A Multimodal AI System for Out-of-Distribution Generalization of Seizure Identification

Yikai Yang, Nhan Duy Truong, Jason K. Eshraghian, Christina Maher, Armin Nikpour, and Omid Kavehei, *Senior Member, IEEE*

***Abstract*—Artificial intelligence (AI) and health sensory data-fusion hold the potential to automate many laborious and time-consuming processes in hospitals or ambulatory settings, e.g. home monitoring and telehealth. One such unmet challenge is rapid and accurate epileptic seizure annotation. An accurate and automatic approach can provide an alternative way to label seizures in epilepsy or deliver a substitute for inaccurate patient self-reports. Multimodal sensory fusion is believed to provide an avenue to improve the performance of AI systems in seizure identification. We propose a state-of-the-art performing AI system that combines electroencephalogram (EEG) and electrocardiogram (ECG) for seizure identification, tested on clinical data with early evidence demonstrating generalization across hospitals. The model was trained and validated on the publicly available Temple University Hospital (TUH) dataset. To evaluate performance in a clinical setting, we conducted non-patient-specific pseudo-prospective inference tests on three out-of-distribution datasets, including EPILEPSIAE (30 patients) and the Royal Prince Alfred Hospital (RPAH) in Sydney, Australia (31 neurologists-shortlisted patients and 30 randomly selected). Our multimodal approach improves the area under the receiver operating characteristic curve (AUC-ROC) by an average margin of 6.71% and 14.42% for deep learning techniques using EEG-only and ECG-only, respectively. Our model's state-of-the-art performance and robustness to out-of-distribution datasets show the accuracy and efficiency necessary to improve epilepsy diagnoses. To the best of our knowledge, this is the first pseudo-prospective study of an AI system combining EEG and ECG modalities for automatic seizure annotation achieved with fusion of two deep learning networks.**

***Index Terms*—Affective computing, autonomous systems, expert systems.**

Manuscript received July 4, 2021; revised December 6, 2021 and February 3, 2022; accepted March 2, 2022. Date of publication March 9, 2022; date of current version July 4, 2022. The work of Omid Kavehei was supported by The University of Sydney through a SOAR Fellowship and Microsoft's support through a Microsoft AI for Accessibility grant. *(Corresponding author: Omid Kavehei.)*

Yikai Yang, Nhan Duy Truong, Christina Maher, and Omid Kavehei are with the School of Biomedical Engineering, and the Australian Research Council Training Centre for Innovative BioEngineering, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: yikai.yang@sydney.edu.au; duy.truong@sydney.edu.au; christina.maher@sydney.edu.au; omid.kavehei@sydney.edu.au).

Armin Nikpour is with the Comprehensive Epilepsy Services, Department of Neurology, The Royal Prince Alfred Hospital, Camperdown, NSW 2050, Australia, and also with the Faculty of Medicine and Health, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: armin@sydneyneurology.com.au).

Jason K. Eshraghian is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor 48109 USA, and also with the School of Medicine, University of Western Australia, Crawley, WA 6009, Australia (e-mail: jasonesh@umich.edu).

This article has supplementary downloadable material available at https://dx.doi.org/10.21227/w85b-jr11, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2022.3157877

I. Introduction

EPILEPSY affects about 1% of people globally, placing it as one of the most common severe neurological disorders worldwide [1]–[5]. Accurate and objective seizure counting plays an integral role in a wide range of clinical diagnoses and management decisions for epilepsy. For example, the International League Against Epilepsy (ILAE) defines epilepsy as a brain disorder with (1) at least two unprovoked seizures that are more than 24 hrs apart, (2) one unprovoked seizure and at least 60% general recurrence risk over the next ten years, or (3) diagnosis of an epilepsy syndrome [6]. Accurate seizure counting in the long-term has important implications for driving the management of epilepsy [6], for instance, being seizure-free for ten years, while off anti-epileptic drugs (AEDs) for at least five, identifies whether epilepsy is considered in remission.

The golden standard of epilepsy diagnosis relies on surface or scalp electroencephalogram (EEG) readings and accurate annotation [7]. Nonetheless, epilepsy misdiagnosis or delayed diagnosis is unfortunately still common and has serious consequences [8], [9]. False positives can lead to the inappropriate prescription of AEDs that result in mistreated or worsening symptoms [9], [10]. A strong case can be made for an assistive EEG interpretation given the difficulties in reading and interpreting EEG. While automated EEG interpretation is an opportunity for seizure identification models to act as clinician support systems [11], [12], alternative data modalities can also be used to developing more robust models.

More specifically, epileptic seizures are known to cause short-term and long-term heart rate disturbances. Ictal tachycardia occurs in over 80% of partial onset-seizures [13], [14], which may precede electrographic or clinical onset [15], [16]. Compared with EEG, electrocardiograms (ECG) are relatively more portable and are routinely recorded simultaneously with EEG [17]. Most successful studies using ECG have focused on identification of seizure onset (early prediction) [18], [19], for instance, using heart-rate variability (HRV) to predict events in children with temporal-lobe epilepsy [15], [20]–[23]. In contrast, ECG for seizure identification has had limited focus as its performance is not comparable with multi-channel EEG.

Despite the lower performance of lone usage, ECG recordings can also extract markers of seizure events that EEG may not pick up [24], [25].

In this paper, we propose a multimodal AI system that demonstrates the state-of-the-art performance of non-patient-specific seizure identification, and provides early evidence of generalization on out-of-distribution datasets across continents with different data acquisition hardware infrastructure. We achieve this by designing a multimodal neural network model that accounts for the independent contributions of EEG and ECG towards the classification of seizure events, in addition to their correlated contribution. We expect this novel network architecture to be capable of improving other multimodal tasks as well. The specific contributions of this paper are summarized below:

- Our proposed non-patient-specific multimodal deep learning model using both EEG and ECG data is shown to achieve state-of-the-art results on the publicly available Temple University Hospital (TUH) dataset (US). We improve the area under the receiver operating characteristic curve (AUC-ROC) by a margin of 2.18% over the use of EEG alone, and by 1.67% when we test it against the previously best performing multimodal neural network that combines EEG and ECG signals together.
- Early evidence of generalization is shown by pseudo-prospectively assessing performance on datasets acquired across hospitals in two different continents without any re-training. The European EPILEPSIAE dataset (30 patients) and the Australian RPAH datasets (a total of 61 patients in two groups) are acquired with different electrophysiological recording infrastructures. The AUC-ROC on the European dataset is 0.8595, while for Australian datasets, the averaged AUC-ROC value for the RPA-selected and RPA-random datasets is 0.8549. The EEG-only approach and the ECG-only approach result in averaged AUC-ROC values of 0.8026 and 0.7485 across the three out-of-distribution datasets, while our proposed model obtains an AUC-ROC value of 0.8564 which respectively improves the two comparative methods by 6.71% and 14.42%.
- Our proposed model maintains its effectiveness when either EEG or ECG data modality is missing (i.e., measurement dropout), which confirms the robustness of our data fusion approach.

The remainder of the paper is organized as follows. The following section provides a background of related work, and Section III describes the features of the datasets used in the models. Section IV introduces the proposed method for automatic seizure identification and the multimodal network. Finally, we discuss the results and conclude the paper.

II. Background

A. Prior Art

Recently, several deep learning techniques have achieved promising results using EEG for non-patient-specific seizure identification on the TUH EEG dataset [26]. Beyond clinical use (in- and out-patients), EEG-based methods are limited as the recording apparatus is typically not designed for ongoing wear and would otherwise cause discomfort. Attempts to reduce the number of EEG channels have yielded limited results. A recent approach saw the Neureka 2020 Epilepsy Challenge accounting for the number of channels in their scoring formula. Despite this, the winner of this challenge relied on a 16-channel EEG and still only managed to achieve 12.37% sensitivity (with one false alarm per 24 hours) [27]. 16-channels are inappropriate for ambient use, and the state-of-the-art result highlights the challenges associated with developing a high-performance EEG-based seizure identification system using a constrained number of electrodes. Portability requires complementary biomarkers to EEG that are already integrated into wearable devices [28]. The work in [29] implemented several machine learning methods (HMM, CNN) that were trained on only a small portion of the TUH dataset using EEG-only, achieving sensitivity ranging between 30.05% to 39.09%, specificity from 96.86% to 73.35%. The low sensitivity may be a bottleneck to practical clinical deployment due to the large number of unidentified seizures, and may be a result of using limited training data. Our proposed method can achieve a balanced result of 75.40% sensitivity and 76.60% specificity, which offers a more practical solution in clinical usage.

On clinical utility, a significant challenge in AI-based seizure identification is the need for models that generalize across patients, recording equipment, and hospitals. Unfortunately, all prior works reported in a recent review [30] on the application of deep learning on seizure identification are retrospective and were only benchmarked on test sets sourced identically to the training set. Such models provide low confidence for deployment in clinics that differ from where the data was gathered. This requires rigorous testing and high performance on out-of-distribution datasets. Despite being a key barrier to deployment, generalization across hospitals is not a common metric that is optimized for due to its associated difficulties.

As an example of a study that aims to overcome generalization, the work in [31] leveraged the abundance of weak annotations that were analyzed by a mixed group of technicians, fellows, students, and epileptologists to train a convolutional neural network (CNN), achieving an AUC-ROC score of 0.78. When generalizing the network to the Stanford hospital dataset, the AUC-ROC score dropped to 0.70. Our recent work in [32] reached an AUC-ROC score of 0.84 using a convolutional long short-term memory (ConvLSTM) network tested on unseen patient data. These two studies are among the few publicly available inter-hospital results of a deep learning algorithm using EEG-only recordings.

These results show the potential to attain specialist-level diagnostic capability that can be used as either a primary diagnostic tool or a secondary decision support system. But not only is there insufficient evidence of a practical out-of-distribution performance; these deep learning methods ignore insight provided by other biomarkers that clinicians have access to. We aim to address this by including simultaneously recorded ECG signals.

B. Novelty and Significance

This study aims to improve the seizure identification rate in adults by combining simultaneously acquired EEG and ECG

TABLE I
DATASET COMPARISON

Dataset Detail	TUH dataset		EPILEPSIAE	RPAH dataset	
Datasets	TUH train	TUH development	EPILEPSIAE	RPA-selected	RPA-random
Hardware recording information					
EEG hardware brand	NicoletOne	NicoletOne	Multiple devices	Compumedics	Compumedics
EEG original sample rate (Hz)	250 – 1000	250 – 1000	256	500	500
ECG original sample rate (Hz)	250 – 1000	250 – 1000	256	500	500
Number of EEG | ECG electrodes	19 | 1	19 | 1	19 | 1	19 | 1	19 | 1
Data information					
Number of files	4141	953	4747	–	–
Number of Sessions	1095	228	–	193	183
Number of Patients	540	46	30	31	30
Number of Files with seizures	746	258	276	–	–
Number of Sessions with seizures	301	94	–	103	84
Number of Patients with seizures	174	36	30	31	30
Number of seizures	2129	650	238	238	147
Background duration (hours)	669.3	149.2	4599.0	2489.3	2585.7
Seizure duration (hours)	43.0	14.6	5.0	6.4	4.5
Total duration (hours)	712.3	163.8	4604.0	2495.7	2590.2

Details of all three datasets are in the Supplementary Information Table I, II, III, IV. All three datasets' EEG arrangement of electrodes we used for testing followed the International 10/20 system setting, which have at least 19 electrodes and the names are described as following: "Fp1," "Fp2," "F7," "F3," "Fz," "F4," "F8," "T3," "C3," "Cz," "C4," "T4," "T5," "P3," "Pz," "P4," "T6," "O1," "O2"; while ECG only have one electrode. EPILPSIAE dataset does not contain session number information, and in the RPAH dataset, each file represents one session. Besides, we resample the signal to the same frequency (250 Hz).

recordings in a fused deep learning system. Our study demonstrates that using both recordings in an appropriately structured multimodal neural network can provide a more robust diagnosis than either measurement alone and also improves upon previously reported state-of-the-art multimodal neural networks applied to this task [33]. To the best of our knowledge, only a set of limited works concatenates EEG and ECG recordings for seizure identification, one, in particular, achieving an AUC-ROC improvement of 0.01 and 0.11, when compared with EEG-only and ECG-only models, respectively [34]–[36]. However, naively concatenating different features in a machine learning model poses several challenges [37], which show even lower performance compared with the EEG-only. The use of more features opens up susceptibility to overfitting; combining heterogeneous sources of data increases the difficulty of feature extraction; the inability to isolate the noise of correlated distributions can increase the bias of the network. These issues can severely harm the out-of-distribution performance on unseen patients and datasets.

III. DATASET

Three datasets are used in this work: the Temple University Hospital (TUH) seizure corpus v1.5.1 [26], EPILEPSIAE [38], and two sets from the Royal Prince Alfred Hospital (RPAH) [32]. All datasets contain surface EEG data from adult individuals living with epilepsy. A comparison of these datasets is summarized in Table I, and the detailed information among these datasets is shown in the Supplementary Information Table I, II, III, IV. Only the TUH dataset is used to train and validate our network. Pseudo-prospective evaluations are performed on the EPILEPSIAE and RPAH datasets (without any re-training). The TUH (US), EPILEPSIAE (Germany), and RPAH (Australia) datasets are from three continents recorded with different data acquisition infrastructures.

For a rigorous evaluation on challenging data, such as a variety of seizure types, and various foci on the brain network inputs to the autonomic nervous system [39], the RPAH dataset is divided into two test sets: one set includes 31 patients with different seizure foci and are shortlisted (selected) by neurologists from a long list of patients. The aim was to consider seizure type and semiology in the expert selection of patients. The second set includes 30 randomly selected patients from the same pool without any prior information. The gold standard is labeled by a group of board-certified neurologists and EEG specialists combined with multiple sources, including professional EEG reading, clinical reports, and video information.

A. TUH Dataset

The world's largest open database, the Temple University Hospital (TUH) seizure corpus v1.5.1 [26], was used for training and validation of our deep learning model. The TUH dataset consists of simultaneously recorded EEG and ECG data. Patients with missing ECG recordings were omitted, leaving 1,095 sessions with 540 patients (174 participants with seizures) in the training set. The training set was randomly split (80/20) for training and validation. After training and parameter tuning, the model is then fixed for all future evaluation, hence a pseudo-prospective analysis can be undertaken.

For testing on the TUH dataset, we used 228 sessions with 46 patients (36 patients with seizures). In the absence of a publicly released labeled test set, we treat the TUH development dataset as the unseen test set to reduce bias in our assessment. This is summarized in the Supplementary Information Table I. To assess clinical utility, strictly no further training, tuning, or model selection took place beyond what we were satisfied with on the TUH train and development sets. This strictly inference-only approach on the out-of-distribution test sets is adopted to emulate a prospective study.

B. EPILEPSIAE Dataset

EPILEPSIAE is the largest epilepsy EEG database in Europe, containing EEG and ECG data from 275 patients [38]. In this work, we analyze scalp-EEG and ECG with 19 electrodes of 30

TABLE II
INFERENCE (AUC) RESULTS COMPARISON

Inference Tests	Retrospective	Pseudo-prospective		
Datasets	TUH development	EPILEPSIAE	RPA-selected	RPA-random
EEG hardware brand	NicoletOne	Multiple devices	Compumedics	Compumedics
Multimodal Input (Section V-A1)				
ConvLSTM (EEG)	0.8270	0.8094	0.7511	0.8473
ResConv (ECG)	0.6344	0.7012	0.7493	0.7950
EmbraceNet (EEG+ECG)	0.8311	0.8517	0.7994	0.8732
Proposed (EEG+ECG)	**0.8450**	**0.8595**	**0.8175**	**0.8923**
EEG-Only Input (Section V-A2a)				
ConvLSTM (EEG)	0.8270	0.8094	0.7511	0.8473
EmbraceNet (MC)	0.7537 (−8.86%)	0.7567 (−6.51%)	0.7083 (−5.70%)	0.7711 (−8.99%)
Proposed (MC)	**0.8297** (+0.33%)	**0.8250** (+1.93%)	**0.7711** (+2.66%)	**0.8510** (+0.44%)
ECG-Only Input (Sec. V-A2b)				
ResConv (ECG)	**0.6344**	0.7012	**0.7493**	**0.7950**
EmbraceNet (ME)	0.6146 (−3.12%)	**0.7214** (+2.88%)	0.7323 (−2.27%)	0.7565 (−4.84%)
Proposed (ME)	0.6218 (−1.99%)	0.7035 (+0.33%)	0.7384 (−1.45%)	0.7722 (−2.87%)

EmbraceNet (MC): Missing ECG information to the EmbraceNet fused network, Proposed (MC): Missing ECG information to the proposed fused network, EmbraceNet (ME): Missing EEG information to the EmbraceNet fused network, Proposed (ME): Missing EEG information to the Proposed fused network. Numbers in brackets indicate the relative difference when compared with AUC value of the corresponding baseline model (ConvLSTM or ResConv).

patients with 238 seizures and 4,604 recording hours in total. The sampling rate of the EEG is 256 Hz. (see Supplementary Information Table II).

C. RPAH Datasets

We have extracted 192 adult in-patient EEG monitoring data from 2011 to 2019 at RPA Hospital in Sydney, and long-listed 111 patients with seizures recorded. In this study, RPAH neurologists assist in shortlisting 31 epilepsy adults with different seizure foci. Specifically, neurologists are asked to select patients with the six most common seizure types, namely generalized, frontal, frontotemporal, temporal, parietal, and unspecified focal epilepsy (see Supplementary Information Table III). The total number of seizures and the mean seizure duration are 238 and 97.2 seconds, respectively. To confirm the reliability of our fused network, a randomized test is also performed where 30 patients from 111 adult patients with seizures are randomly selected without any prior information. (see Supplementary Information Table IV) Note that the ratio of the total seizure duration (time) over the total background data for RPAH data is significantly higher than the curated TUH dataset; hence it creates a highly realistic inference-only evaluation for false positives. This is mainly due to more network exposure to noise and artifacts.

IV. METHODS

Our study on RPAH clinical data is approved by the local Research Ethics Committee (see Ethics Declaration section).

A. Pre-Processing

The EEG and ECG signal segment lengths are chosen to be identical, such that one modality does not have a higher impact than the other [31]. Furthermore, clinicians typically read EEG signals in 10–15 s long pages, which is sufficient for expert EEG-readers to extract relevant features from the pre-seizure phase. To better compare with [31], and as we have domain knowledge that 10-15 s is a sufficient window of time for human experts to extract relevant information. Thus, we chose a 12-second window length.

1) EEG: Prior to being passed into the neural network, eye artifacts are removed, and frequency information is extracted by using independent component analysis (ICA) [40] and short-time Fourier Transform (STFT).

The EEG signals are first split into 12-second segments, and then the ICA algorithm is applied to decompose the signal into several statistically independent components. Blind source separation (BSS) [41] is used in the ICA [40] algorithm to obtain several statistically independent topographic maps. Eq 1 shows the working principle of BSS, where $\mathbf{T} \in \mathbb{R}^{I_t \times I_e}$ is the multi-channel EEG signal, I_t represents the number of samples over time, and I_e is the number of electrodes. After decomposition, $\mathbf{M} \in \mathbb{R}^{I_t \times R}$ contains temporal information of the decomposed signal, $\mathbf{A} \in \mathbb{R}^{I_e \times R}$ contains the topographic weight map, and R is the source number estimation.

$$\mathbf{T} \approx \mathbf{M}\mathbf{A}^{\top} \tag{1}$$

Eye movement information is recorded on the electro-oculography (EOG) channel, which is physically close to the EEG channels labeled 'FP1' and 'FP2'. To remove this artifact, a fully automated approach based on Pearson correlation is used [42]. Independent signal information with a strong correlation with channels 'FP1' and 'FP2' above a given threshold (based on adaptive z-scoring) are removed. STFT is then applied to the clean EEG signals with a 250 sample (or 1 s) window length and 50% overlapping. The direct-current (DC) components refer to the 0-Hz component in the EEG signal. The DC component is susceptible to changes in recording environments (e.g., change in power line current, slight movements of electrodes). The mean value of the post-processed electrode is 0, but it varies across each recording. This DC part is removed as it is known to have no relation to seizure occurrences. The pre-processed dimensionality of the input data becomes $(19 \times 23 \times 125)$, where 19 is the number of electrodes, 23 is

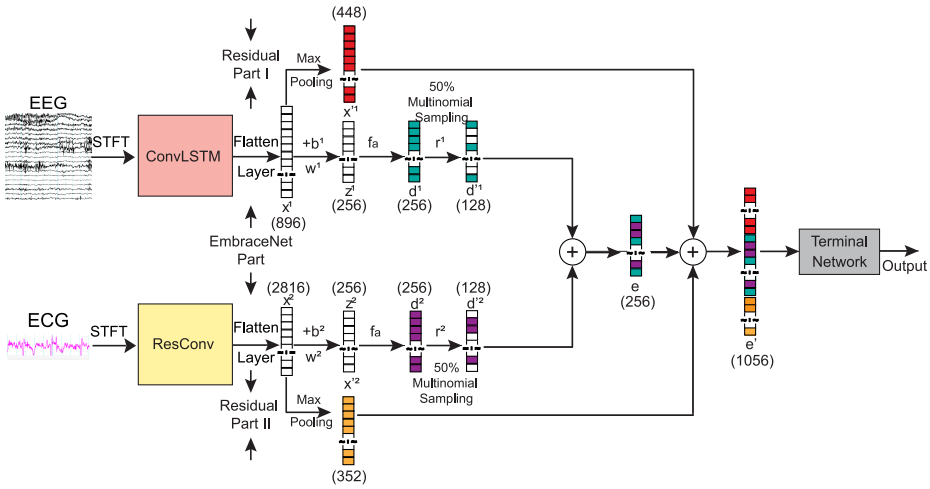

Fig. 1. Proposed multimodal network.

the time sample, and 125 is the range of frequencies. Artifact removal is performed using the MNE Python package [43].

2) ECG: Heart rate variability (HRV) [44] is one of the most common features extracted from ECG for seizure identification. HRV typically requires around (50 or 100 R-R intervals) 60 seconds of raw ECG to have a reliable estimation [45]. This may eliminate the time information within those 60 seconds, which is critical for our ResConv. Therefore, we used deep learning to extract features from STFT of raw ECG signals by itself. As the ECG will ultimately be processed by a CNN, we expect the network can inherently extract relevant information without manual feature engineering. Although raw ECG signals can be directly fed into the neural network, the lack of explicit frequency information makes it difficult for the network to extract essential features. As with the EEG recordings, we apply an STFT to the ECG to translate 12 s segments of raw ECG signals into spectra as input to the neural network. To address differences in sample rates of recording equipment, all ECG signals are re-sampled to 250 Hz. Therefore, a 12 s ECG signal contains a total of 3,000 samples. We used a window length of 250 samples (or 1 s) with 50% overlapping when applying the STFT to transform the data dimensions to $(23 \times 1 \times 126)$. The DC component in the spectrogram is removed, resulting in the dimensionality of $(23 \times 1 \times 125)$.

B. Deep Learning Network

The overall model structure can be separated into three parts. A ConvLSTM network dedicated to the EEG data, a residual CNN for the ECG data, and a fused network that takes the outputs of the individual networks to model the cross-modal representation of the EEG and ECG signals. All three networks are then connected (using both sequential and residual connections) to a terminal network consisting of several dense layers.

1) EEG-ConvLSTM Network: The deep learning network used for training the EEG signal is adopted from our previous work [32]. It is well recognized that the frequency content of EEG signals is crucial for identifying seizures [46], [47]. Different types of seizures may have different brain effects, but frequency evolution [48] is one of the common biomarkers during seizure onset, and the Short-Time Fourier Transform (STFT) is useful for frequency-based feature extraction. Thus, we use the STFT for seizure pre-processing, and use ConvLSTM as convolutions are effective at capturing local temporal dependencies typically on the order of several time-steps (via shared weights), while LSTMs are effective at long-range temporal dependencies (typically on the order of 100 s of time-steps). A ConvLSTM combines the best of both worlds, where the temporal data is the frequency evolution of the spectrogram in our case.

Three deep convolutional long short term memory (ConvLSTM) blocks [49] are combined with three fully-connected layers. The detailed structure is shown in Supplementary Information Fig. 1. The first ConvLSTM layer uses 16 $(n \times 2 \times 3)$ kernels with (1×2) stride, where n represents the number of channels. The next two ConvLSTM blocks both use (1×2) stride and (1×3) kernel sizes. 32 filters and 64 for the ConvLSTM block 2 and 3, respectively. The three ConvLSTM blocks are two fully connected layers with sigmoid activation and output sizes of 256 and 2, respectively.

2) ECG-ResConv Network: Recently, a deep network based on convolutional neural network (CNN)-residual [50] blocks achieved excellent performance on cardiovascular disease classification problems using 12-lead ECG channels [51]. In ECG recordings, localized features (i.e., changes that occur within short ranges of times) seem to carry more significance than long-range disturbances. Therefore, we use convolutions and avoid LSTMs to avoid learning around long-range dependencies that do not exist in the data. However, LSTMs exist to alleviate the vanishing gradient problem in deep networks. The residual connections compensate for this in the ResConv architecture. Another evidence is that a similar architecture has been successfully employed to identify abnormalities in single-lead ECG signals [52]. Thus, we bring this idea to the epileptic seizure identification, as this network is proved to be the best in identifying of ECG abnormalities. Our model fine-tunes this

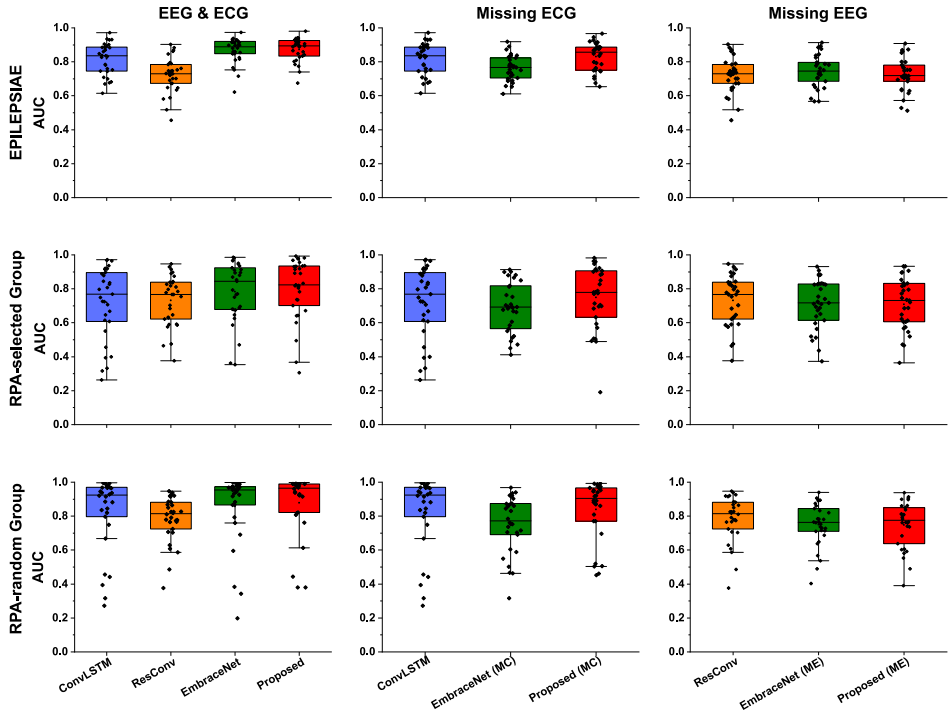

Fig. 2. Pseudo-prospective seizure identification performance comparison. **ConvLSTM**: EEG-only with ConvLSTM network, **ResConv**: ECG-only with ResConv network, **EmbraceNet**: EmbraceNet architecture trained on combined EEG and ECG data, **Proposed**: our approach to combined EEG and ECG data, **EmbraceNet (MC)**: EmbraceNet architecture trained on EEG and ECG data, but missing ECG during inference, **EmbraceNet (ME)**: as above but missing EEG during inference, **Proposed (MC)**: Our proposed approach trained on EEG and ECG data, but missing ECG data during inference, **Proposed (ME)**: As above, but missing EEG data during inference.

ResConv (CNN-residual) network to efficiently and accurately use ECG signals for seizure identification. As shown in the Supplementary Information Fig. 2, the input was first fed into a batch normalization layer, ensuring the input data has zero mean and unit variance to reduce the internal covariate shift [53]. The ReLU activation function was used inside the network [54], and the kernel size for all blocks was (3×1). The residual block was designed with a skip connection combined with two branches, and the down-sampling value in the max-pooling layer was selected to normalize the output sample sizes. The output feature size was halved block-by-block, from 64 to 8, while the number of filters was doubled block-by- block, from 32 to 256. The four residual blocks were flattened, followed by a fully connected layer with sigmoid activation and output dimension of 2. Both the flattened layer and fully connected layer had a 0.5 dropout rate.

To avoid overfitting to the training data, dropout ($p = 0.5$ in flattened and fully-connected layers) and early-stopping (patience of 20 epochs on the validation set) were applied to terminate the training process. In general, the hyperparameters of the architecture, dropout rate, and the overfitting patience parameters were chosen via a combination of grid search and manual tuning. We used a learning rate of 5e-4 and batch size of 32 with the Adam optimizer during training. While during the fine-tuning of the terminal network, the patience for early stopping is set to 5 as this step only updates the terminal layers and validation accuracy/loss did not change much after the first epoch. We implemented our model in Python with Keras 2.0 with a Tensorflow 1.4.0 backend.

C. Proposed Multimodal Network

1) Additive Representation of Multimodal Data Distributions: The proposed network is derived from the calculation of additive signal power from a pair of correlated random variables. Multimodal data obtained from the same target is typically expected to be correlated. As an example, the average power of two superimposed waveforms containing noise content is calculated using the following equation:

$$P_{\text{av}} = P_{\text{av1}} + P_{\text{av2}} + \lim_{\Delta T \to \infty} \frac{1}{T} \int_{-T/2}^{+T/2} 2x_1(t)x_2(t)dt \quad (2)$$

where P_{av1} and P_{av2} denote the average power of time-varying input signals $x_1(t)$ and $x_2(t)$ respectively. The third term is the correlation between the two input signals. The EmbraceNet architecture from [33] models the cross-modal correlation from the integral term, lessening the dependence on the independent contributions of the input signals present in the first two terms of (2).

To address the lack of contribution from the independent input data distributions, the EmbraceNet architecture is stacked with residual connections from the EEG-ConvLSTM network and the ECG-ResConv network (see Fig. 1). This provides a more faithful representation of correlated signals, and also addresses the vanishing gradient problem in deep networks by introducing skipped pathways for gradient backpropagation.

2) Network Structure: The fused network takes flattened vectors $\boldsymbol{x}^{(k)}$ of the independent network models as inputs. In our case, the flattened layers of the EEG and ECG network are

denoted as $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, respectively. The i-th input for both networks are $x_i^{(1)}, x_i^{(2)}$.

a) Cross-modal correlation using EmbraceNet: The cross-modal network that finds a joint representation of the two data modalities originates from the state-of-the-art network "EmbraceNet" [33]. The outputs of the independent EEG and ECG networks are first connected to a pair of dense layers to standardize the input feature vector ($c = 256$ in our tests), reflected in the equation below:

$$z_i^{(k)} = \boldsymbol{w}_i^{(k)} \cdot \boldsymbol{x}_i^{(k)} + b_i^{(k)} \tag{3}$$

where $k = 1$, indicates the EEG network output, and $k = 2$ is the ECG network modality output. $\boldsymbol{w}_i^{(k)}, b_i^{(k)}$ are the weights and biases, respectively.

A nonlinear activation function f_a (e.g. ReLU) is applied to $z_i^{(k)}$, obtaining:

$$d_i^{(k)} = f_a(z_i^{(k)}) \tag{4}$$

Note that $\boldsymbol{d}^{(k)}$ is of dimension c.

Rather than using summation to fuse the vectors, the EmbraceNet model employs an elaborate fusion technique based on a multinomial sampling process:

$\boldsymbol{r}_i = [r_i^{(1)}, r_i^{(2)}]^\intercal$ which is drawn from a multinomial distribution.

$$\mathbf{r_i} \sim \mathbf{Multinomial}(1, \boldsymbol{p}) \tag{5}$$

where $\boldsymbol{p} = [p_1, p_2]^\intercal$ and $\sum_j p_j = 1$, indicating that only one element of the vector $\boldsymbol{r}_i$ is equal to 1, while the rest are 0. The Hadamard product between $\boldsymbol{r}^k$ and $\boldsymbol{d}^k$ is taken to obtain the output $\boldsymbol{d}^{'(k)}$ in EmbraceNet.

$$\boldsymbol{d}^{'(k)} = \boldsymbol{r}^{(k)} \circ \boldsymbol{d}^{(k)} \tag{6}$$

Finally, the output vector is the sum across all elements of $\boldsymbol{d}^{'(k)}$.

$$e_i = \sum_k d_i^{'(k)} \tag{7}$$

b) Residual connections: To include the independent contributions of the input data distributions, residual connections are applied to the outputs of the EEG-ConvLSTM and ECG-ResConv networks (i.e., the inputs to EmbraceNet), $\boldsymbol{x}^k$.

$$x^{'k} = g^k \sum_{i=0}^{c/t} \max(x_i^k, \ldots x_{i+t-1}^k) \tag{8}$$

where c is the dimension of $\boldsymbol{x}^k$, and t is the 1D max pooling stride. The function of g^k represents the max pooling. In our experiments, t is chosen to be a power of 2 which sets the dimensions of $\boldsymbol{x}^{'k}$ close to that of $\boldsymbol{e}$. This is desirable as by setting the dimensions of the latent representations of each data modality to be close to the cross-modal representation, the network capacity for each distribution is normalized before being combined. The practical effect is that the complexity of relationships that are learnable for each data modality are made to be uniform.

Finally, the output of the fused network $\boldsymbol{e}'$ is the combination of both residual networks (Fig. 1, labeled part I: $\boldsymbol{x}^{'1}$ and part II: $\boldsymbol{x}^{'2}$), with the EmbraceNet output e. In our case, the terminal network is three dense layers of sizes 256, 128, and 2, sequentially. The output represents the probability that a seizure is occurring.

D. Performance Metrics

1) AUC-ROC Score: To evaluate the performance of the proposed method for the seizure identification task, we used the area under the Receiver Operating Characteristic curve (AUC-ROC). The AUC-ROC measures the area under the recall vs. the false-positive rate (FPR) plots. Formal definitions of the recall and the FPR are provided below:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{9}$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN+FP}} \tag{10}$$

where TP, TN, FP, and FN represent true positives (correct seizure identification), true negatives (correct non-seizure identification), false positives (incorrect seizure identification), and false negatives (incorrect non-seizure identification).

2) The Wilcoxon Signed-Rank Test: To evaluate the performance of our model, the Wilcoxon signed-rank statistic [55] is used. The obtained p-value provides a metric indicating the significance of performance improvement.

V. Results

A. Test Cases

The following tests are applied to the TUH test set, with a pseudo-prospective study on the EPILEPSIAE and RPAH (selected and randomized) groups.

1) Multimodal Approach: To explore the effectiveness of our proposed network, we compare the following options, using the networks shown in Fig. 1, on EEG and ECG recordings:

- ConvLSTM (EEG only)
- ResConv (ECG only)
- EmbraceNet (EN), prior state-of-the-art for multimodal data [33] (EEG and ECG)
- Proposed (ECG and EEG)

2) Missing Modalities: The above networks are then tested for the cases where either EEG or ECG recordings are unavailable, which may occur due to poor electrode contact or unexpected signal dropout during recording.

a) Missing ECG (MC): The ECG channel is treated as missing and is set to zero. The EEG channel is kept the same, and the same network models described above are used for testing (without retraining). Our proposed approach is compared to EmbraceNet, and the single-modal ConvLSTM EEG-only network.

b) Missing EEG (ME): All EEG inputs are set to zero while ECG recordings are used as per normal. The same comparison as the missing ECG case is made, but with the single-modal ResConv ECG-only network instead.

B. Performance

The distribution of AUC-ROC for the pseudo-prospective out-of-distribution analyses are shown in Fig. 2, tabulated in Table II, and the ROCs under all test scenarios are provided in Supplementary Information Fig. 3.

The AUC-ROC for each individual patient across the EPILEPSIAE, RPA-selected and randomized groups is shown in Fig. 2. The first column depicts the results for the multimodal approach. Combining EEG and ECG data using a fused network (for both EmbraceNet and our proposed method) improves performance across all three out-of-distribution datasets. Our proposed method consistently outperforms EmbraceNet across the three experiments. From Table II, the absolute AUC improvement of our proposed method when compared with EmbraceNet is 0.0139 for the TUH test set, 0.0078 for the ELIEPSIAE set, 0.0181 for the selected RPA group, and 0.0191 for the randomized RPA group.

The absolute margin of the AUC-ROC from our approach improves upon the prior state-of-the-art fusion method [33] on the out-of-distribution datasets by 6.71% for EEG-only, 14.42% for ECG-only, and 1.76% for EmbraceNet's multimodal approach. The AUC-ROC curves across each dataset are shown in Supplementary Information Fig. 3. The p-value derived from the Wilcoxon signed-rank test for our proposed model (EEG+ECG) when compared with EEG only, ECG only and EmbraceNet (EEG+ECG) is ($p < 0.0001$, $p < 0.0001$, $p = 0.0418$) in the EPILEPSIAE dataset, ($p < 0.0001$, $p = 0.0037$, $p = 0.0013$) for the selected RPA group, and ($p = 0.0001, p = 0.0006, p = 0.0003$) RPA-random group. This demonstrates the statistically significant performance of our model for a specified threshold of 0.01.

For the test case of the missing ECG modality (MC), we repeat the above tests when the two fused networks (our proposed and EmbraceNet) are missing ECG information. Compared with EEG-only performance using the ConvLSTM network, the performance of EmbraceNet dropped by 8.86% (TUH), 6.51% (EPILEPSIAE), 5.70% (RPA-selected), and 8.99% (RPA-random). In contrast, our proposed method increases performance of the ConvLSTM EEG-only network by 0.33% (TUH), 1.93% (EPILEPSIAE), 2.66% (RPA-selected), and 0.44% (RPA-random). The p-value under this case for our proposed approach against EmbraceNet is ($p < 0.0001$, $p = 0.0001$, $p < 0.0001$) EPILEPSIAE, RPA-selected and -random datasets, respectively.

The above process was repeated for the missing EEG modality case (ME). When compared with the ECG-only ResConv network, the performance of EmbraceNet with missing EEG information dropped by 3.12% (TUH), 2.27% (RPA-selected), and 4.84% (RPA-random) and improved by 2.88% (EPILEPSIAE). The proposed method dropped by 1.99% (TUH), 1.45% (RPA-selected), and 2.87% (RPA-random), and improved by 0.33% (EPILEPSIAE). The p-values when compared to EmbraceNet (ME) are $p = 0.2644$ (RPA-selected), and $p = 0.3897$ (RPA-random) respectively. Other than the EPILEPSIAE dataset, the decrease in our model's performance is less than the decrease in EmbraceNet's performance, demonstrating it is robust when the EEG modality is missing.

VI. Discussion

For the multimodal case, described in Section V-A1, our proposed model outperforms the AUC-ROC for seizure identification compared to the EEG-only network, ECG-only network, and the EmbraceNet fused approach across all datasets.

The work in [31] uses a 12-second windows convolutional neural network (CNN), achieving an AUC-ROC score of 0.78 on the TUH dataset v1.4.0 (detail comparison with the v1.5.1 and v1.4.0 are shown in the Supplementary Information Table I), and when generalizing the network to the Stanford hospital dataset, the AUC-ROC score dropped to 0.70. Our proposed methods show better results with 0.8450 on the TUH dataset, and an average of 0.8564 using the three out-of-distribution datasets.

Our approach tested on the non-prospective TUH test set improved the AUC-ROC attained using EmbraceNet from 0.8311 to 0.8450. This improvement of 1.67% is the smallest margin of all cases. All pseudo-prospective trials showed a larger improvement other than for EPILEPSIAE (0.0078), thus demonstrating our method's capacity for generalization and potential for use in a clinical setting. The ECG-only network was the only approach to improve upon the TUH test set baseline on pseudo-prospective samples, but came with the cost of large performance variation and significantly lower absolute AUC-ROC. We expect this variation arises from individuals with different seizure origins which variably influence the autonomic nervous system [56]. The table shows that performance on the RPA-random set is better than the RPA-selected by 9.15% when using the proposed method to test. Based on Supplementary Information Table IV, we find that generalized patients and patients with seizure foci on the frontotemporal and temporal lobes typically show higher performance than patients with seizure foci on other locations. There are 20 patients under these groups in the RPAH selected dataset and 23 patients in the RPAH random dataset. Furthermore, seizure identification performance varies from patient to patient, and the mechanism is rather complicated. In our test, based on our observation, we believe it may be caused by different proportions of the seizure foci.

For the ECG-missing case, described in Section V-A2a, we evaluated the AUC performance when all ECG data was omitted. Our results show that the proposed method is still able to marginally improve the performance over the EEG-only network (average improvement of 1.30%), whereas the EmbraceNet approach has a significant drop in performance (average drop of 7.51%). This reflects that our proposed network has high robustness for missing (potentially corrupted) ECG recordings.

Finally, we analyze the case for missing EEG information, discussed in Section V-A2b. Our proposed method experienced a slight drop in performance when compared to the dedicated ECG-only ResConv model on the TUH and RPAH groups, and an increase on the EPILEPSIAE dataset by 0.33%. The performance of EmbraceNet dropped more (in the range of 2.27% and 4.84%), other than for the EPILEPSIAE dataset. Our network is much more stable than EmbraceNet, although this test makes it evident that our network relies more on EEG data to surpass the other networks.

We note that EmbraceNet (ME) and Proposed (ME) are almost worse than using ResConv (ECG). However, the important note

is that the proposed method is robust to missing EEG signals (ME). EmbraceNet and our proposed method still perform comparably to ResConv, trained specifically for ECG. The fusion layer, initially proposed by EmraceNet and then improved by our method, which ensures the model does not rely wholly on an individual modality. In other words, the fusion layer strengthens the feature learning using both EEG and ECG signals. The underlying mechanism for fusing ECG and EEG features using the proposed methods is explained following: The EEG-only embedding of data is designed for patients whose seizures can only be identified by the EEG, which is especially important for patients whose epileptic seizures occur without obvious convulsive behavioral manifestations [57].On the other hand, not all seizures can be identified by the EEG, but some patients may have tachycardia [24], or in other words, life-threatening arrhythmias are amplified during a seizure [57], [58]. Finally, the joint embedding of EEG and ECG enables our model to account for patients who exhibit both biomarkers in the EEG and the tachycardia from ECG during the seizures. In practice, the two are rarely mutually exclusive events, and there may be subtle biomarkers in either or both data modalities. The joint embedding modelled by the EmbraceNet architecture aims to model the joint distribution of EEG and ECG signals that represent the onset of a seizure.

Compared with EEG-only and ECG-only methods, the proposed method can identify more seizures while maintaining the same false-positive rate verified on data that includes a very large amount of background (non-seizure) data. AUC score is a good metric for comparing the general method, while in real seizure identification applications, a certain threshold will be selected to achieve a balanced sensitivity and specificity. For example, in the TUH test set, one of the balance points we apply for clinical usage for the proposed method results in 75.40% sensitivity and 76.60% specificity. In comparison with the EEG-only method, the sensitivity drops to 73.06% to achieve the same specificity, and for the ECG-only method, the sensitivity can only achieve around 39.61%. The total number of seizures in the TUH test set is 650, which means that while maintaining the same false-positive rate, using the proposed fusion methods can identify 15 and 232 more seizures than EEG-only and ECG-only methods in the real-world application. For a long-duration recording clinical dataset like the RPAH dataset, the proposed method will be more helpful, for example, in the RPAH-select dataset, the ratio of background duration over seizure duration will be around 389:1, which means that neurologists need to review 389 seconds of EEG data on average to find only 1 s of seizure information. Finding 10 more seizures would require experts to review 10.5 hours of background EEG information without assistance from the AI. It should be noted that the average seizure duration is 97.2 seconds in the RPAH-select dataset. Due to the small number of seizures relative to the background data, even small performance improvements add up to large savings in terms of time and effort in seizure identification and annotation.

Our results illustrate the proposed model and method are generalizable across different datasets acquired with different equipment. We have also shown the robustness of our network in terms of performance and susceptibility to missing data modalities when compared to the state-of-the-art in data fusion [33]. Our experiments confirm that multimodal data can achieve better performance than either EEG or ECG alone. We envision a potential use of our system in decision support systems in the context of providing a confirmatory secondary reading. In such a case, our proposed system would act as the risk mitigation strategy in supporting clinical readings.

VII. Conclusion

Despite the extensive studies over the past four decades of using EEG in seizure identification, the use of ECG is quite limited and never previously reported in a multimodal deep learning model. Our proposed model and fused modality approach show promise in using EEG and ECG signals together, and were demonstrated to generalize to pseudo-prospective studies. Our analysis shows that a seizure identification system can sustain state-of-the-art performance on out-of-distribution samples, which is a critical feature for clinical translation.

Acknowledgment

The author Yikai Yang would like to thank the Research Training Program (RTP) support provided by the Australia Government.

Ethics declarations

Ethics approval number X19-0323-2019/STE16040 on *Validating epileptic seizure detection, prediction and classification algorithms* approved on 19 September 2019 by the NSW Local Health District (LHD) for implementation in the Comprehensive Epilepsy Services, Department of Neurology, The Royal Prince Alfred Hospital (RPAH).

Data availability

The Temple University Hospital dataset is publicly available at https://www.isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml. The training on the publicly available TUH dataset provides readers with sufficient level of ability to independently confirm some of the results reported. The EPILEPSIAE dataset is available at cost via http://www.epilepsiae.eu/project_outputs/european_database_on_epilepsy. The Royal Prince Alfred Hospital was used under ethics Review Board approval for our use only.

Code Availability

The code used to generate all results in this manuscript can be made available upon request.

References

[1] P. N. Banerjee, D. Filippi, and W. A. Hauser, "The descriptive epidemiology of epilepsy-A review," *Epilepsy Res.*, vol. 85, no. 1, pp. 31–45, 2009.
[2] E. Foster *et al.*, "The costs of epilepsy in Australia: A productivity-based analysis," *Neurol.*, vol. 95, no. 24, pp. e3221–e3231, 2020.
[3] E. Beghi, "Addressing the burden of epilepsy: Many unmet needs," *Pharmacological Res.*, vol. 107, pp. 79–84, 2016.
[4] A. Jacoby, D. Snape, and G. A. Baker, "Epilepsy and social identity: The stigma of a chronic neurological disorder," *Lancet Neurol.*, vol. 4, no. 3, pp. 171–178, 2005.
[5] R. S. Fisher *et al.*, "The impact of epilepsy from the patient's perspective i. descriptions and subjective perceptions," *Epilepsy Res.*, vol. 41, no. 1, pp. 39–51, 2000.

[6] R. S. Fisher *et al.*, "ILAE official report: A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014.

[7] S. R. Benbadis, S. Beniczky, E. Bertram, S. MacIver, and S. L. Moshé, "The role of EEG in patients with suspected epilepsy," *Epileptic Disord.*, vol. 22, no. 2, pp. 143–155, 2020.

[8] M. M. Oto, "The misdiagnosis of epilepsy: Appraising risks and managing uncertainty," *Seizure*, vol. 44, pp. 143–146, 2017.

[9] S. R. Benbadis, "The tragedy of over-read EEGs and wrong diagnoses of epilepsy," *Expert Rev. Neurotherapeutics*, vol. 10, no. 3, pp. 343–346, 2010.

[10] E. Lee-Lane *et al.*, "Epilepsy, antiepileptic drugs, and the risk of major cardiovascular events," *Epilepsia*, vol. 62, no. 7, pp. 1604–1616, 2021.

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

[12] N. D. Truong *et al.*, "Seizure susceptibility prediction in uncontrolled epilepsy," *Front. Neurol.*, vol. 12, 2021, Art. no. 721491.

[13] K. Jansen and L. Lagae, "Cardiac changes in epilepsy," *Seizure*, vol. 19, no. 8, pp. 455–460, 2010.

[14] C. Sevcencu and J. J. Struijk, "Autonomic alterations and cardiac changes in epilepsy," *Epilepsia*, vol. 51, no. 5, pp. 725–737, 2010.

[15] M. Zijlmans, D. Flanagan, and J. Gotman, "Heart rate changes and ECG abnormalities during epileptic seizures: Prevalence and definition of an objective clinical sign," *Epilepsia*, vol. 43, no. 8, pp. 847–854, 2002.

[16] F. Leutmezer, C. Schernthaner, S. Lurger, K. Pötzelberger, and C. Baumgartner, "Electrocardiographic changes at the onset of epileptic seizures," *Epilepsia*, vol. 44, no. 3, pp. 348–354, 2003.

[17] P. Boon *et al.*, "A prospective, multicenter study of cardiac-based seizure detection to activate vagus nerve stimulation," *Seizure*, vol. 32, pp. 52–61, 2015.

[18] J. Pavei *et al.*, "Early seizure detection based on cardiac autonomic regulation dynamics," *Front. Physiol.*, vol. 8, p. 765, 2017.

[19] J. J. Thiagarajan, D. Rajan, S. Katoch, and A. Spanias, "DDxNet: A deep learning model for automatic interpretation of electronic health records, electrocardiograms and electroencephalograms," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.

[20] K. Schiecke, M. Wacker, F. Benninger, M. Feucht, L. Leistritz, and H. Witte, "Advantages of signal-adaptive approaches for the nonlinear, time-variant analysis of heart rate variability of children with temporal lobe epilepsy," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 6377–6380.

[21] I. Osorio and B. Manly, "Probability of detection of clinical seizures using heart rate changes," *Seizure*, vol. 30, pp. 120–123, 2015.

[22] T. De Cooman, C. Varon, B. Hunyadi, W. Van Paesschen, L. Lagae, and S. Van Huffel, "Online automated seizure detection in temporal lobe epilepsy patients using single-lead ECG," *Int. J. Neural Syst.*, vol. 27, no. 7, 2017, Art. no. 1750022.

[23] T. De Cooman, T. W. Kjær, S. Van Huffel, and H. B. Sorensen, "Adaptive heart rate-based epileptic seizure detection using real-time user feedback," *Physiol. Meas.*, vol. 39, no. 1, 2018, Art. no. 014005.

[24] C. A. Galimberti, E. Marchioni, F. Barzizza, R. Manni, I. Sartori, and A. Tartara, "Partial epileptic seizures of different origin variably affect cardiac rhythm," *Epilepsia*, vol. 37, no. 8, pp. 742–747, 1996.

[25] K. Vandecasteele *et al.*, "The power of ECG in multimodal patient-specific seizure monitoring: Added value to an EEG-based detector using limited channels," *Epilepsia*, vol. 62, no. 10, pp. 2333–2343, 2021.

[26] V. Shah *et al.*, "The Temple University Hospital seizure detection corpus," *Front. Neuroinformat.*, vol. 12, p. 83, 2018.

[27] C. Chatzichristos *et al.*, "Epileptic seizure detection in EEG via fusion of multi-view attention-gated U-net deep neural networks," *Proc. IEEE Signal Process. Med. Biol. Symp.*, 2020, pp. 1–7.

[28] J. Engel Jr *et al.*, "Epilepsy biomarkers," *Epilepsia*, vol. 54, pp. 61–69, 2013.

[29] M. Golmohammadi, V. Shah, I. Obeid, and J. Picone, "Deep learning approaches for automated seizure detection from scalp electroencephalograms," in *Proc. Signal Process. Med. Biol.*, 2020, pp. 235–276.

[30] A. Shoeibi *et al.*, "Epileptic seizures detection using deep learning techniques: A review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 11, 2021, Art. no. 5780.

[31] K. Saab, J. Dunnmon, C. Ré, D. Rubin, and C. Lee-Messer, "Weak supervision as an efficient approach for automated seizure detection in electroencephalography," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–12, 2020.

[32] Y. Yang, N. D. Truong, C. Maher, A. Nikpour, and O. Kavehei, "Continental generalization of an AI system for clinical seizure recognition," 2021, *arXiv:2103.10900*.

[33] J.-H. Choi and J.-S. Lee, "EmbraceNet: A robust deep learning architecture for multimodal classification," *Inf. Fusion*, vol. 51, pp. 259–270, 2019.

[34] B. R. Greene, G. B. Boylan, R. B. Reilly, P. de Chazal, and S. Connolly, "Combination of EEG and ECG for improved automatic neonatal seizure detection," *Clin. Neuriophysiol.*, vol. 118, no. 6, pp. 1348–1359, 2007.

[35] D. Cogan, J. Birjandtalab, M. Nourani, J. Harvey, and V. Nagaraddi, "Multi-biosignal analysis for epileptic seizure monitoring," *Int. J. Neural Syst.*, vol. 27, no. 01, 2017, Art. no. 1650031.

[36] P. M. d. C. B. Maia, "NeuroMov: Multimodal approach for epileptic seizure detection and prediction," Ph.D. dissertation, Universidade do Porto, Jul. 2019. [Online]. Available: https://repositorio-aberto.up.pt/bitstream/10216/122327/2/352306.pdf

[37] F. Fürbass *et al.*, "Automatic multimodal detection for long-term seizure documentation in epilepsy," *Clin. Neuriophysiol.*, vol. 128, no. 8, pp. 1466–1472, 2017.

[38] J. Klatt *et al.*, "The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients," *Epilepsia*, vol. 53, no. 9, pp. 1669–1676, 2012.

[39] K. Vandecasteele *et al.*, "Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment," *Sensors*, vol. 17, no. 10, 2017, Art. no. 2338.

[40] P. Comon, "Independent component analysis, a new concept," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[41] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[42] J. Dammers *et al.*, "Integration of amplitude and phase statistics for complete artifact removal in independent components of neuromagnetic recordings," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2353–2362, Oct. 2008.

[43] A. Gramfort *et al.*, "MEG and EEG data analysis with MNE-Python," *Front. Neurosci.*, vol. 7, p. 267, 2013.

[44] M. B. Malarvili and M. Mesbah, "Newborn seizure detection based on heart rate variability," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2594–2603, Nov. 2009.

[45] J. Jeppesen *et al.*, "Seizure detection based on heart rate variability using a wearable electrocardiography device," *Epilepsia*, vol. 60, no. 10, pp. 2105–2113, 2019.

[46] S. J. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *J. Neurol., Neurosurgery Psychiatry*, vol. 76, no. suppl 2, pp. ii2–ii7, 2005.

[47] M. Zijlmans, P. Jiruska, R. Zelmann, F. S. Leijten, J. G. Jefferys, and J. Gotman, "High-frequency oscillations as a new biomarker in epilepsy," *Ann. Neurol.*, vol. 71, no. 2, pp. 169–178, 2012.

[48] R. Q. Quiroga, H. Garcia, and A. Rabinowicz, "Frequency evolution during tonic-clonic seizures," *Electromyogr. Clin. Neuriophysiol.*, vol. 42, no. 6, pp. 323–332, 2002.

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

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[51] A. H. Ribeiro *et al.*, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Commun.*, vol. 11, no. 1, pp. 1–9, 2020.

[52] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, pp. 65–69, 2019.

[53] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[54] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.

[55] R. F. Woolson, "Wilcoxon signed-rank test," *Wiley Encyclopedia Clin. Trials*, pp. 1–3, 2007.

[56] S. G. Mueller, L. M. Bateman, M. Nei, A. M. Goldman, and K. D. Laxer, "Brainstem atrophy in focal epilepsy destabilizes brainstem-brain interactions: Preliminary findings," *NeuroImage: Clin.*, vol. 23, 2019, Art. no. 101888.

[57] A. Bardai *et al.*, "Epilepsy is a risk factor for sudden cardiac arrest in the general population," *PLoS One*, vol. 7, no. 8, 2012, Art. no. e42749.

[58] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," in *Proc. Comput. Methods Prog. Biomed.*, 2016, vol. 127, pp. 144–164.