Intraoperative Electrocorticography Signal Synthesis to Improve the Classification of Epileptiform Tissue

Leonor Almeida[1,*] [a], Sem Hoogteijling[2,3,*] [b], Inês Silveira[1] [c], Dania Furk[1] [d], Irene Heijink[2,3] [e],
Maryse van't Klooster[2] [f], Hugo Gamboa[1] [g], Luís Silva[1] [h] and Maeike Zijlmans[2,3] [i]

[1]*Laboratório de Instrumentação, Engenharia Biomédica e Física da Radiação (LIBPhys-UNL), Departamento de Física, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal*
[2]*Department of Neurology and Neurosurgery, Brain Center, University Medical Center Utrecht, Part of ERN EpiCARE, P.O. box 85500, 3508 GA Utrecht, The Netherlands*
[3]*Stichting Epilepsie Instellingen Nederland (SEIN), The Netherlands*

Keywords: Synthetic Data, Epilepsy, Epileptiform Activity, Epileptogenic Tissue, ioECoG, GAN.

Abstract: Epilepsy surgery is a viable option for treating drug-resistant cases where anti-seizure medications fail, but accurately localizing epileptic tissue remains challenging. This process can be guided by the visual assessment of intraoperative electrocorticography (ioECoG). Data scarcity limits developing machine learning (ML) models for automatic epileptic tissue classification. To address this, we propose a generative model based on Generative Adversarial Networks (GANs) to synthesize realistic ioECoG signals. Our approach identified three distinct ioECoG patterns using Agglomerative Clustering, which guided training individual Deep Convolutional Wasserstein GANs with Gradient Penalty (DCwGAN-GP). Synthetic data (SD) was evaluated across multiple dimensions: fidelity using temporal (e.g., Wasserstein distance (WD)), frequency and time-frequency metrics; diversity through dimensionality reduction; and utility by comparing ML performance with and without SD. It replicated temporal and frequency characteristics of real signals (fidelity), though lacked variability (diversity) due to potential data misclassifications. Specifically, the WD between real and synthetic signals outperformed literature benchmarks (i.e., 0.043 ± 0.025 vs. 0.078). Classifiers trained on a combination of real and SD achieved 88% accuracy, compared to 85% with real data alone. These results demonstrate the potential of SD to replicate real signals, address data scarcity, augment ioECoG datasets, and advance ML-based epilepsy surgery research.

1 INTRODUCTION

Epilepsy is a central nervous system disorder (Shoka et al., 2023) characterized by abnormal brain activity, affecting approximately 50 million people globally (WHO, 2019). While anti-seizure medication is the first line of treatment (Consales et al., 2021), about a third of patients have drug-resistant epilepsy

and do not achieve seizure control through this procedure (Duncan and Taylor, 2023). For these people, epilepsy surgery offers a viable solution, potentially curing seizures and improving their quality of life (Consales et al., 2021). The surgery procedure involves the removal of the epileptogenic tissue, that is, the brain tissue responsible for triggering seizures (Zijlmans et al., 2019). To localize epileptic tissue, pathological electrographic activity in the intra-operative electrocorticogram (ioECoG) can be used, including spikes, sharp waves and ictiform spike patterns or the more recently discovered high frequency oscillations (HFOs) (Fernández and Loddenkemper, 2013) (Greiner et al., 2016) (Wang et al., 2024) (Zweiphenning et al., 2022). This method involves placing electrodes directly on surgically exposed cortex to precisely map the tissue to be removed (Tatum, 2021). However, differentiating epileptiform activ-

[a] https://orcid.org/0009-0003-1185-9750
[b] https://orcid.org/0000-0002-4455-6700
[c] https://orcid.org/0000-0003-0675-3444
[d] https://orcid.org/0009-0008-7139-8665
[e] https://orcid.org/0009-0005-9266-1168
[f] https://orcid.org/0000-0002-6594-8965
[g] https://orcid.org/0000-0002-4022-7424
[h] https://orcid.org/0000-0001-9811-0571
[i] https://orcid.org/0000-0003-1258-5678
* These authors contributed equally to this work.

ity from normal ioECoG patterns remains a challenge (Zweiphenning et al., 2022). Clinicians often rely on visual interpretation, which is prone to human errors and can miss critical information (Li et al., 2020) (Saminu et al., 2022). Also, the typically visible pathological interictal EEG activity may not be optimally predicting the diseased tissue: epileptic spikes are seen over a broader area than just the seizure onset zone and epileptic HFOs can be difficult to discern from artefacts and physiological HFOs. This highlights the need for more reliable ways to distinguish epileptic and non-epileptic tissue.

Recent research shows promise in using ML for automatic detection of interictal epileptiform discharges in intracranial EEG (iEEG) signals (Saminu et al., 2022) by efficiently identifying epilepsy biomarkers such as spikes, improving diagnostic accuracy and assist in epilepsy surgery (Chaibi et al., 2024). A step further is to use ML not to detect short epileptiform events but also background activity that differs between epileptic and non-epileptic tissue (Hoogteijling et al., 2024). Data scarcity limits the effectiveness of ML models, which require large datasets to accurately distinguish epileptiform patterns (Du et al., 2024a). To overcome this, the generation of SD (SD) for medical applications has gained importance not only for augmenting training datasets and therefore improving the performance of ML models in EEG based tasks, but also for addressing patient privacy concerns (Pascual et al., 2021) (Aznan et al., 2019) (Park et al., 2024) (Carrle et al., 2023) (Wu et al., 2024) (Fang et al., 2024) (Nia et al., 2024).

Common data augmentation methods include GANs, Variational Auto-Enconders and Diffusion Models (Liang et al., 2023). While GANs have made significant strides in SD generation, their application in the medical domain, particularly for EEG data, remains challenging due to its high variability, which leads to issues such as mode collapse, complex training dynamics, and instability (Habashi et al., 2023), (Liang et al., 2023). Nevertheless, GANs remains the leading method for SD generation (Liang et al., 2023) (Carrle et al., 2023) (Fan et al., 2024) (Habashi et al., 2023). Over the past decade, several GAN architectures have emerged to address earlier limitations. Conditional GAN (cGAN), introduced in 2014 (Mirza and Osindero, 2014), has been effective in enhancing training stability and reducing mode collapse (Panwar et al., 2020). (Pascual et al., 2021) utilized cGAN for synthesizing 4-second EEG signals, while (Carrle et al., 2023) applied it for depression diagnosis, demonstrating the utility of cGANs in generating targeted SD for specific clinical applications. Similarly, (Aznan et al., 2019) used Deep Convolutional GANs

(DCGAN) to generate 3-second dry-EEG signals for a Steady State Visually Evoked Potential task. Other variations, such as Wasserstein GANs with Gradient Penalty (wGAN-GP), offer even more stable gradients and robust training, resulting in higher-quality SD (Fang et al., 2024) (Nia et al., 2024). Recent advancements have integrated Long Short Term Memory (LSTM) layers (Du et al., 2024b) or U-Net generators (Pascual et al., 2019) (Pan et al., 2023) to better capture time-series characteristics. Auxiliary decoding techniques, such as those used by (Liang et al., 2023) and (Aznan et al., 2020) further improve signal quality by classifying input data alongside distinguishing real from SD.

(Hartmann et al., 2018) pioneered EEG signal synthesis with EEG-GAN, using an improved wGAN-GP for generating single-channel signals. This framework utilized convolutional layers for upsampling and downsampling, and was evaluated using Fréchet Inception Distance, Inception Score, and sliced Wasserstein Distance (WD) (Xu et al., 2022) compared different GAN variants, finding that DCw-GAN performed best for generating multichannel EEG data. In recent work, (Wu et al., 2024) introduced a conditional transformer-based wGAN-GP for synthesizing stereoelectroencephalography (SEEG) epileptic signals. This approach leveraged transformers to capture time dependencies, with evaluation through t-SNE, Cosine Similarity, Jensen-Shannon Distance, and classifier performance, with results outperforming traditional augmentation methods. Similarly, (Du et al., 2024a) proposed a Deep Convolutional Wasserstein GAN with Gradient-Penalty (DCwGAN-GP) to generate time-frequency representations, improving classification performance when combined with real data (RD). Finally, (Cook et al., 2024) presented an architecture for generating EEG signals to predict brain age, comparing cGAN, wGAN, and wGAN-GP models. These were evaluated through the Kolmogorov-Smirnov test and classifier performance, with all models showing improvement when combined SD with RD.

This study aimed to develop a generative model using GANs to produce synthetic ioECoG data that captures both epileptic and non-epileptic characteristics. To achieve this, we performed a clustering analysis on the original dataset to identify distinct data patterns, which were subsequently processed using the DCwGAN-GP generative model. The resulting synthetic signals were evaluated for fidelity, diversity, and utility, ensuring comprehensive assessment across these domains.

The document is structured as follows: Section 2 details the methods, including data handling, gen-

erative process, and evaluation metrics. Section 3 presents the results, followed by their discussion in Section 4. Finally, Section 5 summarizes the conclusions and proposes future work.

2 METHOD

2.1 Sample

As part of the Synthetic and Scalable Data Platform for Medical Empowered AI (AISym4Med) project, conducted in collaboration with the Utrecht Medical Center (UMCU), this study analyzed data from 18 patients who underwent intraoperative electrocorticography (ioECoG)-guided epilepsy surgery at UMCU, Netherlands, from 2008 onward. Specifically, only pre-resection data were utilized in this analysis.

2.1.1 Participants

The ioECoG data, sourced from the UMCU-SEIN RESPect database, is stored in a Brain Imaging Data Structure (BIDS) format (Demuru et al., 2022), with all patients providing informed consent. Included patients had ioECoG sampled at 2048 Hz, electrode placement photos, and were at least 1-year post-surgical seizure free (Engel 1A). Exclusion criteria included prior brain surgery, multiple epileptic foci, or surgeries involving an amygdala-hippocampectomy.

2.1.2 Procedure

During surgery, the ioECoG was recorded using electrode grids or strips with 1 cm interelectrode distance. Recordings were made while propofol anesthesia was paused to prevent suppression of epileptiform activity (Sun et al., 2024). The ioECoG enabled the identification of epileptiform patterns in real time, assisting neurosurgeons by tailoring the extent of the resection. More details about the data collection can be found in (Hoogteijling et al., 2024). The pipeline of this study starts with processing and analysing the raw ioECoG. The signals were clustered into different groups based on specific data patterns. Each cluster is then fed into its own generative model, producing SD that aligns with the structure defined during the clustering stage.

2.2 Signal Processing and Analysis

To preserve the statistical integrity of the ioECoG signals for authentic data generation, they were only trimmed and filtered, following the recommendations of (Delorme, 2023). The signals were shortened at their midpoint, reducing the average length from 360 to 60 seconds, matching the duration used by (van Klink et al., 2014). A 4th-order Butterworth filter was applied (Rasheed and Miften, 2023) (Wu et al., 2024) with a 0.16 Hz low-cut to remove drift artifacts (Miller, 2019) and a 512 Hz high-cut to retain high-frequency oscillations (HFOs) (Zweiphenning et al., 2022). Additionally, an Infinite Impulse Response Notch filter at 50 Hz intervals was used to eliminate powerline noise. Subsequently, an indepth analysis of both the time and frequency domains was performed. Given that the patients were under anesthesia, it was anticipated that specific frequency bands, particularly the delta and alpha bands, would exhibit increased power compared to other frequency band(Shin et al., 2020). Accordingly, the relative power of each frequency band, as well as the dominant frequency, were systematically examined.

2.3 Clustering

An unsupervised clustering analysis was performed to group signals into distinct clusters based on their intrinsic characteristics. This data-driven approach facilitated a deeper understanding of the variability within the dataset and enabled the design of a more tailored generative process. By associating specific labels with individual generative models, this method improved the precision and diversity of SD generation.

2.3.1 Feature Extraction

For the clustering analysis, 33 features were extracted using the TSFEL (v0.1.7) (Barandas et al., 2020), Numpy (v1.26.4) (Harris et al., 2020), and Scipy (v1.13.0) (Virtanen et al., 2020) libraries, covering temporal, frequency, time-frequency, and non-linear domains. Temporal features followed methods from (Rasheed and Miften, 2023) and (Du et al., 2024c), known for clustering epileptic EEG signals. Time-frequency features were based on (Chaibi et al., 2024), while frequency and non-linear features were computed specifically for this study, some including: Relative Band Power, Dominant Frequency, Hurst Exponent and Autocorrelation.

Following feature extraction, the features were standardized using the StandardScaler from Scikit-Learn (v1.4.1.post1) (Pedregosa et al., 2018) and features with a correlation above 95% were removed to reduce redundancy (Zhou et al., 2022). To identify key features, the signals were categorized as resected and non-resected according to the clinicians' labels,

and the mean values of each feature were compared between the groups.

2.3.2 Cluster Analysis

To determine the appropriate number of clusters for the clustering algorithm, the Elbow method and Silhouette analysis were applied. After selecting the most discriminative features and determining the optimal number of clusters, the Agglomerative Clustering algorithm from Scikit-Learn (Pedregosa et al., 2018) with the Ward's method (Kononenko and Kukar, 2007) was used to classify the signals into distinct patterns. The clusters were then labelled based on the signal characteristics and their alignment with the clinician labels (i.e., resected and non-resected).

2.4 Generative Model

This section describes the proposed generative model, DCwGAN-GP, covering the input data, architecture, parameters, and training process. The model was developed using Pytorch (v2.3.1) (Paszke et al., 2019) and an NVIDIA RTX 6000 Ada GPU (NVIDIA Corporation, 2024). To generate SD for each cluster, a separate generative model was created per cluster, following the same architecture and training process, with only the batch size differing for the final cluster.

2.4.1 Data Preparation

After initial processing, the clustered ioECoG signals were further prepared for the generative model. This involved segmenting the 60-second signals into 20-second windows and downsampling them to 512 Hz for computational efficiency. The signals were then normalized and transformed into tensors as input for the model.

2.4.2 Model Architecture

The model architecture, shown in Figure 1, uses 1D CNNs for both the Generator and Critic, trained with a Wasserstein Distance loss and Gradient Penalty. While CNNs are typically utilized for image generation, 1D CNNs are stated to effectively capture patterns in one-dimensional EEG data (Aznan et al., 2019). The Generator processes a latent vector Z of size 100 through a series of 1D Transposed Convolutional Layers, configured with a kernel size of 4, stride of 2, and padding of 1 to upsample the input (Aznan et al., 2019). Batch normalization and ReLU activation are applied between layers. A Tanh activation is used in the final layer to generate synthetic samples, $G(z)$. The Critic processes both real and synthetic samples using 1D Convolutional Layers with

the same configuration of kernel size 4, stride 2, and padding 1. It uses Spectral Normalization for stability (Zhong et al., 2023) and Leaky ReLU activation with alpha 0.2 (Park et al., 2024), except in the final block. The Critic outputs a score distinguishing real from synthetic samples, with the loss back-propagating to improve both networks.

2.4.3 Model Training and Tuning

The training process for the DCwGAN-GP ran for a set number of epochs, processing batches of RD in each iteration. The Critic was trained five times for each Generator update, where each Critic update began with zeroing its gradients to prevent accumulation. The Critic then computed scores for both real $C(x)$ and SD $C(G(z))$, applying the gradient penalty, calculated as equation 1 (Shu et al., 2023):

$$GP = \lambda \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[\left(\|\nabla C(\hat{x})\|_2 - 1 \right)^2 \right] \quad (1)$$

The Critic's loss was then calculated using equation 2 (Shu et al., 2023):

$$L = \mathbb{E}_{x \sim p_g} \left[C(G(z)) \right] - \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[C(x) \right] + GP \quad (2)$$

This loss aimed to maximize the difference between scores for real and SD while penalizing large gradients. The loss was backpropagated, and the Critic's parameters were updated using the Adam optimizer. For the Generator, its gradients were zeroed, and a new batch of SD $G(z)$ was generated using the latent tensor Z. The Critic evaluated this SD, producing a score $C(G(z))$. The generator's loss is then calculated as 3 (Shu et al., 2023):

$$\text{loss}_G = -\mathbb{E}[C(G(z))] \quad (3)$$

The loss was backpropagated, and the Generator's parameters were updated using the Adam optimizer. At the end of each epoch, learning rate schedulers adjusted the learning rates for both models to optimize training progression. Following training, the model underwent hyperparameter tuning over 51 iterations to achieve optimal performance. The final version included six layers in the Generator and five in the Critic. The training parameters are detailed in Table 1. The batch size was set to 40, with exception on the final model, which was reduced to 30.

2.5 Evaluation Metrics

This section outlines the evaluation metrics used to assess the quality of the generated ioECoG signals, categorized into fidelity, diversity and utility.
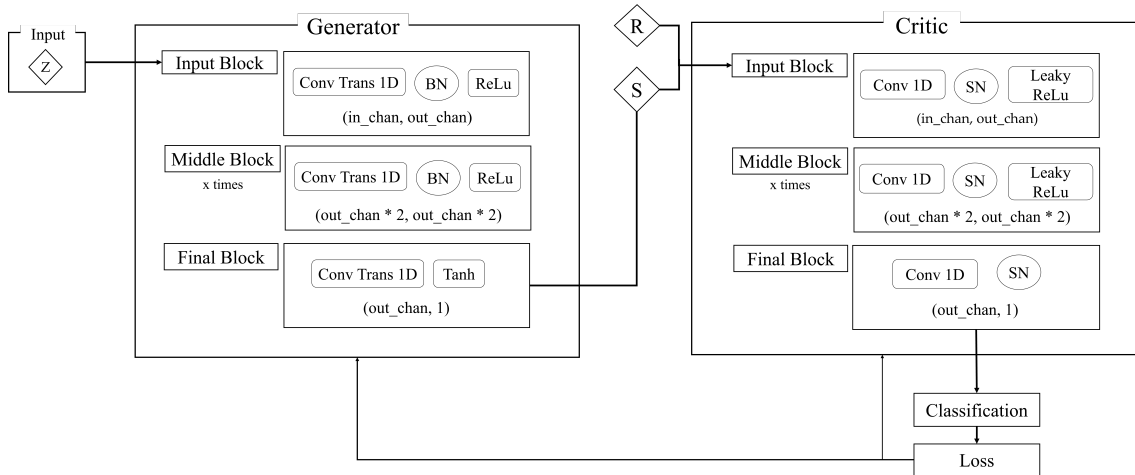
Figure 1: Generative model's architecture. Each Neural Network is comprised of three types of blocks: input, middle and final block. The input and final block are constituted by one layer, while the middle block was constituted by several layers. The number of layers contained in the middle block was defined through the optimization of the model.

Table 1: Training Parameters for DCwGAN-GP.

Training Parameters	
Number of Epochs	2000
Batch Size	40
Learning Rate	0.00005
Critic Interactions	5
Gradient Penalty Lambda	10
Latent Dimension	100
Number of samples	Data points in each cluster

2.5.1 Fidelity Analysis

Fidelity analysis evaluates how closely synthetic signals resemble real ones across time, frequency, and time-frequency domains without replicating them exactly (Naeem et al., 2020) (Figueira and Vaz, 2022). In this dimension, we calculated metrics for three pairings: real-real, real-synthetic, and synthetic-synthetic. The real-real pairing serves as a reference to evaluate the fidelity of the synthetic data. To perform this analysis, each sample was compared with every other sample within its respective pairing group.

In the time domain, the mean and standard deviation (STD) of both real and synthetic signals were evaluated, and the Wasserstein Distance (WD) was calculated to quantify differences in their probability distributions. In the frequency domain, the Power Spectral Density (PSD) was computed for both sets of signals to ensure similar spectral characteristics, and a plot was created to visually compare the power distribution across frequencies. Finally, for the time-frequency domain, scalograms were generated using the Morlet Wavelet Transform to capture dynamic changes in frequency content over time. Similarity between real and synthetic scalograms was quantified using the Pearson Correlation Coefficient, Co-sine Similarity, Structural Similarity Index (SSIM) and Mean Squared Error (MSE). This metrics were calculated with the *Scipy* (Virtanen et al., 2020) and *Scikit-Learn* Image Processing libraries (van der Walt et al., 2014).

2.5.2 Diversity Analysis

To evaluate the diversity of SD in comparison to RD, Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) were employed as dimensionality reduction techniques. These methods enabled a visual comparison of the distribution of synthetic samples against the original data (Jansen, 2020).

2.5.3 Utility Analysis

To evaluate the utility of the SD, several classification models were developed and trained on three datasets: RD, SD, and a combination of both. The aim was to determine if the SD could improve the model's ability to distinguish between clusters, with better performance on the combined dataset indicating the utility of the SD. The statistical significance of the observed differences was then evaluated using a Z-test, which compares the performance metrics between datasets to determine if the observed variations are likely due to chance.

Model Training and Evaluation

The same features used in the clustering process (Section 2.3.1) were applied in the classifiers. The data was split into 70% training and 30% testing sets, using stratified sampling to maintain class proportions. Various ML classifiers were developed using the

Scikit-Learn library (Pedregosa et al., 2018), including Logistic Regression, Support Vector Machines (SVM), Random Forest, Neural Network Multi-Layer Perceptron, Gradient Boosting, and Balanced Random Forest. These models were chosen for their ability to handle imbalanced datasets (Chen et al., 2004), capture complex non-linear patterns (Jaiswal, 2024), and perform well with tabular data (Tuychiev, 2023), with the first two serving as baselines for comparison. A 10-fold Stratified Cross-Validation method was used with the shuffle parameter set to "True" to minimize bias from any inherent order, ensuring the models generalization to unseen data and avoiding overfitting. Performance metrics such as Accuracy, F1-Score, Precision, Recall and AUC were used to compare the models' effectiveness.

3 RESULTS

3.1 Clustering Analysis

Three clusters were chosen for the Agglomerative Clustering algorithm. Based on the alignment with the clinician binary classifications (non-resected and resected tissue) and the data characteristics, the clusters were labelled as: Cluster 0 (Noisy), Cluster 1 (Epileptic), and Cluster 2 (Non-Epileptic). Specifically, Cluster 0 consisted of 28 signals, with 57.14% of the data points classified as non-resected and 42.86% as resected. Cluster 1 included 71 signals, comprising 39.89% non-resected and 61.11% resected data points. Cluster 2 contained 31 signals, of which 83.87% were non-resected and 16.13% were resected. Some overlap occurred in the 3D scatter plots between Clusters 0 and 1, and Clusters 1 and 2. Figure 2 illustrates three representative signals from each cluster.
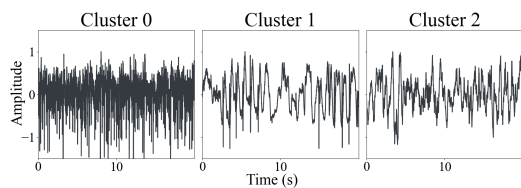

Figure 2: Illustration of three representative signals of each cluster. The signals were normalized, therefore the amplitude is in arbitrary units.

Cluster 0 exhibited irregular and noisy fluctuations with no clear structure. Cluster 1, on the other hand, had more structured patterns, while Cluster 2 showed a level of organization higher than Cluster 0, but its patterns were still not as structured as those

in Cluster 1. Furthermore, the frequency analysis revealed that Cluster 0 had a low PSD with no dominant frequency peaks (as shown in Figure 5), reflecting its noisy nature. In contrast, Cluster 1 demonstrated a strong delta band peak with higher variability in the PSD. Cluster 2 showed consistent delta-band peaks, although with less variability compared to Cluster 1.

3.2 Evaluation of Synthetic Data

3.2.1 Fidelity Analysis

Time Analysis
The mean signals and their respective STD for real and SD across Clusters 0, 1, and 2, depicted in Figure 3, reveal consistent waveform patterns. However, synthetic signals in Cluster 0 display a slightly higher amplitude and increased noise compared to their real counterparts. In turn, the WD analysis illustrated in Figure 4 shows relatively small differences between the data pairings (real-real (RR), real-synthetic (RS) and synthetic-synthetic (SS)). WD is highest for the RR pairing, followed by RS, and lowest for SS. Cluster 1 consistently has the highest WD across all pairings, Cluster 2 is moderate, and Cluster 0 has the lowest values. For the RS pairing, the WD values were 0.043 ± 0.025 for Cluster 0, 0.065 ± 0.045 for Cluster 1, and 0.045 ± 0.036 for Cluster 2.

Frequency Analysis
In the frequency domain depicted in Figure 5, the synthetic signals match closely with the real ones in low-frequency behavior, especially in the delta band. However, the SD exhibit lower variability, as indicated by their lower STD across all clusters.

Time-Frequency Analysis
Figure 6 evaluates the similarity between real and synthetic scalograms across real-real (RR), real-synthetic (RS), and synthetic-synthetic (SS) pairings. Cosine Similarity increased from 0.729 ± 0.037 (RR) to 0.795 ± 0.005 (SS), indicating greater similarity within SD. SSIM and Pearson's Correlation followed a similar pattern, with Cluster 1 and Cluster 2 scoring higher than Cluster 0. The Pearson correlation of cluster 0 improved significantly in SS (+0.275), but remained lower overall compared to other clusters. Cluster 0 consistently exhibited the lowest similarity across all metrics, while Clusters 1 and 2 demonstrated moderate-to-high similarity, particularly in RS pairings, where Pearson's Correlation reached 0.694 ± 0.087 and 0.672 ± 0.063, and Cosine Similarity scored 0.757 ± 0.066 and 0.747 ± 0.051, respectively. MSE showed a clear reduction, decreasing by 0.182
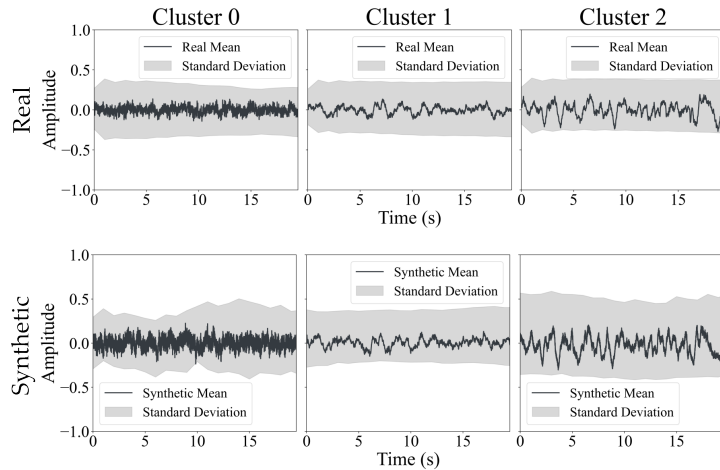
Figure 3: Comparison of the mean and respective Standard Deviation of the signals between real and synthetic clusters. The signals are normalized, therefore the amplitude is in arbitrary units.
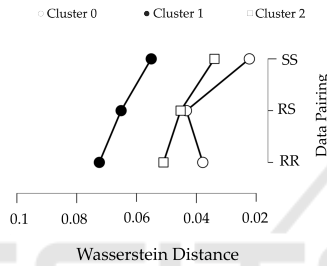

Figure 4: Comparison of the Wasserstein Distance between different data pairing. In this case, this metric is being compared between the real and synthetic (RS) signals, and within the real (RR) and synthetic (SS) signals separately.

± 0.052 from RR to SS, with Cluster 1 achieving the lowest error in RS (0.499 ± 0.078). Despite these trends, the differences between RR and SS pairings remained relatively modest overall, with Cluster 0 again showing the largest variations in SSIM (0.180 ± 0.051) and Pearson's Correlation (0.275 ± 0.119). These results highlight greater internal consistency within SD and the varying performance of clusters, with Clusters 1 and 2 outperforming Cluster 0 across all similarity metrics.

3.2.2 Diversity Analysis

In the PCA results illustrated in Figure 7 (upper row), the SD overlap with the RD, replicating the overall structure but is more concentrated in the center, suggesting less diversity. In the t-SNE plots (second line), synthetic signals are more scattered and less grouped, indicating they fail to capture the local relationships present in the RD.

3.2.3 Utility Analysis

The Logistic Regression model yielded the best results. Table 2 presents improvements in four performance metrics and a slight decline in AUC when trained on combined real and SD (RS) versus training on RD alone (R). However, these differences were not statistically significant. Additional classification results are available in the Appendix.

Table 2: Logistic Regression Scores. Comparison between the scores of the model trained on the combined data (RS) and tested on real (R), with the model trained and tested solely on the RD.

	Train R	Test R	Train RS	Test R
Accuracy	0.89	0.85	0.93	0.88
Precision	0.89	0.73	0.92	0.76
Recall	0.92	0.80	0.94	0.81
F1-Score	0.90	0.75	0.93	0.78
AUC	0.97	0.94	0.99	0.92

4 DISCUSSION

This study aimed to develop a generative model in order to produce synthetic ioECoG data from ioECoG signals of patients undergoing epilepsy surgery.

The analysis of the real ioECoG dataset revealed three distinct clusters, each associated with unique signal patterns. Cluster 1 could be linked to pathological signals, displaying epileptiform activity characterized by spikes and slow-wave complexes, similar to those described by (Li et al., 2020). Moreover, Cluster 1 and Cluster 2, labelled as epileptic and non-epileptic, showed a strong alignment with the clinical
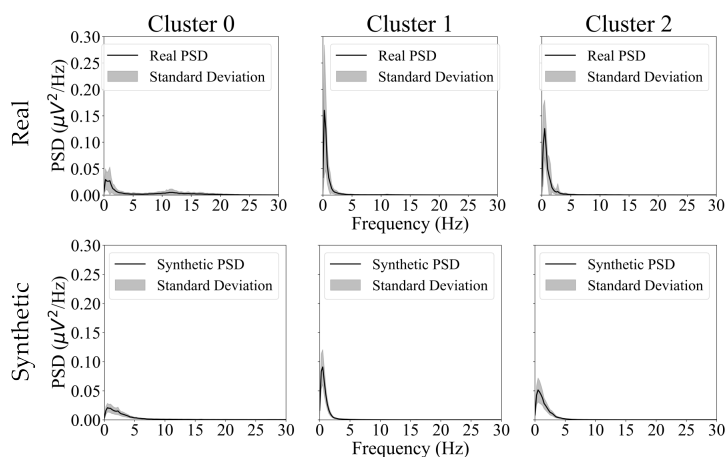
Figure 5: Comparison of the mean PSD and respective STD between real and synthetic clusters.
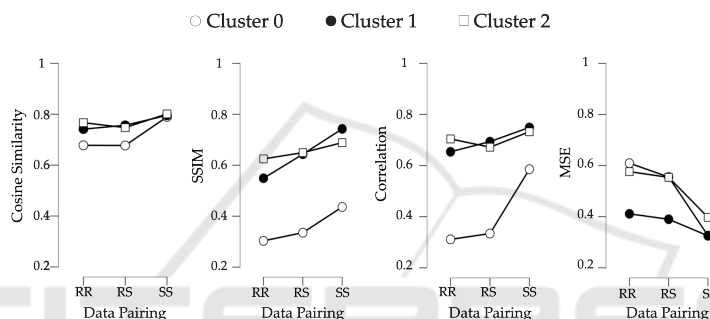

Figure 6: Similarity metrics in the time-frequency domain between different data pairing: real-real (RR), real-synthetic (RS) and synthetic-synthetic (SS).
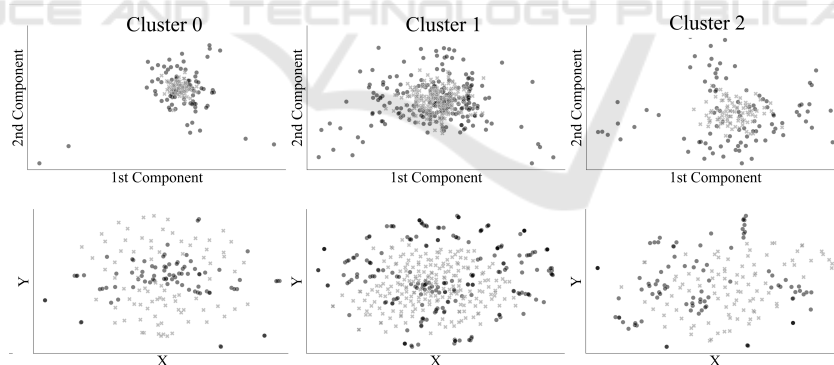

Figure 7: Diversity evaluation. The upper row is correspondent to the PCA evaluation, while the second line is correspondent to the t-SNE evaluation. The columns correspond to each cluster. The SD is represented by black dots, while the RD is represented by gray crosses.

categories, matching 61.1 and 83.9% of the clinician's classifications, respectively. This alignment was also supported by the study of (Gajic et al., 2015), who showcased epileptic signals with more spaced events and with fewer fluctuations, while non-epileptic signals showed more noise and greater variability. In turn, Cluster 0 contained a mix of non-epileptic and epileptic signals, likely due to contamination from

neighboring channels within the same electrode grid, as observed in other intracranial EEG studies (Sindhu et al., 2023). Additionally, this overlap may simply reflect the random nature of noise, which can occur with similar frequency in both types of tissue. Its origin may be attributed to machine artifacts, which can mimic pathological signals and interfere with accurate classification (Nejedly et al., 2020). Having a

noisy cluster could also aid clinical implementation, as it removes the need to exclude channels affected by artifacts, simplifying the process and preserving data integrity. In the frequency domain, all clusters showed a peak in the delta band, as expected during propofol-induced deep sleep (Moini and Piran, 2020), with Cluster 0 also exhibiting an alpha peak that might be related to propofol use (Purdon et al., 2015).

Overall, in Figure 3, the synthetic signals closely match the real ones in amplitude and waveform, indicating that the model effectively captures the global temporal patterns. Minor STD deviations, similar to those reported by (Hartmann et al., 2018), suggest slight differences in variability between real and synthetic signals. In turn, WD scores showed that synthetic signals preserved the similarity trends across clusters, with Cluster 1 having the highest WD probably due to its distinct epileptiform patterns, followed by Clusters 2 and 0, where Cluster 0 displayed the greatest similarity with lower WD values. Slightly lower SS values suggested reduced variability in synthetic signals. RS WD scores in this study seem to outperform benchmarks from the literature, such as 0.078 reported by (Hartmann et al., 2018) and 0.450 by (Xu et al., 2022). Furthermore, the 20-second synthetic time-series offers a more realistic representation compared to the shorter windows commonly used in other studies using scalp EEG data —such as 3 seconds (Aznan et al., 2019), 4 seconds (Pascual et al., 2019), and even as short as 0.05 seconds (Pan et al., 2023)—allowing for improved fidelity and signal diversity.

Figure 5 shows that the synthetic signals closely match the real ones in power distribution, mirroring their trend particularly in the delta band, despite the real signals displaying higher variability (STD). These results align with those reported by (Park et al., 2024) and (Hartmann et al., 2018), where synthetic PSDs showed similar prominent peaks to their real counterparts, with slight power variations. In contrast, (Carrle et al., 2023) found that synthetic PSDs did not fully capture the prominent peaks of RD, indicating a stronger reliability from the generative model in this study. However, the lower STD in synthetic PSDs suggests reduced variability compared to the real signals.

In the evaluation of the time-frequency domain (Figure 6), the lower similarity values in the RR pairing across all metrics reflect greater variability in the RD compared to the higher SS scores, indicating that synthetic signals display higher inter-similarity and fail to capture the full variability of the RD. This is most evident in Cluster 0, which performed worst in

all metrics but showed the largest improvement from RR to SS, particularly in SSIM and Pearson's Correlation. However, the differences between RR and SS pairings are relatively small. Although Cluster 0 had the lowest RS scores, they were consistent with the RR pairing, suggesting that while the SD does not fully capture RD variability, it still maintains strong correlations with the real signals. To the best of our knowledge, this type of analysis was not found in comparable studies (Wu et al., 2024), which focused on signal similarity rather than time-frequency representations.

The limited variability of the synthetic signals is confirmed by the diversity evaluation results. While synthetic signals replicated most real patterns, as shown by the overlap in the PCA plot, their concentration toward the center indicates only partial diversity capture, with the tails of the real distributions being overlooked. Similarly, the t-SNE plots reflect findings from (Wu et al., 2024), though our study showed less overlap, highlighting a lack of diversity in the generated samples. This issue likely stems from misclassifications during clustering, particularly in Cluster 1, introducing unwanted variability in the RD. During training, the Critic consistently labelled these misclassified signals as synthetic, causing the generator to avoid synthesizing them. As a result, the synthetic signals focused on common patterns and lacked the broader variability seen in the RD, especially at the distribution tails.

All classifiers showed improved performance across Accuracy, Precision, Recall, and F1-score when using the augmented ioECoG dataset, highlighting the utility of the generated signals. The Logistic Regression Model showcased the best results when trained on the augmented dataset, but tested only with the real signals. The 3% increase in accuracy aligns with improvements reported in other studies, such as 2.51% (Du et al., 2024a), 5% (Xu et al., 2022), and 6% (Wu et al., 2024). However, the decrease in AUC may indicate that the SD exhibit distributional differences compared to RD, confirming their inability to fully capture the variability and complexity of the RD.

Finally, a comprehensive visual review and validation of all generated signals remains necessary to confirm their suitability for clinical applications.

5 CONCLUSIONS AND FUTURE WORK

This study developed a generative model, DCwGAN-GP, capable of synthesizing realistic ioECoG signals

from epileptic and non-epileptic tissue with the potential to enhance ML models and support epilepsy surgery treatment. The generated signals replicated the key characteristics of the RD across time, frequency, and time-frequency domains, contributing to improved performance in ML classifiers trained to distinguish epileptic, non-epileptic, and noisy signals. The 3% performance improvement demonstrated the utility of the SD in enhancing classifier performance.

While the DCwGAN-GP model demonstrated reliable fidelity in capturing temporal patterns within individual channels, it did not account for inter-channel temporal dynamics or spatial relationships between electrodes. Incorporating techniques such as Long Short-Term Memory (LSTM) networks, Recurrent Neural Networks (RNNs) or Transformers for temporal dependencies, along with graph-based models for spatial relationships, could further enhance the realism and clinical relevance of the generated signals. In addition, clinical validation of each signal should be considered to ensure reliability and clinical applicability.

Future work could focus on improving the ground-truth labelling of signals to enhance accurate distinction of epileptiform activity, thereby strengthening the generative process. Machine Learning and Deep Clustering techniques hold promise for refining pattern recognition within ioECoG data, potentially reducing misclassifications and achieving more nuanced synthetic signals that reflect real-world diversity. Another critical direction involves explicitly modeling spatial and temporal dependencies across signals.

ACKNOWLEDGEMENTS

REFERENCES

Aznan, N. K. N., Atapour-Abarghouei, A., Bonner, S., Connolly, J., Moubayed, N. A., and Breckon, T. (2019). Simulating brain signals: Creating synthetic eeg data via neural-based generative models for im-

proved ssvep classification. In *International Joint Conference on Neural Networks (IJCNN)*. IEEE.

Aznan, N. K. N., Atapour-Abarghouei, A., Bonner, S., Connolly, J. D., and Breckon, T. P. (2020). Leveraging synthetic subject invariant eeg signals for zero calibration bci.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). Tsfel: Time series feature extraction library. *SoftwareX*, 11:100456.

Carrle, F. P., Hollenbenders, Y., and Reichenbach, A. (2023). Generation of synthetic eeg data for training algorithms supporting the diagnosis of major depressive disorder. *Frontiers in Neuroscience*, 17.

Chaibi, S., Mahjoub, C., Ayadi, W., and Kachouri, A. (2024). Epileptic eeg patterns recognition through machine learning techniques and relevant time–frequency features. *Biomedical Engineering / Biomedizinische Technik*, 69:111–123.

Chen, C., Liaw, A., and Breiman, L. (2004). Using random forest to learn imbalanced data. Technical report, Department of Statistics, UC Berkley.

Consales, A., Casciato, S., Asioli, S., Barba, C., Caulo, M., Colicchio, G., Cossu, M., de Palma, L., Morano, A., Vatti, G., Villani, F., Zamponi, N., Tassi, L., Gennaro, G. D., and Marras, C. E. (2021). The surgical treatment of epilepsy. *Neurological Sciences*, 42:2249–2260.

Cook, Z., Sinha, G., Wang, J., Zhao, C., Belacel, N., Doesburg, S., Medvedev, G., Ribary, U., Vakorin, V., and Xi, P. (2024). Enhancing brain age prediction: A generative ai approach for eeg machine learning models. In *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE.

Delorme, A. (2023). Eeg is better left alone. *Scientific Reports*, 13:2372.

Demuru, M., van Blooijs, D., Zweiphenning, W., Hermes, D., Leijten, F., and Zijlmans, M. (2022). A practical workflow for organizing clinical intraoperative and long-term ieeg data in bids. *Neuroinformatics*, 20:727–736.

Du, X., Ding, X., Xi, M., Lv, Y., Qiu, S., and Liu, Q. (2024a). A data augmentation method for motor imagery eeg signals based on dcgan-gp network. *Brain Sciences*, 14:375.

Du, X., Wang, X., Zhu, L., Ding, X., Lv, Y., Qiu, S., and Liu, Q. (2024b). Electroencephalographic signal data augmentation based on improved generative adversarial network. *Brain Sciences*, 14.

Du, Y., Li, G., Wu, M., and Chen, F. (2024c). Unsupervised multivariate feature-based adaptive clustering analysis of epileptic eeg signals. *Brain Sciences*, 14:342.

Duncan, J. S. and Taylor, P. N. (2023). Optimising epilepsy surgery. *The Lancet Neurology*, 22:373–374.

Fan, Y., Wang, B., and Zhang, T. (2024). A dual-discriminator gan for sleep eeg signal synthesis. *Journal Of Bioinformatics and Neuroscience (JBINS)*.

Fang, L., Li, Y., Shao, M., Yu, A., Felemban, B. F., Aly, A. A., Rani, S., and Lyu, X. (2024). Enhancing med-

ical signal processing and diagnosis with ai-generated content techniques. *IEEE Journal of Biomedical and Health Informatics*.

Fernández, I. S. and Loddenkemper, T. (2013). Electro-corticography for seizure foci mapping in epilepsy surgery. *Journal of Clinical Neurophysiology*, 30(6):554–570.

Figueira, A. and Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and gans. *Mathematics*, 10:1–41.

Gajic, D., Djurovic, Z., Gligorijevic, J., Gennaro, S. D., and Savic-Gajic, I. (2015). Detection of epileptiform activity in eeg signals based on time-frequency and non-linear analysis. *Frontiers in Computational Neuroscience*, 9.

Greiner, H. M., Horn, P. S., Tenney, J. R., Arya, R., Jain, S. V., Holland, K. D., Leach, J. L., Miles, L., Rose, D. F., Fujiwara, H., and Mangano, F. T. (2016). Preresection intraoperative electrocorticography (ecog) abnormalities predict seizure-onset zone and outcome in pediatric epilepsy surgery. *Epilepsia*, 57(4):582–589.

Habashi, A. G., Azab, A. M., Eldawlatly, S., and Aly, G. M. (2023). Generative adversarial networks in eeg analysis: an overview. *Journal of NeuroEngineering and Rehabilitation*, 20:40.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Hartmann, K. G., Schirrmeister, R. T., and Ball, T. (2018). Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals. *Electrical Engineering and Systems Science*.

Hoogteijling, S., Schaft, E. V., Dirks, E. H., Straumann, S., Demuru, M., van Eijsden, P., Gebbink, T., Otte, W. M., Huiskamp, G. M., van't Klooster, M. A., et al. (2024). Machine learning for (non-) epileptic tissue detection from the intraoperative electrocorticogram. *Clinical Neurophysiology*, 167:14–25.

Jaiswal, S. (2024). Multilayer perceptrons in machine learning: A comprehensive guide.

Jansen, S. (2020). *Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python*. Packt Publishing, 2nd edition.

Kononenko, I. and Kukar, M. (2007). Chapter 12 - cluster analysis. In Kononenko, I. and Kukar, M., editors, *Machine Learning and Data Mining*, pages 321–358. Woodhead Publishing.

Li, Q., Gao, J., Zhang, Z., Huang, Q., Wu, Y., and Xu, B. (2020). Distinguishing epileptiform discharges from normal electroencephalograms using adaptive fractal and network analysis: A clinical perspective. *Frontiers in Physiology*, 11:828.

Liang, S., Kuang, S., Wang, D., Yuan, Z., Zhang, H., and Sun, L. (2023). An auxiliary synthesis framework for

enhancing eeg-based classification with limited data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:2120–2131.

Miller, K. J. (2019). A library of human electrocorticographic data and analyses. *Nature Human Behaviour*, 3:1225–1235.

Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *ArXiv*.

Moini, J. and Piran, P. (2020). Chapter 6 - cerebral cortex. In Moini, J. and Piran, P., editors, *Functional and Clinical Neuroanatomy*, pages 177–240. Academic Press.

Naeem, M. F., Oh, S. J., Uh, Y., Choi, Y., and Yoo, J. (2020). Reliable fidelity and diversity metrics for generative models. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7176–7185. PMLR.

Nejedly, P., Kremen, V., Sladky, V., Cimbalnik, J., Klimes, P., Plesinger, F., Mivalt, F., Travnicek, V., Viscor, I., Pail, M., Halamek, J., Brinkmann, B. H., Brazdil, M., Jurak, P., and Worrell, G. (2020). Multicenter intracranial eeg dataset for classification of graphoelements and artifactual signals. *Scientific Data*, 7(1):179.

Nia, A. F., Tang, V., Talou, G. M., and Billinghurst, M. (2024). Synthesizing affective neurophysiological signals using generative models: A review paper. *Journal of Neuroscience Methods*, 406.

NVIDIA Corporation (2024). Nvidia rtx 6000 ada generation graphics card. https://www.nvidia.com/en-us/design-visualization/rtx-6000/.

Pan, Y., Li, N., Zhang, Y., Xu, P., and Yao, D. (2023). Short-length ssvep data extension by a novel generative adversarial networks based framework. *Cognitive Neurodynamics*.

Panwar, S., Rad, P., Jung, T.-P., and Huang, Y. (2020). Modeling eeg data distribution with a wasserstein generative adversarial network to predict rsvp events. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28:1720–1730.

Park, J., Mahey, P., and Adeniyi, O. (2024). Improving eeg signal classification accuracy using wasserstein generative adversarial networks.

Pascual, D., Aminifar, A., Atienza, D., Ryvlin, P., and Wattenhofer, R. (2019). Synthetic epileptic brain activities using generative adversarial networks. *arXiv preprint arXiv:1907.10518*.

Pascual, D., Amirshahi, A., Aminifar, A., Atienza, D., Ryvlin, P., and Wattenhofer, R. (2021). Epilepsygan: Synthetic epileptic brain activities with privacy preservation. *IEEE Transactions on Biomedical Engineering*, 68:2435–2446.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information*

Processing Systems 32, pages 8024–8035. Curran Associates, Inc.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay (2018). Scikit-learn: Machine learning in python.

Purdon, P. L., Sampson, A., Pavone, K. J., and Brown, E. N. (2015). Clinical Electroencephalography for Anesthesiologists: Part I: Background and Basic Signatures. *Anesthesiology*, 123(4):937–960.

Rasheed, S. S. and Miften, F. S. (2023). Improve of neonatal seizure detection based on eeg signal using k-mean clustering. In *2023 Al-Sadiq International Conference on Communication and Information Technology (AIC-CIT)*, pages 181–184. IEEE.

Saminu, S., Xu, G., Shuai, Z., Kader, I. A. E., Jabire, A. H., Ahmed, Y. K., Karaye, I. A., and Ahmad, I. S. (2022). Application of deep learning and wt-sst in localization of epileptogenic zone using epileptic eeg signals. *Applied Sciences*, 12:4879.

Shin, H. W., Kim, H. J., Jang, Y. K., You, H. S., Huh, H., Choi, Y. J., Choi, S. U., and Hong, J. S. (2020). Monitoring of anesthetic depth and eeg band power using phase lag entropy during propofol anesthesia. *BMC Anesthesiology*, 20:49.

Shoka, A. A. E., Dessouky, M. M., El-Sayed, A., and Hemdan, E. E.-D. (2023). Eeg seizure detection: concepts, techniques, challenges, and future trends. *Multimedia Tools and Applications*, 82:42021–42051.

Shu, K., Zhao, Y., Wu, L., Liu, A., Qian, R., and Chen, X. (2023). Data augmentation for seizure prediction with generative diffusion model.

Sindhu, K. R., Ngo, D., Ombao, H., Olaya, J. E., Shrey, D. W., and Lopour, B. A. (2023). A novel method for dynamically altering the surface area of intracranial eeg electrodes. *Journal of Neural Engineering*, 20(2):026002.

Sun, D., van 't Klooster, M. A., Ringeling, E. M., Schaft, E. V., van Rijen, P. C., Leijten, F. S., Demuru, M., Robe, P. A., Hoff, R. G., and Zijlmans, M. (2024). Pausing propofol during neurosurgery to record intraoperative electrocorticography is feasible;10 years of clinical experience. *Clinical Neurophysiology*, 167:84–91.

Tatum, W. O., editor (2021). *Handbook of EEG Interpretation*. Springer Publishing Company.

Tuychiev, B. (2023). A guide to the gradient boosting algorithm. https://www.datacamp.com/tutorial/guide-to-the-gradient-boosting-algorithm. Accessed: 2024-09-07.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in python. *PeerJ*, 2:e453.

van Klink, N., van 't Klooster, M., Zelmann, R., Leijten, F., Ferrier, C., Braun, K., van Rijen, P., van Putten, M., Huiskamp, G., and Zijlmans, M. (2014). High frequency oscillations in intra-operative electrocorticog-

raphy before and after epilepsy surgery. *Clinical Neurophysiology*, 125(11):2212–2219.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Wang, Z., Guo, J., van 't Klooster, M., Hoogteijling, S., Jacobs, J., and Zijlmans, M. (2024). Prognostic value of complete resection of the high-frequency oscillation area in intracranial eeg: A systematic review and meta-analysis. *Neurology*, 102(9):e209216.

WHO (2019). *Epilepsy: a public health imperative*. World Health Organization.

Wu, X., Zhang, D., Li, G., Gao, X., Metcalfe, B., and Chen, L. (2024). Data augmentation for invasive brain-computer interfaces based on stereo-electroencephalography (seeg). *Journal of Neural Engineering*.

Xu, Y., Yang, J., and Sawan, M. (2022). Multichannel synthetic preictal eeg signals to enhance the prediction of epileptic seizures. *IEEE Transactions on Biomedical Engineering*, 69:3516–3525.

Zhong, H., Yu, S., Trinh, H., Lv, Y., Yuan, R., and Wang, Y. (2023). Fine-tuning transfer learning based on dc-gan integrated with self-attention and spectral normalization for bearing fault diagnosis. *Measurement*, 210:112421.

Zhou, H., Wang, X., and Zhu, R. (2022). Feature selection based on mutual information with correlation coefficient. *Applied Intelligence*, 52:5457–5474.

Zijlmans, M., Zweiphenning, W., and van Klink, N. (2019). Changing concepts in presurgical assessment for epilepsy surgery. *Nature Reviews Neurology*, 15(10):594–606.

Zweiphenning, W., van 't Klooster, M. A., van Klink, N. E. C., Leijten, F. S. S., Ferrier, C. H., Gebbink, T., Huiskamp, G., van Zandvoort, M. J. E., van Schooneveld, M. M. J., Bourez, M., Goemans, S., Straumann, S., van Rijen, P. C., Gosselaar, P. H., van Eijsden, P., Otte, W. M., van Diessen, E., Braun, K. P. J., Zijlmans, M., Bloemen-Carlier, E. M., Cibulková, V., de Munnink, R., van der Salm, S., Eijkemans, M. J., van Eck, J. M. O., Velders, A., van Asch, C. J., Zwemmer, J., van Regteren-van Griethuysen, R., Smeding, H., van der Berg, L., de Bresser, J., de Kort, G. A., and Dankbaar, J.-W. (2022). Intraoperative electrocorticography using high-frequency oscillations or spikes to tailor epilepsy surgery in the netherlands (the hfo trial): a randomised, single-blind, adaptive non-inferiority trial. *The Lancet Neurology*, 21:982–993.

APPENDIX

Table 3: Comparison of classification models trained on combined data (RS) and tested on real data (R) versus models trained and tested solely on RD.

Metric	Models				
	Random Forest	**SVM**	**Neural Network**	**Gradient Boosting**	**Balanced RF**
Accuracy	tR: 1.0 TR: 0.87	tR: 0.89 TR: 0.85	tR: 1.0 TR: 0.77	tR: 1.0 TR: 0.85	tR: 0.94 TR: 0.84
	tRS: 1.0 TR: 0.87	tRS: 0.95 TR: 0.85	tRS: 1.0 TR: 0.83	tRS: 1.0 TR: 0.80	tRS: 0.97 TR: 0.78
Precision	tR: 1.0 TR: 0.63	tR: 0.89 TR: 0.69	tR: 1.0 TR: 0.63	tR: 1.0 TR: 0.63	tR: 0.94 TR: 0.65
	tRS: 1.0 TR: 0.63	tRS: 0.95 TR: 0.63	tRS: 1.0 TR: 0.61	tRS: 1.0 TR: 0.72	tRS: 0.96 TR: 0.60
Recall	tR: 1.0 TR: 0.64	tR: 0.92 TR: 0.70	tR: 1.0 TR: 0.66	tR: 1.0 TR: 0.63	tR: 0.96 TR: 0.66
	tRS: 1.0 TR: 0.64	tRS: 0.97 TR: 0.63	tRS: 1.0 TR: 0.62	tRS: 1.0 TR: 0.75	tRS: 0.98 TR: 0.60
F1-Score	tR: 1.0 TR: 0.64	tR: 0.90 TR: 0.70	tR: 1.0 TR: 0.64	tR: 1.0 TR: 0.63	tR: 0.95 TR: 0.66
	tRS: 1.0 TR: 0.64	tRS: 0.95 TR: 0.63	tRS: 1.0 TR: 0.62	tRS: 1.0 TR: 0.72	tRS: 0.98 TR: 0.60

Abbreviations: tR: train on real data; TR: test on real data; tRS: train on real + synthetic data; RF: Random Forest.