Epileptic EEG Classification Using Synchrosqueezing Transform and Machine Learning

Ozlem Karabiber Cura
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Cigli, Izmir, Turkey
ozlem.karabiber@ikcu.edu.tr

Aydin Akan
Dept. of Electrical and Electronics Eng.
Izmir University of Economics
Balcova, Izmir, Turkey
akan.aydin@ieu.edu.tr

***Abstract*—Epilepsy is one of the neurological diseases that occur incidences worldwide. The electroencephalography (EEG) recording method is the most frequently used clinical practice in the diagnosis and monitoring of epilepsy. Many computer-aided analysis methods have been developed in the literature to facilitate the analysis of long-term EEG signals. In the proposed study, the patient-based seizure detection approach is proposed using a high-resolution time-frequency (TF) representation named Synchrosqueezed Transform (SST) method. The SST of two different data sets called the IKCU data set and CHB-MIT data set are obtained, and Higher-order joint TF (HOJ-TF) based and Gray-level co-occurrence matrix (GLCM) based features are calculated using these SSTs. Using some machine learning methods such as Decision Tree (DT), k-Nearest Neighbor (kNN), and Logistic Regression (LR), classification processes are conducted. High patient-based seizure detection success is achieved for both the IKCU data set (94.25%) and the CHB-MIT data set (95.15%).**

***Keywords*—EEG, SST,Time-Frequency Analysis, patient-based seizure detection.**

I. Introduction

Epilepsy is a neurological disease that occurs as recurrent seizures as a result of the unexpected electrical activities of brain nerve cells. It is the second most common nervous system disease after stroke, affecting approximately 0.6-0.8% of the world's population [1], [2]. The EEG method, in which the electrical activity of the brain is recorded, is the most frequently used clinical practice by neurologists in the diagnosis and monitoring of epilepsy because of fast, inexpensive, and easily accessible. However, the examination of long-term EEG records by experts is very time-consuming and exhausting [3]. Therefore, many advanced signal processing methods have been introduced using the EEG signal to facilitate seizure detection and prediction, and efficient performance results have been obtained. [2], [4].

Many TF analysis methods have been proposed for the more effective analysis of non-stationary and nonlinear EEG signals. Wavelet transform (WT) and its derivative approaches [2], [5], [6], [7], Short-Time Fourier transform (STFT) [8], Wigner–Ville distribution (WVD) [9] have yielded highly effective results in epileptic seizure detection and prediction. TF representations of EEG signals were obtained by using different TF analysis methods such as CWT [10], Hilbert-Huang transforms (HHT)[11], and these TFRs were used to obtain various texture features to achieve high seizure detection or prediction performance. In addition to the studies using advanced signal processing methods, deep learning approaches have been used successfully in this field especially in recent years [12], [13], [14]. Using traditional TF methods such as STFT or CWT, to obtain high-resolution TF localization is not possible. In order to overcome this disadvantage and to obtain TF representation with high resolution, a new STFT or CWT based TF analysis method, SST, has been developed [15], [16], [17].

In this paper, the high-resolution SST based epileptic seizure detection approach is introduced using two different data set. HOJ-TF moments based and GLCM based features are calculated using the magnitude of SST and classification processes are performed using various classifiers such as DT, kNN, and LR.

II. Materials and Methods

In the proposed study, patient-based seizure detection approached is performed using the SST method. TF representation of pre-seizure (or inter-seizure) and seizure EEG segments are obtained using the SST approach and HOJ-TF moment-based and GLCM based features are computed. Using various classifiers such as DT, kNN, and LR, classification processes are carried out and performance evaluation results are compared.

*1) **EEG data sets**:* Two different EEG data set, the IKCU data set and CHB-MIT data set, are utilized for this study.

• EEG data set-1 (IKCU data set): This data set consists of EEG signals of 16 epilepsy patients (5 Female; 11 Male, the average age is 37.3∓7) and includes a total of 507 seconds pre-seizure and 892 seconds seizure EEG signals. EEG signals were recorded from 18 different channels (Fp1-F7, F7-T1, T1-T3, T3-T5, T5-O1, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F8, F8-T2, T2-T4, T4-T6, T6-O2, Fp2-F4 F4-C4, C4-P4, P4-O2) using the 10-20 electrode system with a 10 ms sampling period at Izmir Katip Celebi University School of Medicine, Department of Neurolog. Since epileptic seizures in the data set are known to be predominant in the Left Temporal and Frontal lobes, 10 channels (Fp1-F7, F7-T1, T1-T3, T3-T5, Fp1-F3, Fp2-F8, F8-T2, T2-T4, T4-T6, Fp2-F4) in the Temporal and Frontal lobes are used from these 18 channels. Ethical approval has been obtained to use these EEG signals in the study (numbered 2969 and dated 08.08.2019). Pre-seizure and seizure EEG signals are divided into non-overlapping, 1s duration EEG segments.

• Epileptic EEG data set-2 (CHB-MIT data set): Another examined data set in our proposed study is public Children's Hospital Boston, CBH-MIT, epileptic EEG data set [18]. This data set consists of epileptic EEG signals of 24 pediatric patients recorded from 23 or 18 channels with a sampling frequency of 256 Hz. For our experiment, temporal and frontal lobe-weighted 10 channels (FP1-F7, F7-T7, T7-P7, FP1-F3, T7-FT9, FP2-F8, F8-T8, T8-P8, FP2-F4, FT10-T8) of 23 patients (chb01-chb23) are selected to be parallel to the IKCU data set. Non-overlapping, 1s duration EEG segments are obtained from the seizure EEG signals, and 10-minute inter-seizure EEG signals which end the minimum 1 hours before the onset of the seizure signals.

*2) **Synchrosqueezing Transform**:* Synchrosqueezing Transform that is a member of the TF reassignment methods (RM) class, is proposed to achieve highly localized TFRs of non-stationary processes. In the literature, CWT-based or STFT-based SST approaches have been proposed, but in our proposed study STFT-based SST (FSST) approach is used to achieve high seizure detection performance [15], [17], [19].

In the first step of SST, STFT of EEG segment $x(t)$ in the frequency domain is computed;

$$X(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\xi)W(\omega-\xi)e^{j\xi t}d\xi. \qquad (1)$$

Here, $X(\xi)$ and $W(\xi)$ are the Fourier Transforms of EEG segment $x(t)$ and the window function $w(t)$, respectively. $X(t,\omega)$ denotes the STFT of the corresponding EEG segment.

Instantaneous frequency (IF) $\omega_0(t,\omega)$ information, which is generally neglected in STFT, is estimated by computing the derivative of the STFT $X(t,\omega)$ with respect to time for the reassignment in SST.

$$\omega_0(t,\omega) = -j\frac{\partial_t X(t,\omega)}{X(t,\omega)} \qquad (2)$$

Using the calculated IF, ω_0(t, ω), STFT coefficients that have the same frequency information are gathered where they should be present to achieve improved energy concentration. Hence, using the synchrosqueezing operator $\int_{-\infty}^{\infty}\delta(\eta-\omega_0(t,\omega))d\omega$, SST of the EEG segment $T(t,\eta)$ is formulated as,

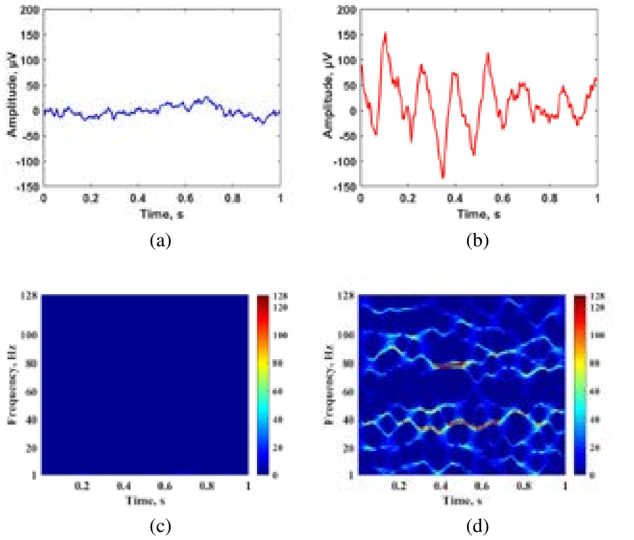
$$T(t,\eta) = \int_{-\infty}^{\infty} X(t,\omega)\delta(\eta-\omega_0(t,\omega))d\omega \qquad (3)$$

An example TF representations of 1s duration inter-seizure and seizure EEG segments of CHB-MIT data set obtained using the SST approach are demonstrated in Fig. 1.

*3) **Feature Extraction**:* HOJ-TF moment and GLCM based features are calculated using TFR obtained by SST to obtain two different feature sets.

• **HOJ-TF moment** based features $< t^i\ \omega^j >$; $i,j = 1,\ldots,4$. are computed using magnitude square ($S(t,\omega)$) of SST ($T(t,\eta)$).

$$< t^i\ \omega^j > = \int\int t^i\ \omega^j\ S(t,\omega)\ dt\ d\omega, \quad i,j = 1,\ldots,4. \qquad (4)$$



Şekil 1: 1 s. duration, (a) inter-seizure, (b) seizure EEG; magnitude SST of (c) inter-seizure, (d) seizure EEG segments of CHB-MIT data set.

Log normalization, $\log(\frac{<t^i\ \omega^j>}{i!j!})$, $\quad i,j = 1,\ldots,4$, is applied to reduce dynamic range of joint TF moments. Thus, 1×16 log-normalized HOJ-TF moment based feature vector is obtained for each pre-seizure, inter-seizure, and seizure EEG segment.

• TFR obtained by SST is considered as a gray-level image. The **Gray Level Co-occurrence Matrix (GLCM)**, one of the texture descriptors, is obtained by calculating joint probability distributions of two neighboring image pixel pairs with a specific position that includes of distance (d) and direction (θ) information. By using position information ($\Delta = (\theta, d)$) of two image pixel pair, the GLCM of this TF image is expressed as follow;

$$G_\Delta(i,j) \qquad (i,j = 0,1,...,N_g-1) \qquad (5)$$

Here, i,j denotes the intensity values of two pixels, N_g indicates the gray level number of the image. In our experiments, GLCM matrices of each EEG segments are calculated for position $\Delta = (\theta = 0^o, d = 1)$. Second-order statistical features, i.e., contrast, correlation, energy, and homogeneity are computed using the obtained GLCM matrix [10], [20]. Thus, 1×4 GLCM feature vector is calculated for each EEG segment.

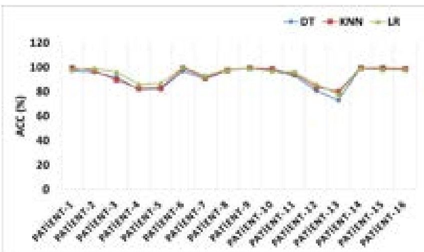
*4) **Classification and Performance Evaluation**:* In the classification processes, various machine learning algorithms such as DT kNN, and LR are used for classification. **DT** is a successful machine learning method that uses tree-like structures such as branches or nodes, and the training process is realized by learning a series of decision-rules [8]. In our experiment, the coarse tree structure in which the maximum number of the split is equaled to 4 is used. In another high-performance machine learning approach, **kNN**, the distance between the sample to be classified and k neighbor is calculated. The sample to be classified is assigned as an element to the class in which low-distance neighbors are common [10]. In our experiment, k is chosen as 10, and Euclid distance is used as the distance measurement. **LR** is one of the most widely used statistical classification methods in binary classifications, producing binary output such as yes/no, 1/0 [8]. Statistical metrics such as Accuracy (ACC), Recall (REC), False positive rate (FPR) are used to evaluate the performance of the classifiers [8], [21].

$$ACC = \frac{TP + TN}{TP + FN + FP + TN}$$
$$REC = \frac{TP}{TP + FN}, \qquad (6)$$

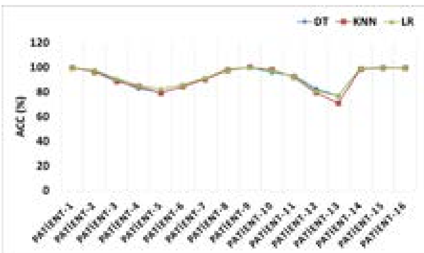
Here, true-positive (TP) is the number of correctly classified seizure samples, true-negative (TN) is the number of correctly classified pre-seizure (or inter-seizure) samples. Additionally, false-positive (FP) is the number of missclassified pre-seizure (or inter-seizure) samples. False-negative (FN) denotes the number of missclassified seizure samples [8], [21].

III. Results and Discussion

In this study, the patient-based seizure detection approach is performed using two different data set, the IKCU data set and the CHB-MIT data set, and the SST method. HOJ-FT moment-based and GLCM based features are calculated using the magnitude square of TFRs that are achieved using the SST method.



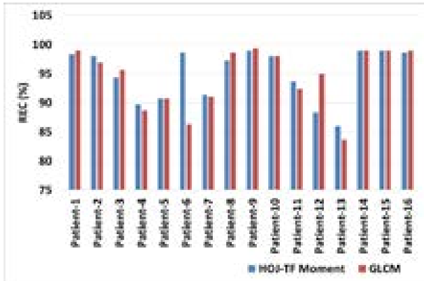
(a)



(b)

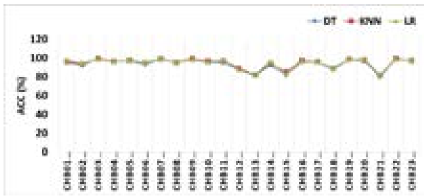
Şekil 2: Accuracy values obtained using (a) HOJ-TF moment based, (b) GLCM based feature sets of IKCU data set.

The patient-based seizure detection results of the SST approach obtained using the IKCU data set are given in Figs 2 and 3. The accuracies of patient-based seizure detection obtained from three classifiers using the HOJ-TF moment based feature set and the GLCM-based feature set are given in Figures Figs 2a and 2b, respectively. The LR classifier provides the highest performance with an average of 94.25 % and 92.3 % ACC values in both HOJ-TF moment-based and GLCM based feature sets. High average REC (higher than 85%) values are obtained using three classifiers for all patients except patients 4, 5, 12, and 13 (given in Fig. 3) from both HOJ-TF moment-based and GLCM based feature set of the IKCU data set.

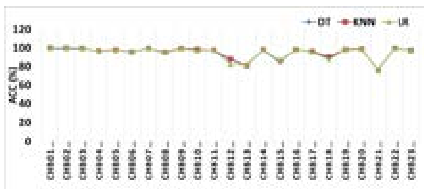


(a)

Şekil 3: Change of patient-based recall values of IKCU data set.

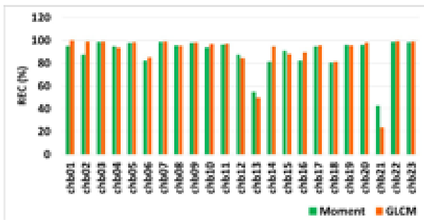


(a)



(b)

Şekil 4: Accuracy values obtained using (a) HOJ-TF moment based, (b) GLCM based feature sets of CHB-MIT data set.



(a)

Şekil 5: Change of patient-based recall values of CHB-MIT data set.

Performance evaluation results of SST based seizure detection approach obtained using the CHB-MIT data set are given in Figs.4 and 5. For all patients except chb12, chb13, chb15, and chb21, high accuracy values (higher than 95%)

are achieved from all three classifiers using both HOJ-TF moment-based and GLCM based feature set (given in Fig. 4). The kNN classifier yields high average accuracy values calculated using all patient ACC values, for both HOJ-TF moment-based (94.47%) and GLCM based (95.15%) feature sets. High recall values (greater than 90%) that indicate the rate of seizure segment correctly detected are obtained from all patients except chb06, chb12, chb13, chb18, chb21 patients (shown in Fig.5).

IV. CONCLUSIONS

In this study, the patient-based seizure detection approach is performed on two different data sets using SST, the new high-resolution TF analysis method. HOJ-TF moment-based and GLCM based feature sets are computed using the magnitude square of obtained SSTs obtained from the pre-seizure and seizure EEG segments of the IKCU data set and inter-seizure and seizure EEG segments of the CHB-MIT data set. High patient-based accuracy and recall values are achieved using both HOJ-TF moment-based and GLCM based feature sets of two data set.

Many seizure detection studies have been conducted in the literature using the CHB-MIT data set [10], [14], [12], [13], [21]. While CNN based studies yield lower average REC values (in study [14]= 85% REC; in study [12]= 71.45% REC), in proposed SST based seizure detection approach is to provide higher average REC value (90.30%). In another study [10] in which GLCM-based features are calculated the lower REC value (70.19%) is obtained than the proposed SST-based seizure detection study. On the other hand, ın studies [13], [21] are achieved higher performance values than our proposed study.

Considering these evaluations, it is seen that the proposed SST approach can be used successfully in detecting seizures. The future goal of our study is to perform a seizure prediction study using the same method and the CHB-MIT data set.

KAYNAKÇA

[1] S. K. Rout and P. K. Biswal, "An efficient error-minimized random vector functional link network for epileptic seizure classification using VMD," *Biomedical Signal Processing and Control*, vol. 57, p. 101787, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809419303684

[2] M. Sharma, R. B. Pachori, and U. R. Acharya], "A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension," *Pattern Recognition Letters*, vol. 94, pp. 172 – 179, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865517300995

[3] D. Bhati, M. Sharma, R. B. Pachori, and V. M. Gadre, "Time–frequency localized three-band biorthogonal wavelet filter bank using semidefinite relaxation and nonlinear least squares with epileptic seizure EEG signal classification," *Digital Signal Processing*, vol. 62, pp. 259 – 273, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105120041630210X

[4] U. R. Acharya, S. V. Sree, S. Chattopadhyay, W. Yu, and P. C. A. Ang, "Application of recurrence quantification analysis for the automated identification of epileptic EEG signals," *International journal of neural systems*, vol. 21, no. 03, pp. 199–211, 2011.

[5] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 205–211, 2007.

[6] A. B. Das, M. I. H. Bhuiyan, and S. S. Alam, "Classification of EEG signals using normal inverse gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection," *Signal, Image and Video Processing*, vol. 10, no. 2, pp. 259–266, 2016.

[7] C. Sun, H. Cui, W. Zhou, W. Nie, X. Wang, and Q. Yuan, "Epileptic seizure detection with EEG textural features and imbalanced classification based on easyensemble learning," *International journal of neural systems*, vol. 29, no. 10, p. 1950021, 2019.

[8] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time–frequency analysis," *IEEE transactions on information technology in biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.

[9] N. A. Khan and S. Ali, "A new feature for the classification of non-stationary signals based on the direction of signal energy in the time–frequency domain," *Computers in biology and medicine*, vol. 100, pp. 10–16, 2018.

[10] K. Samiee, S. Kiranyaz, M. Gabbouj, and T. Saramäki, "Long-term epileptic EEG classification via 2d mapping and textural features," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7175 – 7185, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417415003218

[11] K. Fu, J. Qu, Y. Chai, and Y. Dong, "Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM," *Biomedical Signal Processing and Control*, vol. 13, pp. 15 – 22, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809414000457

[12] K. Ayodele, W. Ikezogwo, M. Komolafe, and P. Ogunbona, "Supervised domain generalization for integration of disparate scalp EEG datasets for automatic epileptic seizure detection," *Computers in Biology and Medicine*, p. 103757, 2020.

[13] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, "Epileptic seizure detection based on EEG signals and CNN," *Frontiers in neuroinformatics*, vol. 12, p. 95, 2018.

[14] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Machine learning for healthcare conference*, 2016, pp. 178–190.

[15] T. Oberlin, S. Meignen, and V. Perrier, "Second-order synchrosqueezing transform or invertible reassignment? towards ideal time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1335–1344, 2015.

[16] A. Ahrabian, D. Looney, L. Stanković, and D. P. Mandic, "Synchrosqueezing-based time-frequency analysis of multivariate data," *Signal Processing*, vol. 106, pp. 331–341, 2015.

[17] S. Mamli and H. Kalbkhani, "Gray-level co-occurrence matrix of fourier synchro-squeezed transform for epileptic seizure detection," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 87–99, 2019.

[18] "Public Data set: CHB-MIT EEG dataset," https://physionet.org/content/chbmit/1.0.0/, accessed: 2020-05-16.

[19] G. Yu, M. Yu, and C. Xu, "Synchroextracting transform," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 10, pp. 8042–8054, 2017.

[20] Y. Chen, H. Li, L. Hou, and X. Bu, "Feature extraction using dominant frequency bands and time-frequency image analysis for chatter detection in milling," *Precision Engineering*, vol. 56, pp. 235–245, 2019.

[21] D. P. Dash, M. H. Kolekar, and K. Jha, "Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and hidden markov model," *Computers in Biology and Medicine*, vol. 116, p. 103571, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482519304251