in Equation (1).

If \( \psi(t) \) is almost analytic (if supported only one side of the frequency axis), obtained conversion may have features such as lack of aliasing, directionality and shift invariance as in Fourier transform in which complex base functions are analytic [31]. For \( \psi(t) \) to be approximately analytic, it is required that one wavelet basis must be approximately equal to Hilbert transform of the other wavelet basis [32]. As seen in Equation (3), \( \psi_g(t) \) is designed to provide the Hilbert transformation of \( \psi_h(t) \) [32]:

\[ \psi_g(t) = H[\psi_h(t)] \]

To ensure the condition in Equation (3), there should be approximately half-sample distance between the low-pass analysis filters in real and virtual trees [33]:

\[ g_0[n] = h_0[n - 0.5] \]

To perform these operations, there are various filter design methods in literature such as q-shift solution and common factor solution. In this study, q-shift filter which is proposed by Kingsbury [31], was used in filter design. Detailed information about q-shift filtering can be obtained from the reference #31.

B. Complex-valued neural networks (CVANN)

First studies in the literature that present and apply CVANN algorithm were examined and presented in this section. The CVANN algorithm was first introduced with the announcement of the complex least mean squares (LMS) algorithm by Widrow, McCool and Ball [34]. A complex-valued learning algorithm for signal processing application was published by Kim and Guest [35]. Another version of CVANN, incorporating a different activation function, was presented by Georgiou and Koutsougeras [36]. Nitta [37] reports an extensive study of CVANN. The results obtained from this study are as follows: a) Two hyper-surfaces which intersect orthogonally, create the decision boundary of a single complex-valued neuron, and four equal sections are formed in the decision region. When both the absolute values of real and imaginary parts of the net inputs to all hidden neurons are sufficiently large, an orthogonal intersection of three layered complex-valued neural network will be realized by the decision boundaries for real and imaginary parts of an output neuron [38]. b) Average learning speeds of complex BP algorithm and real BP algorithm are different, with the former being faster. The standard deviation of the learning speed of the complex BP and the real BP is different, with the latter being larger. Therefore, the complex-valued neural network and the related algorithm are natural for learning of complex-valued patterns [38].

CVANN algorithm can be used with multilayered neural networks where weights, threshold values, inputs and outputs all are complex numbers (see Fig. 3). As can be seen from Fig. 3, the CVANN structure is actually a model of ANN that processes complex numbers. In terms of this model, complex-valued vectors are provided as inputs for CVANN. Complex-valued output vectors are obtained by having complex values in the weight values situated between the input layer and the hidden layer and between the hidden layer and the output layer, and by using the activation function in line with the equation provided in (5). CVANN is initiated with the help of the transmission of complex input signals or data through the connection, where each connection has an associated weight that enhances the signal that is transmitted; where the received signals is transformed by each neuron (the sum of the input multiplied by the connection weight) through an activation function and, in turn, the output signal is determined. Later, error values are minimized by updating the weight and threshold values of the network through a complex back-propagation (BP) algorithm. This is presented with the help of formulas in mathematical modelling section.

1) The advantages of CVANN algorithm: When real-valued ANNs are used in solving problems with complex-valued numbers, ANNs should be applied separately for real and imaginary parts. However, when CVANNs are applied to the same problem, the operation time decreases due to the ability to operate directly. In addition, the ratio of accuracy increases [37]. The use of complex-valued input/output, weights and activation functions enables an increase in the functionality and performance of single neurons and the neural network, as well as a decrease in training time [39]. Fig. 2 can be examined as a simple example to observe the advantages of a complex-valued neural network. The example shows that a two-input network is reduced to one input in the case of complex-valued neural network use. This provides the opportunity to substantially reduce the degree of complexity in large networks and allows speedy training. Here; \( z = a + ib \) and \( w = w_1 + iw_2 \). What provides the advantage is the fact that complex numbers can carry two-dimensional information in a single dimension due to having both real and imaginary parts. As mentioned above, such a situation will result in the decrease of the network size and in speedy training.

In addition to the advantages that are mentioned above, there are other advantages of complex-valued neural networks compared to real-valued neural networks. These advantages are high level functionality, better plasticity and greater flexibility. They learn faster and retain better generalisations [40]. The skill of a single neuron in complex-valued neural networks to learn input/output mappings that can be differentiated as nonlinear in the real domain shows greater flexibility. These neurons have the skill to learn without generating higher degree inputs and progressing to a higher dimensional space.
Fig. 2. The representation of neural network with one input and one weight value in the complex plane which was normally realized in the real plane with 2 inputs and 2 weight values (RVN: Real-valued neuron, CVD: Complex-valued neuron). Here; $z = a + ib$ and $w = w1 + iw2$.

Nitta et al.’s study [41] can be examined to see the advantages of CVANN more clearly. This study shows that the XOR problem which cannot be solved using two-layered real-valued neural networks can easily be solved by using two-layered CVANN.

2) The mathematical model of CVANN algorithm: Brief details about the mathematical model of CVANNs are provided below [37], [42].

Let’s assume sigmoid function is selected as the layer neuron and the XOR problem which cannot be solved using two-layered network. Let’s assume sigmoid function is selected as the layer neuron.

In Equation (5), $W_{nm}$ is the complex-valued connection weight between $n$ neuron and $m$ neuron. $l_m$, is the complex-valued input signal of the $m$ neuron, and $\theta_n$ is the complex-valued threshold of $n$ neuron. In order to obtain the complex-valued output signal, $Y_n$ active value is transformed into two components as real and imaginary.

$$Y_n = x + iy = z$$

Here, $l$ stands for the value of $\sqrt{-1}$. When various output functions of each neuron are considered, the output function ($f_c$) can be defined using the following equation:

$$f_c(z) = l_0(x) + i.l_0(y)$$

Where $f_0(u)$ is called the activation function of neural network. Let’s assume sigmoid function is selected as the activation function. In this case, $f_0(u) = 1/(1 + exp(-u))$, $u \in R$ (R denotes the set of real numbers), the real and imaginary parts of an output of a neuron mean the sigmoid functions of the real part $x$ and imaginary part $y$ of the net input $z$ to the neuron, respectively.

The CVANN used in this study is composed of three layers (input, hidden and output layers). Fig. 3 presents the three layered CVANN structure used in the study. Here, $W_{ml}$, is the weight between the input layer neuron $l$ and the hidden layer neuron $m$, $V_{nm}$ is the weight between the hidden layer neuron $m$ and the output layer neuron $n$, $\theta_m$ is the threshold value for the hidden layer neuron $m$ and $\lambda_n$ is the threshold value for the output layer neuron $n$. $l_1, H_m, O_n$ are the output values for the input layer neuron $l$, the hidden layer neuron $m$ and the output layer neuron $n$, respectively. Similarly, $U_m$ and $S_n$ are the active values of the hidden layer neuron $m$ and the output layer neuron $n$, respectively.

$$U_m = \sum_l W_{ml}l + \theta_m$$

$$S_n = \sum_m V_{nm}H_m + \lambda_n$$

The study prefers a square error function. The square error for p pattern can be expressed as:

$$E_p = (1/2) \sum_n |I_n - O_n|^2 = (1/2) \sum_n |\delta_n|^2$$

Here $N$ is the number of neurons in the output layer, $(\delta_n^2 = T_n - O_n)$, is the error between $O_n$, obtained by $n$ output layer neuron and $T_n$, the target output. The square error also can be rewritten as:

$$E_p = (1/2) \sum_n [|Re(T_n) - Re(O_n)|^2 + |Im(T_n) - Im(O_n)|^2]$$

The learning rule for the complex-valued back propagation model is defined below to minimize the square error $E_p$ [43]. The arrangement of weights and threshold values are designed according to the equations provided below (where $\eta > 0$, $\eta$ is a small learning constant):

$$\Delta V_{nm} = -\eta \frac{\partial E_p}{\partial |V_{nm}|} = i.\eta \frac{\partial E_p}{\partial |Im(V_{nm})|}$$

$$\Delta \lambda_n = -\eta \frac{\partial E_p}{\partial |\lambda_n|}$$

$$\Delta W_{ml} = -\eta \frac{\partial E_p}{\partial |W_{ml}|} = i.\eta \frac{\partial E_p}{\partial |Im(W_{ml})|}$$

$$\Delta \theta_m = -\eta \frac{\partial E_p}{\partial |\theta_m|}$$

Expressions from Equation (14) to Equation (17) can be rewritten as provided below:

$$\Delta V_{nm} = H_m \Delta \lambda_n$$

$$\Delta \lambda_n = \eta(Re[\delta_n^2](1 - Re(O_n))Re(O_n) + i.1.\eta|m\delta_n|^2)(1 - Im(O_n))Im(O_n))$$

$$\Delta W_{ml} = I_l \Delta \theta_m$$

$$\Delta \theta_m = \eta \sum_n \left[ Re[\delta_n^2](1 - Re(O_n))Re(O_n) + i.\eta|m\delta_n|^2(1 - Im(O_n))Im(O_n)) \right]$$

The study prefers a square error function. The square error for p pattern can be expressed as: 

$$2168-2194 \ (c) \ 2013 \ IEEE. \ Personal \ use \ is \ permitted, \ but \ republication/reshublication \ requires \ IEEE \ permission. \ See \ http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
3) Activation functions: One of the difficulties experienced in using back propagation algorithm in the complex plane is the selection of the appropriate activation function [44]. There is no consensus in literature regarding which activation function should be selected. Activation function selection depends on input training data type and the case. The current study investigates the most efficient activation function for the problem. Literature proposes different complex activation functions. In this study, four different activation functions were used in order to identify the activation functions having positive impact on the results. Effective ones of these activation functions were determined by trial and error method. Brief information about these activation functions are presented below.

Sigmoid function has been adapted to CVANNs by Birx & Pipenberg [45] and Benvenuto & Piazza [46] as seen in the equation:

\[
 f(z) = \frac{1}{1 + e^{-(Rez^2)}} + i\frac{1}{1 + e^{-(Imz^2)}}
\]

The study employs the expression in Equation (22) for the complex logsig activation function. In addition, tanh function with singular points in each \( z = (n + 1/2)i\pi, n \in \mathbb{Z} \) has also been adapted to CVNNs as a real-imaginary type activation function.

\[
 f(z) = \text{tanh}(Rez) + i\text{tanh}(Imz)
\]

The current study also investigates the effects of Mexican hat and Haar wavelet activation functions used in complex-valued wavelet neural networks on the results. The study has employed these functions in the complex plane as explained below:

\[
 f(z) = \psi_{\text{Mexhat}} = (1 - aRez^2)e^{-(Rez^2)/2} + i(1 - Imz^2)e^{-(Imz^2)/2}
\]

\[
 f(z) = \psi_{\text{Haar}} = (1 - Rez^2)e^{-(Rez^2)/2} + i(1 - Imz^2)e^{-(Imz^2)/2}
\]

4) Summary of CVANN algorithm:

Initialisation: Assign all weight and threshold values as small complex-valued numbers bigger than zero.

Presentation of input and output (target): Present the complex-valued input vectors \( I_1, I_2, I_3, ..., I_N \) and the corresponding complex-valued output (target) vectors \( T_1, T_2, T_3, ..., T_N \) to the network (\( N \) is the number of patterns that will be used in the training).

Calculation of the actual output: Use Equation (11) to calculate actual output (\( O_n \)).

Calculation of error value: Calculate error value using Equation (12) based on the obtained output and the target output values.

Changing weights and thresholds: Update weights and thresholds by using the formulas in Equations (18)-(21). Continue until the error is minimised.

V. APPLICATION AND THE EXPERIMENTAL RESULTS

The study aims to increase the productivity of the CVANNs by using DTCWT in the classification of EEG data and statistical feature extraction. The block diagram of the proposed structure is given in Fig. 4. In order to extract the features of data, 2 separate vector matrices are created with the size of (100 segments×4096 samples) for both data set A (healthy) and data set E (epileptic activity condition). Sample numbers of each column of A and E data sets with the size of 100×4096 are decreased by using 3 different levels of DTCWT. 5 statistical features are extracted from new data set with reduced size and are classified with the help of CVANN. These statistical features are presented in Table I. In Equation (26-31), \( x_n = 1, 2, 3, ..., n \) is a time series, \( N \) is the number of data points, \( AM \) is the mean of the sample. If the number of values is odd then Equation (30) is used for Median. If number of values is even, then Equation (31) is used for Median (where \( N = \) number of items). A sample of a feature table is presented in Table II.

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum value</td>
<td>( \text{min}_n { x_n } ) (26)</td>
</tr>
<tr>
<td>Maximum value</td>
<td>( \text{max}_n { x_n } ) (27)</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>( \sqrt{\frac{\sum_{n=1}^{N}(x_n - AM)^2}{N-1}} ) (28)</td>
</tr>
<tr>
<td>Arithmetic mean</td>
<td>( \frac{1}{N} \sum_{n=1}^{N} x_n ) (29)</td>
</tr>
<tr>
<td>Median</td>
<td>( \left( \frac{N+1}{2} \right)^{th} ) (30)</td>
</tr>
<tr>
<td>Median</td>
<td>( \left( \frac{N}{2} \right)^{th} ) \text{value} + ( \frac{N}{2} ) \text{th value} ) / 2 (31)</td>
</tr>
</tbody>
</table>

In the sample given, third level DTCWT was applied to data of normal subjects and epilepsy patients. In this way, data set consisting of 4096 samples in each segment decreased to 2048 samples. In final situation, 2048 complex valued samples were obtained. Each sample consists of real and complex coefficients. For these coefficients, arithmetic mean, median, standard deviation, minimum and maximum values were calculated individually. For example, suppose that arithmetic mean of the real values of 2048 samples in a segment is \( x \) and arithmetic mean of the imaginary values is \( y \). In this case, the value of the arithmetic mean feature will be \( x + yi \) for this sample. Other feature values have also been calculated in this way. In literature, statistical features were used in real valued wavelet transforms for dimension reduction. However, such an application has not been observed in DTCWT applications. In addition, last column of Table II shows the output values. The output value for epilepsy patients is (0+i), the output value for healthy people is (1+i).
Since the study employs DTCWT on the data in three different levels (first, second and third levels), there are three problems with different input data. For ease, these three problems are called DTCWT-CVANN-1, DTCWT-CVANN-2 and DTCWT-CVANN-3 respectively in terms of level values. Operations undertaken during the implementation phase are listed below:

A. Experimental Results

A computer with Intel(R) Core™ i7-2670QM 2.20 GHz microprocessor and 8 GB RAM is used for the solution of problems. Parameter set displayed in Table III is selected for parameter values for CVANN. The parameter values of the complex-valued neural network were obtained by trial-and-error.

The first step is the calculation of root mean square error (RMSE) values obtained through the best parameter results. RMSE is used to determine the error ratio between assessment values and model estimates and the RMSE values close to zero shows increase in the ability of the model for estimation. RMSE values are presented in Table IV. It is seen that the best results are obtained by DTCWT-CVANN-1 method. The other methods are also found to be effective.

The second step of the study consists of the calculation of statistical analysis values. Table IV displays the obtained results according to performance evaluation criteria. Best results for average classification accuracy are obtained using DTCWT+CVANN-1 and DTCWT+CVANN-3 algorithm for epileptic seizure detection.

Although DTCWT+CVANN-1 and DTCWT+CVANN-3 provide the same accuracy rates, its RMSE values are different. RMSE value obtained by DTCWT+CVANN-1 method is lower. This results shows that obtained output values and target output values are closer and it points to the fact that the most effective method is DTCWT+CVANN-1.

The classification accuracy rates obtained in this study and in previous studies on the same data set are compared and the analysis results are presented in Table V. Only the studies that used the same data set were utilized in order to achieve a reliable and fair comparison. In this sense, only the database used in Subasi and Ercelebi’s study [10] differs. The reason for including this study in the comparison table is related to the fact that the authors have used the real-valued version of the proposed method in their studies. It is possible this way to compare the success rates of real and complex-valued classifiers.

Generally, hold-out and k-fold cross validation methods are used in the selection of training and test data. As shown in comparison tables (Table V and Table VI), the authors have performed the analysis by one of these two methods. For a fair comparison, data selection in this study was performed using two methods. As shown in Table V, in the classification of A-E problem, 100% accuracy rate was obtained with the proposed method. The results show that the proposed method produced better results compared to those of the other studies. With regard to important matter such as medical diagnoses and diagnostic systems, even a 0.1% increase in accuracy rates is very important. In addition, the proposed method seems to give good results when the sensitivity and specificity values are analyzed. It is observed that Tzallas et al. have also achieved 100% accuracy rate by using time–frequency analysis and artificial neural network methods on the same dataset.

Table V shows that Subasi and Ercelebi’s [10] studies are impressive. It is seen that the accuracy rates obtained in the current study is higher than those of studies including real-valued wavelet and neural networks which is an indicator that
the effect of complex-valued classifiers is higher. Subasi and Ercelebi’s study [10] does not provide information regarding the computation time. We have mentioned that complex classifiers have important effects on the time parameter. The same experiment was repeated to compare the computation times of the real-valued classifiers and the proposed complex-valued classifiers.

Real-valued wavelet transform (WT) and neural networks (ANN) were used on the same data to compare computation times. The back propagation feed-forward neural network was preferred for neural networks, and various models were tried for wavelet transform. The best one was selected to be the wavelet transform that used 4th degree Daubechies wavelet. 92.5% classification accuracy was obtained while using the real-valued WT-ANN structure. The result obtained is similar to the result in study # [10] and the computation time was found to be 126.4 seconds. For the same problem, computation times were found to be 45.6 seconds, 40.2 seconds and 30.9 seconds for the DTCWT+CVANN-1 model, the DTCWT+CVANN-2 model and the DTCWT+CVANN-3 model respectively. For a fair comparison, all the experiments were carried out on the same computer using the same programming language.

As can be seen, the use of DTCWT+CVANN models significantly reduces computation time compared to real-valued classifiers. The difference in the computation times of DTCWT+CVANN-1, DTCWT+CVANN-2 and DTCWT+CVANN-3 models is related to the change in the numbers of input samples provided to the CVANN input. The DTCWT+CVANN-3 model with less input has a lower computation time.

Peker and Sen [25] used complex-valued methods for classifying EEG data. There are significant differences between the proposed method used in this study and the proposed methods in study #25. In study #25, real-valued attributes were obtained to represent EEG data, and these attributes were converted to a complex number format. In this study, complex-valued attributes are obtained directly from EEG signals thanks to complex wavelet transform.

It can be concluded from the above results that the hybrid system combining the DTCWT + CVANN, obtains highly accurate results in classifying the possible epileptic seizure of patients.

The advantages of DTCWT and CVANN were mentioned separately in Sections IV.A and IV.B.1. It is observed that the DTCWT-CVANN hybrid model obtained by combining the two models achieves the targeted success in terms of results. It is known that the WT-ANN approach produces effective results in the classification of EEG data [7]. However, as mentioned in the first part of the study, it is known that complex numbers play an important role in signal processing. In line with this idea, the proposed method generates better results than does the real-valued WT-ANN model in terms of both accuracy rates and computation times. It should also be mentioned that methods other than this can also provide high classification accuracies. Certainly there are researchers who are keen to obtain high accuracy ratios. However, in addition to achieving high accuracy values, the current study completely implemented complex-valued methods with regard to EEG data for the first time, and obtained successful results. In this sense, it is believed that the current study will act as a good guide to researchers in the field.

Up to this point, experiments were performed on two classes (A and E). Analysis phase has been extended in order to see the efficiency of the algorithm better. The effect of the proposed method on more complex problems was also investigated. To this end, the proposed method has been applied to 4 different experiments.

Experiment 1: Three sets from the data set were used and they were classified into three different classes. These classes are as follows: Healthy ones (A), epilepsy patients in seizure-free intervals (D) and epilepsy patients (E). Notation is simplified as A-D-E.

Experiment 2: All sets in the data set were used and they were classified into two different classes: A, B, C and D sets were included in the non-seizure class and E set was included in the seizure class. Notation is simplified as ABCD-E. This classification problem is more close to clinical applications.

Experiment 3: Four sets from the data set were used and they were classified into two different classes. Non-seizure sets excluding healthy with eyes closed (A, C, D) were employed in the first class and epilepsy patients with epileptic seizure (E) was used in the second class. Notation is simplified as ACD-E.

Experiment 4: Five sets from the data set were used and they were classified into three different classes. These classes are as follows: Healthy ones (AB), epilepsy patients in seizure-free intervals (CD) and epilepsy patients with epileptic seizure (E). Notation is simplified as AB-CD-E.

The results obtained with the proposed method for the 4 different experiments are presented in Table VI. In the table, there are also results obtained in the literature on the same dataset. A brief evaluation of the results obtained is presented below.

For experiment 1 (A-D-E), higher accuracy rate was obtained with the proposed method compared to the studies in the literature. 99.3% accuracy rate was obtained with 10-fold cross validation method and 99.04% accuracy rate was obtained with the hold-out data selection method.

For experiment 2 (ABCD-E), 99.33% accuracy rate was obtained with the hold-out data selection method. This result is 0.27% less than the result in the study of Orhan et al. [49].

In other experiments, classification was performed with the accuracy rates higher than the rates in the study of Orhan et al. [49].

For experiment 3 (ACD-E), 98.37% accuracy rate was obtained with 10-fold cross validation. Kumar et al. who used Fuzzy approximate entropy and SVM methods reported the closest value (98.15%) to this result.

For experiment 4 (AB-CD-E), 98.28% accuracy rate was obtained with 10-fold cross validation. Tzallas et al. who used time–frequency analysis and artificial neural network methods reported the closest value (97.72%) to this result.
This experiment shows that the proposed method can be applied successfully to more complex problems. All the methods and stages that were applied in the article are presented in detail. This may give the idea that a complex and computation burden method is applied. However, there is no substantial difference between the method used and the methods given in Table V and Table VI in terms of simplicity and computational load. The proposed method is a two-stage one, such as the studies presented in Table V and Table VI. These phases are feature extraction and classification phase.

It is used in order to determine the parameter values in the initial stage. The difference from the other studies in terms of the proposed method is the use of complex numbers instead of real values. Complex numbers do not impose a mathematical burden. As presented in the last paragraph of Chapter II, it is known that complex numbers play an important role in the fundamentals of signal processing. Therefore, the use of complex numbers provides a more natural way of achieving a solution with regard to signal processing. When computing time is analyzed, it is observed that the complex-valued classifier classifies faster than the real-valued classifier. As a result, the method applied is a fast one with a low calculation load.

VI. CONCLUSION

This study proposes a hybrid model for neurologists to help them analyze EEG signals for epilepsy diagnostic purposes with high rates of accuracy.

These models can be used in the automatic diagnosis of epileptic activities from EEG signals. The prominent parts of the study are listed below:

- The effect of complex-valued classifiers shown to have a positive impact on classification accuracy of EEG signal data. It is projected that such high levels of prediction accuracy rates can also be obtained in different biomedical signal processing fields.

- The study investigates the effects of different levels of complex-valued wavelet transformation, the results and comparative analyses of findings for those different levels are presented using popular statistical analysis metrics.

- The reported measures include the required time for obtaining the experimental results. In our future studies, we would like to minimize this duration by realizing the system using graphic processors. This will be possible by utilizing a new technology in the form of CUDA programming. Also future studies will make use of a visual interface for the model for ease of model building and. The use of such a visual interactive interface would increase the model’s intuitiveness applicability.

- In order to determine the success of the method used, the preferred data sets in previous studies were used. For future studies it is intended that this method will be implemented on EEG data involving larger and more continuous data.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>Data selection</th>
<th>Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Srinivasan et al. [8]</td>
<td>Time &amp; frequency domain feature-Recurrent neural network</td>
<td>Hold out (%60 training-%40 test)</td>
<td>99.60%</td>
<td>99.4%</td>
<td>99.8%</td>
</tr>
<tr>
<td>Subasi and Ercelebi [10]</td>
<td>WT+ANN</td>
<td>Hold out (%60 training-%40 test)</td>
<td>92%</td>
<td>91.6%</td>
<td>91.4%</td>
</tr>
<tr>
<td>Subasi [11]</td>
<td>Discrete wavelet transform-Mixture of expert model</td>
<td>Hold out (%60 training-%40 test)</td>
<td>94.5%</td>
<td>95%</td>
<td>94%</td>
</tr>
<tr>
<td>Kannathal et al. [12]</td>
<td>Entropy measures-Adaptive neuro-fuzzy inference system</td>
<td>Hold out (%60 training-%40 test)</td>
<td>92.22%</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Kannathal et al. [13]</td>
<td>Chaotic measures-Surrogate data analysis</td>
<td>N/A</td>
<td>90.00%</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Tzallas et al. [14]</td>
<td>Time frequency analysis-ANN</td>
<td>Hold out (%50 training-%50 test)</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Polat et al. [15]</td>
<td>Fast Fourier transform-Decision tree</td>
<td>10-fold cross-validation</td>
<td>98.72%</td>
<td>99.4%</td>
<td>99.31%</td>
</tr>
<tr>
<td>Acharya et al. [22]</td>
<td>WPD-PCA-GMM</td>
<td>10-fold cross-validation</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
</tr>
<tr>
<td>Acharya et al. [47]</td>
<td>Entropies + HOS + Higuchi FD + Hurst exponent + FC</td>
<td>10-fold cross-validation</td>
<td>99.7%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Acharya et al. [20]</td>
<td>HOS cumulants from WPD coefficients+SVM</td>
<td>3-fold cross-validation</td>
<td>98.5%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Guo et al. [19]</td>
<td>WT–RWE-MLPNN</td>
<td>Hold out (%50 training-%50 test)</td>
<td>95.20%</td>
<td>98.17%</td>
<td>92.12%</td>
</tr>
<tr>
<td>Our work</td>
<td>DTCWT+CVANN-1</td>
<td>10-fold cross-validation</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Our work</td>
<td>DTCWT+CVANN-2</td>
<td>Hold out (%60 training-%40 test)</td>
<td>99.5%</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>Our work</td>
<td>DTCWT+CVANN-3</td>
<td>10-fold cross-validation</td>
<td>99%</td>
<td>98.5%</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hold out (%60 training-%40 test)</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

TABLE V
A COMPARISON OF THE RESULTS OBTAINED BY OUR METHOD AND OTHERS’ METHODS FOR A-E CLASSIFICATION PROBLEM
Our work

Orhan et al. [48] WT+ k-means clustering+ MLPNN AB-CD-E Hold out (%50 training-%50 test) 96.67% 97.98% 94.12 % 99.58% 94.12 % 97.98% 92.45 %

Guo et al. [49] Genetic algorithm+ KNN A, D, E N/A 93.5% N/A N/A

Peker and Sen [25] Real and complex-valued features + CVANN A, D, E 10-fold cross-validation 97.01% 97% 98% 96% 96%

Our work

DTCWT+CVANN-1 A, D, E Hold out (%60 training-%40 test) 99.04 % 99.26% 97.75% 98.01% 97.05%

Tzallas et al. [14] Time–frequency analysis-artificial neural network ABCD-E Hold out (%50 training-%50 test) 97.73% 99.05% 94.20%

Orhan et al. [48] WT+ k-means clustering+ MLPNN ABCD-E Hold out (%50 training-%50 test) 99.60% 100% 98.04%

Guo et al. [50] Discrete wavelet transform-line length feature (LLF)-MLPNN ABCD-E N/A 97.77% 98.61% 94.6%

Kumar et al. [51] Fuzzy ApEn-SVM ABCD-E N/A 97.38% 98.1% 94.4%

Our work

DTCWT+CVANN-2 ABCD-E 10-fold cross-validation 99.15% 100% 97.89%

Kumar et al. [51] Fuzzy ApEn, SVM ACD-E N/A 98.15% 98.8% 96.2%

Guo et al. [50] Discrete wavelet transform-LLF-MLPNN ACD-E N/A 97.75% 98.55% 95.61%

Ocak [52] WT - ApEn ACD-E N/A 96% N/A N/A

Our work

DTCWT+CVANN-3 ACD-E 10-fold cross-validation 98.37% 99.05% 96.67%

Tzallas et al. [14] Time–frequency analysis-ANN AB-CD-E Hold out (%50 training-%50 test) 98.8% 92.8% 93%

Orhan et al. [48] WT+ k-means clustering+ MLPNN AB-CD-E Hold out (%50 training-%50 test) 95.6% 92.38% 93.8%

Our work

DTCWT+CVANN-1 AB-CD-E 10-fold cross-validation 98.28% 99.43% 98.35%


