Unsupervised seizure identification on EEG

Ikay Yıldız[*],

Rachael Garner,

Matthew Lai,

Dominique Duncan

Laboratory of Neuro Imaging, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, 2025 Zonal Avenue, Los Angeles, CA 90033, United States

Abstract

Background and Objective: Epilepsy is one of the most common neurological disorders, whose development is typically detected via early seizures. Electroencephalogram (EEG) is prevalently employed for seizure identification due to its routine and low expense collection. The stochastic nature of EEG makes manual seizure inspections laborsome, motivating automated seizure identification. The relevant literature focuses mostly on supervised machine learning. Despite their success, supervised methods require expert labels indicating seizure segments, which are difficult to obtain on clinically-acquired EEG. Thus, we aim to devise an unsupervised method for seizure identification on EEG.

Methods: We propose the first fully-unsupervised deep learning method for seizure identification on raw EEG, using a variational autoencoder (VAE). In doing so, we train the VAE on recordings without seizures. As training captures non-seizure activity, we identify seizures with respect to the reconstruction errors at inference time. Moreover, we extend the traditional VAE training loss to suppress EEG artifacts. Our method does not require ground-truth expert labels indicating seizure segments or manual feature extraction.

Results: We implement our method using the PyTorch library and execute experiments on an NVIDIA V100 GPU. We evaluate our method on three benchmark EEG datasets: (i) intracranial recordings from the University of Pennsylvania and the Mayo Clinic, (ii) scalp recordings from the Temple University Hospital of Philadelphia, and (iii) scalp recordings from the Massachusetts Institute of Technology and the Boston Children's Hospital. To assess performance, we report accuracy, precision, recall, Area under the Receiver Operating Characteristics Curve (AUC), and p-value under the Welch *t*-test for distinguishing seizure vs. non-seizure EEG windows. Our approach can successfully distinguish seizures from non-seizure activity, with up to 0.83 AUC

[*]Corresponding author. ilkyyldz95@gmail.com (. Yıldız).

on intracranial recordings. Moreover, our algorithm has the potential for real-time inference, by processing at least 10 s of EEG in a second.

Conclusion: We take the first successful steps in deep learning-based unsupervised seizure identification on raw EEG. Our approach has the potential of alleviating the burden on clinical experts regarding laborsome EEG inspections for seizures. Furthermore, aiding the identification of early seizures via our method could facilitate successful detection of epilepsy development and initiate antiepileptogenic therapies.

Keywords

1. Introduction

Epilepsy is one of the most common neurological disorders, affecting over 70 million people worldwide [1]. Epilepsy development is typically identified via seizures, involving uncontrolled jerking movements or momentary losses of awareness due to abnormal excessive or synchronous activities in the brain [2]. The degraded quality of life for patients strongly motivates seizure identification, as early seizures have been shown as prognostic markers for later epileptogenic development [3]. Successful identification of early seizures can facilitate early detection of epilepsy development, and in turn, can initiate antiepileptogenic intervention and therapies that can remarkably improve the quality of life for patients and their caregivers. To this end, electroencephalogram (EEG) recordings received particular attention for identifying seizure segments on EEG [3], due to their routine and low expense collection compared to, e.g., neuroimaging. Seizure on EEG is defined as generalized spike-wave discharges at three per second or faster, and clearly evolving discharges of any type that reach a frequency of four per second or faster [4].

Despite their volume and richness, EEG recordings are prevalently known to contain many artifacts other than seizure due to movement, physiological activity such as perspiration, and measurement hardware [5]. The stochastic nature of EEG makes seizure identification via manual inspections laborsome and difficult, leading to significant variability across clinical labels of different experts [6]. This challenge motivated the recent literature to focus on automated identification of epileptic seizures on EEG as a promising complement to manual inspection.

The literature on automated seizure identification on EEG is vast (c.f. Section 2), focusing mostly on well-established *supervised* machine learning methods. Supervised methods employ both spatiotemporal feature extraction followed by classification algorithms [7,8,9,10], as well as deep learning algorithms applied on raw time-series without feature extraction [11,12,13,14]. Despite their success, these methods require expert labels indicating EEG segments that contain seizures, which are difficult to obtain due to the stochastic nature of clinically-acquired EEG [6].

Unsupervised machine learning methods that do not rely on labeled data have not yet been widely explored. A few shallow machine learning models, including K-means

clustering, Hierarchical clustering, and Gaussian mixture models have been applied on both raw EEG [15], as well as extracted features [16,17,18]. Recently, You et al. [19] implemented an unsupervised deep learning method for seizure identification on EEG, albeit requiring manual feature extraction prior to training. They preprocess EEG to extract time-frequency spectrogram images and train a generative adversarial network (GAN) [20] on the spectrograms that do not contain seizures. As the GAN is trained with non-seizure activity, test spectrograms that significantly differ from the spectrograms generated by GAN are identified to contain seizures.

To the best of our knowledge, an unsupervised deep learning method that *does not* require manual feature extraction has not yet been studied. To this end, we make the following contributions:

- Our main contribution is to propose the first *fully-unsupervised* deep learning method for seizure identification on *raw* EEG. To this end, we train a variational autoencoder (VAE) [21] on EEG recordings that *do not* contain seizures. As training captures non-seizure activity, we identify seizures with respect to (w.r.t.) the median of reconstruction errors at inference time.

- We extend the traditional training objective of VAE with a sparsity-enforcing loss function to suppress EEG artifacts, motivated by similar applications in other domains including probabilistic background modeling [22], and anomaly detection in energy-time series [23].

We validate the seizure identification performance of our method on three publicly available benchmark EEG datasets: (i) intracranial recordings from the University of Pennsylvania and the Mayo Clinic, (ii) scalp recordings from the Temple University Hospital of Philadelphia, and (iii) scalp recordings from the Massachusetts Institute of Technology and the Boston Children's Hospital. Our VAE-based unsupervised approach can successfully distinguish between non-seizure vs. seizure windows and consistently outperforms clustering. Particularly on intracranial recordings, we attain *0.83 Area under the Receiver Operating Characteristics Curve (AUC)*, outperforming state-of-the-art *supervised* methods. We further demonstrate that our algorithm has the potential of performing real-time inference, as it can compute seizure evidence scores over at least 10 s of EEG in a second.

Our method establishes the first successful steps in deep learning-based unsupervised seizure identification, without the need for ground-truth expert labels indicating seizure segments or manual feature extraction prior to seizure identification. Thus, our approach is amenable to save time and effort for both clinical experts, as well as the scientists that devise automated identification methods to aid experts. Aiding the identification of early seizures via our method could facilitate successful and early detection of epilepsy development, and in turn, initiate clinical trials of antiepileptogenic therapies.

The remainder of this paper is organized as follows. Section 2 introduces the literature on automated seizure identification on EEG. Section 3 formalizes our VAE-based unsupervised method. Section 4 presents our experimental results. Section 5 discusses our results in

relation to the recent literature. We make concluding remarks and present future directions in Section 6.

2. Related work

The literature on automated seizure identification on EEG is vast, we refer the reader to the review by Boonyakitanont et al. [24] for more details. A significant body of works focus on extracting spatiotemporal features on EEG via, e.g., wavelet transform [7,25,26], Fourier transform [8,9], power spectra [10], and reconstructed phase space images [27]. Extracted features are used to train *supervised* machine learning algorithms to identify whether a given EEG segment contains a seizure or not. These algorithms employ both shallow models, including support vector machines, decision trees, and nearest neighbour methods, as well as deep learning models, including convolutional neural networks (CNN).

Deep learning-based supervised seizure identification methods lately dominated the literature [28,29], reducing the need for manual feature extraction. Deep models further improved in combination with long-short term memory (LSTM) networks to aid time-series modeling [11], adversarial training to generalize identification across patients [14], autoencoder-based feature extraction [30], attention mechanisms [13], and transformer architectures that improve predictions and interpretability [31,12,32]. In Section 5, we expand upon the state-of-the-art supervised methods applied on the same datasets we employ and compare our approach with them.

All in all, the literature on automated seizure identification often focused on supervised machine learning, with well-established methods over several benchmark EEG datasets. Despite their success, these methods require expert labels indicating EEG segments that contain seizures, which are difficult to obtain due to the stochastic nature of EEG [6]. Meanwhile, few methods employed unsupervised learning via K-means clustering, Hierarchical clustering, and Gaussian mixture models, on both raw EEG [15], as well as extracted features [16,17]. Charupanit et al. [18] applied hierarchical clustering on High Frequency Oscillations (HFOs), which correlate with epilepsy development and have been found to be prone to false detections. Overall, unsupervised seizure identification methods that do not rely on labeled data have not yet been widely explored compared to supervised learning, except for a few shallow machine learning models.

Recently, You et al. [19] implemented an unsupervised deep learning method for seizure identification on EEG, albeit requiring manual feature extraction prior to training. They preprocess EEG to extract spectrogram images and trains a GAN on the spectrograms that do not contain seizures. For each spectrogram at testing time, they have to search for the latent GAN input that leads to the smallest loss value, and use the corresponding generated spectrogram for seizure identification. As the GAN is trained with non-seizure activity, test spectrograms that significantly differ from the spectrograms generated by GAN are successfully identified to contain seizures. We differ from You et al. [19] by applying a fully-unsupervised VAE on *raw* EEG. Our seizure identification metric is based on reconstruction errors made by the VAE, which is trained on non-seizure activity and does not require a sophisticated minimax optimization such as GAN training. Moreover,

we implement the method by You et al. [19] (c.f. Section 4.4) on publicly available EEG datasets and observe that the training objective diverges.

3. Problem formulation

We consider a dataset of N EEG recordings, each collected from M electrode channels and consisting of T time points. Formally, we denote each EEG recording by $X^{(i)} \in \mathbb{R}^{M \times T}$, for $i \in [1, ..., N]$. Our aim is to design an unsupervised method that does not rely on ground-truth expert labels during learning and can identify the existence of seizures in a given EEG recording. To this end, we employ a variational autoencoder (VAE) neural network architecture [21], trained with a sparsity-enforcing loss function to suppress EEG artifacts (c.f. Section 3.2). Our main contribution is to propose the first fully-unsupervised deep learning method that can identify seizures on raw EEG. Note that our method naturally generalizes to EEG recordings comprising different number of time points and channels; we refer the reader to our preprocessing setup in Section 4.2.

3.1. Variational autoencoder

A VAE extracts low-dimensional stochastic latent features $z \in \mathbb{R}^D$ that govern the generation of all samples in a given dataset [21]. Latent features are sampled from a Gaussian distribution with diagonal covariance, in which standard deviation varies by each of the D dimensions. A VAE contains an encoder network $g_\phi : \mathbb{R}^{M \times T} \to \mathbb{R}^{2 \times D}$ and a decoder network $p_\theta : \mathbb{R}^D \to \mathbb{R}^{M \times T}$, with trainable parameters $\phi \in \mathbb{R}^{d_g}$ and $\theta \in \mathbb{R}^{d_p}$, respectively. The encoder receives a sample $X^{(i)} \in \mathbb{R}^{M \times T}$ and predicts the mean $\mu \in \mathbb{R}^D$ and standard deviation $\sigma \in \mathbb{R}^D_+$ of the Gaussian distribution generating latent features. The decoder samples a latent feature $z \in \mathbb{R}^D$ from the predicted Gaussian distribution $\mathcal{N}(\mu, \text{diag}(\sigma))$, with diag denoting diagonalization, and reconstructs a data sample $\widehat{X}^{(i)} \in \mathbb{R}^{M \times T}$. Traditional VAE training aims for input and reconstructed samples to have the same probability distribution generated from the latent features. Particularly, given an input $X^{(i)}$, a traditional VAE is trained by minimizing the following objective w.r.t. $\phi \in \mathbb{R}^{d_g}, \theta \in \mathbb{R}^{d_p}$:

$$\frac{1}{2} \sum_{j=1}^{D} \left(\sigma_j^2 + \mu_j^2 - \log \sigma_j^2 - 1 \right) - \frac{1}{L} \sum_{l=1}^{L} \log p_\theta \left(X^{(i)} \mid z^{(l)} \right), \tag{1}$$

in which $l \in [1, ..., L]$ is the index of each latent feature $z^{(l)}$ sampled from $\mathcal{N}(\mu, \text{diag}(\sigma))$.

The second term in Eq. (1) performs maximum-likelihood estimation of parameters θ under the generative model $p_\theta(X \mid z)$, while the first term enforces the encoder distribution g_ϕ to be similar to the prior distribution $\mathcal{N}(\mu, \text{diag}(\sigma))$ of the latent features. As sampling is not a continuous operation within the training process, Kingma and Welling [21] employ the reparametrization trick: They first sample an auxiliary variable $\epsilon^{(l)}$ from the standard Gaussian $\mathcal{N}(0, I)$, and reparametrize $\epsilon^{(l)}$ to obtain $z^{(l)} = \mu + \sigma \otimes \epsilon^{(l)}$, where represents elementwise product.

3.2. Sparsity-enforcing loss function

As also discussed in Section 1, EEG recordings are prevalently known to contain many artifacts other than seizures due to movement, physiological activity such as perspiration, and measurement hardware [5]. To this end, we assume that VAE reconstructions introduce additive error w.r.t. inputs, where these errors follow a Laplace distribution with identity covariance. It is well-known that maximum likelihood estimation in this setting is equivalent to minimizing the ℓ_1-norm of the difference between input and reconstructed recordings [33], which is a standard technique for outlier and artifact suppression [34]. Motivated by this observation, we replace the second term of the traditional VAE loss function (1) by the ℓ_1-norm of the reconstruction error. The resulting training objective of our VAE architecture is:

$$\frac{1}{2} \sum_{j=1}^{D} \left(\sigma_j^2 + \mu_j^2 - \log \sigma_j^2 - 1 \right) - \frac{1}{L} \sum_{l=1}^{L} \left\| \widehat{X}^{(i,l)} - X^{(i)} \right\|_1, \tag{2}$$

where $\widehat{X}^{(i,l)}$ is the reconstruction from the latent feature sample $z^{(l)}$ for each $l \in [1, \dots, L]$. Note that VAE training with sparse reconstruction errors has been motivated and employed in other domains, including probabilistic background modeling [22], and anomaly detection in energy-time series [23].

3.3. Seizure identification

We aim to employ the trained VAE to distinguish between EEG recordings that contain seizures and those which do not. Thus, we train our VAE architecture on recordings that *do not* contain seizures, using the sparsity-enforcing loss function (2) to suppress EEG artifacts. This allows for the learned latent features to capture non-seizure activity rather than seizures [19]. In real-life applications, EEG data with no seizure activity can be easily augmented with recordings from healthy individuals, which are much more commonly accessible compared to patients experiencing seizures.

At inference time, each reconstruction from the trained VAE is compared with the corresponding input recording. As training captures non-seizure activity, recordings with no seizures are expected to be reconstructed with low error. Meanwhile, a larger reconstruction error w.r.t. the input recording indicates evidence for a seizure. Our overall unsupervised identification algorithm is summarized in Algorithm 1. We explain our exact metric for seizure identification in Section 4.5.

4.   Experiments

4.1. Datasets

We evaluate our method on three publicly available benchmark EEG datasets collected at: (i) the University of Pennsylvania and the Mayo Clinic [35], (ii) the Temple University Hospital of Philadelphia (TUH) [36], and (iii) the Massachusetts Institute of Technology (MIT) and the Boston Children's Hospital [37].

The UPenn dataset contains 1 s long EEG recordings of 8 patients, acquired intracranially at $500 - 5000$ Hz from a maximum of $M = 72$ channels. 1307 recordings correspond to consistent seizures. The total duration of non-seizure recordings is 7164 s and seizure recordings is 653 s.

The TUH dataset contains continuous EEG recordings of 10,874 patients, acquired on the scalp with 250 Hz sampling rate from a maximum of $M = 38$ channels. 1229 seizure recordings were labeled w.r.t. their start and end times. The total duration of non-seizure recordings is 49,922 s and seizure recordings is 2600 s.

The MIT dataset contains continuous EEG recordings of 24 patients, acquired on scalp with 256 Hz sampling rate from a maximum of $M = 38$ channels. 198 seizure recordings were labeled w.r.t. their start and end times. The total duration of non-seizure recordings is 40,800 s and seizure recordings is 2889 s.

4.2. Preprocessing

EEG recordings are typically preprocessed before the application of any analysis [19] to eliminate the powerline noise at 60 Hz. We first unify the sampling rates in each dataset by downsampling to the smallest sampling rate across all recordings. Then, we filter the recordings via a 4th order Butterworth bandpass filter with frequency range 0.5–50 Hz.

Algorithm 1

Our unsupervised seizure identification algorithm. We employ a variational autoencoder architecture comprising an encoder network $g_{\boldsymbol{\phi}}: \mathbb{R}^{M \times T} \to \mathbb{R}^{2 \times D}$ and a decoder network $p_{\boldsymbol{\theta}}: \mathbb{R}^{D} \to \mathbb{R}^{M \times T}$, with trainable parameters $\boldsymbol{\phi} \in \mathbb{R}^{d_g}$ and $\theta \in \mathbb{R}^{d_p}$, respectively. We train our architecture on EEG recordings that *do not* contain any seizures, employing a sparsity-enforcing loss function to suppress EEG artifacts (c.f. Section (3)). As training captures non-seizure activity, we identify seizures w.r.t. the reconstruction errors at inference time.

1:     **procedure** TRAINING ($\boldsymbol{X}^{(i)} \in \mathbb{R}^{M \times T}$ for $i \in$ non-seizure training recordings, $g_{\boldsymbol{\phi}}$, $p_{\boldsymbol{\theta}}$)

2:         Initialize trainable parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$

3:         **repeat**

4:             Sample recording $\boldsymbol{X}^{(i)}$ from non-seizure training recordings

5:             Sample auxiliary variables $\epsilon^{(l)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), l \in \{1, \ldots, L\}$

6:             Reparametrize to obtain latent features $\boldsymbol{z}^{(l)} = \boldsymbol{\mu} + \boldsymbol{\sigma} \otimes \epsilon^{(l)}, l \in \{1, \ldots, L\}$

7:             Compute sparsity-enforcing loss (2) and its gradients w.r.t. $\sim \boldsymbol{\phi}$ and $\boldsymbol{\theta}$

8:             Update trainable parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ via Adam optimization

9:         **until** Loss value (2) converged

10:     **return** Trained $g_{\boldsymbol{\phi}}$, Trained $p_{\boldsymbol{\theta}}$

11:     **end procedure**

1:     **procedure** INFERENCE ($\boldsymbol{X}^{(i)} \in \mathbb{R}^{M \times T}$ for $i \in$ test recordings, Trained $g_{\boldsymbol{\phi}}$, Trained $p_{\boldsymbol{\theta}}$)

2:         **repeat**

3:        Sample recording $\boldsymbol{X}^{(i)}$ from test recordings

4:        Sample auxiliary variables $\epsilon^{(l)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), l \in \{1, ..., L\}$

5:        Reparametrize to obtain latent features $\boldsymbol{z}^{(l)} = \boldsymbol{\mu} + \boldsymbol{\sigma} \otimes \epsilon^{(l)}, l \in \{1, ..., L\}$

6:        Compute $\widehat{\boldsymbol{X}}^{(i, l)}$ for each $\boldsymbol{z}^{(l)}$

7:        Compute decoder reconstruction $\widehat{\boldsymbol{X}}^{(i)}$ by averaging $\widehat{\boldsymbol{X}}^{(i, l)}$ over $l \in \{1, ..., L\}$

8:        Compute seizure evidence score (3) w.r.t.~the reconstruction error between $\boldsymbol{X}^{(i)}$ and $\widehat{\boldsymbol{X}}^{(i)}$

9:     **until** All recordings are tested

10:    **return** Seizure evidence scores for all test recordings

11:    **end procedure**

To attain samples with the same length, we extract sliding windows over each recording, where each window contains T time points and overlaps with its consecutive window by 50%. We choose T based on the shortest seizure segment in each dataset, so as not to omit even the shortest seizures. In doing so, $T = 500$ for UPenn, $T = 462$ for TUH, and $T = 1536$ for MIT. This process results in 14,329 windows with non-seizure activity and 1307 windows with seizure activity for UPenn, 54,264 windows with non-seizure activity and 2826 windows with seizure activity for TUH, and 13,600 windows with non-seizure activity and 963 windows with seizure activity for MIT. In real-life applications, a typical minimum seizure window length can be declared by clinical experts, as in UPenn that directly provides 1 second-long seizure recordings. The input window length of the VAE architecture would be accordingly updated, without a technical change in our approach.

Moreover, we aim to consistently attain $M \times T$ size windows, while not disregarding any channels with potential seizure activity. Thus, to attain samples with the same number of M channels, we reuse data from other channels for the recordings that have missing data at certain channels, compared to the recording with the largest number of channels in each dataset. Again, in real-life applications, clinical experts can determine which and how many channels to employ or discard for seizure identification.

Finally, we apply min-max normalization on all windows to aid the convergence of gradient-based training [38]. As a result, each preprocessed dataset contains windows of the form $\boldsymbol{X}^{(i)} \in [0, 1]^{M \times T}, i \in [1, ..., N]$.

4.3. Network architecture and training

Convolutional neural networks have been successfully applied on raw multivariate time-series data alternative to, e.g., recurrent neural networks, in many domains including seizure identification on EEG [14]. Motivated by this, we employ convolutional encoder and decoder networks depicted in Fig. 1. We initialize all weights via Xavier initialization [39] and all biases as 0.01. For regularization, we apply dropout with probability 0.3 [40]. We train our VAE architecture on non-seizure windows via the sparsity-enforcing loss function (2) using Adam optimization [41] for 200 epochs, where $L = 5$. To further emphasize the reconstruction quality of VAE [22], we multiply the first term in Eq. (2) by 0.8. We repeat

our experiments for D varying in $\{16, 64, 256\}$ and learning rate varying in $\{1e^{-2}, 1e^{-3}, 1e^{-4}\}$, and report the results corresponding to the best seizure identification AUC.

4.4. Competing methods

Following the literature on unsupervised seizure identification methods [15], we implement two clustering algorithms as competing methods on raw EEG. We first reduce the dimension of all EEG windows in the test set to 3 using the t-Distributed Stochastic Neighbor Embedding (t-SNE) [42] algorithm. Then, we apply K-means clustering and hierarchical clustering [43] on the resulting windows with two clusters indicating non-seizure and seizure.

We further implement the unsupervised deep learning method by You et al. [19]. Following EEG window extraction described in Section 4.2, we construct a two-sided power spectral density spectrogram for each channel in each window via short-time Fourier transform. Each spectrogram is augmented with the mean of the two sides, resulting in three spectrograms for each channel in each window. Resulting spectograms are scaled to the range $[-1, 1]$ via min-max normalization. We employ a discriminator with 4 convolutional layers, each followed by batch normalization and leaky relu activation, except for the final layer with sigmoid activation. Moreover, we employ a generator with 4 convolutional transpose layers, each followed by batch normalization and leaky relu activation, except for the final layer with tanh activation. We train the discriminator and the generator in alternating turns via cross-entropy loss [20] on the spectograms extracted from non-seizure windows, using Adam optimization with learning rate of 0.0 0 02. Unfortunately, the generator loss consistently diverges and the GAN cannot be properly trained on our datasets. Note that GAN training is commonly difficult to stabilize and converge due to minimax optimization [44], as also reported by You et al. [19].

4.5. Experiment setup

We employ 5-fold cross-validation to partition the non-seizure windows into train and test sets, where each fold contains 80% of all non-seizure windows for training. The remaining 20% of non-seizure windows, along with all seizure windows are used for testing. For each test window i, we compute the reconstruction $\widehat{X}^{(i)}$ by averaging $\widehat{X}^{(i,l)}$ over $l \in \{1, ..., L\}$.

To evaluate the seizure identification performance of the trained VAE, we use the pointwise absolute error between each input window and the corresponding reconstruction. As training does not involve seizures, non-seizure windows are expected to have much lower absolute error compared to seizure windows. We employ the median reconstruction error over the time points in each EEG window, due to its success as a performance metric in anomalous activity identification [45]; we also observe in our experiments that the resulting identification performance is higher than, e.g., using mean error. Finally, the seizure evidence score for each window i is calculated as the maximum of the median absolute error over electrode channels:

$$\max_{m \in [1, M]} \text{median}_{t \in [1, T]} \left| X_{m,t}^{(i)} - \widehat{X}_{m,t}^{(i)} \right|, \tag{3}$$

where subscript m, t represents the value at channel m and time point t.

We compute AUC and p-value under the independent two sample Welch t-test [46] for distinguishing seizure vs. non-seizure windows w.r.t. the evidence score (3). To further compute binary decision metrics, we threshold the score at the value for which the geometric mean of recall and true negative rate is maximum [47]. Using the respective threshold, we calculate accuracy, precision, and recall for binary identification of seizure vs. non-seizure windows. We report all prediction performance metrics as averaged over the test folds, along with the corresponding standard deviations. In real-life applications, decision thresholds may be determined by clinical experts with respect to the desired trade-off between false positives and negatives.

4.6. Execution environment and scalability

All experiments are implemented using the Python 3.6 programming language and the PyTorch library. Training and evaluations are executed on an NVIDIA V100 GPU with 32 GB memory. Our code will be made publicly available as a Github repository upon publication.

Each training epoch over UPenn, TUH, and MIT training sets takes 0.9, 1.7, and 9.4 min, respectively, proportionally and modestly scaling with respect to their relative training set sizes and window dimensions. Having trained our VAE architecture over non-seizure windows, inference stage of our algorithm (c.f. Algorithm 1) takes at most 0.1 s to process a 1 second-long EEG window over the test sets of all datasets. This implies that our method has the potential of performing *real-time inference*, by determining seizure evidence scores over at least 10 s of EEG in a second.

4.7. Seizure identification performance

Table 1 shows the average seizure identification performance metrics of our method vs. t-SNE followed by K-means and hierarchical clustering, along with the corresponding standard deviations. Fig. 2 visualizes the corresponding distributions of the performance metrics of our method over the 5 test folds.

Our VAE-based unsupervised identification method can successfully distinguish between non-seizure vs. seizure windows, with up to *0.83 AUC* on UPenn. Clustering on raw EEG windows *cannot* capture the complex evolution of EEG and identifies all windows as non-seizure; this is indicated by the fact that the statistical difference between seizure evidence score distributions over non-seizure vs. seizure windows attains a p-value of 1.0. As the distribution of non-seizure vs. seizure windows is severely imbalanced, identifying all windows as non-seizure may lead to well above 0.5 accuracy. This is precisely why we also report AUC, which does not require a binary identification threshold. Our method consistently and significantly outperforms clustering over all datasets and performance metrics, further motivating our more sophisticated VAE-based identification method.

Note that the performance difference on MIT and TUH vs. UPenn stems from their acquisition differences; MIT and TUH being collected on scalp compared to

UPenn collected intracranially degrades identification performance due to having limited spatiotemporal resolution and more artifacts [48].

4.8. Seizure identification examples

We visualize example EEG windows and corresponding seizure identifications from UPenn in Fig. 3. We make positive and negative identification decisions using the evidence score threshold described in Section 4.5. Agreeing with the clinical descriptions of seizure, true seizure-positive windows in Fig. 3a contain high-frequency spikes and waves evolving with large amplitude [2]. Meanwhile, true non-seizure windows in Fig. 3b attain significantly less amplitude changes and spikes compared to true positive windows. Rarer spikes as in Patient 2 seizure window from Fig. 3c may be neglected due to employing median absolute error (3). As spikes may indicate both seizure-related behaviour, as well as artifacts such as loose electrode placement or bad conductivity [5], this design choice establishes a trade-off between artifact suppression and successful seizure identification.

Note that the seizure patterns cannot be successfully identified w.r.t. only large amplitude or high frequency, motivating a more complex approach such as ours. For instance, the bottom right seizure window in Fig. 3c have the same amplitude range as the non-seizure windows in Fig. 3a, while the seizure windows on the right in Fig. 3d have subtle and similar spikes with lower frequency such as the non-seizure windows in Fig. 3b and 3d.

4.9. Latent features

To illustrate the discriminative features learned by our VAE architecture on UPenn, we apply t-SNE on the latent mean vectors predicted from all windows and project them onto 3-dimensional space. Fig. 4 shows the resulting latent means w.r.t. each pair of the 3 dimensions for seizure (red) vs. non-seizure (blue). Agreeing with the performance on UPenn in Table 1, the latent features captured by our method can distinguish between non-seizure vs. seizure windows.

5. Discussion

Following the literature on unsupervised seizure identification methods [15], we presented two clustering algorithms applied on raw EEG and illustrated that our VAE-based unsupervised identification method significantly outperforms both. As virtually all recent methods on seizure identification are supervised, we also discuss supervised methods in this section and compare them with our approach. Table 2 summarizes the experimental results of the state-of-the-art supervised methods applied on our datasets.

To begin with intracranial recordings, Sun et al. [30] and Zhu and Shoaran [10] employ the UPenn dataset for supervised seizure identification. Sun et al. [30] extract features from raw EEG using forecasting via an echo state network, which is an extension of recurrent neural networks. Resulting features are used for seizure identification via a support vector machine (SVM), which attains 0.81 AUC over 2 patients. Zhu and Shoaran [10] extract power spectrum features, which are transformed to remove patient-specific content by unsupervised adversarial learning across patients. Patient-independent features are used for cross-patient seizure identification via a decision-tree, which attains 0.71 AUC. Our

VAE-based unsupervised identification method outperforms both Sun et al. [30] and Zhu and Shoaran [10], by attaining 0.83 AUC and 0.79 Accuracy on UPenn. In doing so, our method does not require ground-truth seizure labels for training and manual feature extraction prior to seizure identification.

Focusing on scalp recordings, Zhang et al. [14] and Li et al. [13] employ the TUH dataset for supervised seizure identification. Zhang et al. [14] extract patient-independent features on raw EEG via adversarial learning across patients, paired with an attention mechanism to weigh importance over channels. Extracted features are used to train a CNN architecture for seizure identification, which attains 0.8 Accuracy over 14 patients. Li et al. [13] employ another attention-infused classification architecture called Squeeze and-Excitation Network on raw EEG, which weighs importance over channels and across time. The resulting architecture attains 0.92 Accuracy for seizure identification. Having noted the success of these deep learning approaches and attention mechanism on raw EEG, Zhang et al. [14] and Li et al. [13] discard a considerable number of short seizures that do not occur at many channels. Zhang et al. [14] discard recordings in which seizures occur on less than 12 channels or seizures last less than 250 s. Meanwhile, the shortest seizure duration in TUH is 1.84 s and there are recordings in which seizures occur on only one or a few channels [49]. Li et al. [13] discard seizure events that last less than 4 s and consider only 20 common channels across all recordings, although we find that recordings contain up to 38 channels. Despite the more limited dataset considered, our VAE-based identification method attains similar accuracy to Zhang et al. [14], while not requiring any label supervision in training.

Last but not least, Mehla et al. [8] and Chakrabarti et al. [11] employ the scalp dataset from MIT for supervised seizure identification. Mehla et al. [8] extract features from raw EEG using vector norms computed from Fourier intrinsic band functions. Extracted features are used by an SVM classifier for seizure identification that attains 0.99 Accuracy. Chakrabarti et al. [11] apply a channel-independent LSTM classifier on raw EEG and also attains 0.99 Accuracy. Having noted the success of these deep learning approaches with and without manual feature extraction, Mehla et al. [8] and Chakrabarti et al. [11] balance the distribution of non-seizure and seizure windows prior to data partitioning and algorithm development. This process not only aids predictions by removing overfitting due to severe class imbalance [50], but also hinders applicability in real-life where the distribution of non-seizure vs. seizure windows is unknown.

All in all, our novel unsupervised approach attains state-of-the art seizure identification performance on intracranial recordings. We also recognize that the performance of our method over scalp recordings, particularly MIT, is not as high as the well-established supervised methods applied on the same datasets. That said, our method requires *no ground-truth labels and manual feature extraction* for training, saving time and effort for both clinical experts, as well as the scientists that devise automated identification methods to aid experts. In doing so, our method naturally benefits from no expert supervision, unlike supervised learning. Moreover, our approach does not disregard any channels or windows based on seizure length or class distribution, making it less restrictive for real-life applications.

6. Conclusions and future work

We propose the first fully-unsupervised deep learning method for seizure identification on raw EEG, employing a VAE architecture. Our method captures the non-seizure activity without ground-truth seizure labels and manual feature extraction for training, saving time and effort for both clinical experts, as well as the scientists that devise automated identification methods to aid experts. Following training, we identify seizure activity based on the reconstruction errors of VAE. Our method can successfully distinguish between non-seizure vs. seizure windows and consistently outperforms clustering. Particularly on intracranial recordings, we attain 0.83 AUC, outperforming state-of-the-art *supervised* methods. Moreover, our approach has the potential of performing real-time inference, as it can compute seizure evidence scores over at least 10 s of EEG in a second.

Aiding the identification of seizures via our method could facilitate early and successful detection of epilepsy development, as early seizures can be prognostic markers for later epileptogenic development [3]; this could in turn initiate successful clinical trials of antiepileptogenic therapies. Moreover, our method is designed not only to differentiate non-seizure vs. seizure windows, but more generally to differentiate anomalous activities on EEG. For example, when trained on EEG collected from healthy patients, our method can be applied to identify other epileptic activities such as periodic discharges. Overall, our unsupervised approach is not limited to seizure identification, and can thus, be easily generalized to other applications involving anomalous activity detection on multivariate time-series data such as EEG.

Further improvements on training and architecture designs are amenable to improve the identification performance, particularly over scalp recordings. For instance, finding a more sophisticated encoder-decoder architecture via neural architecture search, and a training procedure via, e.g., learning rate scheduling, are likely to aid performance [50]. Meanwhile, these design choices are beyond our main contribution in establishing the first fully-unsupervised deep learning method for seizure identification on raw EEG, and require a more extensive search for training and model optimization. Thus, we present our method as a novel proof-of-concept and leave potential experimental improvements as future work.

Beyond identification, unsupervised longitudinal prediction of seizures [51] also remains an open direction that would also aid identification performance, as our current approach does not capture the time stamps of EEG windows. Given the EEG history of a patient, extending our method to predict when the patient will experience seizures via, e.g., incorporating recurrent units to capture the evolution of latent features, is a promising future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

[1]. Thijs RD, Surges R, O'Brien TJ, Sander JW, Epilepsy in adults, Lancet 393 (10172) (2019) 689–701. [PubMed: 30686584]

[2]. Engel J, Seizures and epilepsy, volume 83, Oxford University Press, 2013.

[3]. Vespa PM, Shrestha V, Abend N, Agoston D, Au A, Bell MJ, Bleck TP, Blanco MB, Claassen J, Diaz-Arrastia R, et al. , The epilepsy bioinformatics study for anti-epileptogenic therapy (EpiBioS4Rx) clinical biomarker: study design and protocol, Neurobiol. Dis. 123 (2019) 110–114. [PubMed: 30048805]

[4]. Hirsch L, LaRoche S, Gaspard N, Gerard E, Svoronos A, Herman S, Mani R, Arif H, Jette N, Minazad Y, et al. , American clinical neurophysiology society standardized critical care EEG terminology: 2012 version, J. Clin. Neurophysiol. 30 (1) (2013) 1–27. [PubMed: 23377439]

[5]. Saba-Sadiya S, Chantland E, Alhanai T, Liu T, Ghassemi MM, Unsupervised EEG artifact detection and correction, Front. Digit. Health 2 (2021) 57.

[6]. Zhang T, Chen W, LMD based features for the automatic seizure detection of EEG signals using SVM, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (8) (2016) 1100–1108. [PubMed: 27662677]

[7]. Amin HU, Yusoff MZ, Ahmad RF, A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques, Biomed. Signal Process. Control 56 (2020) 101707.

[8]. Mehla VK, Singhal A, Singh P, Pachori RB, An efficient method for identification of epileptic seizures from EEG signals using fourier analysis, Phys. Eng. Sci. Med. (2021) 1–14. [PubMed: 33559037]

[9]. Ramos-Aguilar R, Olvera-López JA, Olmos-Pineda I, Sánchez-Urrieta S, Feature extraction from EEG spectrograms for epileptic seizure detection, Pattern Recognit. Lett. 133 (2020) 202–209.

[10]. Zhu B, Shoaran M, Unsupervised domain adaptation for cross-subject few-shot neurological symptom detection, in: Proceedings of the 10th International IEEE/EMBS Conference on Neural Engineering (NER), 2021, pp. 181–184, doi: 10.1109/NER49283.2021.9441235.

[11]. Chakrabarti S, Swetapadma A, Pattnaik PK, A channel independent generalized seizure detection method for pediatric epileptic seizures, Comput. Methods Progr. Biomed. 209 (2021) 106335, doi: 10.1016/j.cmpb.2021.106335.

[12]. Kostas D, Aroca-Ouellette S, Rudzicz F, Bendr: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data, (2021) arXiv preprint arXiv:2101.12037.

[13]. Li Y, Liu Y, Cui W-G, Guo Y-Z, Huang H, Hu Z-Y, Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (4) (2020) 782–794. [PubMed: 32078551]

[14]. Zhang X, Yao L, Dong M, Liu Z, Zhang Y, Li Y, Adversarial representation learning for robust patient-independent epileptic seizure detection, IEEE J. Biomed. Health Inform. 24 (10) (2020) 2852–2859. [PubMed: 32071011]

[15]. Chakrabarti S, Swetapadma A, Pattnaik PK, Samajdar T, Pediatric seizure prediction from EEG signals based on unsupervised learning techniques using various distance measures, in: Proceedings of the 1st International Conference on Electronics, Materials Engineering and Nano-Technology, 2017, pp. 1–5, doi: 10.1109/IEMENTECH.2017.8076983.

[16]. Belhadj S, Attia A, Adnane AB, Ahmed-Foitih Z, Taleb AA, Whole brain epileptic seizure detection using unsupervised classification, in: Proceedings of the 8th International Conference on Modelling, Identification and Control, 2016, pp. 977–982, doi: 10.1109/ICMIC.2016.7804256.

[17]. Birjandtalab J, Pouyan MB, Nourani M, Unsupervised EEG analysis for automated epileptic seizure detection, in: Proceedings of the First International Workshop on Pattern Recognition, volume 10011, International Society for Optics and Photonics, 2016, p. 100110M.

[18]. Charupanit K, Sen-Gupta I, Lin JJ, Lopour BA, Detection of anomalous high-frequency events in human intracranial EEG, Epilepsia Open 5 (2) (2020) 263–273. [PubMed: 32524052]

[19]. You S, Cho BH, Yook S, Kim JY, Shon YM, Seo DW, Kim IY, Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear EEG using an anomaly-detecting generative adversarial network, Comput. Methods Progr. Biomed. 193 (2020) 105472.

[20]. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[21]. Kingma DP, Welling M, Auto-encoding variational (2013) bayes, arXiv preprint arXiv:1312.6114.

[22]. Farnoosh A, Rezaei B, Ostadabbas S, DeepPBM: deep probabilistic background model estimation from video sequences (2019) arXiv preprint:1902.00820.

[23]. Pereira J, Silveira M, Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention, in: Proceedings of the International Conference on Machine Learning and Applications, IEEE, 2018, pp. 1275–1282.

[24]. Boonyakitanont P, Lek-Uthai A, Chomtho K, Songsiri J, A review of feature extraction and performance evaluation in epileptic seizure detection using EEG, Biomed. Signal Process. Control 57 (2020) 101702.

[25]. Carrera EV, Quinga F, Analysis of epileptic seizure predictions based on intracranial EEG records, in: Proceedings of the IEEE Colombian Conference on Communications and Computing, IEEE, 2018, pp. 1–5.

[26]. Radman M, Moradi M, Chaibakhsh A, Kordestani M, Saif M, Multi-feature fusion approach for epileptic seizure detection from EEG signals, IEEE Sens. J. 21 (3) (2020) 3533–3543.

[27]. Ilakiyaselvan N, Khan AN, Shahina A, Deep learning approach to detect seizure using reconstructed phase space images, J. Biomed. Res. 34 (3) (2020) 240. [PubMed: 32561702]

[28]. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (2018) 270–278. [PubMed: 28974302]

[29]. Zhao W, Zhao W, Wang W, Jiang X, Zhang X, Peng Y, Zhang B, Zhang G, A novel deep neural network for robust detection of seizures using EEG signals, Comput. Math. Methods Med. 2020 (2020).

[30]. Sun L, Jin B, Yang H, Tong J, Liu C, Xiong H, Unsupervised EEG feature extraction based on echo state network, Inf. Sci. 475 (2019) 1–17.

[31]. Eldele E, Ragab M, Chen Z, Wu M, Kwoh CK, Li X, Guan C, Time-series representation learning via temporal and contextual contrasting, (2021) arXiv preprint arXiv:2106.14112.

[32]. Mohsenvand MN, Izadi MR, Maes P, Contrastive representation learning for electroencephalogram classification, in: Proceedings of the Machine Learning for Health, PMLR, 2020, pp. 238–253.

[33]. Tibshirani R, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B (Methodol.) 58 (1) (1996) 267–288.

[34]. Wright J, Ganesh A, Rao SR, Peng Y, Ma Y, Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization, in: Proceedings of the Neural Information Processing Systems, volume 58, 2009.

[35]. UPenn MayoClinic, Upenn and Mayo Clinic's seizure detection challenge, 2014, https://www.kaggle.com/c/seizure-detection/.

[36]. Obeid I, Picone J, The temple university hospital EEG data corpus, Front. Neurosci. 10 (2016) 196, doi: 10.3389/fnins.2016.00196. [PubMed: 27242402]

[37]. Shoeb AH, Application of machine learning to epileptic seizure onset detection and treatment, Massachusetts Institute of Technology, 2009 Ph.D. thesis.

[38]. Ioffe S, Szegedy C, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: Proceedings of the International Conference on Machine Learning, PMLR, 2015, pp. 448–456.

[39]. Glorot X, Bengio Y, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[40]. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[41]. Kingma DP, Ba J, Adam: a method for stochastic optimization, (2014) arXiv preprint arXiv:1412.6980.

[42]. Van der Maaten L, Hinton G, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).

[43]. Bishop CM, Pattern recognition, Mach. Learn. 128 (9) (2006).

[44]. Ham H, Jun TJ, Kim D, Unbalanced GANs: pre-training the generator of generative adversarial network using variational autoencoder (2020) arXiv preprint arXiv:2002.02112.

[45]. Hochenbaum J, Vallis OS, Kejariwal A, Automatic anomaly detection in the cloud via statistical learning (2017) arXiv preprint arXiv:1704.07706.

[46]. Peck R, Olsen C, Devore JL, Introduction to Statistics and Data Analysis, Cengage Learning, 2015.

[47]. Fawcett T, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.

[48]. Ramantani G, Maillard L, Koessler L, Correlation of invasive EEG and scalp EEG, Seizure 41 (2016) 196–200. [PubMed: 27324839]

[49]. Harati A, Lopez S, Obeid I, Picone J, Jacobson M, Tobochnik S, The TUH EEG corpus: a big data resource for automated EEG interpretation, in: Proceedings of the Signal Processing in Medicine and Biology Symposium, IEEE, 2014, pp. 1–5.

[50]. Goodfellow I, Bengio Y, Courville A, Deep Learning MIT press, 2016.

[51]. Dissanayake T, Fernando T, Denman S, Sridharan S, Fookes C, Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals, IEEE Sens. J. 21 (7) (2021) 9377–9388.
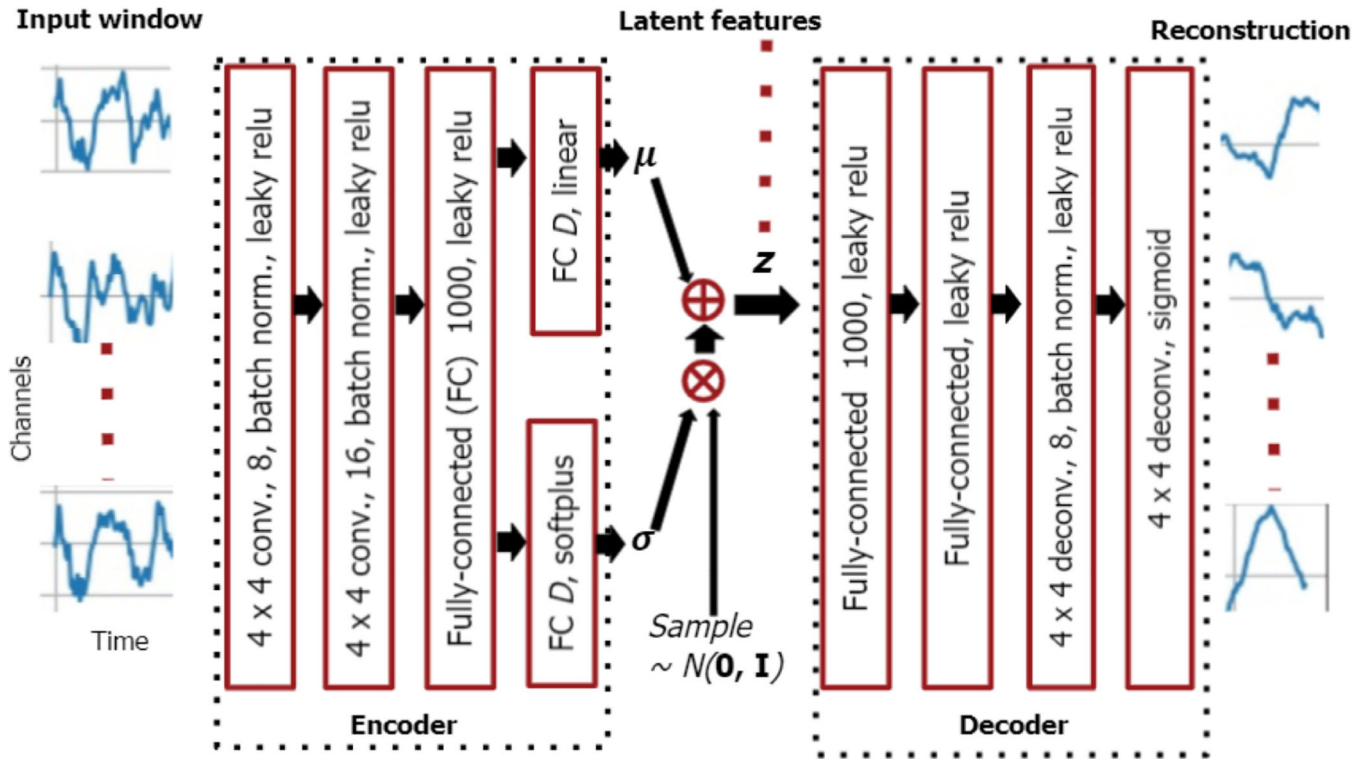
Fig. 1.
Our Variational Autoencoder architecture. The encoder contains convolutional (conv.), batch-normalization (batch norm.), and fully-connected (FC) layers for latent feature extraction, while the decoder contains convolutional transpose (deconv.) and FC layers for upsampling and reconstruction [50]. Conv. and deconv. layers apply 4×4 convolutional filters, with the number of filter channels written next to the filter size. FC layers are described via their output dimension. For each layer, the activation function is written in the end of the corresponding description.
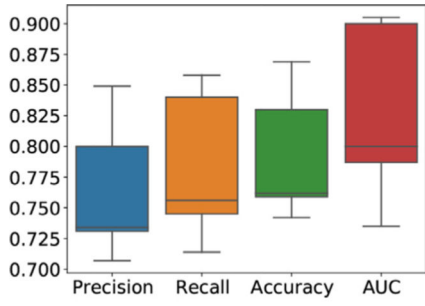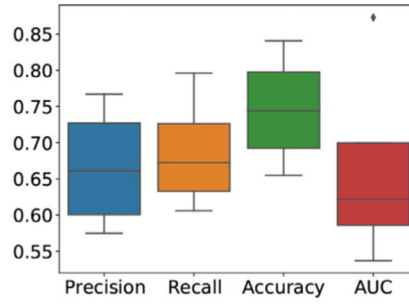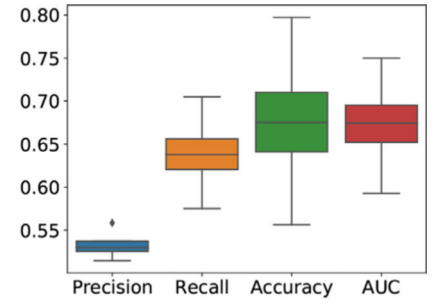
(a) UPenn   (b) TUH   (c) MIT

Fig. 2.

Distributions of the seizure identification performance metrics of our VAE-based unsupervised method. For each metric, the line inside each box indicates the median, upper and lower limits of each box indicate the upper and lower quartiles, and upper and lower limits of each vertical line indicate the maximum and minimums attained over the 5 test folds.
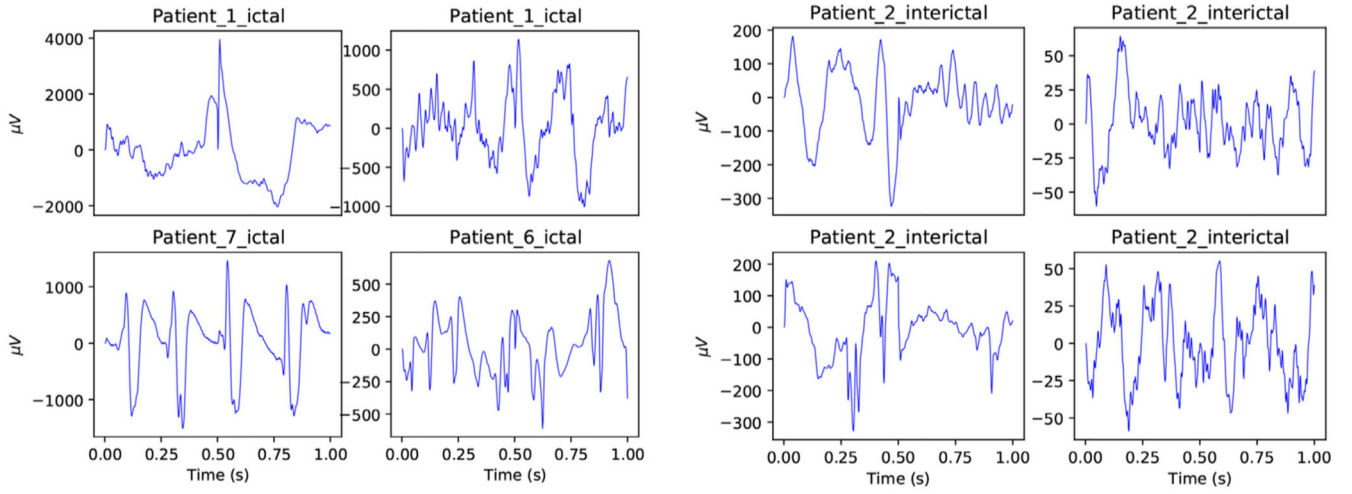
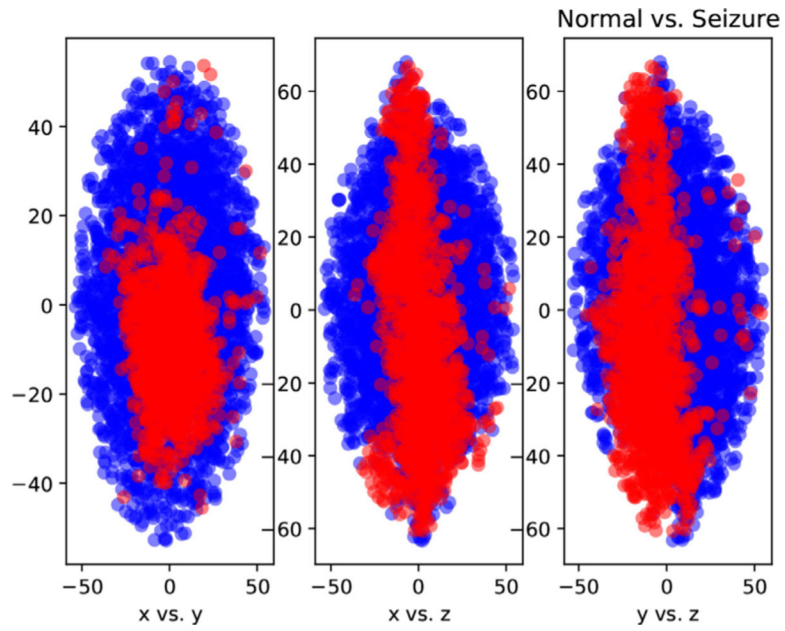(a) Correctly Identified Seizure Windows (True Positive)

(b) Correctly Identified Non-seizure Windows (True Negative)
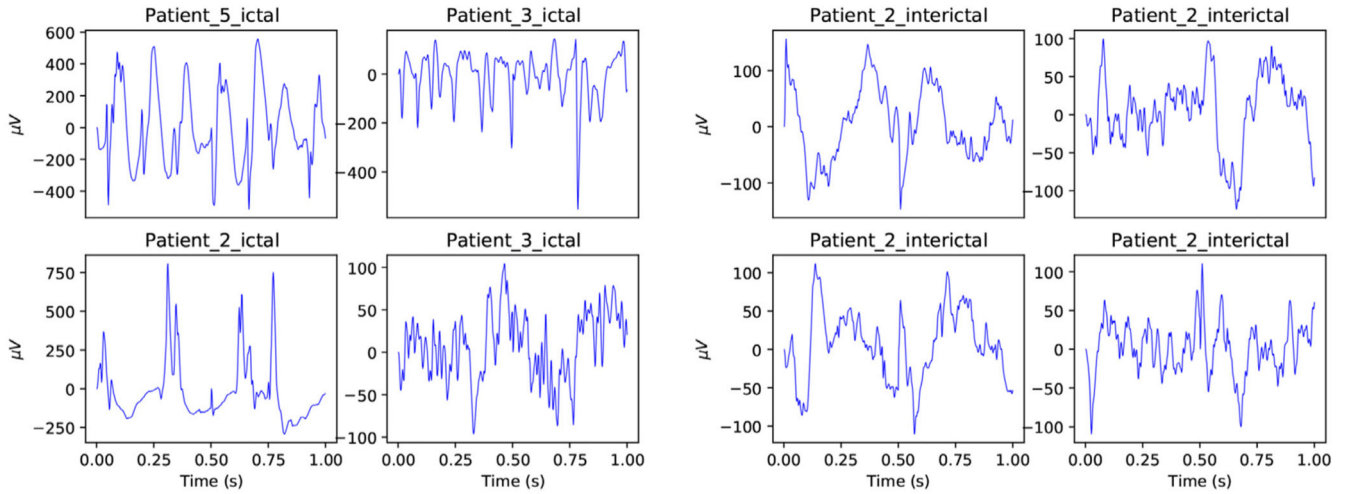
(c) Falsely Identified Seizure Windows (False Negative)

(d) Falsely Identified Non-seizure Windows (False Positive)

Fig. 3.
Example EEG windows and corresponding seizure identifications on UPenn.

Fig. 4.
Latent means predicted from seizure (red) vs. non-seizure (blue) windows on UPenn w.r.t. each pair of 3 dimensions. Dimension is reduced from $D = 64$ to 3 using the t-SNE algorithm.

Table 1

Seizure identification performance metrics of our method (VAE) vs. t-SNE followed by K-means and hierarchical clustering. We report all prediction performance metrics as averaged over the 5 test folds, along with the corresponding standard deviations. Lower p-values indicate higher significance in identification performance.

Dataset	Method	Precision	Recall	Accuracy	AUC	p-Value $\downarrow$
UPenn	**VAE**	**0.76 ± 0.05**	**0.78 ± 0.05**	**0.79 ± 0.04**	**0.83 ± 0.06**	**$1e^{-106}$**
	K-means	0.33 ± 0.0	0.5 ± 0.0	0.5 ± 0.0	0.56 ± 0.0	1.0
	Hierarchical	0.33 ± 0.0	0.5 ± 0.0	0.67 ± 0.0	0.5 ± 0.0	1.0
TUH	**VAE**	**0.67 ± 0.08**	**0.69 ± 0.08**	**0.75 ± 0.08**	**0.67 ± 0.13**	**$1e^{-102}$**
	K-means	0.17 ± 0.0	0.5 ± 0.0	0.35 ± 0.0	0.57 ± 0.0	1.0
	Hierarchical	0.17 ± 0.0	0.5 ± 0.0	0.35 ± 0.0	0.45 ± 0.0	1.0
MIT	**VAE**	**0.54 ± 0.02**	**0.64 ± 0.05**	**0.68 ± 0.09**	**0.68 ± 0.06**	**$1e^{-21}$**
	K-means	0.33 ± 0.0	0.5 ± 0.0	0.5 ± 0.0	0.59 ± 0.0	1.0
	Hierarchical	0.33 ± 0.0	0.5 ± 0.0	0.67 ± 0.0	0.53 ± 0.0	1.0

Table 2

Experimental results of the state-of-the-art supervised seizure identification methods applied on the same datasets we employ.

Dataset	Study	Methodology	Seizure Identification Performance
UPenn	Sun et al. [30]	Echo State Network → SVM	0.81 AUC
	Zhu and Shoaran [10]	Power spectra → Patient-independent adversarial learning	0.71 Accuracy
TUH	Zhang et al. [14]	Patient-independent adversarial feature extraction → CNN	0.8 Accuracy
	Li et al. [13]	Spectral-temporal Squeeze and-Excitation Network	0.92 Accuracy
MIT	Mehla et al. [8]	Fourier function norms → SVM	0.99 Accuracy
	Chakrabarti et al. [11]	Channel-independent LSTM	0.99 Accuracy