Epileptic Seizure Classification With Symmetric and Hybrid Bilinear Models

Tennison Liu ⬦, Nhan Duy Truong ⬦, *Student Member, IEEE*, Armin Nikpour,
Luping Zhou, *Senior Member, IEEE*, and Omid Kavehei ⬦, *Senior Member, IEEE*

Abstract—Epilepsy affects nearly 1% of the global population, of which two thirds can be treated by anti-epileptic drugs and a much lower percentage by surgery. Diagnostic procedures for epilepsy and monitoring are highly specialized and labour-intensive. The accuracy of the diagnosis is also complicated by overlapping medical symptoms, varying levels of experience and inter-observer variability among clinical professions. This paper proposes a novel hybrid bilinear deep learning network with an application in the clinical procedures of epilepsy classification diagnosis, where the use of surface electroencephalogram (sEEG) and audiovisual monitoring is standard practice. Hybrid bilinear models based on two types of feature extractors, namely Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are trained using Short-Time Fourier Transform (STFT) of one-second sEEG. In the proposed hybrid models, CNNs extract spatio-temporal patterns, while RNNs focus on the characteristics of temporal dynamics in relatively longer intervals given the same input data. Second-order features, based on interactions between these spatio-temporal features are further explored by bilinear pooling and used for epilepsy classification. Our proposed methods obtain an F1-score of 97.4% on the Temple University Hospital Seizure Corpus and 97.2% on the EPILEPSIAE dataset, comparing favourably to existing benchmarks for sEEG-based seizure type classification. The open-source implementation of this study is available at https://github.com/NeuroSyd/Epileptic-Seizure-Classification.

Index Terms—Epilepsy, epileptic seizure classification, EEG, bilinear models, deep learning.

I. INTRODUCTION

THE International League Against Epilepsy (ILAE), defines epilepsy as "a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures" [1], [2]. Epilepsy attacks come in different types and are treated

differently. Critical treatment and prognosis procedures all rely and start with the correct identification of epileptic seizure type. The classification of seizure type is made primarily on clinical grounds based on demographic and medical history variables and is supported by EEG and radiographic studies. Long-term electroencephalogram monitoring with video recording (video-EEG) is the most common supporting method of seizure classification [3]. Generally, epilepsy can be successfully treated with anti-epileptic medication. Around 60–70% of people diagnosed with seizure will gain seizure control with medication. Surgery is another viable medical option for certain conditions of epileptic seizures. The correct seizure type diagnosis is vital in order to select the appropriate drug therapy and to provide information regarding the prognosis [4]. This paper focuses on deep learning tools for automatic epileptic seizure type classification.

A. Diagnostic Challenges in Seizure Classification

An accurate clinical diagnosis requires a thorough history from the patient and observers, which can be compromised by inaccurate and inadequate patient and witness history [5]. Overlapping clinical features also plays a contributory role in incorrect differentials as focal and generalized seizure disorders show overlap of both clinical and EEG symptoms [6]. Recent studies have shown that focal and generalized epilepsy are often difficult to differentiate even by experienced neurologists [7]. Clinical diagnosis is further complicated by the variable expression of epilepsy since the manifestation of same classes of the epilepsy can be quite varied between different patients and also for an individual patient over time [6].

When a diagnosis cannot be reliably reached on clinical grounds, the video-EEG has been shown to be indispensable to confirm epilepsy diagnosis. In many conditions, including infantile spasms, myoclonic epilepsy and idiopathic generalized epilepsy, the video-EEG may specifically confirm or support a correct diagnosis [7]. Video-EEG monitoring involves patients staying in epilepsy monitoring units for several days where natural or induced seizure events are recorded. Neurologists then visually examine these video-EEG records, resulting in a tedious and time-consuming process particularly where hour's or day's worth of EEG needs to be reviewed visually. In many countries, there is a shortage of neurologists and other EEG professionals - manual examination of EEG records delays the diagnosis process and occupies technicians in a field already facing shortages [8]. The inherent subjectivity in visual

examination also contributes to variability in clinical interpretation based on the EEG readers level of expertise. The presence of signal artifacts further complicates the reader's ability to accurately identify key bio-markers. The time-consuming nature of clinical EEG diagnosis and its variability could be greatly improved with an automated seizure classification system that assists professionals.

B. Machine Learning Based Methods

The uncertainty inherent in contemporary seizure diagnostic tools motivated this studys investigation into the use of novel deep learning algorithms for seizure diagnosis. Recent work has demonstrated the potential of using classical machine learning techniques and deep learning frameworks on related EEG and epilepsy problems. The authors in [9] applied the K-Nearest Neighbors (KNN) classifier for seizure detection, achieving 93% for area under the curve. [10] performed detection of epileptic seizures using Support Vector Machine on scalp-EEG data, achieving 96% accuracy. The study in [11] demonstrated the viability of applying deep learning algorithms to automatic seizure detection, achieving 94% specificity using deep residual learning. For the more challenging task of seizure prediction, [12] applied Convolutional Neural Networks (CNNs) on frequency domain data obtained through a Short-Time Fourier Transform to achieve a sensitivity of 81.4% on patient-specific seizure prediction task.

The aforementioned techniques mainly focus on the problems of seizure detection, which is defined as identifying a seizure as it occurs, and seizure prediction, which aims to predict seizures in advance. This study seeks to augment automated seizure advisory systems by introducing seizure type classification. Specific to the problem of seizure type classification, [13] applied text mining on electronic patient records for a 3-class classification task, achieving a F1-score of 71.05%. [14] utilized ontology-based and genetics-based machine learning algorithms on ictal symptoms for a 2-class seizure type classification task, with both algorithms achieving 77.8% accuracy. Recently, [15] applied KNN and Extreme Gradient Boosting to sEEG to classify between 7-classes of seizures, achieving F1-scores of 90.7% and 70.7% respectively. The above mentioned methods can all be viewed as variations of classical machine learning algorithms, while [16] proposed SeizureNet, an ensemble architecture composed of three DenseNets and extracts features from sEEG, achieving F1 of 98.4% on 7-class classification problem. The studies [17] and [18] applied Neural Memory Networks and VGG-19 based models, respectively, to sEEG data to classify between 8 seizure classes.

Apart from the CNN models, the classical machine learning algorithms described above are limited by the potential to recognize and capture complex patterns in multi-channel EEG signal. These classical techniques are also limited in ability to exploit hierarchical structures in natural signals and learning from raw data without a priori feature selection [19]. Recently, deep learning based methods [12], [20] have shown to outperform classical methods for learning more discriminative features from the EEG data compared to the hand crafted features.

TABLE I
SUMMARY OF DATASETS

Dataset features	TUH	EPILEPSIAE
No. of patients (sEEG)	314	30
No. of seizure recordings	2012	276
No. of seizure classes	8	4

Dataset	Seizure type	No. of seizures events	Total recording time (minutes)
TUH	Focal Non-Specific	992	1224
	Generalised Non-Specific	415	567
	Simple Partial	44	26
	Complex Partial	342	540
	Absence	99	14
	Tonic	67	21
	Tonic Clonic	50	80
	Myonic	3	22
Total		**2012**	**2494**
EPILEPSIAE	Complex Partial	139	186
	Unclassified	66	50
	Simple Partial	49	56
	Secondarily Generalised	22	57
Total		**276**	**349**

This study is motivated by the potential of applying deep learning architectures to the problem of seizure type classification. Experiments are conducted using CNNs, RNNs and Bilinear Networks to uncover complex patterns from multidimensional EEG signals and perform this classification task. The main contribution of this paper is in demonstrating the effectiveness of second-order statistics of CNN and RNN features and bilinear pooling in diagnostic classification tasks. Additionally, our study demonstrates that hybrid dual-stream architecture outperforms symmetric bilinear models by exploiting the interactions of two explicit feature types. Lastly, the predictive performance of the proposed method and the generalization ability establish new benchmarks in seizure type classification.

II. PROPOSED METHOD

A. Dataset

Table I summarizes the datasets being used in this work: the Temple University Hospital (TUH) Seizure Corpus dataset v1.4.0 and the EPILEPSIAE dataset. The TUH dataset contains surface EEG data of 817 sessions, of which 305 are sessions that contain seizures, resulting in a total of 2012 seizure events recorded [21]. There are eight classes of seizures identified, the total number of seizure per type and the total recording time of each seizure is shown in Table I. The TUH dataset consists of EEG recordings of various sampling rates, ranging from 250 Hz to 512 Hz. Recordings consist of 24 to 36 channels of signal data as well as an annotation channel.

The EPILEPSIAE dataset was also employed in this study, for the sole purpose to test the generalization ability of our models. The dataset contains sEEG recordings from 30 patients, resulting in 276 seizure events recorded [22]. Identified epileptic seizures are classified into four types, with total number of seizures per type shown in Table I. All recordings have a sampling rate of 256 Hz and contain 19 channels.
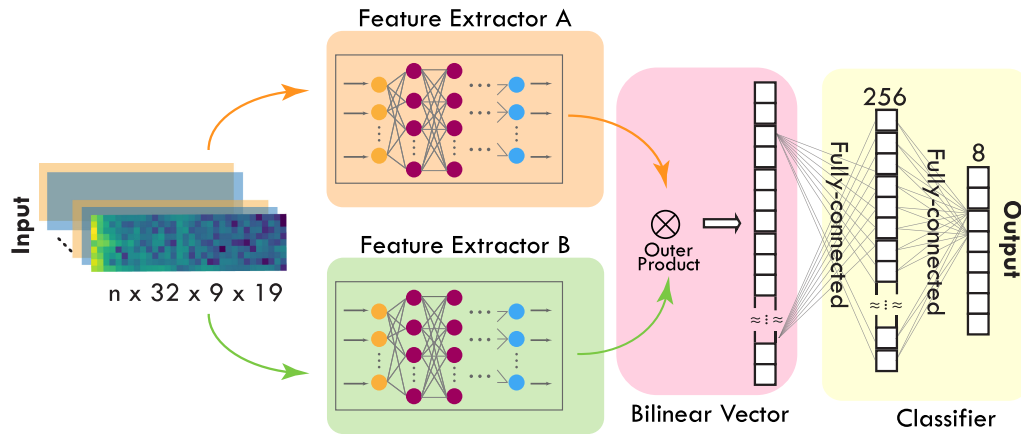
Fig. 1.   Bilinear network.

B. Pre-Processing

As the TUH data contains various sampling rates, all samples were re-sampled to 250 Hz to ensure uniform input dimensions to the neural network. 19 channels common to all recordings were selected and rearranged based on the proximity of electrodes in the 10–20 layout of scalp EEG electrodes.

This study considers the use of Short-Time Fourier Transform (STFT) to exploit the frequency domain representation of raw EEG signal. In the proposed feature engineering method, an STFT is performed on 1-second samples with 50% window overlap using a 64-point FFT cosine analysis window. The intensity values were then calculated by taking $\log_{10}$. After this procedure, the dimension of each training sample becomes $(32, 9, 19)$ where 32 is the number of frequency points, 9 is the number of time steps and 19 is the number of EEG channels.

Following the onset of most seizures, a subset of EEG channels develops rhythmic activity that is typically composed of multiple frequency components [10]. The spectral structure and manifestations across multiple channels and across time are thus important in characterizing features of EEG data. STFT, being a time-frequency analysis method, effectively captures the time-varying frequency and phase structure of epileptic EEG signals [23].

C. Bilinear Models

This study proposes the use of bilinear models for the purpose of seizure-type classification. Bilinear models have been shown to be exceptionally effective at fine-grained recognition and differentiating between similar objects (e.g., classification of different dog breeds) [24]. The bilinear model consists of two feature extractors whose outputs are multiplied using matrix outer product at each location, then pooled to obtain a high level descriptor. This architecture can model local pairwise feature interactions as well as localize discriminative parts in a translationally invariant manner. The detailed topology of bilinear model is shown in Fig. 1. In this section, we discuss why CNN and RNN are suitable feature extraction models for our bilinear architecture. In our experiments, these models are pre-trained on the same dataset, and then extracted and used in

the bilinear structure. This architecture is modular as the 'feature extractor' in Fig. 1 can be replaced interchangeably with either CNN or RNN models. As the EEG patterns of certain epileptic seizure classes, particularly focal and generalized seizures, share an overlap of EEG artifacts, bilinear models would effectively discriminate these events.

This study considers the use of symmetric and hybrid bilinear model. Symmetric networks are initialized with identical base feature extractors. This paper experiments with bilinear-CNN models (B-CNN) composed of identical pre-trained CNN blocks and bilinear-RNN models (B-RNN), composed of pre-trained Convolutional-LSTM (ConvLSTM) blocks. Hybrid bilinear models divorce this symmetry by employing different feature extractors. The hybrid model employs both CNN and RNN as feature extractors, where one network (ConvLSTM) extracts temporal features and the other (CNN) models spatial features, which are combined into second-order statistics through bilinear pooling.

In the bilinear architecture, the outputs of base feature extractor (i.e., the output from the last convolution or ConvLSTM layer) are multiplied using outer product to derive bilinear features for each location in the output. The two feature extractors extracts features of size $O \times M$ and $O \times N$ respectively, where O is the output dimensions ($width \times height$), and M and N are the feature dimensions for each location. The outer matrix product of the two feature vectors ($M \times 1$ and $N \times 1$) for each location results in a bilinear feature of size $(M \times N)$. For the outer product operation to be compatible between the two extracted features, both feature outputs must have matching output dimensions O. Special care is taken while designing the networks to ensure the output sizes match. Calculating the outer product over each location thus produces O bilinear features of $(M \times N)$ bilinear features. Bilinear features across all O locations are aggregated through sum pooling (bilinear pooling) to achieve a final output size of $M \times N$. This matrix is then reshaped into a bilinear vector of size $MN \times 1$, normalized and fed into a fully-connected classifier.

Both the CNN and ConvLSTM proposed in this study produce features of output dimension 12 and a 64 dimension feature at each location. The bilinear operation and pooling thus results in
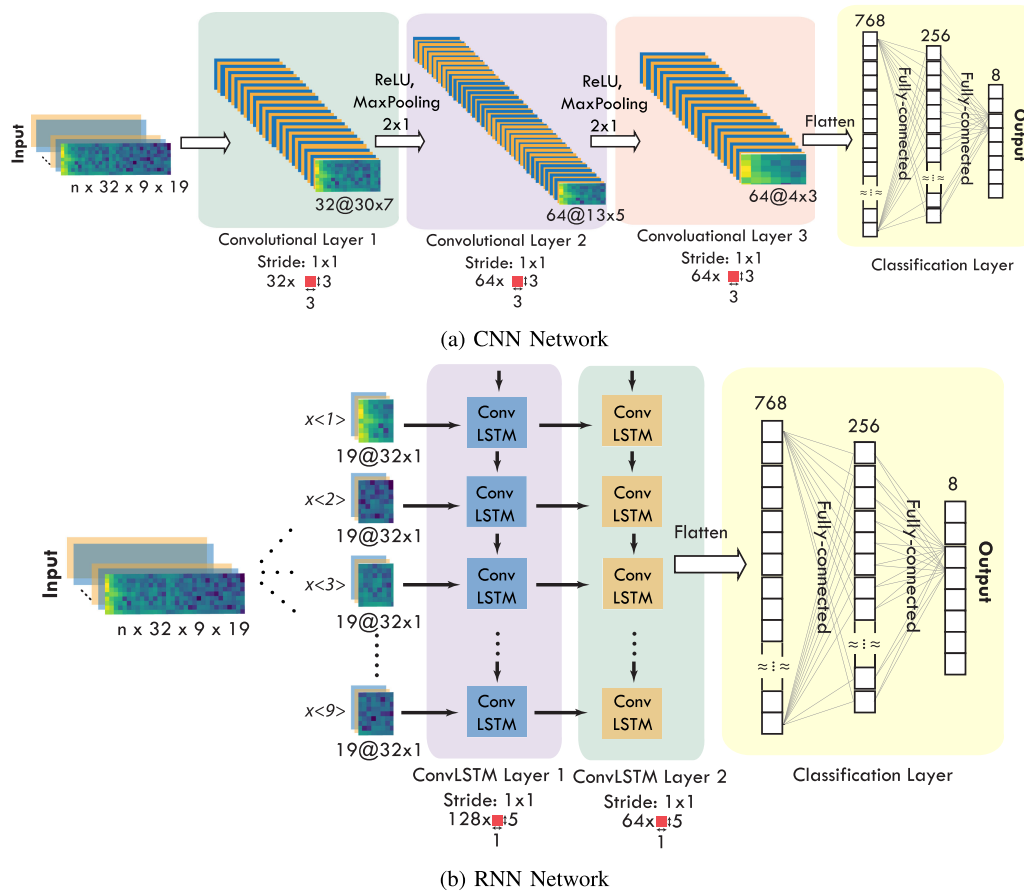
(a) CNN Network

(b) RNN Network

Fig. 2. Base feature extractors.

a bilinear vector of dimensions $(64 \times 64) \times 1$. Predictions are obtained through the last layer, a softmax activation layer, with each node corresponding to the probability that the input signal belongs to a particular seizure class. The following sections will explore the base models in more detail and highlight the two-step training procedure employed.

1) Convolutional Neural Network: Convolutional Neural Networks (CNN) have been successfully applied to many aspects of biomedical research based on physiological signals and medical image analysis (see [12], [18]). The CNN employed in this study includes two components, a feature extractor comprising three convolutional blocks and a fully-connected dense classifier. The topology of the CNN used in this study is highlighted in Fig. 2(a).

2) Recurrent Neural Network: Recurrent Neural Networks (RNN) were designed to work with sequence prediction problems, with Long-Short Term Memory (LSTM) models demonstrating significant promise in time-series classification tasks. This study uses a variation of LSTM networks, the Convolutional LSTM (ConvLSTM), first proposed by [25]. ConvLSTM replaces matrix multiplication performed in traditional LSTM cells with convolution operations, leveraging parameter sharing and sparsity of connection of data. The RNN employed in this study is composed of two components, a feature extractor composed of two ConvLSTM layers and a fully-connected dense

classifier. The topology used in this study is shown in detail in Fig. 2(b).

D. Training and Evaluation

The TUH dataset suffers from uneven class distribution (see Table I) and the class with the lowest count, absence seizures, has 14 minutes of recording. The EPILEPSIAE dataset is relatively more even but still suffers from class imbalance. Under these circumstances, accuracy alone will be unlikely to select the best performing model. The F1-score is thus used as the metric to evaluate the performance of our trained models [28].

To tackle class imbalance problem, weights of the classes are incorporated into the training of the classifier (i.e., giving higher weighting to minority classes and lower weights to majority classes). Early stopping is employed as a regularization technique to reduce over-fitting during the training process. The method monitors validation loss and terminates the training process if validation loss does not improve over 10 epochs.

Stratified five-fold cross-validation is used to robustly evaluate the performance of proposed algorithms. The dataset is randomly split into five-folds, where each fold maintains the proportional distribution of classes in the entire dataset. During training, the model is estimated on four folds and evaluated on the remaining fifth fold (test set). Within the four-folds, the

---

Algorithm 1: Bilinear Model Training and Evaluation.

---

1 Create five stratified folds from dataset;
2 **for** *fold k_i in 5 folds* **do**
3   assign fold k to test set;
4   assign remaining folds to training set;
5   train CNN, RNN on training set;
6   extract trained feature extractor layers with learned weights from CNN, RNN and insert into bilinear network;
7   train bilinear and classification layers of bilinear network on training set;
8   fine-tune the entire model;
9   evaluate F1 on test set;
10 **end**
11 calculate average F1-score over all folds;

---

data is split 80–20 into train and validation sets. The overall performance of the model is based on the average test F1-score over all folds.

The neural networks (CNN and RNN) are trained with the Adam optimization algorithm with batch sizes of 32 and over 200 epochs. The bilinear models rely on pre-trained feature extractors. The training and evaluation process for bilinear models are illustrated in Algorithm 1. In each fold, the CNN and RNN are first trained on the training set. Subsequently, the feature extraction layers, namely convolutional and recurrent layers are inserted with learned weights into the bilinear networks. The bilinear models are then trained on the same training set using a two-step process (lines 7 and 8 of Algorithm 1). Firstly, only the bilinear pooling and dense layers were trained for 50 epochs with early stopping. After which, the entire model, including the pre-trained feature extract layers, is fine-tuned through back-propagation through 100 epochs with early stopping.

As both the outer product and sum-pooling are differentiable matrix operations and the bilinear model remains a directed acyclic graph and parameters can be trained in an end-to-end fashion using backpropagation. The details of gradient propagation through bilinear layers is clarified in work [24]. Additionally, the bilinear pooling operation creates bilinear vector of large dimensions, significantly increasing the number of trainable parameters. By loading pre-trained layers and employing the two-step training procedure, training time is reduced significantly.

It is important to note that the proposed architectures for each respective neural network topology was trained and fine-tuned solely on the TUH dataset. The models, without any adjustments to their architectures, were then trained on the EPILEPSIAE dataset. This was done to evaluate the generalization ability of our algorithms.

III. Results

In this section, we test our approach on both datasets. All models were implemented using Python 3.5 and Keras 2.0 with the Tensorflow 1.4.0 back-end. The models were run on an NVIDIA K80 graphic cards. On the TUH dataset, each training, validation and testing epoch completed in roughly

100, 20 and 20 seconds respectively. On EPILEPSIAE, a significantly smaller dataset, the training, validation and testing stage of each epoch completed in roughly 20, 5 and 5 seconds, respectively. Five-fold cross-validation was performed and the average F1-scores were reported. Table III summarizes seizure type classification results on the two datasets. On the TUH dataset, the CNN and RNN alone do remarkably well, achieving F1-score of 95.50% and 95.80% on the STFT data, respectively. The symmetric bilinear models further improved the classification performance by achieving 96.70% and 96.90% and the hybrid model achieves the best performance with F1-score of 97.40%.

The STFT based inputs and our proposed algorithms achieved similar levels of performance on the EPILEPSIAE dataset, achieving F1-score of 87.3% and 89.0% on the base CNN and RNN models, respectively. The B-CNN model achieved 93.7% while the B-RNN achieved 94.9%. Once again, the hybrid model achieves the best performance of 97.0%. It is important to note that our approach works comparably on two separate datasets without any modifications of the algorithm for the EPILEPSIAE dataset.

Table II describes a benchmark of recent seizure classification approaches. In some cases, it is challenging to compare different approaches as each method is tested on a different dataset with differing number of seizure classes. In the domain of deep learning based methods, the authors in [16] achieved an average F1-score of 98.4% through an ensemble of three DenseNet-based CNN's, "SeizureNet," trained on the TUH dataset. This proposed architecture contains 45.94 million parameters, compared to 1.2 million parameters in the hybrid bilinear structure proposed in this study. The higher complexity of "SeizureNet" leads to an inference time of 32 ms as compared to 1.2 ms for our proposed hybrid network. This work also omitted myoclonic seizures from the classification task due to the low numbers of samples of myoclonic class in the TUH dataset. Both [27] and [18] proposed solutions to the 8-class classification problem on the TUH dataset, achieving 84.06% and 94.05%. As can be seen from Table III, the bilinear models proposed in this study achieve better performance on the 8-class classification problem. Non-deep learning methods, including the KNN proposed in [15] and the Support Vector Machine proposed in [26], demonstrate reasonable performance (90.7% and 91.4%) but were achieved through extensive feature engineering that is not desirable. Our algorithms demonstrated very high classification accuracy with minimum feature engineering. We also demonstrated that our algorithms do not overfit to a specific dataset and can generalize well across two different datasets.

IV. Discussion

Figs. 3 and 4 highlight the classification performance of the bilinear algorithms on each seizure class over the two datasets. For the TUH dataset, classification performance across different seizure types is generally comparable, with the exception of absence seizures (ABSZ). This can be attributed to the low count of absence seizures in the TUH dataset, with only 14 minutes worth of recording for the algorithms to learn from. On the EPILEPSIAE dataset, the bilinear models do slightly worse on

TABLE II
COMPARISON WITH OTHER MACHINE LEARNING BASED CLASSIFICATION METHODS [1]

Method	Dataset	Data Type	Seizure classes	Performance
Support Vector Machine [26]	TUH	sEEG	3 - focal non-specific, generalized non-specific, tonic clonic	Accuracy: 91.4%
K-Nearest Neighbors [15]	TUH	sEEG	7 - focal non-specific, generalized non-specific, simple-partial, complex-partial, absence, tonic, tonic-clonic	F1: 90.70%
Extreme Gradient Boosting [15]	TUH	sEEG	As above	F1: 70.70%
Ensemble CNN's [16]	TUH	sEEG	As above	F1: 98.40%
CNN [18]	TUH	sEEG	8 - focal non-specific, generalized non-specific, simple partial, complex partial, absence, tonic, tonic clonic, myoclonic	Accuracy: 84.06%
Neural Memory Networks [17]	TUH	sEEG	As above	F1: 94.50%
Hybrid Bilinear (this work)	TUH	sEEG	As above	**F1: 97.40**%
Hybrid Bilinear (this work)	EPILEPSIAE	sEEG	4 - unclassified, simple partial, complex partial, secondarily generalized	**F1: 97.00**%
Text Mining [13]	Private Dataset*	Text	3 - complex focal seizure, simple focal seizure, generalized convulsive	F1: 71.05%
Ontology-Based Algorithm [14]	Private Dataset†	Text	2 - temporal lobe, extra-temporal lobe	Accuracy: 77.80%
Genetics-Based Algorithm [14]	Private Dataset†	Text	As above	Accuracy: 77.80%
CNN [27]	Private Dataset‡	Video	2 - mesial temporal lobe, extra-temporal lobe	Accuracy: 92.1%

*: Leiria-Pombal Hospital Dataset, †: Carlo Besta Neurological Institute, Niguarda Ca Granda Hospital and San Paolo Hospital Dataset, ‡: Brisbane Mater Hospital Dataset
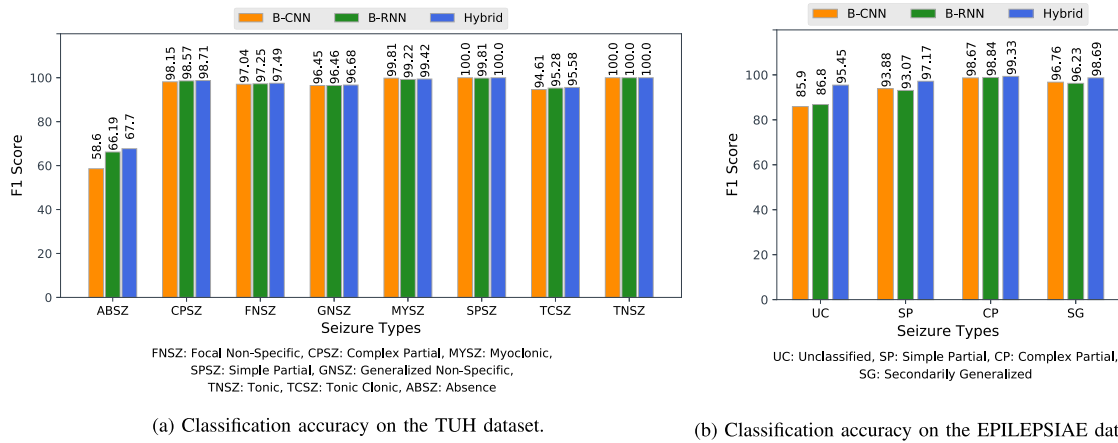

(a) Classification accuracy on the TUH dataset.

(b) Classification accuracy on the EPILEPSIAE dataset.

Fig. 3. Classification over seizure classes.

TABLE III
SUMMARY OF EPILEPTIC SEIZURE TYPE CLASSIFICATION RESULTS
(F1-SCORE %)

Dataset	Pre-processing	CNN	RNN	B-CNN	B-RNN	Hybrid
TUH	STFT	95.5	95.8	96.7	96.9	**97.4**
EPILEPSIAE	STFT	87.3	89.0	93.7	94.9	**97.0**

unclassified (UC) seizures. Like absence seizures, unclassified seizures have the shortest recording time in the EPILEPSIAE dataset. Additionally, a number of different seizure activities and difficult EEG patterns could have been generally labelled as 'unclassified', making it hard to make reliable predictions for that class. It should also be noted that hybrid bilinear networks perform significantly better than symmetric network when predicting on these difficult classes.

Information extracted from EEG signals in frequency and time domains has been widely used in biomedical classification tasks. The performance of models trained on FFT-processed inputs is similar to those achieved using STFT. However, the FFT proposed by [15] in a preceding study contains several flaws. Firstly, the FFT method forfeits valuable temporal information that is critical to highly non-stationary EEG signals. Secondly, the FFT truncates the spectral information to the first 24 frequency bands. While this is a common method of pre-processing to help traditional machine learning algorithms to learn, it is unnecessary in the context of deep learning, where algorithms can learn from more complex signals. Finally, it is not clear why the pre-processing yielded 20 channels as there are only 19 common recording EEG channels.

The bilinear models compare favourably to existing benchmarks (Table II). However, the basic CNN and RNN already

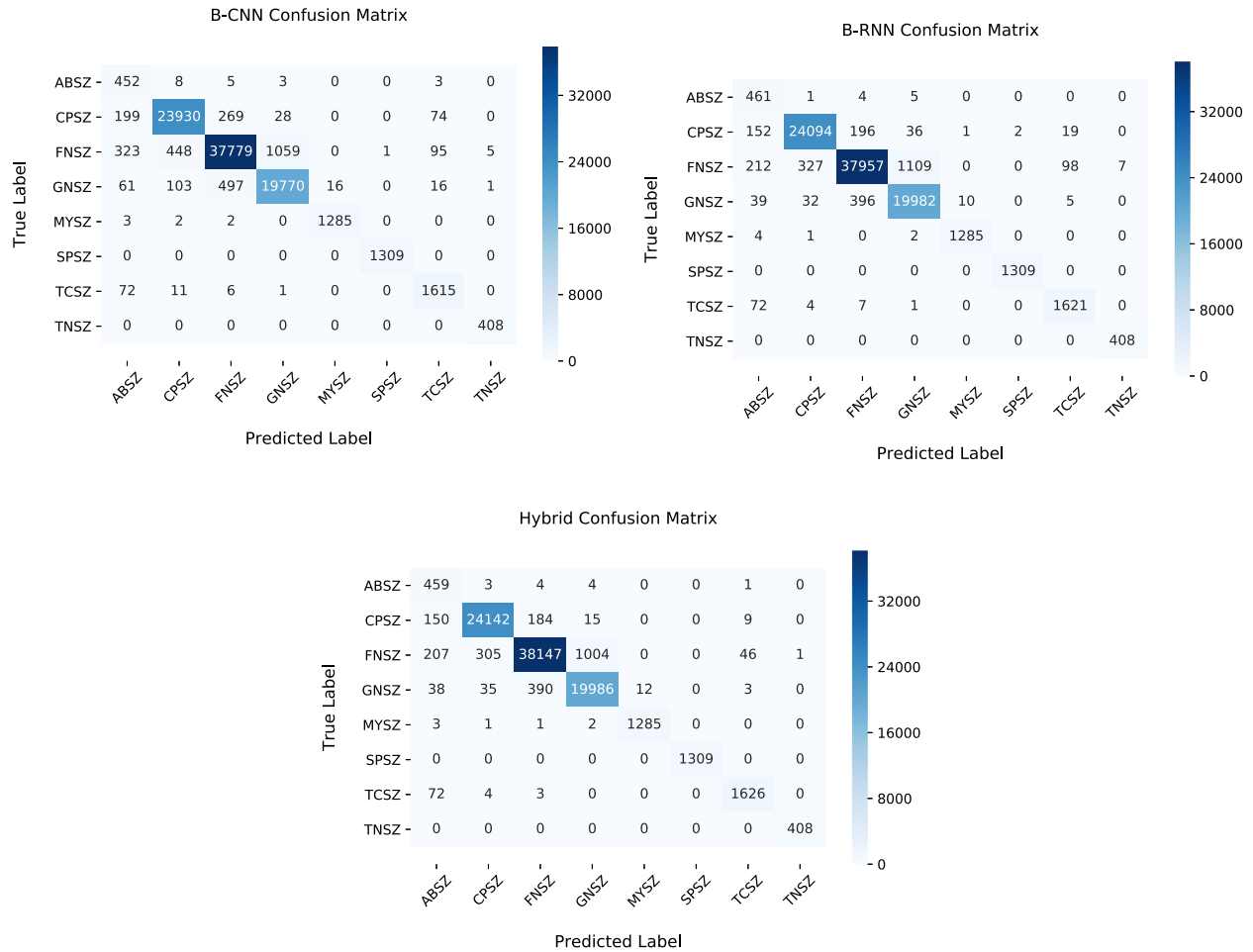[1]Only the best performing model was selected from each study.

Fig. 4.   Confusion matrices of seizure type classification for the symmetric and hybrid bilinear models testing with the TUH dataset.

achieve very high classification performances. A possible explanation is that these networks were carefully designed and fine-tuned for the problem. We performed a Mann-Whitney U Test to compare the performance of the hybrid networks against symmetric architecture across 100 folds. The p-value $= 0.025$ indicates that the difference is statistically significant, demonstrating the ability for hybrid networks to outperform symmetric architectures. The hybrid network outperforms single stream networks by more than 2%, which is also statistically significantly different than the best single stream network (p-value $= 0.022$). Furthermore, the CNN and RNN do not generalize well onto a different dataset, evident in the decreased performance on the EPILEPSIAE dataset. The advantages of the hybrid network is more salient on the EPILEPSIAE dataset, where hybrid architecture achieved significantly better performance over symmetric networks. The hybrid bilinear network achieves similar-level performance on the EPILEPSIAE and TUH datasets, highlighting the ability for the model to generalize. The EPILEPSIAE dataset is smaller with only 349 minutes of recording compared to 2494 minutes in the TUH dataset, demonstrating the ability for hybrid models to learn from smaller amount of data.

The bilinear models work well as the bilinear vector obtained by multiplying the output of feature extractors can effectively model interactions between pairs of local features. Thus, they can discriminate well when the input data is similar and can thus differentiate similar EEG artifacts in different seizure classes. Intuitively, the hybrid bilinear model leverages the unique strengths of two different types of deep networks and considers all pairwise interactions between spatial and temporal features. As the bilinear architecture is modular, different feature extracting methods can be exchanged to achieve optimal performance. For example, hand-crafted feature extractors that are trained to recognize specific EEG artifacts can be combined to deliver improved performance.

As Table I indicates, the number of patients is less than the total number of seizure events recorded on both the TUH and EPILEPSIAE dataset. This gives rise to the potential for signals from one patient to be present in both training and validation/testing datasets, leading to potential bias. This is a drawback of our work as well as existing works in literature for seizure classification. However, our purpose in this work is to confirm whether an automated seizure classification based on EEG is possible or not. The study presents a step towards automated seizure diagnostic tools that can assist neurologists in diagnosing epileptic seizure types. This tool can be extended to also detect and classify between epileptic and non-epileptic seizures, forming an integrated diagnostic tool that can greatly

improve the speed, accuracy and reliability of epilepsy diagnosis. Research has also shown that alternative types of sensory data, including electrocardiogram (ECG), blood oxygen levels, temperature, glucose levels as well as patient variables such as age and gender all influence the manifestation of epilepsy [29]–[32]. The inconsistent performance of EEG-based models on edge cases (e.g., absence seizures) further highlights the need for multi-modal systems and future work can focus on integrating these additional sensory data to provide more accurate diagnosis.

V. CONCLUSION

Classification of epileptic seizures has been a challenge for neurologists diagnosing epilepsy, prescribing treatment and arriving at a prognosis. The automated seizure classification system proposed in this study can assist clinical professionals in diagnosing the disease, reducing time and potentially improves accuracy and reliability. This paper proposes a novel hybrid bilinear model and demonstrates the capability of using these models with minimal feature engineering to classify seizures based on sEEG data. The hybrid bilinear architecture achieves a significant improvement in classification accuracy, achieving 97.4% and 97.0% on the TUH and EPILEPSIAE datasets, establishing new benchmarks in performance. The number of parameters in our models is just over one million with the inference time of around 1.2 ms on an NVIDIA K80 GPU. The model also generalizes well across different datasets and minority seizure types. Future works can focus on improving the performance of this study, by incorporating sensory data-fusion techniques using multi-modal data, primarily audiovisual.

ACKNOWLEDGMENT

REFERENCES

[1] R. S. Fisher *et al.*, "Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.

[2] R. S. Fisher *et al.*, "Operational classification of seizure types by the international league against epilepsy: Position paper of the ILAE commission for classification and terminology," *Epilepsia*, vol. 58, no. 4, pp. 522–530, 2017.

[3] D. E. Friedman and L. J. Hirsch, "How long does it take to make an accurate diagnosis in an epilepsy monitoring unit?" *J. Clin. Neurophysiol.*, vol. 26, no. 4, pp. 213–217, 2009.

[4] M. M. Goldenberg, "Overview of drugs used for epilepsy and seizures: Etiology, diagnosis, and treatment," *Pharmacy Therapeutics*, vol. 35, no. 7, pp. 392–415, 2010.

[5] T. U. Syed *et al.*, "Can semiology predict psychogenic nonepileptic seizures? A prospective study," *Ann. Neurol.*, vol. 69, no. 6, pp. 997–1004, 2011.

[6] C. Panayiotopoulos, "Clinical aspects of the diagnosis of epileptic seizures and epileptic syndromes," in *The Epilepsies: Seizures, Syndromes and Management*, Oxfordshire, UK: Bladon Medical Publishing, 2005.

[7] C. Panayiotopoulos, "Optimal use of the eeg in the diagnosis and management of epilepsies," in *The Epilepsies: Seizures, Syndromes and Management*, Oxfordshire, UK: Bladon Medical Publishing, 2005.

[8] A. Burton, "How do we fix the shortage of neurologists?" *Lancet Neurol.*, vol. 17, no. 6, pp. 502–503, 2018.

[9] P. Fergus, A. Hussain, D. Hignett, D. Al-Jumeily, K. Abdel-Aziz, and H. Hamdan, "A machine learning system for automated whole-brain seizure detection," *Appl. Comput. Informat.*, vol. 12, no. 1, pp. 70–89, 2016.

[10] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 975–982.

[11] M. Golmohammadi, I. Obeid, and J. Picone, "Deep residual learning for automatic seizure detection," in *Proc. 26th Conf. Intell. Sys. Mol. Biol.*, Chicago, USA, Jul. 2018.

[12] N. D. Truong *et al.*, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Netw.*, vol. 105, pp. 104–111, 2018.

[13] L. Pereira, R. Rijo, C. Silva, and M. Agostinho, "ICD9-based text mining approach to children epilepsy classification," *Procedia Technol.*, vol. 9, pp. 1351–1360, 2013.

[14] Y. Kassahun *et al.*, "Automatic classification of epilepsy types using ontology-based and genetics-based machine learning," *Artif. Intell. Medicine*, vol. 61, no. 2, pp. 79–88, 2014.

[15] S. Roy, U. Asif, J. Tang, and S. Harrer, "Machine learning for seizure type classification: Setting the benchmark," 2019, *arXiv:1902.01012*.

[16] U. Asif, S. Roy, J. Tang, and S. Harrer, "SeizureNet: Multi-spectral deep feature learning for seizure type classification," 2019, *arXiv:1903.03232*.

[17] D. Ahmedt-Aristizabal, T. Fernando, S. Denman, L. Petersson, M. J. Aburn, and C. Fookes, "Neural memory networks for seizure type classification," 2019, *arXiv:1912.04968*.

[18] N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "A convolutional neural network based framework for classification of seizure types," in *Proc. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2019, pp. 2547–2550.

[19] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[20] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, 2014.

[21] V. Shah, E. von Weltin, S. Lopez, J. R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The Temple University Hospital seizure detection corpus," *Frontiers Neuroinformat.*, vol. 12, 2018, doi: 10.3389/fninf.2018.00083.

[22] M. Ihle *et al.*, "EPILEPSIAE - A European epilepsy database," *Comput. Methods Programs Biomedicine*, vol. 106, no. 3, pp. 127–138, 2012.

[23] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in eegs using time–frequency analysis," *IEEE Trans. Inf. Technol. Biomedicine*, vol. 13, no. 5, pp. 703–710, Sep. 2009.

[24] T.-Y. Lin, A. Roychowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 1449–1457.

[25] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[26] I. R. D. Saputro, N. D. Maryati, S. R. Solihati, I. Wijayanto, S. Hadiyoso, and R. Patmasari, "Seizure type classification on EEG signal using support vector machine," *J. Phys.: Conf. Ser.*, vol. 1201, no. 1, 2019, Art. no. 012065.

[27] D. Ahmedt-Aristizabal, K. Nguyen, S. Denman, S. Sridharan, S. Dionisio, and C. Fookes, "Deep motion analysis for epileptic seizure classification," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2018, pp. 3578–3581.

[28] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Proc. Eur. Conf. Inf. Retrieval*, 2005, pp. 345–359.

[29] S. Yao *et al.*, "A wearable hydration sensor with conformal nanowire electrodes," *Adv. Healthcare Mater.*, vol. 6, no. 6, 2017, Art. no. 1601159.

[30] D. Cogan, M. Nourani, J. Harvey, and V. Nagaraddi, "Epileptic seizure detection using wristworn biosensors," in *Proc. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2015, pp. 5086–5089.

[31] Y. H. Kwak, W. Kim, K. B. Park, K. Kim, and S. Seo, "Flexible heartbeat sensor for wearable device," *Biosensors Bioelectronics*, vol. 94, pp. 250–255, 2017.

[32] R. Di Giminiani *et al.*, "A wearable integrated textile EMG and muscle oximetry system for monitoring exercise-induced effects: A feasibility study," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, 2018, pp. 1–5.