A LIGHTWEIGHT AUTOENCODER MODEL FOR EPILEPTIC SEIZURE DETECTION

Raja Mourad
Univ Rennes, Inserm
LTSI - UMR 1099
Signals and Control Systems
Lebanese International University
Tripoli, Lebanon
raja.mourad@liu.edu.lb

Ahmad Diab
Signals Processing
Department of Biomedical Engineering
Lebanese International University
Tripoli, Lebanon
ahmad.diab@liu.edu.lb

Zaher Merhi
Department of Computer
and Communication Engineering
Lebanese International University
Tripoli, Lebanon
zaher.merhi@liu.edu.lb

Mohamad Khalil
Faculty of Engineering and Azm
Center for Research in Biotechnology and its applications
Lebanese University
Tripoli, Lebanon
mohamad.khalil@ul.edu.lb

Régine Le Bouquin Jeannès
Univ Rennes, Inserm
LTSI - UMR 1099
F-35000 Rennes, France
Rennes, France
regine.le-bouquin-jeannes@univ-rennes.fr

***Abstract*—Epilepsy is a neurological disorder marked by recurrent and unpredictable disturbances in brain activity, leading to epileptic seizures (ES). One approach of diagnosing these seizures involves analyzing ElectroEncephaloGraphic (EEG) signals, which are highly dynamic and complex. Recently, researchers have been implementing Machine Learning (ML) Deep Learning (DL) techniques to perform automatic seizure detection of these EEG signals. ML models require extensive feature extraction techniques compared with DL ones. Despite the promising results of DL models in the detection task, their computational cost is still a challenge. Moreover, DL models require a large amount of training data which is not always available. Hence, in this paper, a hybrid channel-wise lightweight automatic seizure detection model based on vanilla autoencoders and K-NN classifier is introduced. The model is able to generate almost real-like data that can be used as data augmentation technique and increasing the dataset spread for DL models. Mean Square Error (MSE), Pearson Correlation Coefficient (PCC), and Power Spectral Density (PSD) similarity were used to evaluate the quality of the generated signal. Compared to other methods, this model provided a competitive average reconstruction loss of 0.4515, an average precision of 93.69%, and a low parameter count of 41.152k. The lightweight nature of this model could potentially be used in wearable clinical devices for real-time seizure detection.**

***Keywords*—Epilepsy, Detection, EEG, Deep Learning, Autoencoder**

I. Introduction

Epilepsy is the fourth most common neurological disorder, affecting around 50 million people worldwide [1]. It is characterized by unprovoked seizures caused by sudden irregular neuronal activity in the brain and can lead to serious health issues, including death [2]. Electroencephalography (EEG) is a widely used method for monitoring and diagnosing epilepsy due to its affordability and high temporal resolution [3]. EEG records brain activity using multiple electrodes placed on different areas of the scalp or directly on the surface of the brain, known as scalp EEG and intracranial EEG respectively. Most studies focus on scalp EEG because this technique is non-invasive and easier to obtain [4]. Interpreting EEG signals requires trained neurologists, but manually analyzing long-term recordings is time-consuming and labor-intensive [1]. As a result, automatic EEG seizure detection has become a key area of research in neuroinformatics.

In recent years, numerous studies have focused on developing automatic seizure detection systems using machine learning (ML) classifiers. This process typically involves two main stages: feature extraction and classification. Feature extraction plays a crucial role in identifying distinctive EEG patterns, enabling the classifier to differentiate between relevant and less relevant features more effectively. Most studies employ extensive feature extraction techniques in order to detect seizures, which can be classified into time-domain, frequency-domain, and time-frequency domain features. The design of these handcrafted methods, despite its efficiency in extracting very useful features, often requires the expert knowledge to derive interpretations. For example, the short-time Fourier transform (STFT) is a common method to analyze EEG signals and extract time-dependent frequency components from the EEG signals, where these components can serve as features for ML classifiers. In addition, time-domain features were also known for their effectiveness in seizure detection such as energy, min, max, skewness, kurtosis, and line-length [5]. More recently, deep learning (DL) approaches have arisen and shown promising results in the field of epileptic seizure detection, due to their ability to learn embedded features from the signal without any feature extraction technique [6]. In EEG seizure detection, deep learning models like convolutional neural networks (CNN) and autoencoders (AE) have demon-

This study is supported by the Lebanese International University

strated superior robustness in extracting features compared to handcrafted methods, leading to improved detection performance [7]. Therefore, this study focuses on deep learning-based approaches for EEG seizure detection using AE.

Despite the promising results obtained by DL models in EEG epileptic seizure detection task, there are still unsolved challenges that should be addressed. They require high computational resources which potentially limit their application in wearable clinical devices. In addition, DL models require huge amount of data to be trained efficiently, which is not always available.

To address the aforementioned challenges, a semi-supervised lightweight hybrid DL/ML model based on AE and K-Nearest Neighbors (K-NN) is proposed in this contribution. Specifically, we used AE along with short time segments from raw EEG data to extract features which then were trained using K-NN classifier to perform classification. Hence, the main contributions of this paper are:

- To significantly decrease the computational complexity of the AE model by implementing a channel-wise detection model rather than a multi-channel one.
- To be able to generate epilepsy-like new data for the use in data augmentation.

The remainder of the paper is divided as follows: Section II discusses some of the AE-related works done recently. Section III introduces the methodology steps to build our model. Section IV shows the results obtained and Section V concludes this paper and suggests some future prospects.

II. Related Works

Epileptic seizure detection using EEG signals has evolved from traditional feature extraction methods to DL approaches. Early methods relied on handcrafted features like STFT, wavelet transform, and power spectral density (PSD) analysis, requiring expert knowledge and extensive preprocessing. With the rise of DL, models like CNN and recurrent neural network (RNN) improved classification accuracy but remained computationally expensive, limiting their use in real-time applications. In recent years, AEs have gained attention for their ability to extract meaningful representations and generate synthetic EEG data, reducing the need for large labeled datasets. In this context, Abdelhameed et al. [8] developed a technique based on convolutional variational autoencoder (VAE) to derive features from EEG data, aiming to eliminate the need for a manual feature extraction step prior to model training. The features produced were then used with a supervised classifier for seizure detection. Also, Daoud et al. [9] evaluated two approaches focused on automatic feature extraction. The first approach considered a deep convolutional AE to extract features, which were then classified by a multilayer perceptron (MLP). The second approach was an unsupervised pipeline combining a deep convolutional VAE with K-Means clustering based on the latent features.

Many studies on seizure detection primarily use AEs or VAEs to extract features, which are then utilized with supervised methods to classify EEG data. However, a small number of approaches depend solely on AEs and VAEs for the entire seizure detection process, employing only reconstruction error metrics for classification. Huang et al. [10] applied AEs for feature extraction in epilepsy detection, comparing this approach against traditional principal component analysis (PCA). They evaluated signal reconstruction using three metrics: original-to-reconstructed signal ratio (ORSR), mean squared error (MSE), and cosine similarity (CS), and found these metrics to be sensitive indicators of epilepsy. They also incorporated permutation importance and Shapley additive explanations (SHAP) [11] to enhance model interpretability, which validated the effectiveness of AE-based feature extraction over PCA.

Despite the advancements in DL-based seizure detection methods, existing approaches face several challenges that limit their practicality in real-world applications. Many AE and VAE-based models rely on multi-channel EEG inputs, increasing computational complexity and making real-time deployment difficult, particularly in resource-constrained environments such as wearable devices. Additionally, while these models effectively extract latent features, they often require large datasets for training, which are not always available in medical applications. Moreover, while supervised classifiers improve detection accuracy, they introduce additional computational complexity, making them less suitable for low-power medical devices. Given these limitations, this study proposes a lightweight channel-wise AE model combined with a K-NN classifier to reduce computational complexity. By using a channel-wise approach, the model minimizes parameter requirements and enables real-time seizure detection, making it a more efficient and practical solution for embedded medical applications.

III. Methodology

This section presents the different steps to implement the AE approach for epileptic seizure detection and EEG signal generation. The section starts with the dataset used to perform the analysis, goes on with the preprocessing pipeline to ensure artefact-free and interpretable dataset ready for training, then presents the model architecture and ends with the training procedure.

A. Dataset

CHB-MIT open access scalp EEG dataset was used for this work. It was collected from the Children's Hospital of Boston [12]. In this dataset, EEG multi-channel signals were recorded upon 23 patients with intractable seizure, with 5 males and 18 females from age 2 to 22. The sampling rate of each channel is 256Hz with 16-bit resolution. The beginning and end of seizure periods were annotated by experts through visual inspection. The majority of the data were recorded with 23 electrodes placed according to the 10-20 international system. In this experiment, only patients with 23 channels were utilized.

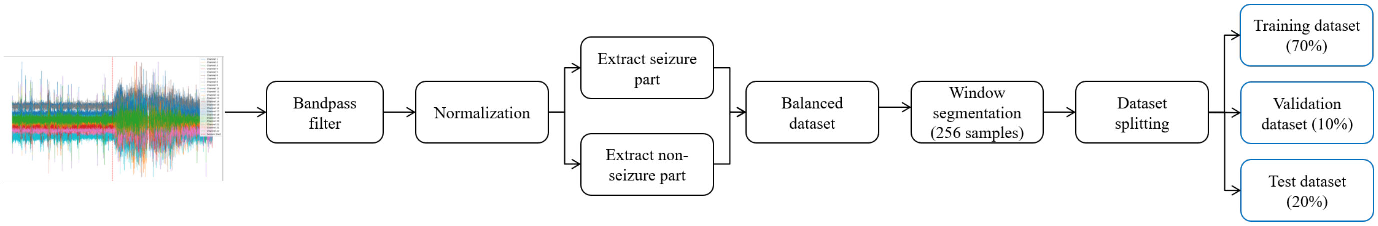

Fig. 1: Preprocessing pipeline

B. Preprocessing

Preprocessing organizes the EEG signal for effective training while also removing artifacts. Figure 1 illustrates the preprocessing pipeline implemented on the CHB-MIT dataset. The EEG recordings of the dataset are first filtered by a Butterworth bandpass filter to maintain frequencies between 0.5-40 Hz. Then, each recording was normalized per patient and per channel by converting them to follow a standard normal distribution with zero mean and unit variance. This step is essential to standardize the dataset and helps improve the training efficiency. The standard score of each data, z, sample is given by:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where μ and σ are the mean and standard deviation of the input data, respectively.

After normalizing the data, an equal number of seizure and non-seizure segments were extracted to ensure dataset balance. These segments were then further divided using a 1-second non-overlapping moving window. The length of the sliding window is usually selected between 1 and 30 s [1]. Finally, the segmented windows were split into 70% training, 10% validation, and 20% testing ensuring no data leakage between patients (the training dataset has no common patient with the validation and testing datasets).

C. Model Architecture

In this work, a separate AE model was implemented for each individual channel, c, with $c = (1, 2, \ldots, 23)$. Figure 2 (a) shows the proposed AE model for epileptic seizure detection. Each channel from each window was separated and passed through a separate AE model. The learned latent spaces of each channel were concatenated and passed through a K-NN classifier to perform binary classification. AE models have 256 input neurons, 128 hidden layer neurons, and 64 latent space neurons. The number of hidden neurons and latent space neurons was tuned via trial and error to reach the lowest reconstruction loss.

The AE objective function is expressed as the MSE between the input EEG samples and the reconstructed samples. The goal of the AE is to minimize the objective function as much as possible, so that the reconstructed signal closely approximates the original signal. The loss function of the individual AE model is given by:

$$\mathcal{L}(x, \hat{x}) = MSE(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{2}$$

where N is the number of data points in the input (or output layer), x_i and $\hat{x}_i$ are the original channel input value and the reconstructed channel output value, respectively. As for the K-NN classifier, K=10 was chosen as the nearest neighbors' factor. Training this classifier was done among the concatenated latent features of the 23 EEG channels.

For generating new data, Figure 2 (b) provides a detailed illustration of the generation process. Test signals pass through the trained encoders of all channels, after which random Gaussian noise is added to the latent space vector. This perturbation modifies the features, resulting in data that are different from the original test data.

D. Training

The AE model was implemented using TensorFlow and Keras libraries. Adam's optimizer was used to minimize the loss function with a learning rate of 0.0001. In addition, L2-regularization and linear dropout layers (30% dropout) were used to prevent overfitting. Linear activation functions for each layer were used and an early stopping mechanism was implemented to stop the training if the model did not improve. The number of epochs was selected to 50 along with 32 batch size

IV. Results and Discussion

The evaluation metrics of the testing stage were split into two categories:

- Signal quality metrics: they evaluate how well the generated signals look like the original input signals. These include:
  o MSE: it measures the average squared difference between the original and generated signals. A lower MSE indicates better reconstruction, and is given by Equation (2).
  o Pearson correlation coefficient (PCC): it measures the linear relationship between original and generated signals and is given by:

$$PCC = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(\hat{x}_i - \bar{\hat{x}})^2}} \tag{3}$$

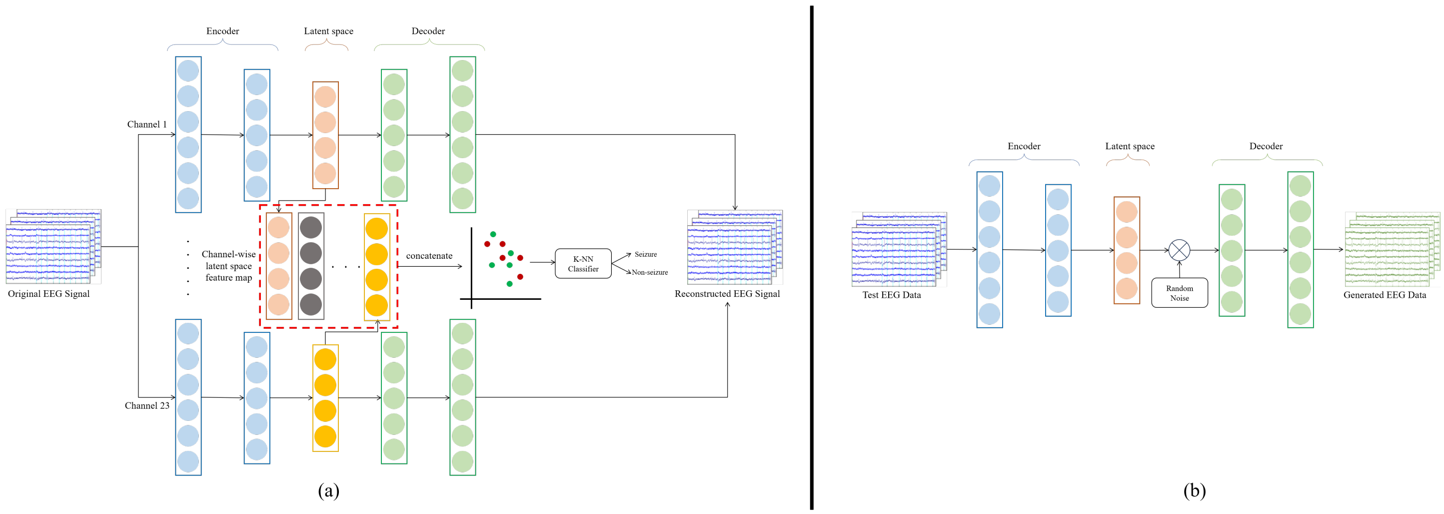

Fig. 2: (a) Proposed model architecture. (b) AE model for signal generation approach

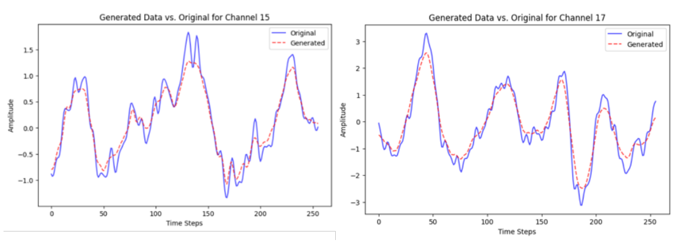

Fig. 3: Generated vs original EEG signal for channels 15 and 17

TABLE I: Signal generation quality metrics and K-NN classification report

Metric	Average Value (for all channels)
Signal Quality Metrics	
MSE	0.4515
PCC	0.9580
Classification Metric	
Precision	96.69%

where N is the number of samples in each window, x_i and $\hat{x}_i$ are the original and reconstructed EEG data samples respectively, while $\bar{x}$ and $\bar{\hat{x}}$ are the mean values of the original and the generated signals, respectively.

- Classification metrics: we consider here the precision since we focus on detecting the true positives (i.e., seizure events) and is given by:

$$precision = \frac{TP}{TP + FP} \tag{4}$$

where TP is the true positives, and FP is the false positives. Table I presents the results of AE-based EEG data classification and generation approach, evaluating both signal reconstruction quality and classification performance. The results obtained by the proposed approach indicate a strong ability to generate new EEG signals with high fidelity, as reflected by the PCC of 0.9580, which suggests a near-perfect linear correlation between the original and generated signals. This high correlation demonstrates that the AE effectively captures and retains key signal features. In addition, the MSE of 0.4515 suggests that the generated signal differ in shape from the original test signal, indicating a successful generation of new and original-like signals. Figure 3 illustrates two generated EEG signals from different channels.

Table II shows a comparison between our approach and other methods in the literature. The methods were chosen based on their explicit reporting of computational complexity. The proposed model achieved significantly lower computational resource consumption (of 41.15k parameters) than other DL-based epileptic seizure detection models. Compared to complex architectures like STFT+CNN and FFT, wavelet packet decomposition (WPD)+2D/3D-CNN, the AE-based approach drastically reduces the computational parameters, making it more feasible for real-time and embedded applications. Even when compared to raw-data-based models, such as deep convolution autoencoder (DCAE)+Bi long-short-term memory (LSTM) and 1D-CNN approaches, this AE approach still maintains a much smaller amount of memory usage. This suggests that the model retains essential EEG features while using significantly fewer parameters. In terms of classification performance, the proposed model achieved a precision of 93.69%. Given the lightweight nature of the model, the precision is acceptable. However, this result also highlights the inherent trade-off between precision and computational resource consumption. In models with higher computational complexity, it is often possible to achieve slightly higher precision by utilizing more complex architectures and larger parameter sets. However, these approaches can be impractical for real-time or embedded applications due to their high memory requirements. Thus, while the precision may be slightly lower than that of more resource-intensive models, the trade-off is justified by the model ability to operate efficiently in real-world scenarios.

TABLE II: Comparison with different methods

Work	Method	Precision (%)	Parameters
[14]	STFT+CNN	-	143.7M
[15]	FFT, WPD+ 2D-CNN, 3D-CNN	99.14	28.8M
[16]	Raw data + DCAE + BiLSTM	98.86	139.6k
[17]	Raw data + 1D-CNN	99.78	113.8k
[18]	Raw data + 1D-CNN	99.62	106.8k
This work	**Raw data + AE**	**93.69**	**41.15k**

V. Conclusion

This study proposed a lightweight AE-based model for epileptic seizure detection, designed to balance precision with computational efficiency. Using channel-wise autoencoders for feature extraction and a K-NN classifier for classification, the model achieved a precision of 93.69%, demonstrating its effectiveness in distinguishing seizure and non-seizure events. Additionally, the model's ability to generate EEG-like signals was validated through a high PCC of 0.9580 and an MSE of 0.4515, confirming the realistic nature of the generated data that could be used in models where the dataset is limited. A key advantage of this approach is its reduced computational complexity, requiring only 41.15k parameters, significantly fewer than traditional deep learning models, making it well-suited for real-time applications in resource-constrained environments such as wearable and embedded medical devices.

Despite promising results in generating new similar-original signals, there is still room for improvement, particularly in terms of classification precision. Although the model precision is good, further enhancement could be achieved without affecting its lightweight nature. Fine-tuning the latent space representations within the autoencoder could optimize the feature extraction process, leading to better classification without significantly increasing computational requirements. Moreover, pruning techniques could be employed to remove unnecessary connections or neurons in the network, further reducing the model size and improving its speed without losing much in terms of classification precision. These approaches allow for continued improvements in classification performance while preserving efficiency, making the model even more suitable for real-time and embedded applications, where computational resources are often limited.

References

[1] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A Multi-View Deep Learning Framework for EEG Seizure Detection," IEEE J. Biomed. Health Inform., vol. 23, no. 1, pp. 83–94, Jan. 2019, doi: 10.1109/JBHI.2018.2871678.

[2] S. Tong and N. V. Thankor, Quantitative EEG Analysis Methods and Clinical Applications. Artech House, 2009.

[3] L. S. Vidyaratne and K. M. Iftekharuddin, "Real-Time Epileptic Seizure Detection Using EEG," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 25, no. 11, pp. 2146–2156, Nov. 2017, doi: 10.1109/TNSRE.2017.2697920.

[4] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. Abd El-Samie, "EEG seizure detection and prediction algorithms: a survey," EURASIP J. Adv. Signal Process., vol. 2014, no. 1, p. 183, Dec. 2014, doi: 10.1186/1687-6180-2014-183.

[5] M. K. Siddiqui, R. Morales-Menendez, X. Huang, and N. Hussain, "A review of epileptic seizure detection using machine learning classifiers," Brain Inform., vol. 7, no. 1, p. 5, May 2020, doi: 10.1186/s40708-020-00105-1.

[6] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.

[7] "Bio-Signal Complexity Analysis in Epileptic Seizure Monitoring: A Topic Review." Accessed: Feb. 21, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/18/6/1720

[8] A. M. Abdelhameed and M. Bayoumi, "Semi-Supervised EEG Signals Classification System for Epileptic Seizure Detection," IEEE Signal Process. Lett., vol. 26, no. 12, pp. 1922–1926, Dec. 2019, doi: 10.1109/LSP.2019.2953870.

[9] H. Daoud and M. Bayoumi, "Deep Learning Approach for Epileptic Focus Localization," IEEE Trans. Biomed. Circuits Syst., vol. 14, no. 2, pp. 209–220, Apr. 2020, doi: 10.1109/TBCAS.2019.2957087.

[10] "A Novel Epilepsy Detection Method Based on Feature Extraction by Deep Autoencoder on EEG Signal." Accessed: Oct. 15, 2024. [Online]. Available: https://www.mdpi.com/1660-4601/19/22/15110

[11] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Accessed: Oct. 15, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[12] J. Guttag, "CHB-MIT Scalp EEG Database." Jun. 09, 2010. [Online]. Available: https://physionet.org/content/chbmit/1.0.0/

[13] P. Baldi, "Autoencoders, Unsupervised Learning, and Deep Architectures," in Proceedings of ICML Workshop on Unsupervised and Transfer Learning, JMLR Workshop and Conference Proceedings, Jun. 2012, pp. 37–49. Accessed: Feb. 20, 2025. [Online]. Available: https://proceedings.mlr.press/v27/baldi12a.html

[14] B. Zhang et al., "Cross-Subject Seizure Detection in EEGs Using Deep Transfer Learning," Comput. Math. Methods Med., vol. 2020, p. 7902072, 2020, doi: 10.1155/2020/7902072.

[15] X. Tian et al., "Deep Multi-View Feature Learning for EEG-Based Epileptic Seizure Detection," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 27, no. 10, pp. 1962–1972, Oct. 2019, doi: 10.1109/TNSRE.2019.2940485.

[16] A. Abdelhameed and M. Bayoumi, "A Deep Learning Approach for Automatic Seizure Detection in Children With Epilepsy," Front. Comput. Neurosci., vol. 15, p. 650050, 2021, doi: 10.3389/fncom.2021.650050.

[17] S. Qiu, W. Wang, and H. Jiao, "LightSeizureNet: A Lightweight Deep Learning Model for Real-Time Epileptic Seizure Detection," IEEE J. Biomed. Health Inform., vol. 27, no. 4, pp. 1845–1856, Apr. 2023, doi: 10.1109/JBHI.2022.3223970.

[18] X. Wang, X. Wang, W. Liu, Z. Chang, T. Kärkkäinen, and F. Cong, "One dimensional convolutional neural networks for seizure onset detection using long-term scalp and intracranial EEG," Neurocomputing, vol. 459, pp. 212–222, Oct. 2021, doi: 10.1016/j.neucom.2021.06.048.