DOI 10.1109/ICHI52183.2021.00015

Imbalanced EEG Analysis Using One-Shot Learning with Siamese Neural Network

Munawara Saiyara Munia, Seyyed MohammadSaleh Hosseini, Mehrdad Nourani
Predictive Analytics and Technologies Laboratory
The University of Texas at Dallas
Richardson, TX, USA
{MunawaraSaiyara.Munia, SeyyedMohammadsaleh.Hosseini, nourani}@utdallas.edu

Jay Harvey, Hina Dave
Department of Neurology and Neurotherapeutics
The University of Texas Southwestern Medical Center
Dallas, TX, USA
{jay.harvey, Hina.Dave}@utsouthwestern.edu

***Abstract*—Epilepsy is a socially-stigmatizing chronic neurological condition. Limited availability of seizure Electroencephalogram (EEG) data makes the application of machine learning techniques for epileptic seizure detection very challenging. In this work, an efficient algorithmic procedure is proposed to facilitate the learning and classification of epileptic seizures from imbalanced EEG data. We designed an end-to-end architecture by combining local binary pattern with Siamese convolutional neural network. We used local binary pattern due to its capability to capture distinguishable morphological characteristics in the EEG signal. Siamese convolutional neural network was used since it can learn a similarity metric using an extremely small number of training samples for seizure episodes. With availability of a very small amount of training (seizure) samples, the effectiveness of the proposed method was verified by comparing the Siamese convolutional neural network with a baseline convolutional neural network. The proposed architecture outperforms the baseline model and achieves an average of 11.66% increase in F1-measure.**

***Index Terms*—Epileptic Seizure Detection, One-shot Learning, Siamese Neural Network, Local Binary Pattern (LBP), Convolutional Neural Network (CNN), Electroencephalogram (EEG).**

I. Introduction

A. Motivation and Background

Epilepsy is an acute neurological disorder, affecting roughly 0.8% of the world's population [1]. It is characterized by recurring spontaneous seizures. According to the World Health Organization (WHO), around five million people are diagnosed with epilepsy each year, and between four and ten per 1000 people suffer from active epilepsy (i.e. repetitive seizures or with the need for treatment) [2]. Despite pharmacological therapy [3], around 30% of patients still suffer from recurrent seizures; hence, efficient algorithms are highly desired for the detection of epileptic seizure and early alert. Most of the previous work on seizure detection assumes that brain activity in patients with epilepsy has two discrete states (seizure and non-seizure), and these states are detectable by exploiting Electroencephalogram (EEG). The emergence of EEG as the prime tool to monitor the brain activities of patients with epilepsy has drawn a raft of researchers' attention to a great degree owing to its high temporal resolution, affordability, portability, and the capability to capture neural activities in a non-invasive way [4], [5].

The progression of deep learning has ignited remarkable interest in leveraging deep learning methods for healthcare applications, including epileptic seizure detection. During the employment of deep learning techniques in biomedical domain, a major challenge in some applications is lack of sufficient labeled training samples. Acquiring ample labeled training samples (e.g. EEG) for each patient and manual labeling of the data by medical experts is immensely time-consuming and costly. Moreover, biophysiologic data like epileptic seizure EEG is typically imbalanced due to its rarity. Other than the difficulties stemming from the seizure dynamics and having a patient-specific nature, seizure detection is extremely challenging using only a small number of ictal examples [6]. Traditional supervised-learning paradigm fails to perform well when there is not enough labeled data for each class for a certain given task in order to train a reliable model. Additionally, repeated retraining of the model is necessary if the total number of classes is dynamically changing, resulting in a very costly training process; therefore, an efficient algorithm is desired that has the capability to learn from limited seizure EEG data and can generalize well to unseen seizure data.

To cope with the above-mentioned issues related to imbalanced datasets, researchers have proposed different techniques: oversampling and undersampling techniques [7], [8], active learning methods [9], cost-sensitive methods [10],

2575-2634/21/$31.00

generation of synthetic samples using generative models [11]–[14] etc. There is also a branch in deep learning developed to efficiently deal with lack of labeled training samples, called one-shot/few-shot learning [15]. One-shot learning is a variant of deep metric learning which requires that a model learns how to categorize an object, after being exposed to it either once or just a few times by leveraging the information it has known about categories. Siamese neural network, in particular, is a deep learning model that can be used for one-shot/few-shot learning.

B. Prior Work

Siamese network (a.k.a. twin neural network) [16] are based on a similarity function. It processes two inputs in parallel and detects if the two inputs are from the same class or not. In Siamese network, there are two parallel neural networks which have the same parameters and shared weights. Each of these networks takes a different input, one being the actual data and the other one the reference data (e.g. a known class). Afterwards, a feature vector which is of fixed-length is generated for each of the input data points. The hypothesis is that if the two input data points are from the same class, then the feature vectors should also be similar, resulting in a small difference between the feature vectors and a high similarity score output [16]. The opposite should also happen if the inputs are from different classes. This dual input strategy used in Siamese neural networks causes a massive increase in the number of training examples, allowing us to exploit deep learning methods designed for small dataset and prevent severe overfitting. Due to their capability to successfully learn from limited data, Siamese networks were utilized in several biomedical applications [17]–[20] in recent years.

The local binary pattern (LBP) proposed by Ojala et al. [21] is a nonparametric image operator that has been extensively leveraged by different image analysis tasks, such as face recognition, text classification, and image retrieval [22]–[24]. LBP is a texture descriptor which allows a system to efficiently capture local structures of a 2D image. LBP is of particular interest due to its discriminative power, simple representation, and efficient computation. These features make LBP a strong choice for real-time applications that require image analysis [25]. It can effectively capture intensity in a pixel's local neighborhood. In this method, we create a binary code for each of the pixels in an image. In order to obtain the corresponding binary code of a pixel (also called the center pixel), we compare the pixel's value with its surrounding pixels' values. The center pixel is used as a threshold for deciding each bit's value in the binary code created for the pixel. We then convert the binary code into its decimal equivalent, also called the LBP code. These LBP codes represents the unique local structural patterns of the image. After the completion of the computation of all the LBP codes, these codes are then used to form a histogram which graphically denotes the structural distribution of patterns across the image. 1D-LBP was first proposed by Chatlani et al. [26], where the authors extracted LBP features from a one-dimensional speech signal and showed that 1D-LBP can be used as an effectual symbolization method to extract unique morphological patterns in non-stationary signals like speech signal.

C. Main Contribution

In this work, we develop an algorithmic procedure for epileptic seizure detection from EEG data that takes advantage of 1D-LBP and Siamese neural network in order to address the aforesaid challenges. The main contribution of this work is two-fold. First, we utilize the idea of EEG data symbolization using 1D-LBP to transform raw EEG data from each electrode into 1D-LBP histogram representation. The rational is that 1D-LBP captures the relational aspects between the consecutive values of raw experimental measures (e.g. in a time series) by converting them into a series of discrete codes. Each LBP code symbolizes a unique wave shape, which solely represents the increase or decrease of amplitudes and is independent of magnitude. Second, we design a regularized Siamese CNN that performs automatic deep feature extraction on the 1D-LBP histogram data and learns a similarity metric to differentiate between seizure and non-seizure classes.

To the best of our knowledge, this is the first time that for epileptic seizure detection task, one-shot learning with Siamese neural networks is used in order to address the issue of imbalanced training dataset. The Siamese CNN performs binary classification of seizure and non-seizure EEG by analyzing the occurrence of common LBP patterns in pairs of feature vectors of similar or different classes (seizure and non-seizure classes), which bears resemblance to classical visual EEG interpretation [27]. For validation, we demonstrate that our model can successfully learn from a small number of samples (e.g. very short episodes of seizure) and still achieve superior classification accuracy. Our algorithm can learn from on an average of 77.04 seconds of seizure samples from 23 patients in CHB-MIT database [28] and shows high generalization on detection of unseen seizures. The model was validated with 5-fold cross validation. We empirically compare the Siamese CNN model with a baseline CNN model in terms of generalization capability to new patients when very little training samples are incorporated in the training dataset. Our experimentation shows that Siamese CNN with One-shot learning outperforms typical deep learning model that cannot deal well with scarce training samples.

II. Background on 1D Local Binary Pattern

1D-LBP is a symbolization technique which was first proposed by Chatlani et al. [26], where the authors showed that it can be deemed a promising approach for extracting unique and discriminatory features in non-stationary speech signals; hence, 1D-LBP can be well-suited for analyzing non-stationary EEG signals as well. The basic procedure to calculate 1D-LBP is very much like texture operator [21]. It, nonetheless,

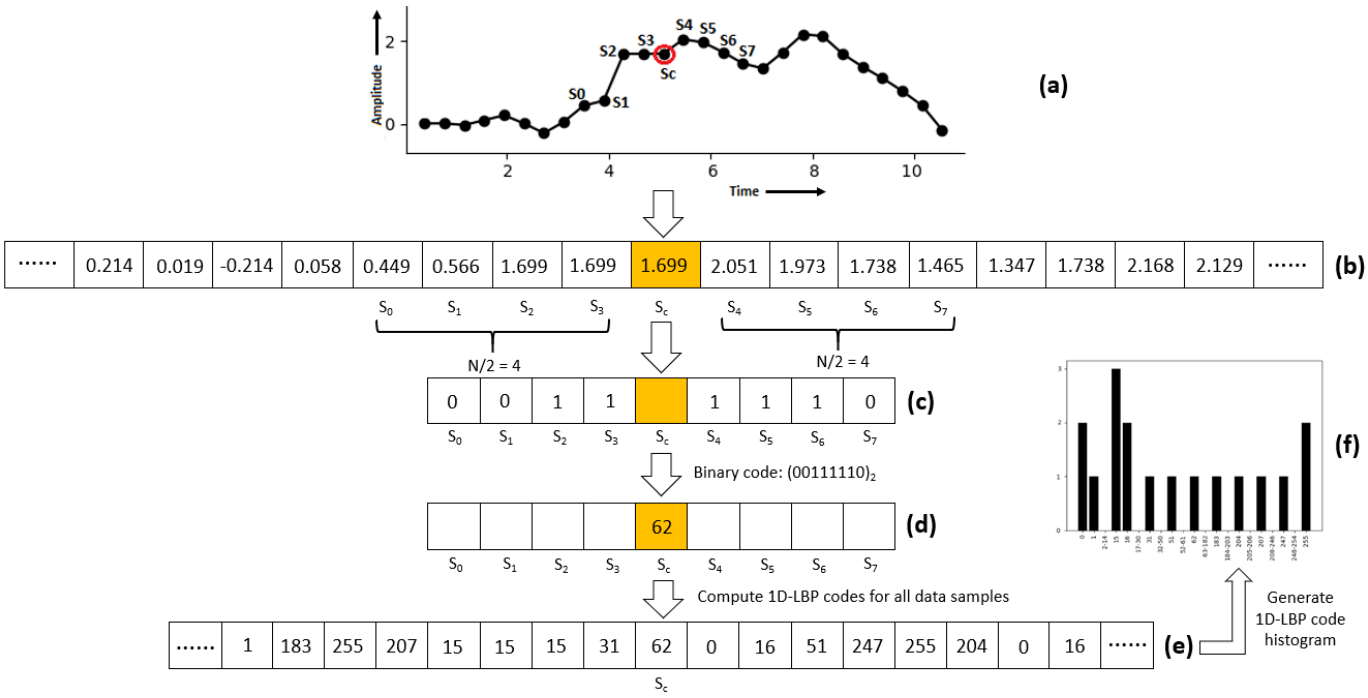

Fig. 1: The steps for generating 1D-LBP code: (a) A sub-signal denoted by S_0-S_7 from a segment of an EEG signal is taken, (b) corresponding amplitudes of S_0-S_7 (c) neighboring samples are compared with the center sample, and an 8-bit binary code is generated after performing the thresholding (d) the decimal value is calculated (e) 1D-LBP code is computed for all data samples of the EEG signal segment, and (f) 1D-LBP code histogram is generated (here, x-axis denotes 2^N possible LBP values, and the y-axis denotes the frequency of the LBP values that have occurred in the signal).

analyzes the samples in the vicinity in time series data in a sequential manner. 1D-LBP's mathematical formulation is akin to 2D-LBP's [29]; however, in lieu of the pixel values for the pixels grid, the amplitude of each sample of the EEG signal is used. 1D-LBP computation can be mathematically shown as follows:

$$t_i = A_i - A_c \tag{1}$$

$$F(t_i) = \begin{cases} 1 & \text{for } t_i \geq 0 \\ 0 & \text{for } t_i < 0 \end{cases} \tag{2}$$

$$LBP(S_c) = \sum_{i=0}^{N} F(t_i) \times 2^i; (1 \leq c \leq N_{samples}) \tag{3}$$

where,
A_i = amplitude of ith neighboring sample S_i
A_c = amplitude of the center sample S_c
N = total number of considered neighbors, $\frac{N}{2}$ neighbors from each side of the center sample
S_c = the current center sample which the 1D-LBP is being computed for
$N_{samples}$ = total number of samples in each window
$F(t_i)$ = Sign function, which converts the differences to an N-bit binary code

The steps of calculating 1D-LBP of EEG signal are as follows:

1) At a given signal position S_c, N neighboring samples from the signal are considered in order to compute the LBP value for the center sample S_c. $\frac{N}{2}$ of these neighboring samples come before the center sample, and $\frac{N}{2}$ come after it. For example, for $N = 8$, the four neighboring samples taken before and after S_c will be S_0, S_1, S_2, S_3 and S_4, S_5, S_6, S_7, respectively, as shown in Fig. 1(a).

2) The amplitudes of all the neighboring samples A_0, A_1, A_2, A_3, A_4, A_5, A_6, and A_7 are compared with the amplitude of the center sample (A_c). The difference between the amplitude of the ith neighbor sample (A_i) and the amplitude of the center sample (A_c) is known as the decision variable t_i, as shown by equation (1). Depending on t_i's value, we perform thresholding of these values to produce a binary number, which gives us the 1D-LBP code, as shown in equation (2). If the neighboring sample's amplitude A_i is greater than or equal to the center sample's amplitude A_c, then we assign 1 for the value $F(t_i)$, else we assign 0; thus, a binary LBP code of length N for a window is formed.

3) The above calculation is sequentially conducted for each one of the samples in the time series data sequentially for the entire time series. At each step, $N+1$ successive samples are considered resulting in an N-bit binary

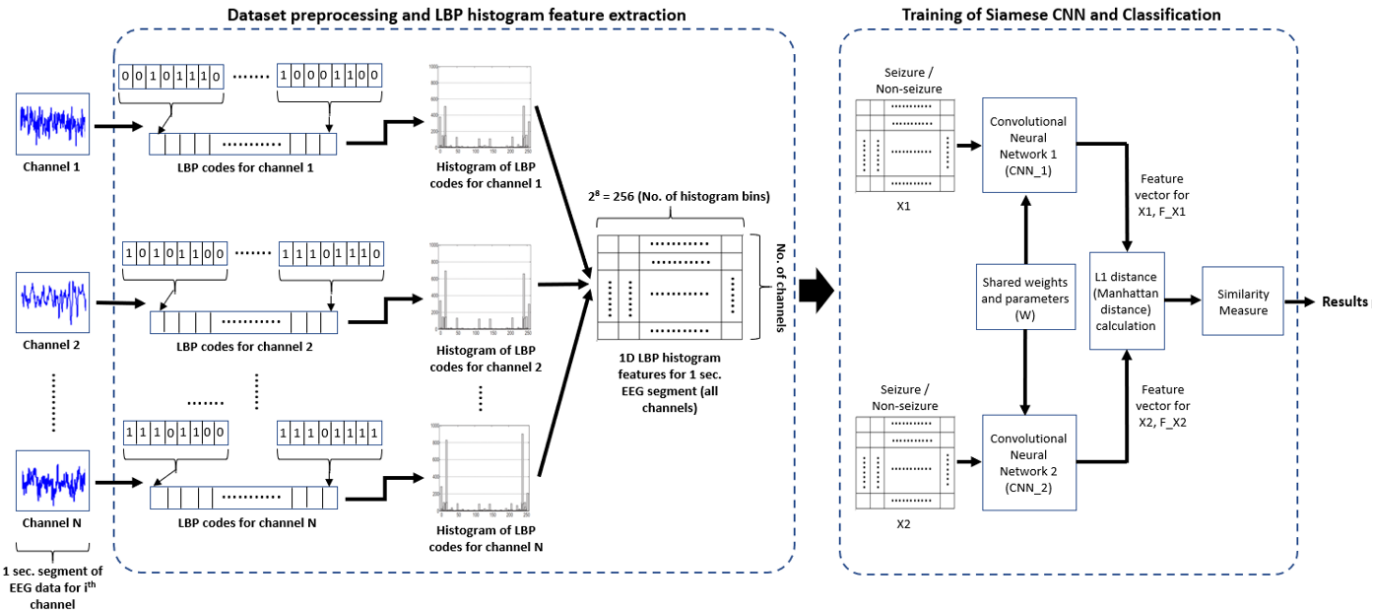

Fig. 2: Proposed Architecture

value for each sample. This value is converted to a decimal number causing LBP values ranging from 0 to $2^N - 1$, which represents the local structural information around S_c [22].

Fig. 1 depicts the method of applying LBP to 1D EEG time series signal. Here, S_c is the center sample and $N = 8$. As shown in Fig. 1, after using equations (1) and (2), the binary code for S_0 to S_3 is $(0011)_2$, and it is $(1110)_2$ for S_4 to S_7, which result in the binary representation $(00111110)_2$. Hence, S_c has an LBP value of $(62)_{10}$, which is then used in the same index as the center sample. The LBP values are calculated for all of the samples in the entire signal in a similar way.

EEG signal in time domain is a series of amplitude values. The morphological characteristics of seizure and non-seizure samples are considerably different. In this work, we characterize raw EEG data as the occurrence of some local binary patterns using 1D-LBP histogram, where x-axis denotes 2^N possible LBP values, and the y-axis denotes the frequency of the LBP values that have occurred in the signal. As every LBP value gives a piece of information in regard to the wave morphology in a local neighborhood, it helps capture different patterns in the original signal and summarizes the structural distribution of patterns in the signal. Existence of repetitive common local patterns symbolizes the existence of seizure in a sample; thus, the difference between the structural distribution of patterns in seizure and non-seizure samples can serve as unique and distinguishable features of the EEG signal.

III. Methodology

A. Proposed Architecture

Fig. 2 shows our proposed method which is divided into two main steps: 1) Dataset preprocessing and LBP histogram feature extraction, 2) Training of the Siamese CNN and classification. The details of each step are discussed next.

1) Dataset Preprocessing (1D-LBP Histogram Feature Extraction): The sampling frequency of the EEG signals used in this work is 256 Hz. In the preprocessing step, at first the EEG data is segmented into 1-second non-overlapping windows each of which consists of 256 samples. For each channel in a window, a 1D-LBP code with length $N = 8$ is calculated for each sampling point [26]. Here, $N = 8$ means the 1D-LBP code for the sampling point is calculated through comparing the current sample's amplitude to its four consecutive left neighbors and four consecutive right neighbors; therefore, the information about local variations of the signal is captured. The procedure is carried out for each sample in the window. The resultant LBP codes are used for symbolization of the EEG signal of an electrode. We can have 2^N different LBP codes which are later used to generate the LBP histogram feature. In the histogram, the number of bins is 2^N, and the y-axis of the histogram shows how many times each LBP code has occurred in a 1-second window (256 samples). These histogram features from each channel are then combined and fed to the next stage for training and classification. Fig. 3 illustrates 1D-LBP histograms for a 1-second seizure episode and a 1-second non-seizure episode. As shown, the LBP codes have distinctive distributions for the non-seizure and seizure episodes. During a non-seizure episode, the LBP codes appear to be randomly distributed in

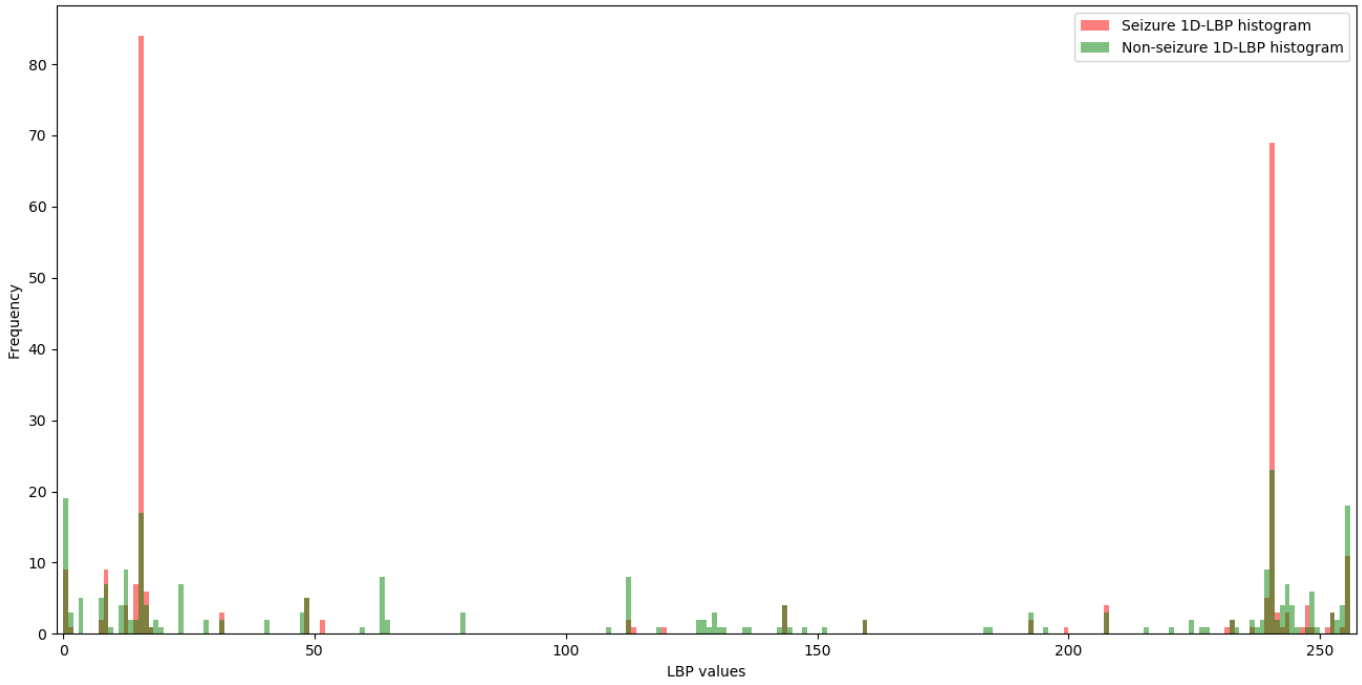

Fig. 3: Seizure vs. Non-Seizure LBP code histogram of channel 1 from Chb01 (MIT-CHB dataset) for 1-second seizure (red) and 1-second non-seizure (green) episodes

the histogram. Conversely, a polarized histogram is formed during the seizure episode, as rhythmic signals are usually observed during a seizure. This implies that distribution of the LBP codes is an important indicator for distinguishing seizure vs. non-seizure states.

2) Training of the Siamese CNN and Classification: In the training step, at first, we construct a new binary dataset to train and validate the Siamese CNN. In the dataset, each example (X_1, X_2, Y) consists of a pair of EEG data samples, X_1 and X_2 and a binary label Y. Label Y is defined as follows:

$$Y = \begin{cases} 1, & \text{if } X_1, X_2 \in C_S \text{ or } X_1, X_2 \in C_{NS} \\ 0, & \text{otherwise} \end{cases}$$

where C_S and C_{NS} are Seizure and Non-seizure classes, respectively. These pairs are randomly generated from all the patients in the training data, which are later fed as input into the Siamese CNN. We generate three different types of pairs for feeding into the Siamese CNN: $S-S$, $S-NS$ and $NS-NS$, where S and NS refer to seizure and non-seizure types. If we have only 30 one-second seizure samples and 600 one-second non-seizure samples, then we can have $\binom{30}{2} + \binom{600}{2} + (30 * 600) = 198,135$ possible pairs for training the Siamese CNN. Hence, the dual input technique of the Siamese CNN significantly increases the total training samples and facilitates the training with small datasets.

We then learn the similarity metric by embedding the EEG data into a vector space and comparing the resulting vectors with L1 distance. The Siamese CNN structure consists of two-sister (or twin) CNN networks with shared parameters and hyper-parameters (denoted as "W" in Fig. 2). These twin CNNs (CNN_1 and CNN_2) aim to extract the essential features (feature vectors of fixed length) of the two EEG data samples. The weights and parameters are shared between the twin CNNs as both of them need to learn the same embedding function to extract the deep features of the EEG samples. Initially, two input data samples $X1$ and $X2$ are passed through CNN_1 and CNN_2 respectively. Both $X1$ and $X2$ can either be seizure (S) or non-seizure (NS) EEG sample. After going through a series of convolution blocks followed by a fully connected layer with sigmoid activation function, the feature vectors F_X1 and F_X2 (for data points $X1$ and $X2$, respectively) are extracted to perform similarity computation. For EEG samples with the same label (both S or both NS), we expect the features of the two EEG samples to be more similar, while we expect that EEG samples with different labels (one S and one NS) should have dissimilar features.

After the feature vectors are generated, the L1 distance (also called Manhattan distance) [30] between the two feature vectors is calculated. Contrary to regular CNNs used for standard classification, a conjoined module (denoted as "*Similarity measure*" module in Fig. 2) is utilized to compute the similarity between the outputs of the two convolutional branches. In this module, a fully-connected layer with sigmoid

activation function flattens the feature maps into a single vector. Then the result is passed through a fully connected layer with the sigmoid function which outputs a similarity score between 0 and 1. Note that during test (classification), $X2$ are known or reference S/NS samples in our database and $X1$ is what we want to classify using the trained Siamese model.

TABLE I: Siamese CNN Architecture and Hyperparameters. Here, KS = Kernel Size, PS = Pool Size, Conv_M = Mth Convolutional layer, Maxpool_N = Nth Maxpooling layer, FC_P = Pth Fully Connected layer.

Layer	Filters	KS	PS	Strides
Conv1 + ReLU	64	3		1
Maxpool1			2	1
Conv2 + ReLU	128	3		1
Maxpool2			2	1
Conv3 + ReLU	128	3		1
Maxpool3			2	1
Conv4 + ReLU	256	3		1
Flatten				
FC_1 + Sigmoid				
Lambda				
FC_2 + Sigmoid				

The architecture of each convolutional neural network along with its hyper-parameters is shown in Table I. The CNN architecture in this work was inspired by the architecture proposed in [15], however, the architecture and hyper-parameters were modified based on our data. Note that fine tuning of the Network's hyper-parameters was performed by a strict hyper-parameter sweep over development dataset. At the beginning, the weights of each convolutional layers are initialized by a set of random variables which have Gaussian distribution with a mean of 0. We also use a normal distribution in order to initialize the biases, which are from both convolutional layers and fully connected layers, yet with a mean of 0.5.

The validation of the model is performed using M-way one-shot learning technique, in which one unseen validation EEG data sample is compared to M different EEG data samples. Out of these M data samples, only one of them belongs to the class of the test data sample. After performing the comparison, we get M similarity scores, $S = S_1, S_2, S_3, \ldots, S_M$. If the model is properly trained, the similarity score of the pair in which both the data samples belong to the same class should be the highest. In this case, the classification is considered as a correct classification; otherwise, it is considered as an incorrect classification. This procedure is repeated for k trials. In our work, we have empirically chosen M = 5 and k = 250.

IV. Experimental Results

A. Dataset

To train and evaluate our proposed model, we used publicly-available CHB-MIT Scalp EEG Database [28]. The EEG was collected from 23 pediatric subjects at the Children's Hospital Boston, with a total EEG duration of 916 h. For each subject, the continuous EEG recording was divided into digitized data, with each file duration between one to four hours. The signals were obtained using international 10-20 electrode placement system, with 23 EEG channels. The sampling rate of the digitized EEG signal was 256 Hz. Although for majority of the subjects the data was recorded using 23 channels, in some subjects it was changed as required. Hence, in this work, we have considered only 18 channels which are common among all patients. To ensure uniformity, we have considered only 1 hour of Non-Seizure data (immediately preceding the seizure). Table II lists the total number of seizures, total seizure duration, minimum seizure duration and maximum seizure duration for every patient.

TABLE II: Patient details

Patient	Number of seizures	Total seizure duration (seconds)	Minimum seizure duration (seconds)	Maximum seizure duration (seconds)
Chb01	7	449	28	102
Chb02	3	175	10	83
Chb03	7	409	48	70
Chb04	4	382	50	117
Chb05	5	563	97	121
Chb06	10	163	13	21
Chb07	3	328	87	144
Chb08	5	924	135	265
Chb09	4	280	63	80
Chb10	7	454	36	90
Chb11	3	809	23	753
Chb12	27	1016	12	98
Chb13	12	547	18	71
Chb14	8	177	15	42
Chb15	20	2012	32	206
Chb16	10	94	7	15
Chb17	3	296	89	116
Chb18	6	323	31	69
Chb19	3	239	78	82
Chb20	8	302	30	50
Chb21	4	203	13	82
Chb22	3	207	59	75
Chb23	7	431	21	114

B. Evaluation Metrics

The performance of the Siamese CNN for each patient is evaluated using the below metrics:

$$Sensitivity(Sen) = \frac{TP}{TP + FN}$$
$$Specificity(Spec) = \frac{TN}{TN + FP}$$
$$Accuracy(Acc) = \frac{TP + TN}{TP + TN + FP + FN}$$
$$F - Measure(F1) = \frac{2 \times Prec \times Rec}{Prec + Rec}$$

where $Precision(Prec) = \frac{TP}{TP+FP}$, $Recall(Rec) = \frac{TP}{TP+FN}$, TP is the number of true positives, FP is the

TABLE III: Performance of Siamese CNN vs. Baseline CNN (using mean seizure duration). Here, N_{S_train} = Total seizure samples used for training, where each sample = 1 second and total seizure samples = mean seizure duration in seconds (rounded to the nearest integer), N_{S_test} = Seizure samples used for testing, each sample = 1 second.

Patients	N_{S_train}	N_{S_test}	**Siamese CNN**				**Baseline CNN**			
			Sen (%)	***Spec (%)***	***Acc (%)***	***F1 (%)***	***Sen (%)***	***Spec (%)***	***Acc (%)***	***F1 (%)***
Chb01	64	385	84.0	96.0	90.0	89.36	66.67	91.11	78.89	69.75
Chb02	58	117	90.4	99.2	94.8	94.56	78.26	81.52	79.89	82.24
Chb03	58	351	63.2	88.0	75.6	72.14	64.03	63.19	63.61	62.76
Chb04	95	287	100	96.8	98.4	98.42	72.55	75.8	70.31	73.55
Chb05	113	450	85.6	96.8	91.2	90.67	71.43	67.86	69.64	70.42
Chb06	16	147	94.39	64.8	79.6	82.22	64.10	61.54	62.82	63.04
Chb07	109	219	98.4	78.4	88.4	89.45	67.48	75.46	71.47	72.89
Chb08	185	739	89.6	72.0	80.8	82.35	86.52	85.87	86.20	85.47
Chb09	70	210	99.2	98.4	98.8	98.8	86.22	89.78	88.00	87.02
Chb10	65	389	93.6	97.6	95.6	95.51	82.4	96.87	90.75	89.22
Chb11	270	539	97.6	96.8	97.2	97.21	82.9	83.0	86.26	83.91
Chb12	38	978	53.6	78.4	66.0	61.18	59.98	55.1	60.49	59.0
Chb13	46	501	88.0	80.8	84.4	84.94	77.2	73.9	76.43	75.04
Chb14	22	155	68.0	76.8	72.4	71.12	70.13	73.2	73.8	71.53
Chb15	101	1911	88.8	95.19	92.0	91.73	79.25	84.51	81.88	82.15
Chb16	9	85	70.39	67.2	68.8	69.29	69.85	66.84	68.21	67.33
Chb17	99	197	92.0	76.8	84.4	85.5	81.33	78.02	80.56	79.94
Chb18	54	269	92.8	96.0	94.4	94.3	80.16	78	78.19	79.27
Chb19	80	159	93.6	90.4	92.0	92.12	83.09	82.5	84.2	84.69
Chb20	38	264	99.2	96.0	97.6	97.63	86.89	84.42	85.77	85.5
Chb21	51	152	94.39	51.2	72.8	77.63	46.3	54.09	52.6	48.55
Chb22	69	138	96.8	56.8	76.8	80.66	70.1	65.94	70.67	68.02
Chb23	62	369	89.6	72.0	80.8	82.35	66.33	81.95	75.54	69.53
Average	**77.04**	**391.78**	**87.96**	**83.58**	**85.73**	**86.04**	**73.61**	**76.10**	**75.48**	**74.38**

number of false positives, TN is the number of true negatives, and FN is the number of false negatives. In this work, the seizure class is considered as the positive class, and the non-seizure class is considered as the negative class.

C. Experimental Setup

For dataset preprocessing, building the Siamese CNN model, and performing different experiments, we have utilized the Python libraries Scikit-learn, Numpy and Keras. The Siamese CNN model was optimized using Adam optimizer with a learning rate of 0.0005. Binary cross-entropy was used as the loss function. We have trained the model for 5000 iterations with batches of 64 training data pairs passed to the Siamese CNN for each iteration. We have also validated the model with 5-way one-shot learning in every 200 iterations. During the validation process, the accuracy of the model was computed over 250 trials.

D. Results and Discussion

In this work, we have used data of 23 patients from CHB-MIT dataset [28]. For maintaining uniformity, we consider only 1 hour of non-seizure data from each patient. To assess the effectiveness of the proposed algorithm, specificity, sensitivity, accuracy, and F1-measure are used. We trained the Siamese CNN by using m seconds $(9 \leq m \leq 270)$ of seizure for each patient, as listed in second column of Table III, and we tested the model with the remaining seizure data. For each patient, we selected these m seconds of seizure samples arbitrarily from all the seizure data of the patient.

We trained the Siamese CNN in three different scenarios. In the first scenario, for each patient, the Siamese CNN is trained using seizure data whose duration is equal to the *shortest seizure's duration* for that patient. The second and third scenarios are the same as the first one, except that in the second scenario, we used training seizure data whose duration is equal to *the mean of all the seizures* for a specific patient, and in the third scenario, we consider the *longest seizure duration*. The minimum, mean, and maximum seizure durations for each patient are listed in Table II and III. Table III contains the results achieved for each patient in the second scenario (mean seizure duration) mentioned above. Note that as we wanted to demonstrate the effectiveness of our network with small amount of training samples, we intentionally chose the training seizure sample duration to be equal to the mean seizure duration for each patient.

From Table III, we observe that our Siamese CNN learns from on an average of 77.04 seconds of seizure data, and achieves an average sensitivity, specificity, accuracy and F1-measure value of 87.96%, 83.58%, 85.73% and 86.04% respectively, with 5-fold cross-validation. We also compare our results with a baseline CNN, which has the same architecture as each branch of the Siamese CNN. 5-fold cross validation was utilized in this case as well. Table III includes the comparative results of Siamese CNN and the baseline CNN. The average Sensitivity, Specificity, Accuracy and F1-measure values for all 23 patients for the baseline CNN are 73.61%, 76.10%,

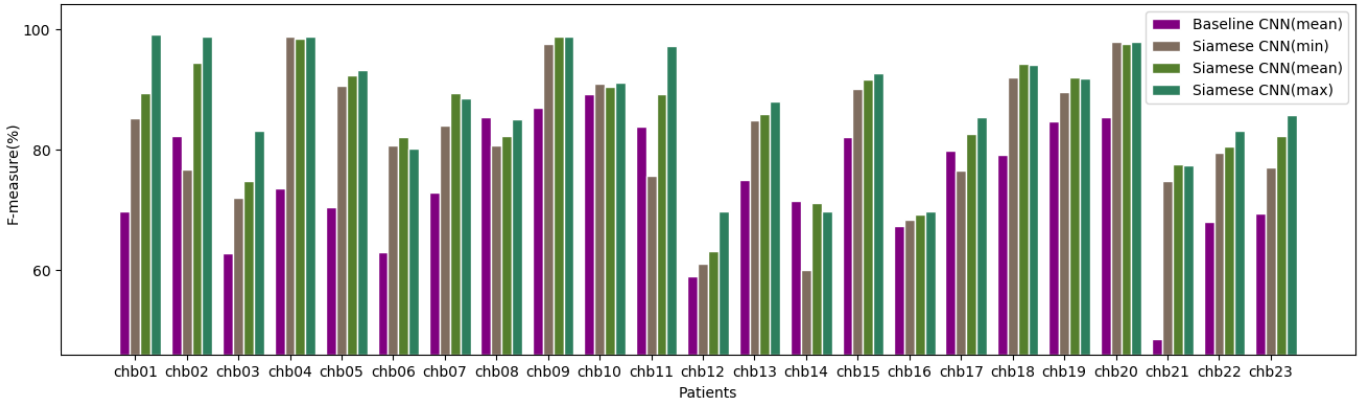

Fig. 4: Effect of increasing the training data on F1-measure

75.48%, and 74.38%, respectively. While in 21 out of 23 cases, the Siamese CNN outperformed the baseline CNN, for patients Chb08 and Chb14, the Siamese CNN performed slightly lower than the baseline CNN. The pattern of EEG signal can be different for different types of seizures, and all these seizure types need to be learned by the Siamese CNN for better classification accuracy. Hence, we hypothesize that lack of similar types of seizure resulted in the slightly lower F1-measure value in these two patients while classifying using the Siamese CNN. Based on Table III, the Siamese CNN has proven itself to be capable of producing excellent classification performance using substantially small amount of training data.

TABLE IV: Effect of number of neighbors considered for 1D-LBP code generation

Patients	**F1-measure (%)**		
	6-bit 1D-LBP	***8-bit 1D-LBP***	***10-bit 1D-LBP***
Chb01	90.61	89.36	85.94
Chb02	97.16	94.56	91.84
Chb07	86.83	89.45	83.21

Effect of Increasing Training Data: Fig. 4 shows the effect of increasing training data size on F1-measure. In general, as the number of training samples increases, the F1-measure expectedly increases. For some patients, nonetheless, increasing the number of training samples results in no further improvement. Note that the maximum seizure duration for each patient can vary significantly. For instance, for Chb16, the maximum seizure duration is 15 seconds, whereas for Chb08, the maximum seizure duration is 265 seconds. To see the impact for various cases, we decided to intentionally consider only one seizure for training the Siamese CNN for each patient; thus, we did not increase the training data size even if the maximum seizure duration for a specific patient is very short (such as the case for Chb16).

Impact of Number of Neighbors: For three randomly chosen patients (Chb01, Chb02 and Chb07), we also performed an experimentation to show the effect of increasing and decreasing the number of neighbors that we consider for generating the LBP codes. The results are shown in Table IV. We can see that, for Chb01 and Chb02, 6-bit 1D-LBP yields the highest F1-measure compared to 8-bit and 10-bit 1D-LBPs, and for Chb07, 8-bit 1D-LBP yields the highest F1-measure. For 10-bit 1D-LBP, the training time becomes considerably higher as the feature set becomes exponentially larger.

We also see decrease in the F1-measures as the number of neighbors (N) considered is increased. One possible explanation can be that as we increase the number of neighbors to generate the LBP codes, more complex local patterns are introduced due to the non-stationary characteristics of EEG data, which makes it more challenging for the Siamese CNN to learn the similarity metric. Moreover, considering the inherent noise scale in EEG signals, having a larger number of neighbors to compute the 1D-LBP only captures and mixes the impact of noise, which can also degrade the performance of the model. Hence, we decided to limit the computation to $N \leq 8$ neighbors to generate the 1D-LBP. For small number of neighbors (e.g. 6 neighbors), smaller and less complicated patterns are captured. As a result, we might end up having similar patterns for both Seizure and Non-seizure EEG samples, which reduces the discriminative capability of LBP. Therefore, even though for some patients, 6-bit 1D-LBP yields better F1-measures, compared to its 8-bit and 10-bit counterpart in our work, we decided to stay with 8-bit 1D-LBP as proposed by other researchers [26].

Effect of Using All Training Data: For 5 randomly selected patients, we also conducted another experiment to show how Siamese CNN performs compared to the baseline CNN if we use all the seizure samples for each patient. From Table V, we can see that while the F1-measure of the baseline CNN has improved significantly for all the patients, for the Siamese CNN we haven't observed significant improvement in the F1-measure. The reason is that due to the dual input

technique, the number of possible training pairs becomes exponentially large, and the training time becomes extremely long if we want to train the Siamese CNN on all possible training pairs. In general, the performance improvement is very small (1-4% absolute increase in F1-measure) and may not justify huge increase in training time (up to 18 times more) of Siamese CNN.

TABLE V: Effect of using all seizure data on F1-measure. Here, N_{S_train} = Seizure data samples (seconds) used for training, B_{CNN} = Baseline CNN, S_{CNN} = Siamese CNN, and S_{CNN_mean} = Siamese CNN trained with mean seizure duration of corresponding patient.

Patients	**F1-measure (%)**			
	N_{S_train}	B_{CNN}	S_{CNN}	S_{CNN_mean}
Chb01	449	92.43	93.82	89.36
Chb07	328	85.21	90.47	89.45
Chb14	177	76.68	72.98	71.12
Chb19	239	93.02	93.51	92.12
Chb22	207	94.61	83.23	80.66

CONCLUSION

Learning from imbalanced data (i.e. very limited seizure samples) has always impeded the performance of machine learning and deep learning techniques in the epileptic seizure detection task. In this work, we have proposed an efficient methodical approach for classification of epileptic seizures using very limited seizure data, by constructing a Siamese convolutional nueral network and combining 1D local binary pattern with it. We have used CHB-MIT Scalp EEG database [28] to train and evaluate the Siamese CNN. The performance of the Siamese CNN was evaluated by conducting a binary classification of seizure and non-seizure EEG samples. The efficiency of the Siamese CNN was further demonstrated by comparing it with a baseline CNN. The promising results show that the Siamese CNN can successfully learn from very limited seizure samples and produce significant improvement in the classification performance.

REFERENCES

[1] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007.

[2] "Epilepsy," https://www.who.int/news-room/fact-sheets/detail/epilepsy, 2019, [Online; accessed 20-Jun-2019].

[3] D. Schmidt and M. Sillanpää, "Evidence-based review on the natural history of the epilepsies," *Current opinion in neurology*, vol. 25, no. 2, pp. 159–163, 2012.

[4] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in biology and medicine*, vol. 100, pp. 270–278, 2018.

[5] I. Ullah, M. Hussain, H. Aboalsamh *et al.*, "An automated system for epilepsy detection using eeg brain signals based on deep learning approach," *Expert Systems with Applications*, vol. 107, pp. 61–71, 2018.

[6] W. Stacey, M. Le Van Quyen, F. Mormann, and A. Schulze-Bonhage, "What is the present-day eeg evidence for a preictal state?" *Epilepsy research*, vol. 97, no. 3, pp. 243–251, 2011.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[8] R. C. Holte, L. Acker, B. W. Porter *et al.*, "Concept learning and the problem of small disjuncts." in *Proc. Int'l J. Conf. Artificial Intelligence*. Citeseer, 1989, pp. 813–818.

[9] S. Ertekin, J. Huang, L. Bottou, and L. Giles, "Learning on the border: active learning in imbalanced data classification," in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 2007, pp. 127–136.

[10] G. M. Weiss, "Mining with rarity: a unifying framework," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 7–19, 2004.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[12] S. Harada, H. Hayashi, and S. Uchida, "Biosignal generation and latent variable analysis with recurrent generative adversarial networks," *CoRR*, vol. abs/1905.07136, 2019. [Online]. Available: http://arxiv.org/abs/1905.07136

[13] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional lstm-cnn generative adversarial network," *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[14] M. S. Munia, M. Nourani, and S. Houari, "Biosignal oversampling using wasserstein generative adversarial network," in *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, 2020, pp. 1–7.

[15] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2. Lille, 2015.

[16] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems*, pp. 737–737, 1994.

[17] A. Mehmood, M. Maqsood, M. Bashir, and Y. Shuyuan, "A deep siamese convolution neural network for multi-class classification of alzheimer disease," *Brain sciences*, vol. 10, no. 2, p. 84, 2020.

[18] Z. Alaverdyan, J. Jung, R. Bouet, and C. Lartizien, "Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: application to epilepsy lesion screening," *Medical image analysis*, vol. 60, p. 101618, 2020.

[19] A. Patane and M. Kwiatkowska, "Calibrating the classifier: siamese neural network architecture for end-to-end arousal recognition from ecg," in *International Conference on Machine Learning, Optimization, and Data Science*. Springer, 2018, pp. 1–13.

[20] S. Shahtalebi, A. Asif, and A. Mohammadi, "Siamese neural networks for eeg-based brain-computer interfaces," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 442–446.

[21] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[22] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[23] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*. Springer, 2004, pp. 469–481.

[24] V. Takala, T. Ahonen, and M. Pietikäinen, "Block-based methods for image retrieval using local binary patterns," in *Scandinavian conference on image analysis*. Springer, 2005, pp. 882–891.

[25] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer vision and image understanding*, vol. 115, no. 4, pp. 541–558, 2011.

[26] N. Chatlani and J. J. Soraghan, "Local binary patterns for 1-d signal processing," in *2010 18th European Signal Processing Conference*. IEEE, 2010, pp. 95–99.

[27] K. Schindler, H. Gast, M. Goodfellow, and C. Rummel, "On seeing the trees and the forest: Single-signal and multisignal analysis of periictal intracranial eeg," *Epilepsia*, vol. 53, no. 9, pp. 1658–1668, 2012.

[28] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[29] L. Wang and D.-C. He, "Texture classification using texture spectrum," *Pattern recognition*, vol. 23, no. 8, pp. 905–910, 1990.

[30] "Taxicab geometry," https://en.wikipedia.org/wiki/Taxicab_geometry, 2021, [Online; accessed 17-Feb-2021].