Detection of Epileptic Seizures Based-on Channel Fusion and Transformer Network in EEG Recordings

José Yauri*[1], Manuel Lagos[2], Hugo Vega-Huerta[3], Percy De-La-Cruz-VdV[3]
Gisella Luisa Elena Maquen-Niño[4], Enrique Condor-Tinoco[5]
Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain[1]
Dpt. Mathematics and Physics, Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, Peru[2]
Dpt. Computer Science, Universidad Nacional Mayor de San Marcos, Lima, Peru[3]
Dpt. of Computing and Electronics, Universidad Nacional Pedro Ruíz Gallo, Lambayeque, Peru[4]
Dpt. Engineering and Information Technology, Universidad Nacional José María Arguedas, Apurimac, Peru[5]

***Abstract*—According to the World Health Organization, epilepsy affects more than 50 million people in the world, and specifically, 80% of them live in developing countries. Therefore, epilepsy has become among the major public issue for many governments and deserves to be engaged. Epilepsy is characterized by uncontrollable seizures in the subject due to a sudden abnormal functionality of the brain. Recurrence of epilepsy attacks change people's lives and interferes with their daily activities. Although epilepsy has no cure, it could be mitigated with an appropriated diagnosis and medication. Usually, epilepsy diagnosis is based on the analysis of an electroencephalogram (EEG) of the patient. However, the process of searching for seizure patterns in a multichannel EEG recording is a visual demanding and time consuming task, even for experienced neurologists. Despite the recent progress in automatic recognition of epilepsy, the multichannel nature of EEG recordings still challenges current methods. In this work, a new method to detect epilepsy in multichannel EEG recordings is proposed. First, the method uses convolutions to perform channel fusion, and next, a self-attention network extracts temporal features to classify between interictal and ictal epilepsy states. The method was validated in the public CHB-MIT dataset using the k-fold cross-validation and achieved 99.74% of specificity and 99.15% of sensitivity, surpassing current approaches.**

Keywords*—Epilepsy; epilepsy detection; EEG; EEG channel fusion; convolutional neural network; self-attention

I. Introduction

Epilepsy is a neurological disease that disturbs the normal functionality of the brain [1]. Epilepsy provokes sudden seizures in the subject, which go from subtle loss of gaze to violent convulsions of the body and extremities, often jointly with fainting, salivation up to the subject's unconscious [2]. Epileptic seizures are produced by a sudden abnormal activity of neurons. The cause that fires such abnormality is still unknown [3]. Recurrence of seizures disrupt the patient's daily activity and damages his personal life, even acquiring additionally psychological illness, such as depression, anxiety, and schizophrenia [4]. Furthermore, patients with epilepsy are often excluded and stigmatized by society [5].

Epilepsy can affect any people without condition of age, gender, or race [6]. According to the World Health Organization (WHO) [7], there are more than 50 million of people suffering of epilepsy around the world, and 80% of patients live in low income countries, facing difficulties in accessing medical services and treatments in order to alleviate the undesired symptoms of epilepsy [8]. As a results, epilepsy has become in a public health problem for many governments and it deserves to make efforts to improve the quality of life of millions of patients with epilepsy [9].

Since the invention of the electroencephalogram (EEG) in 1929, EEG has widely used to study the brain functionality and its associated diseases [10]. Thereby, EEG has become in the standard medical device to detect and diagnose epilepsy due to its easy to use, non-invasive nature, non-age restriction, and real-time sensing features [1]. An EEG records the electric potential generated by neurons while interacting with each others. To do that an EEG employs an array of electrodes which are tied to the head scalp. As a result, an EEG recording provides multiple time-varying signals, one signal for each electrode [1], [2], [11]. As illustration, Fig. 1 shows an EEG recording of three signals from three electrodes.

In order to diagnose epilepsy using EEG, the EEG recording of a patient is analyzed by the neurologist, who performs a visual exploration of signals and searching for spikes, sharp, and slow wave patters that characterize an epileptic seizure [12]. However, epileptic disease can vary widely and shows a wide range of symptoms in patients. As a result, the traditional visual analysis of EEG recordings to diagnose epilepsy is a time-consuming process and quite prone to misdiagnose [13], [14]. Misdiagnosis may lead to maltreatment with undesirable consequences for the patient [15]. Thus, a proper diagnosis of epilepsy is very important in order to provide proper treatment.

Over the past decade, many studies have been carried out with the aim of developing automatic systems for the detection of epilepsy [16], [17], [18] and towards the prediction of seizure episodes [19], [20], [21], [22]. Most studies exploit machine learning (ML) and deep learning (DL) algorithms to build classification models capable of detecting seizure patterns in EEG records. While ML employs hand-crafted features, DL has the capability to learn automatically a rich set of features from training data, offering a more flexible feature space for modeling [23].

Despite the recent advances, detecting epileptic seizures still defies current methods and there are still many unresolved problems. This work addressees two major questions that are

*Corresponding authors

stated below:

1) An EEG device has several electrodes for more accurate medical diagnosis. Although there are methods for merging multiple EEG signals, the method that best suits for an optimal combination of EEG information coming from multiple electrodes is still undefined.
2) The epileptic seizure detection is an unbalanced classification problem in essence, with long hours of normal states (or non-seizure episodes) and a few seconds of abnormal states (seizure episodes). Most existing approaches use certain sampling criteria to balance the number of samples in the training and testing set; however, the effect of using the full dataset on classification performance remains unknown.

So, the main contribution of this study is twofold:

1) An improved EEG channel fusion method for an optimal combination of information from multiple EEG signals, while increasing the classification performance. The proposed classification model firstly uses convolutions to merge EEG channels and increase the representativeness of input signals. Next, a self-attention transformer extracts temporal features of the fused signal to improve the classification performance.
2) The use of a data augmentation method and a weighted loss function that enables the use of large and unbalanced EEG datasets, while improving classification performance.

The remainder of this paper is organized as follows. Section II exposes the background about epilepsy, as well as, the related work. Section III summarizes the methods employed to detect epileptic seizures. Section IV presents the results achieved and provides an analysis of the results compared to previous work. Finally, Section V enlists the findings of this work and the forthcoming investigations.

II. Background

The EEG is the standard device to detect and diagnose epilepsy and other brain diseases [1]. An EEG records the electrical activity of the brain for a certain interval of time (minutes, hours, days) and results in a recording file for further visual analysis by the neurologist [24]. To overcome the time-consuming and visual demand process of the traditional analysis of EEG recordings, many automatic methods have been proposed; being the majority of current methods based on DL algorithms [16], [17], [18]. In this way, the detection of epileptic seizures is commonly stated as supervised classification problem of two classes or binary classification [25], [23].

In order to build a classification model, an enough EEG data should be available. However, typically, researchers do not use all EEG data to avoid unbalance between classes and to reduce the computational burden. Instead, they use specific segment of signals from an EEG recording as training. In such manner, it is common that selected segments correlate the phases of epilepsy experienced by the patient. According to the process of epileptogenesis [19], [26], [1], a patient faces four phases of epilepsy: interictal, preictal, ictal, and postictal. The ictal phase is the seizure episode or attack episode, and the other phases are located in temporal reference to this state. Thereby, the interictal phase is the state a few hours away of a seizure and is considered as the normal state of the patient; the preictal phase is the state of the minutes preceding a seizure; and the postictal phase is the state of the minutes after a seizure. It is worth mentioning that there is still non-consensus in the duration of such stages due to the variability of the epilepsy disease, with the exception of the seizure state [27], [28], [29]. Fig. 1 shows the four phases of epilepsy in an EEG recording of three channels. Note that the seizure (ictal) phase duration is too short in the recording and this makes the EEG data very unbalanced. In addition, because of the diversity of epilepsy among patients, seizure patterns are too diverse and are the main challenge for learning algorithms. Fig. 2 illustrates the seizure segment which is red shaded, showing variability of the signals between EEG channels.

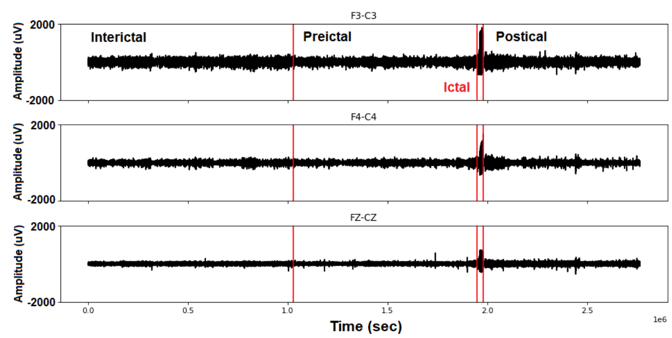

Fig. 1. Epilepsy phases in a long time EEG recording. For convenience, only three channels are plotted.

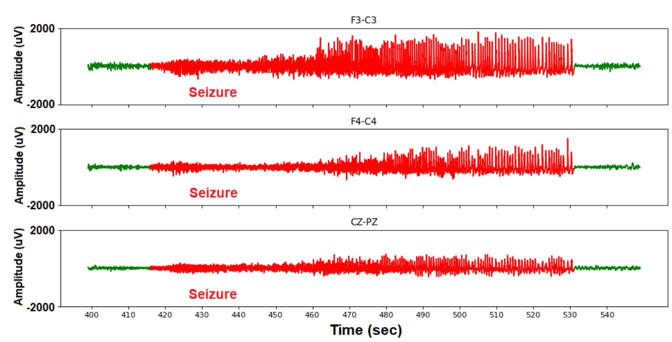

Fig. 2. The epilepsy seizure stage into an EEG recording. The seizure is red shaded, whereas the non-seizure parts are green shaded. For convenience, only three channels are plotted.

Aiming to build a classifier for seizure detection, most researchers use interictal and ictal signals as input data [26], [30], [31], [32], [33], and other investigators use preictal and ictal stages as input data [34], [35], which is also used for researchers that intend to predict a seizure attack [28], [22]. Either using interictal and ictal or preictal and ictal to develop a seizure detector, the classifier is trained to learn how to discriminate between normal and abnormal signals, or non-seizure and seizure segments. On the other hand, tanking into account the classification performance, although the authors have reported high accuracies in their model performance, the majority of the results are not reproducible due to the lack of consensus on the selection of the portion of the signals used for model training and testing.

Dealing with the problem of epilepsy detection, another issue arises during data selection for training a model: the selected signals are too unbalanced because the patients stay many hours normally (interictal phase), but stay just a few seconds of a seizure (ictal phase) [36]. The high unbalance of classes often cripples any learning algorithm [25]. To overcome class imbalance, researchers have proposed undersampling the majority class, oversampling the minority class, and some data augmentation. It is common to find methods which combine majority class undersampling with increasing minority class data [18]. For data augmentation, the method of sliding a window with overlap has provided great results [37], [34], [31] when compared to the generation of new samples by a specialized model [36].

Another problem that hinders epilepsy detection is that EEG recordings are inherently multichannel data. This is because EEG employs an array of electrodes to record the brain activity in many different locations of the head at a given time. So, an EEG recording consists of spatio-temporal sample points recorded by each electrode. For medical diagnostics of epilepsy, EEG headsets with as many electrodes as possible are preferred in order to reach a higher performance in detection, e.g., an EEG with 19 electrodes arranged according to the international 10/20 system [38]. On the other hand, for non-medical applications, an EEG with fewer electrodes is enough [39], e.g, mental fatigue detection in drivers.

In medical diagnosis using EEG, simultaneous EEG signals increase the visual effort of the neurologist and make it prone to misdiagnosis. With the goal to develop robust automatic system for epilepsy detection, the multichannel issue, also named spatial filtering [40], should be addressed. While some researchers have searched for the most discriminative channel that allows the best classification, a few researchers have proposed a specific method to combine multichannel EEG signals. The former strategy consists of evaluating EEG channels, one by one, and selecting the channel that provides the best performance [20]. This procedure might be slow and the trained model relies heavily on the domain of application, e.g., epilepsy detection [41], [42], mental fatigue detection [43], and active brain computer interfaces (BCI) [40]. The latter strategy consists of designing a specific method that carries out the combination of multiple EEG signals. The main advantage of these methods is that they are more scalable and independent of the application domain [35].

More specifically, although the combination of EEG channels might be performed using learnable DL-based models, there are some mathematical transformations to merge multiple EEG channels into single channel, such as the common spatial pattern (CSP) [44] and the Choquet fuzzy integral [45]. Actually, the CSP is still widely used and actively studied to overcome the limitation of the original CSP [46], [47], [33].

On the other hand, some researchers have leveraged the latest developments in DL and have proposed methods to discriminate between non-seizure and seizure segments while combining multiple EEG channels. Usually, these methods are based on convolutional neural networks (CNN) and long short-term memory (LSTM) neural networks, which are able to learn both spatial and temporal features from training data [48], [34], [35]. As a results, different DL-based models have been proposed for epilepsy detection, such as based only in CNN models [26], [31], only in LSTM models [35], or CNN-LSTM hybrid models [42], including the plethora of LSTM variations like the bidirectional-LSTM and nested-LSTM [49], [50], [51]. Recently, the self-attention transformer [52] has also been introduced to classify EEG signals due to its ability to capture long-term temporal dependencies analogously to the LSTM network [53]. In spite of recent advances, the question of how to combine several EEG signals whilst increasing classification performance remains unanswered.

Because our approach proposes to take advantage of recent advances in DL, this work has looked at the latest approaches focused on detecting epilepsy using DL and the largest EEG database CHB-MIT [54]. As follows, we sum up the most important related works which have established the current state of the art (SoA). Furthermore, this study takes into account the works that employ interictal and ictal signals as input data due to two reasons. First, interictal and ictal signals are used by the majority of studies as input data source. Second, interictal and ictal segments seem the correct way to discriminate between normal and abnormal states of an epileptic patient due to higher performance that it provides rather than other sources of data.

One of the first works that applied DL towards seizure detection is the work of Zhou et al. [26]. The authors use CNN to detect seizure at a level of patient. The model consists of a 2D convolution layer, an activation function, a 2D max-pooling, and followed by a fully connected (FC) layer for classification. The input data is extracted from the interictal and ictal signals of epilepsy, and next, they were split into time windows of 1 sec. Two experiments were carried out using two different data sources: using time-domain signals and using 2D-spectrogram. Spectrogram is computed using the fast Fourier transform (FFT) for each time window and its channels, and next, they are concatenated along the depth. Assessing the model performance in the CHB-MIT dataset, the use of spectrogram outperforms the time-domain input data, 97.5% against 62.3% of accuracy, respectively. However, because the model is too simple, the high gain may result from data preparation rather than from the data source used (e.g., spectrogram images are normalized between 0 and 1, while time-domain signals not). Also, no information is provided about the data selection process, neither about the treatment of unbalancing between interictal and ictal samples. In contrast, recent approaches mainly use time domain signals, but increasing the model complexity.

Then, Hossain et al. [34] proposed a specialized EEG channel fusion layer before temporal feature extraction for seizure detection. In a cross-patient scheme, input data belongs to preictal and ictal stages, which are split into time windows of 2 sec, 80% overlap. The model contains four CNN blocks. Each block consists of a convolution, an activation function, and a max-pooling. However, the first block is slightly different: first, a convolution operates along time, and next, a convolution operates along channels, both performing feature extraction, then an activation function is applied. Assessing in the CHB-MIT dataset, the model achieved a sensitivity of 90%, a specificity of 91.65%, and an accuracy of accuracy of 98.05%. Despite the reported high performance, the data selection is unclear and and number of seizures recognized is unknown.

Later, Gao et al. [55] proposed to classify image spectrogram of EEG signals by using transfer learning. The authors stated a classification problem of four classes: interictal (selected from two hours away from the ictal), preictal I (selected from 30 min before the ictal), preictal II (chosen 10 min before the seizure), and ictal. First, only 11 patients where selected from the CHB-MIT dataset. Then, signals are cleaned using the discrete wavelet transform (DWT) and split into time windows of 4 sec. Next, time-domain signals are converted to spectrogram images. A data augmentation of ictal signals is employed using a sliding window with 50% overlap. The authors use three pretrained models from image classification task, Inception-ResNet-v2, Inception-v3, and ResNet152, whose outputs are fed to two FC layers of 1,024 and 512 neurons that are used for classification. Validation is performed in a hold-out cross-validation, 70:30, achieving a sensitivity of 95.8% and a specificity of 99.3% detecting the ictal state.

Then, Li et al. [30] proposed a hybrid architecture of CNN and nested LSTM networks. The model consists of three 1D-CNN layers and 100 nested cells of LSTM. Data was carefully selected from interictal and 135 seizures, which are split into time window of 4 sec. Each time window is reshaped in such a way that EEG channels are as features to be processed by the 1D-CNN, and then, after feature extraction, output features are fed to a FC layer of 50 neurons before classification. The model achieved 95.42% of sensitivity, 95.29% of specificity, and 95.29% of accuracy in a 10-fold cross validation. Although the achieved metrics are higher, this is because seizures to be detected have been carefully selected and reduced to 135.

Next, Wang et al. [31] proposed to classify interictal and ictal signals using only a 1D-CNN model. Selected signals is split into time windows of 2 sec and a data augmentation is applied just to ictal segments with 50% overlap. The model architecture consists of two CNN heads, like an ensemble, whose outputs are fed into two FC layers of 256 and 128 neurons before classification. The model validation is performed in a k-fold cross validation scheme, but at level of seizures in the dataset. Working with 145 seizures, the model achieved in average 88.14% of sensitivity, 99.62% of specificity, and 99.54% of accuracy. It is noticeable that 1D CNN alone does not provide a reliable sensitivity to detect epileptic seizures.

Later, Abdelhameed et al. [32] proposed a 2D autoencoder (AE), together with a LSTM network to classify interictal and ictal signals. Data from 16 patients are selected according to the age criterion within the CHB-MIT dataset. Next, the whole dataset is standardized at once, and later, data is split into time window of 4 sec. Then, each window is normalized to 0-1 to ensure reconstruction by the AE. The AE module consists of four layers of conventional 2D-CNN for encoding and decoding. The classification module employs the latent encoded vector as input data and consists of a LSTM network, followed by a FC layer of 256 neurons. The method was assessed in a 10-fold cross validation, achieving 98.72±0.77% of sensitivity, 98.86±0.53% of specificity, and 98.79±0.53% of accuracy. Despite reported the performance is too high, the standardization of the whose dataset before data splitting is not according to ML practices [25].

As exposed above, CSP still is used for EEG channel fusion and the proposal of Li et al. [33] have reported recently high performances in 5-fold cross validation. The authors used the empirical mode decomposition (EMD) to increase the signal-to-noise ratio before to apply CSP. Next, a support vector machine (SVM) is trained using the variance of signals as input features. The method achieved in average, 97.34% of sensitivity, 97.50% of specificity, and 97.49% of accuracy. Despite the sensitivity is higher, the number of detected seizures is just 131, which have no explanation of their selection criteria.

More recently, the self-attention transformer has been introduced in many areas and is widely used for natural language processing (BERT, GPT-3), image classification (vision transformer-ViT), and others applications [56]. In this way, Pan et al. [53] proposed a transformer model to detect epilepsy. Dataset is prepared as non-epilepsy and epilepsy segments, and next, data is split into time windows of 4 sec, with 50% overlap. Also, to ensure balanced samples, an undersampling of the majority class is carried out. Finally, only three EEG channels are fed to the transformer encoder, whose outputs are send to a FC layer for classification. Evaluation in a 5-fold cross validation, the model achieved 94.96% of sensitivity, 93.97% of specificity, and 94.46% of accuracy. Despite the high performance, the authors do not provide sufficient information about what EEG channels are use as input data, and which phase of epilepsy is considered as non-epilepsy and how is trained the model with a fair five thousand samples of each class.

It should be mentioned that previous studies have used certain sampling criteria to balance the number of samples in each class, regardless of whether or not they have used any method of data augmentation. However, the joint use of data augmentation and weighted loss function methods has not been fully explored and remains also as an open question.

This work addresses two main issues outlined above: combining EEG channels as well as the joint use of data augmentation and weighted loss functions in order to increase the classification performance towards epileptic seizure detection.

III. Our Approach

Fig. 3 presents the general pipeline to detect epileptic seizures in EEG recordings. First, a brief description of the dataset is furnished. Then, the preprocessing methods used are described. Next, the neural network model that performs EEG channels fusion and involves a transformer network is presented. Finally, the classification step is performed and the model performance is validated.

As follows, a deep description of each step of the pipeline is provided.

A. EEG Dataset

The EEG data used in this study comes from the CHB-MIT public dataset [54] and contains almost 980 hours of EEG recordings and 198 seizures. The dataset was collected from 23 pediatric patients with incurable epilepsy, 3–22 age. The

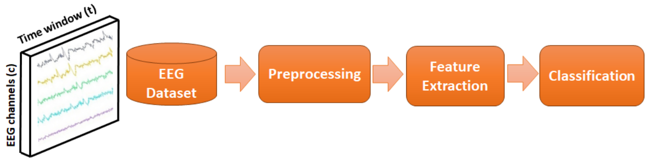

Fig. 3. The epileptic seizure detection pipeline.

recording chb21 was obtained 1.5 years later from the same patient chb01, and because seizure patterns are different, this is treated as a new patient's recording.

Recordings are stored at the sampling frequency of 256 Hz and different EEG devices with different number of electrodes/channels were employed during recording of data, however, EEG recordings of 23 channels are the most common in the dataset. To release the dataset, longtime recordings were usually split into one hour long recording, and sometimes into two or four hours long recording. EEG recordings that contain seizures are referred to as seizure records; otherwise, non-seizure records. As ground truth (GT), the dataset provides the start and end for each seizure in the seizure record.

B. Preprocessing

In this stage, two main processes are performed: data selection and data windowing.

The former process, data selection, aims to select the EEG recordings and their signals to be used to train a model. In previous studies [16], [17], [18], researchers have found that interictal and ictal signals are the best ones to discriminate between NON-SEIZURE and SEIZURE sample data (see Fig. 1 to illustrate about these epilepsy phases). In addition, researchers have used recordings of 23 channels to overcome the diversity of EEG montages in the CHB-MIT dataset. In this work, we also use interictal and ictal signals among the EEG recordings of 23 channels which involve 181 seizures. Interictal data consist of signals two hours away of a seizure (we name non-seizure signals). On Further, ictal data consists of each seizure signals from all patients (we name seizure signals). Moreover, to reduce computations, signals were downsampled to 128 Hz because it does not affect the classification performance [57]. Moreover, no filtering technique is used as in previous studies [26], [32], [20].

The latter process, data windowing, aims to split the selected EEG segments into small processable time windows. A time window is the sampling unit and is used as input data of the model. This work uses a time window of 1 sec. Besides, in order to mitigate the unbalancing between non-seizure and seizure samples, data augmentation is applied to seizure signals. The data augmentation strategy consists in sliding a window with 80% overlap. Fig. 4 delineates the approach of data splitting and data augmentation employed in this work.

C. Neural Network Architecture

Fig. 5 depicts the proposed neural architecture for epilepsy detection. The neural model receives the time windows as input data. Input data is in time-domain window. Then, it performs

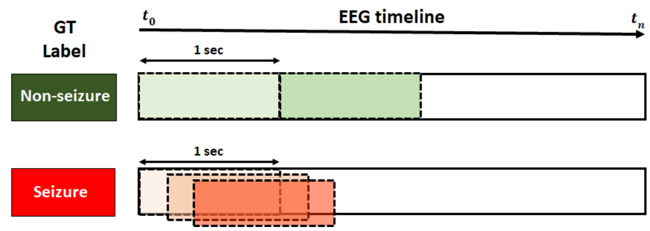

Fig. 4. Data windowing of non-seizure and seizure signals. The image in the top row illustrates the splitting of non-seizure data. The image in the bottom row shows the simultaneous splitting and augmentation of the seizure data.

EEG channel fusion and extracts features for classification. Finally, the model predicts outputs in the form of two categorical data: either non-seizure (interictal) or seizure (ictal) class.

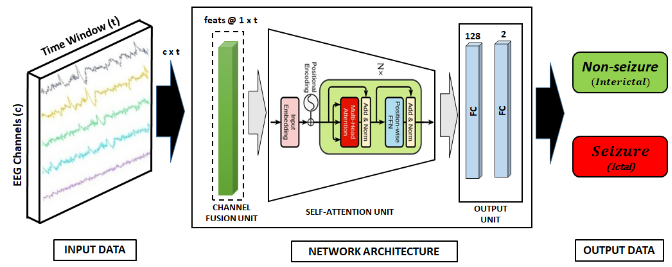

Fig. 5. The neural network architecture.

In brief, the neural network consists of three basic units. First, the Channel Fusion Unit fuses the information coming from different EEG channels into a single channel signal. Second, the Self-attention Unit extracts temporal features based on the previous single channel signal. Third, the Output Unit combines the learned representations for a successful classification.

In this work, the Channel Fusion Unit is inspired by the study of Hossain et al. [34], however, with a slight variation. In our case, once EEG channels are combined, two additional convolutions still refine the spatial features using a small kernel size to diminish the computation burden. The Self-attention Unit is based on the classical transformer architecture proposed by Vaswani et al. [52] and processes the enhanced single channel signal to learn long temporal dependencies of the signal. Ultimately, the Output Unit consists of two layers of fully connected neurons and performs classification.

Table I details the neural network architecture and its parameters. The input data consists of a 2D matrix $X^{c \times t}$, where c is the number of EEG channels (23) and t is the number of time points in the time window (128).

D. Classification

After model training, it is capable to classify non-seizure versus seizure sample units. To predict the label class, the latest layer of the model employs the Softmax activation function to estimate the probability distribution of the input data [25].

TABLE I. DETAILS OF THE PROPOSED NEURAL NETWORK.

Unit	Layer	Output size	Parameters
-	Input	23x128	-
Channel Fusion Unit	Conv 11	23x128x16	kernel 1x3, map 16, None
	Conv 12	1x128x16	kernel 23x1, map 16, BN, Relu
	Conv 2	1x64x256	kernel 1x3, map 256, BN, Relu Maxpool
	Conv 3	1x60x256	kernel 1x3, map 256, BN, Relu
Self-attention Unit	Transformer	60x256	n_heads=8, n_layers=4 d_model=256
	AvgPool	1x256	-
Output Unit	FC 1	128	ReLU, Dropout=0.5
	FC 2	2	Softmax

E. Experimental Design

In order to evaluate the feasibility of the proposed neural architecture and to compare its performance with related work, the model is validated using the k-fold cross validation (k=5). The k-fold cross validation was widely employed in previous studies to validate their model performance [16], [17], [18]. However, to ensure a fair assessment and comparison of model performance, in addition to the accuracy, the sensitivity, specificity, precision, and F1-score should be used as validation metrics, because detecting epilepsy is an extremely unbalanced classification problem [25].

IV. RESULTS AND DISCUSSION OF RESULTS

The proposed model is implemented in the Python 3.9 environment and the Pytorch 1.13 deep learning framework and runs in a computer desktop with a GPU NVIDIA GeForce RTX 2070 Super. The hyper-parameters used to train the model are: the Adam optimizer, the cross-entropy loss function which should use normalized weights depending on the proportion of samples from each class in the training set (the larger the number of samples of the class, the smaller the weight, and conversely), the batch size of 128, the learning rate of 1e-4, and 150 number of epochs for model training.

The model performance is assessed using the 5-fold cross validation, and the achieved results are shown in Table II. The results follow the format of the average plus/minus the standard deviation. The proposed model achieved 99.74 ± 0.08 of sensitivity and 99.15 ± 0.1 of specificity detecting epilepsy patterns, with high precision and great F1-score.

TABLE II. CLASSIFICATION PERFORMANCE OF THE PROPOSED MODEL USING 5-FOLD CROSS VALIDATION IN THE CHB-MIT DATASET.

Classifier	Sensitivity	Specificity	Precision	F1-score	Accuracy
This work	99.74±0.08	99.15±0.1	97.66±0.71	98.4±0.32	99.68±0.06

To ensure an equitable and fair comparison of our achieved results against related work, we selected the most recent SoA methods that use interictal and ictal signals as input data and validate their results using the k-fold cross-validation scheme. Table III summarizes the performance reported by several SoA studies.

On the other hand, there are studies that have presented some specific EEG channel fusion approaches, like Hossain et al. [34] and Chakrabarti et al. [35] for epilepsy detection using preictal and ictal signals, and the work of Gao et al. [58] for fatigue detection in drivers. Because the source of data to train and test their models differs from ours, we implemented such models and trained them using their own suggested hyperparameters for a fair comparison of performance. These studies are highlighted with an * in Table III.

After reviewing Table III, it can be seen that our model is significantly better than models that do not perform channel fusion, being the work of Abdelhameed et al. [32] the only one that comes close to our results, however, the authors just worked with 86 seizures from 16 subjects.

On the other hand, comparing our results against approaches that specifically perform EEG channel fusion, first, it is interesting to observe that the study of Li et al. [33] achieves a higher sensitivity and specificity of almost 97% working with 131 seizures. As this work uses CSP/EMD to fuse EEG channels, we can deduce from this that working with all EEG channels or searching for the best single channel is not good enough for the model, even using the most advanced transformer architecture like in the study of Pan et al. [53]. As a result, fusion of EEG channels before feature extraction can be quite advantageous.

Next, we compare our results against approaches that perform specialized EEG channel fusion based on neural networks. In this way, the study of Hossain et al. [34] implements the channel fusion before the feature extraction, whereas the study of Gao et al. [58] and Chakrabarti et al. [35] implement the channel fusion after feature extraction. Again, it is noted that channel fusion approaches work better than non-channel fusion approaches. Taking into account the achieved performances in descending order, they go from Chakrabarti et al. [35], Gao et al. [58], to Hossain et al. [34]. It seems that EEG channel fusion also is feasible before and after temporal feature extraction. However, our approach, that simultaneously fuses EEG channels and enhance spatial features at input data level before feature extraction, and next, leverages a simple self-attention transformer, outperforms all these approaches and provides the highest sensitivity and specificity to classify between non-seizure and seizure EEG signals.

V. CONCLUSION

In this work, a new approach to detect epileptic seizures has been described. The method is based on a specific channel fusion layer that optimally combines multiple EEG channels into a single channel and enhances spatial features. Then, a simple self-attention transformer is employed to extract temporal features in order to improve the classification performance.

The feasibility of the method was validated in the public CHB-MIT EEG dataset using 5-fold cross validation. In a highly unbalanced dataset and assessing 181 seizures from 24 patients, the proposed model achieved 99.74±0.08 of specificity, 99.15±0.1 of sensitivity, 97.66±0.71 of precision, 98.4±0.32 of F1-score, and 99.68±0.06 of accuracy. Comparing with current SoA methods, the proposed method surpasses them considerably.

In the course of future work, further studies are still needed on new methods of merging EEG channels, especially those

TABLE III. BENCHMARKING THE PERFORMANCE OF THE PROPOSED MODEL AND RELATED WORK IN THE CHB-MIT DATABASE.

Author	Method	Total seizures	Sensitivity	Specificity	Precision	F1-score	Accuraccy
Zhou et al. [26]	2D-CNN	-	-	-	-	-	97.5
Gao et al. [55]	Transfer Learning	-	95.8	99.3	-	-	96.9
Li et al. [30]	CNN-nested LSTM	135	95.42	95.29	-	-	95.29
Wang et al. [31]	1D-CNN	145	88.14	99.62	-	-	99.54
Abdelhameed et al. [32]	AE-2D-CNN - LSTM	86	98.72±0.77	98.86 ± 0.53	98.86±0.53	98.79 ± 0.53	98.79±0.53
Li et al.[33]	CSP/EMD-SVM	131	97.34	97.50	-	-	97.49
Pan et al.[53]	Transformer	-	94.96	93.97	-	-	94.46
Hossain et al. [34] *	CNN	181	91.44±1.32	96.86±0.67	76.17±3.51	83.05±1.62	96.33±0.48
Chakrabarti et al. [35] *	LSTM	181	98.29±0.38	99.52±0.07	95.75±0.6	97±0.24	99.4±0.05
Gao et al. [58] *	CNN-FC	181	97.7±0.18	99.49±0.04	95.43±0.32	96.55±0.12	99.32±0.02
This work	Channel Fusion-Transformer	181	99.15±0.1	99.74±0.08	97.66±0.71	98.4±0.32	99.68±0.06

* After training the model architecture from scratch because the original studies use different data sources.

of a linear nature, as they are easier to understand by humans. Furthermore, new data augmentation techniques are needed and generative neural networks may provide an improvement over existing ones.

REFERENCES

[1] G. Cascino, J. I. Sirven, and W. O. Tatum, *Epilepsy*, 2nd ed. Wiley, 2021.

[2] M. Z. Koubeissi and N. J. Azar, *Epilepsy Board Review : A Comprehensive Guide*, 2017.

[3] S. Shorvon, R. Guerrini, and M. Cook, *Oxford Textbook of Epilepsy and Epileptic Seizures.* OUP Oxford, 2012.

[4] M. Gandy, A. C. Modi, J. L. Wagner, W. C. LaFrance, M. Reuber, V. Tang, K. D. Valente, L. H. Goldstein, K. A. Donald, G. Rayner, and R. Michaelis, "Managing Depression and Anxiety in People with Epilepsy: A Survey of Epilepsy Health Professionals by the ILAE Psychology Task Force," *Epilepsia Open*, vol. 6, no. 1, pp. 127–139, 2021.

[5] N. F. Bandstra, C. S. Camfield, and P. R. Camfield, "Stigma of Epilepsy," *Canadian Journal of Neurological Sciences*, vol. 35, no. 4, pp. 436–440, 2008.

[6] J. G. Burneo, L. Black, R. Martin, O. Devinsky, S. Pacia, E. Faught, B. Vasquez, R. C. Knowlton, D. Luciano, W. Doyle, S. Najjar, and R. I. Kuzniecky, "Race/Ethnicity, Sex, and Socioeconomic Status as Predictors of Outcome After Surgery for Temporal Lobe Epilepsy," *Archives of Neurology*, vol. 63, no. 8, pp. 1106–1110, 2006.

[7] World Health Organization, "Epilepsy," Available on line at: https://www.who.int/health-topics/epilepsy, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[8] C. Espinosa-Jovel, R. Toledano, A. Aledo-Serrano, I. García-Morales, and A. Gil-Nagel, "Epidemiological Profile of Epilepsy in Low Income Populations," *Seizure*, vol. 56, pp. 67–72, 2018.

[9] Center for Disease Control and Prevention, "Epilepsy Is a Public Health Problem," Available on line at: https://www.cdc.gov/epilepsy/communications/infographics/cdc-epilepsy-text.htm, 9 2022. [Online]. Available: https://www.cdc.gov/epilepsy/communications/infographics/cdc-epilepsy-text.htm

[10] R. İnce, S. S. Adanır, and F. Sevmez, "The Inventor of Electroencephalography (EEG): Hans Berger (1873–1941)," *Child's Nervous System*, vol. 37, no. 9, pp. 2723–2724, 2021.

[11] V. S. Wasade and M. V. Spanaki, *Understanding Epilepsy: A Study Guide for the Boards*. Cambridge University Press, 2019.

[12] N. M. Jadeja, *How to Read an EEG*. Cambridge University Press, 2021.

[13] W. O. Tatum, "Mistaking EEG Changes for Epilepsy," in *Common Pitfalls in Epilepsy: Case-Based Learning*, D. Schmidt, W. O. Tatum, and S. Schachter, Eds. Cambridge: Cambridge University Press, 2018, ch. 2, p. 25–42.

[14] U. Amin and S. R. Benbadis, "The Role of EEG in the Erroneous Diagnosis of Epilepsy," *Journal of Clinical Neurophysiology*, vol. 36, no. 4, pp. 294–297, 2019.

[15] E. R. Somerville, "Some Treatments Cause Seizure Aggravation in Idiopathic Epilepsies (especially absence epilepsy)," *Epilepsia*, vol. 50, no. SUPPL. 8, pp. 31–36, 2009.

[16] M. K. Siddiqui, R. Morales-Menendez, X. Huang, and N. Hussain, "A Review of Epileptic Seizure Detection using Machine Learning Classifiers," *Brain Informatics*, vol. 7, no. 1, pp. 1–18, 2020.

[17] J. Prasanna, M. S. P. Subathra, M. A. Mohammed, R. Damaševičius, N. J. Sairamya, and S. T. George, "Automated Epileptic Seizure Detection in Pediatric Subjects of CHB-MIT EEG Database - A Survey," *Journal of Personalized Medicine*, vol. 11, no. 10, p. 1028, 2021.

[18] A. Shoeibi, M. Khodatars, N. Ghassemi, M. Jafari, P. Moridian, R. Alizadehsani, M. Panahiazar, F. Khozeimeh, A. Zare, H. Hosseini-Nejad, A. Khosravi, A. F. Atiya, D. Aminshahidi, S. Hussain, M. Rouhani, S. Nahavandi, and U. R. Acharya, "Epileptic Seizures Detection Using Deep Learning Techniques: A Review," *International Journal of Environmental Research and Public Health 2021, Vol. 18, Page 5780*, vol. 18, no. 11, p. 5780, 2021.

[19] S. M. Usman, M. Usman, and S. Fong, "Epileptic Seizures Prediction Using Machine Learning Methods," *Computational and Mathematical Methods in Medicine*, vol. 2017, p. 9074759, 2017.

[20] S. Toraman, "Automatic Recognition of Preictal and Interictal EEG Signals using 1D-Capsule Networks," *Computers & Electrical Engineering*, vol. 91, p. 107033, 2021.

[21] R. Hussein, S. Lee, and R. Ward, "Multi-Channel Vision Transformer for Epileptic Seizure Prediction," *Biomedicines*, vol. 10, no. 7, p. 1551, 2022.

[22] Z. Altaf, M. A. Unar, S. Narejo, M. A. Zaki, and Naseer-u-Din, "Generalized Epileptic Seizure Prediction using Machine Learning Method," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, 2023.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[24] L. Hu and Z. Zhang, *EEG Signal Processing and Feature Extraction*. Springer Singapore, 1 2019.

[25] K. Murphy, *Machine Learning : A Probabilistic Perspective*. Cambridge: MIT Press, 2012.

[26] 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.

[27] A. Shoeb and J. Guttag, "Application of Machine Learning to Epileptic Seizure Detection," in *ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 975–982.

[28] S. M. Usman, S. Khalid, R. Akhtar, Z. Bortolotto, Z. Bashir, and H. Qiu, "Using Scalp EEG and Intracranial EEG Signals for Predicting Epileptic Seizures: Review of Available Methodologies," *Seizure*, vol. 71, pp. 258–269, 2019.

[29] J. Zhou, L. Liu, Y. Leng, Y. Yang, B. Gao, Z. Jiang, W. Nie, and Q. Yuan, "Both Cross-Patient and Patient-Specific Seizure Detection Based on Self-Organizing Fuzzy Logic," *International Journal of Neural Systems*, vol. 32, no. 6, 2022.

[30] Y. Li, Z. Yu, Y. Chen, C. Yang, Y. Li, X. Allen Li, and B. Li, "Automatic Seizure Detection using Fully Convolutional Nested LSTM," *International Journal of Neural Systems*, vol. 30, no. 4, 2020.

[31] X. Wang, X. Wang, W. Liu, Z. Chang, T. Kärkkäinen, and F. Cong, "One Dimensional Convolutional Neural Networks for Seizure Onset Detection using Long-term Scalp and Intracranial EEG," *Neurocomputing*, vol. 459, pp. 212–222, 2021.

[32] A. Abdelhameed and M. Bayoumi, "A Deep Learning Approach for Automatic Seizure Detection in Children With Epilepsy," *Frontiers in Computational Neuroscience*, vol. 15, p. 29, 2021.

[33] C. Li, W. Zhou, G. Liu, Y. Zhang, M. Geng, Z. Liu, S. Wang, and W. Shang, "Seizure Onset Detection Using Empirical Mode Decomposition and Common Spatial Pattern," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 458–467, 2021.

[34] M. Shamim Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, "Applying Deep Learning for Epilepsy Seizure Detection and Brain Mapping Visualization," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 15, no. 1s, 2019.

[35] S. Chakrabarti, A. Swetapadma, and P. K. Pattnaik, "A Channel Independent Generalized Seizure Detection Method for Pediatric Epileptic Seizures," *Computer Methods and Programs in Biomedicine*, vol. 209, p. 106335, 2021.

[36] B. Gao, J. Zhou, Y. Yang, J. Chi, and Q. Yuan, "Generative Adversarial Network and Convolutional Neural Network-based EEG Imbalanced Classification Model for Seizure Detection," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 1–15, 2022.

[37] Z. Wei, J. Zou, J. Zhang, and J. Xu, "Automatic Epileptic EEG Detection using Convolutional Neural Network with Improvements in Time-domain," *Biomedical Signal Processing and Control*, vol. 53, p. 101551, 2019.

[38] M. Tacke, K. Janson, K. Vill, F. Heinen, L. Gerstl, K. Reiter, and I. Borggraefe, "Effects of a Reduction of the Number of Electrodes in the EEG Montage on the Number of Identified Seizure Patterns," *Scientific Reports*, vol. 12, no. 1, p. 4621, 2022.

[39] J. LaRocco, M. D. Le, and D. G. Paeng, "A Systemic Review of Available Low-Cost EEG Headsets Used for Drowsiness Detection," *Frontiers in Neuroinformatics*, vol. 14, p. 553352, 2020.

[40] N. Tiwari, D. R. Edla, S. Dodia, and A. Bablani, "Brain Computer Interface: A Comprehensive Survey," *Biologically Inspired Cognitive Architectures*, vol. 26, pp. 118–129, 2018.

[41] S. Ammar and M. Senouci, "Seizure Detection with Single-channel EEG using Extreme Learning Machine," in *17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, STA 2016 - Proceedings*. Institute of Electrical and Electronics Engineers Inc., 2017, pp. 776–779.

[42] G. Xu, T. Ren, Y. Chen, and W. Che, "A One-Dimensional CNN-LSTM Model for Epileptic Seizure Recognition Using EEG Signal Analysis," *Frontiers in Neuroscience*, vol. 14, p. 1253, 2020.

[43] P. Li, W. Jiang, and F. Su, "Single-channel EEG-based Mental Fatigue Detection Based on Deep Belief Network," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc., 2016, pp. 367–370.

[44] Y. Wang, S. Gao, and X. Gao, "Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2005. Conf Proc IEEE Eng Med Biol Soc, 2005, pp. 5392–5395.

[45] S. A. Ludwig, "Performance Analysis of Data Fusion Methods Applied to Epileptic Seizure Recognition," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 12, no. 1, pp. 5–17, 2022.

[46] L. Zhang, C. Zhang, H. Higashi, J. Cao, and T. Tanaka, "Common Spatial Pattern Using Multivariate EMD for EEG Classification," in *APSIPA ASC 2011 - Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2011*. APSIPA, 2011, pp. 244–248.

[47] R. Fu, Y. Tian, P. Shi, and T. Bao, "Automatic Detection of Epileptic Seizures in EEG Using Sparse CSP and Fisher Linear Discrimination Analysis Algorithm," *Journal of Medical Systems 2020 44:2*, vol. 44, no. 2, pp. 1–13, 2020.

[48] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[49] D. D. Chakladar, S. Dey, P. P. Roy, and D. P. Dogra, "EEG-based Mental Workload Estimation Using Deep BLSTM-LSTM Network and Evolutionary Algorithm," *Biomedical Signal Processing and Control*, vol. 60, p. 101989, 2020.

[50] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, and Q. Yuan, "Scalp EEG Classification using Deep Bi-LSTM Network for Seizure Detection," *Computers in Biology and Medicine*, vol. 124, 2020.

[51] G. Liu, L. Tian, and W. Zhou, "Patient-Independent Seizure Detection Based on Channel-Perturbation Convolutional Neural Network and Bidirectional Long Short-Term Memory," *International Journal of Neural Systems*, vol. 32, no. 6, 2021.

[52] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All You Need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[53] H. Pan, S. Gong, F. Dong, and L. Jiang, "Epilepsy Seizure Detection using Transformer," *Highlights in Science, Engineering and Technology*, vol. 1, pp. 325–329, 2022.

[54] A. Shoeb and J. Guttag, "CHB-MIT Scalp EEG Database," 2010. [Online]. Available: https://physionet.org/content/chbmit/1.0.0/

[55] Y. Gao, B. Gao, Q. Chen, J. Liu, and Y. Zhang, "Deep Convolutional Neural Network-based Epileptic Electroencephalogram (EEG) Signal Classification," *Frontiers in Neurology*, vol. 11, p. 375, 2020.

[56] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, "Efficient Transformers: A Survey," *ACM Comput. Surv.*, vol. 55, no. 6, pp. 1–28, 2022.

[57] L. A. Moctezuma and M. Molinas, "EEG Channel-Selection Method for Epileptic-Seizure Classification Based on Multi-Objective Optimization," *Frontiers in Neuroscience*, vol. 14, p. 593, 2020.

[58] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, and S. Zuo, "EEG-Based Spatio-Temporal Convolutional Neural Network for Driver Fatigue Evaluation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2755–2763, 2019.