Shiva M. Varnosfaderani, Rihat Rahman, Nabil J. Sarhan, Levin Kuhlmann, Eishi Asano, and Mohammad Alhawari. A Two-Layer LSTM Deep Learning Model for Epileptic Seizure Prediction. In Proceedings of the 3rd IEEE International Conference on Artificial Intelligence Circuits & Systems (AICAS 2021), June 2021. DOI: 110.1109/AICAS51828.2021.9458539.

A Two-Layer LSTM Deep Learning Model for Epileptic Seizure Prediction

Shiva Maleki Varnosfaderani[1], Rihat Rahman[2], Nabil J. Sarhan[1], Levin Kuhlmann[3], Eishi Asano[4], Aimee Luat[4], and Mohammad Alhawari[1]

[1]Department of Electrical and Computer Engineering, Wayne State University, Detroit, USA

[2]Department of Computer Science, Wayne State University, Detroit, USA

[3]Department of Data Science and AI, Faculty of Information Technology, Monash University, Clayton, Australia

[4]Department of Pediatrics and Neurology, Wayne State University, Detroit, USA

***Abstract*— We propose an efficient seizure prediction model based on a two-layer LSTM using the Swish activation function. The proposed structure performs feature extraction based on the time and frequency domains and uses the minimum distance algorithm as a post-processing step. The proposed model is evaluated on the Melbourne dataset and achieves the highest Area Under Curve (AUC) score of 0.92 and the lowest False Positive Rate (FPR) of 0.147 compared to previous work while having sensitivity and accuracy of 86.8 and 85.1, respectively. The proposed system has a low number of trainable parameters, and thus reducing the complexity of resource-constrained applications.**

***Index Terms*—Deep learning, classification, epilepsy seizure prediction, feature extraction, iEEG, LSTM, Melbourne dataset.**

I. Introduction

According to the World Health Organization (WHO), more than 65 million people worldwide suffer from epilepsy, where 3.4 million are living in the United States [1]. Epilepsy is known as a seizure disorder and is diagnosed after a person has at least two seizures that are not caused by some known medical conditions [2]. Epileptic patients have a high risk of Sudden Unexpected Death in Epilepsy (SUDEP), a leading cause of death in people with uncontrolled seizures [3].

Electroencephalography (EEG) plays a vital role in diagnosing epilepsy by monitoring the patients' neural activities. EEG signals can be classified into two main categories, based on the electrodes' placements: *scalp EEG* (sEEG) and *intracranial EEG* (iEEG). sEEG can be recorded from the scalp, while iEEG requires invasive electrodes. EEG signals can be classified into four main categories: *preictal*, *ictal*, *postictal*, and *interictal*, which refer to the time before seizure occurrence, the onset time of the seizure, the time immediately after the seizure, and other periods, respectively [4].

Commercial wearable and implantable devices can monitor EEG in real-time to detect seizure occurrences. For example, the *Responsive Neurostimulation System* (RNS) is an implantable device that can detect the occurrence of focal seizures and provide stimulation to reduce the effect of seizures [5]. *Epihunter Pro* [6], *Nelli* [7], and *Epiltel* [8] are wearable devices that can detect and record seizure activity. In contrast, *Embrace Watch* does not use EEG signals; it detects possible convulsive seizures by measuring sympathetic nervous system activity. There is a need not only to detect seizures but also to predict them before they happen so that the patients can take precautions, stop certain activities (such as driving, swimming alone, or climbing ladders), and/or reduce their effects.

This work was supported by Richard Barber Interdisciplinary Research Program. (Corresponding author: Shiva Maleki Varnosfaderani.)

Deep Learning (DL) methods have been recently used to predict epilepsy with high accuracy [9]–[16]. Although *Convolutional Neural Networks* (CNNs) can be used to classify the data [10], [13], [14], [16], their high computational resources and memory bandwidth requirements limit their applications in resource-constrained devices. In [17], the *Common Spatial Pattern* (CSP) is applied to extract 324 (18×18) features out of 23,040 data points to reduce the number of trainable parameters in the CNN; however, the model has low sensitivity. Due to the capability of learning long-term dependencies, *Long Short-Term Memory Networks* (LSTMs) are recently used in epilepsy prediction applications [9], [11], [12]. In [11], 643 features are extracted from each 5-second segment in the time and frequency domains and then fed to a two-layer LSTM with 128 memory units in each layer, resulting in a large model with more than 500,000 trainable parameters. Several solutions are proposed to reduce the complexity and power consumption of DL networks, including *channel selection* [9]. Most epilepsy prediction architectures are *patient-specific*, where the model is trained and tested using data from the same patient. In contrast, *patient-independent* models allow the training to be based on certain patients and the testing on others. The imbalance data problem occurs in epilepsy prediction systems when the number of interictal samples is much larger than the number of preictal samples, which could lead to a biased model toward interictal samples [18]. To overcome this problem, several techniques can be used to generate synthetic EEG data, including *Overlapping Sampling* [10], *Generative Adversarial Network* (GAN) [13], and *Deep Convolutional Generative Adversarial Networks* (DCGAN) [15].

This paper presents a generalized seizure prediction algorithm based on a two-layer LSTM of 16 memory units followed by dropout, dense, and output layers. The model utilizes the recently proposed *Swish* activation function [19], which does not suffer from the vanishing gradient problem. To address the *data drift* problem, the model employs handcrafted features in both the time and frequency domains. We consider both patient-specific and patient-independent applications; for the latter, we employ normalization. In this study, patient-specific and patient-independent models are trained by an equal number of interictal and preictal samples to overcome the problem of imbalanced data. We present and analyze alternative implementations of the overall solution, including different activation functions, application of normalization, and post-processing in both the patient-specific and patient-independent cases. We analyze three post-processing methods: *arithmetic mean*, *k-of-n*, and *Minimum Distance* (MD). We show that the latter dramatically reduces the number of false predictions. The proposed model provides high-efficiency and low complexity, which is preferable for resource-constrained applications. Under the Melbourne dataset, the model achieves the highest AUC of 0.92 and lowest FPR of 0.147 with sensitivity and accuracy of 86.8 and 85.1, respectively.

The main contributions of this paper can be summarized as follows: (i) considering both patient-specific and patient-independent applications and applying normalization to enhance the performance of the latter, (ii) applying and analyzing the Swish activation function for the first time in seizure prediction, (iii) proposing a simple post-processing technique based on minimum distance to reduce the number of false predictions, (iv) addressing the problem of data drift by utilizing hand-crafted features in the time and frequency domains, and (v) achieving a low number of trainable parameters, thereby leading to better suitability for high-efficiency, low-energy, real-time wearable applications.

The remainder of the paper is organized as follows. Section II presents the proposed system, including pre-processing, feature extraction, classification, and post-processing. Section III shows the performance evaluation methodology, including the dataset, training and testing methods, and main metrics. Section IV presents and analyzes the main results. Finally, conclusions are drawn.

II. Proposed Model

Figure 1 shows the architecture of the proposed model, which comprises *pre-processing*, *feature extraction*, *two-layer LSTM classifier*, and post-processing. The feature extraction is based on the time and frequency domain, whereas post-processing is based on the minimum distance method.

A. Pre-processing

The Melbourne dataset contains some files with low and constant amplitude values across all channels, called *dropout*, caused by machine-related errors. The files comprising the dropout data are removed, while small intervals of dropout data are ignored.

Each 10-minute raw iEEG clip is split into 3-second non-overlapping segments, and then each 35-sequential segment is placed in one group. Each preictal group is assigned to one, and each interictal group is assigned to zero. Considering the 5-sequential LSTM's output, N clips (i.e., 10-minute files in the dataset), 35 sequential segments, 1200 data-points in a 3-second segment, and 16 number of channels, the resulted data dimension is $5\times\ N\times 35\times 1200\times 16$.

We apply normalization to scale the data into a certain range with maximum and minimum values to avoid data scattering. As in [20], each 3-second segment is mapped to the interval [a, b] according to

$$X_k=\frac{b-a}{T_k-t_k}x_k-\frac{aT_k-bt_k}{T_k-t_k}, \tag{1}$$

where X_k is the output of the 3-second segment after normalization, x_k is the input 3-second segment, and T_k and t_k are the maximum and minimum values of x_k, respectively. Note that normalization is only used when evaluating the system for the patient-independent case.

B. Feature Extraction

Performing feature extraction of the EEG signal lowers the complexity and computational cost of the system compared to processing the raw EEG data [21]. Due to significant changes in the EEG data from the interictal to preictal states, time and frequency domain features provide necessary information for epilepsy prediction. The features are extracted from each 3-second iEEG segment with resulted data dimension of $5N\times 35\times 256$, where 35 is the number of sequential segments, and 256 is the number of features.

The time-domain features include energy distribution, deviation, peak-to-peak values, and the number of zero-crossings, as well as four statistical moments: *mean*, *variance*, *skewness*, and *kurtosis* [11]. The energy distribution is given by

$$E_n=\frac{1}{M}\Sigma_{m=1}^{M}x_i^2, \tag{2}$$

where M is the length of the segment.

The total number of extracted features in the time-domain is $8\times k$, where k is the number of channels. Note that the mean, peak-to-peak, and deviation are excluded from the features when the normalization step is applied since they have the same values over all the segments.

The frequency-domain features include the spectral power intensity (PSI), which is extracted from the EEG data for 8 frequency bands from 0.1 Hz to 180 Hz using PyEEG library [22], including 0.1-4 Hz, 4-8 Hz, 8-12 Hz, 12-30 Hz, 30-50 Hz, 50-70 Hz, 70-100 Hz, and 100-180 Hz.

C. Classification

The extracted features are standardized and provided as inputs to the LSTM Network. A two-layer LSTM is proposed, which comprises 16 memory units followed by dropout, dense, and output layers, as shown in Figure 1. To avoid overfitting, dropout is applied to the first and second layers with the factor of 10% and 30%, respectively. The cross-entropy loss function is selected as the cost function, and the Adaptive Moment Estimation (Adam) is used as an optimizer. The Swish function is used for all activations to provide the best performance:

$$Swish=x\times Sigmoid=\frac{x}{1+e^{-x\beta}}. \tag{3}$$

We select $\beta=1$ [19].

D. Post-processing

To decrease the number of false predictions, post-processing is applied to assign an output to a group of sequential outputs from the network. We analyze three post-processing methods *arithmetic mean*, *k-of-n*, and *Minimum Distance* (MD).

III. Performance Evaluation Methodology

As discussed earlier, we use the Melbourne dataset, consisting of long-term iEEG data from 15 patients with refractory focal epilepsy. A subset of three patients (Patients 3, 9, and 11) with the lowest seizure prediction performance from the NeuroVista trial is publicly available via Epilepsyecosystem.org. The data is recorded using (4×4) invasive strip electrodes, placed on the focal hemisphere of the patients to record a total duration of 1,155 hours with a sampling rate of 400 Hz. The duration of interictal and preictal segments is 1-hour, split into six 10-minute files [24]. Table I summarizes the details of this dataset.

To overcome the imbalanced dataset problem, patient-specific and patient-independent models are trained by an

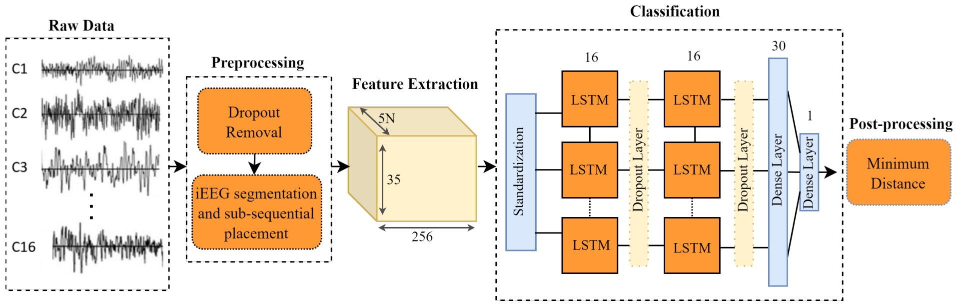

Fig. 1. The proposed two-layer LSTM Model for epileptic seizure prediction.

TABLE I
DETAILS OF THE MELBOURNE DATASET [23]

Dataset Features	***Patient 1***	***Patient 2***	***Patient 3***
Gender	Female	Female	Female
Age (years)	22	51	50
Recording Period (days)	559	393	374
Total Recording (hours)	173	507	475
# of Seizures	390	204	545
# of train files (interictal)	570	1836	1908
# of train files (preictal)	256	222	255
# of test files (interictal)	16	18	18
# of test files (preictal)	46	279	188
# of Channels	16	16	16
Sampling Rate (Hz)	400	400	400

equal number of interictal and preictal samples from training clips in the Melbourne dataset. During the evaluation, all public test files excluding dropout clips are considered. The model is trained by 149,100 3-second samples and tested by 42,875 3-second samples. To investigate the generality of the proposed model, the leave-one-patient-out cross-validation (LOOCV) technique is used in the patient-independent model, where the model is trained by the samples from two patients and tested by the test files of the excluded patient.

The considered metrics include AUC, sensitivity, FPR, and accuracy, where the last three metrics are computed using a threshold of 0.5.

IV. RESULTS PRESENTATION AND ANALYSIS

Table II compares the performance of various alternative implementations of the proposed solution in both patient-specific and patient-independent models, including the choice of activation functions, application of normalization, and the post-processing method. The results indicate that the best AUC is obtained by the MD and arithmetic mean methods for the patient-specific models with Swish and Sigmoid activation functions, respectively. However, the Swish function has better performance than Sigmoid by having a higher AUC Score and accuracy of 0.92 and 85.1, respectively, as well as a lower FPR of 0.147. This improvement can be attributed to its smoothness property, which produces better optimization and generalization in the training step [25]. Applying MD dramatically decreases the number of false predictions to 0.147, thereby leading to high accuracy and AUC score. This improvement is due to assigning an output with the highest similarity to preictal or interictal samples to the 5-sequential network's outputs.

Despite having lower performance in the patient-specific model, applying the normalization has improved the AUC in the patient-independent for all the three models, based on the arithmetic mean, MD, and k-of-n methods. This is due to removing the scattering in the data, and thus the effect of the maximum and minimum values, which sometimes are caused by wrong recordings, are removed. Note that the best AUC score is obtained by the model using MD method in the post-processing step, while normalization decreases the number of trainable parameters from $20,125$ to $17,053$ by reducing the number of features.

Table III summarizes the performance of the proposed model, compared to the most recent methods that used the Melbourne dataset. In the Melbourne University Seizure Prediction Competition on Kaggle, the first-place model, which deployed more than 3000 hand-crafted features and 11 different classification models, achieved an average AUC Score of 0.853 and 0.807 on the public and private test files, respectively [24]. However, for resource-constrained real-time applications, computing more than 3000 features may not be practical. In [14], a prediction model based on a multi-scale CNN architecture was proposed, where downsampling by a factor of two, segmentation, and STFT are used in the pre-processing step, and a multi-scale CNN is used in the classification step. In [16], a model based on semi-dilated CNN architecture is used, while Continuous Wavelet Transform (CWT) is used to convert the signal to an image. In [15], a prediction model based on Convolutional Epileptic Seizure Predictor (CESP) is used with a DCGAN to generate synthetic EEG data. The STFT and one-class support vector machine (SVM) are used to convert one-dimension data to image and evaluate the performance. By introducing the MD method as a post-processing step and applying Swish as an activation function, our proposed structure obtains the highest AUC score of 0.92 compared to previous work, with an enhancement of 3.7%. Extracting the most important iEEG hand-crafted features plays a pivotal role in simplifying our structure and makes it a suitable option for resource-constrained real-time applications with a lower number of trainable parameters of 20,125 compared to [15] and [16].

V. CONCLUSION

We have proposed an epileptic prediction model based on hand-crafted features and a two-layer LSTM with the Swish activation function. The experimental results and comparison with previous work demonstrate that our proposed model is efficient, suitable, and reliable for resource-constrained applications of seizure prediction by achieving the highest AUC of 0.92 and the lowest number of reported trainable parameters. The performance of the system is improved and the FPR is decreased by applying the Swish activation function for the first time in epilepsy prediction systems and proposing a novel post-processing method based on the highest similarity to interictal and preictal samples for assigning the final output. Further, applying normalization in the preprocessing step improved the performance of the patient-independent model

TABLE II
ANALYSIS OF ALTERNATIVE IMPLEMENTATIONS OF THE PROPOSED SOLUTION IN PATIENT-SPECIFIC AND PATIENT-INDEPENDENT MODELS

Alternative Implementation/Model	Patient	*AUC*			*Sen%*			*FPR* (h^{-1})			*Accuracy*		
		Mean	MD	*k-of-n*	Mean	MD	*k-of-n*	Mean	MD	*k-of-n*	Mean	MD	*k-of-n*
Swish-Swish Activation Functions/ Patient-Specific	**Pat1**	0.80	0.907	0.88	66.7	66.7	100	0.28	0.08	0.24	71.4	89.3	78.5
	Pat2	0.91	0.94	0.87	100	100	100	0.32	0.14	0.26	69	86.5	75
	Pat3	0.919	0.914	0.859	100	93.8	100	0.50	0.221	0.28	55.2	79.4	74.4
	Mean	**0.876**	**0.92**	**0.87**	**88.9**	**86.8**	**100**	**0.367**	**0.147**	**0.26**	**65.2**	**85.1**	**76**
Sigmoid-Sigmoid Activation Functions/ Patient-Specific	**Pat1**	0.747	0.613	0.82	66.7	66.7	100	0.32	0.36	0.36	67.9	44.3	67.9
	Pat2	0.96	0.91	0.85	100	100	100	0.48	0.26	0.30	53.8	75	71.2
	Pat3	0.918	0.864	0.831	93.8	93.8	93.8	0.297	0.268	0.275	72.7	75.2	74.5
	Mean	**0.875**	**0.796**	**0.834**	**86.8**	**86.8**	**97.9**	**0.366**	**0.296**	**0.312**	**64.8**	**64.8**	**71.2**
Swish-Swish Activation Functions+ Normalization/ Patient-Specific	**Pat1**	0.64	0.773	0.69	66.7	100	66.7	0.36	0.36	0.28	67.9	67.9	71.4
	Pat2	0.96	0.92	0.92	100	100	100	0.3	0.18	0.16	71.1	82.7	84.6
	Pat3	0.879	0.90	0.839	81.3	93.8	100	0.221	0.208	0.322	78.2	80.6	70.9
	Mean	**0.826**	**0.864**	**0.816**	**82.7**	**97.9**	**88.9**	**0.294**	**0.249**	**0.254**	**72.4**	**77.1**	**75.6**
Swish-Swish Activation Functions/ Patient-Independent	**Pat1**	0.853	0.867	0.593	100	100	66.7	0.44	0.2	0.48	60.7	82.1	53.4
	Pat2	0.71	0.74	0.82	100	100	100	0.30	0.28	0.24	71.1	73.1	76.9
	Pat3	0.773	0.63	0.56	75	68.8	56.3	0.349	0.49	0.442	66.1	52.7	55.8
	Mean	**0.779**	**0.746**	**0.658**	**91.7**	**89.6**	**74.3**	**0.363**	**0.323**	**0.387**	**66**	**69.3**	**62**
Swish-Swish Activation Functions+ Normalization/ Patient-Independent	**Pat1**	0.84	0.907	0.94	100	100	100	0.44	0.16	0.12	60.7	85.7	89.3
	Pat2	0.83	0.80	0.70	100	100	50	0.56	0.30	0.09	46	71.1	88.4
	Pat3	0.688	0.697	0.712	65.1	62.5	62.5	0.497	0.295	0.201	52.1	73.9	78.2
	Mean	**0.786**	**0.801**	**0.784**	**88.4**	**87.5**	**70.8**	**0.499**	**0.252**	**0.137**	**52.9**	**76.9**	**85.3**

TABLE III
COMPARATIVE RESULTS OF THE PROPOSED MODEL WITH RECENT SEIZURE PREDICTION METHODS

Ref.	*Preprocessing and Feature Extraction*	*Classification*	*AUC Score Public/Private*	*Sensitivity %*	*FPR (h^{-1})*	*Accuracy %*	*# of Trainable Parameters*
[24]	Spectral power, distribution statistics, AR error, fractal dimensions, other features	Extreme gradient, boosting, SVM	0.853/0.807	-	-	-	-
[14]	Downsampling by a factor of 2, STFT	Multi-Scale CNN	0.840/-	87.85	-	-	-
[16]	Continuous wavelet transform (CWT)	Semi-dilated CNN	0.883/-	89.52	-	-	69,129,612
[15]	STFT	CESP	-	92.87	0.15	-	2,200,000
[26]	Non-negative matrix factorization	SVM	-	69	.76	-	-
This Work	Time and frequency domain features	Two-Layer LSTM	0.92/-	86.8	0.147	85.1	20,125

by removing the scattering in the data, which opens the door for more innovations in preprocessing techniques to generalize epileptic prediction structures.

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/epilepsy/.

[2] *Neurological Disorders: Public Health Challenges*. World Health Organization, 2006.

[3] T. Tomson, T. Walczak, M. Sillanpaa, and J. W. A. S. Sander, "Sudden unexpected death in epilepsy: A review of incidence and risk factors," *Epilepsia*, vol. 46, no. s11, pp. 54–61, 2005.

[4] M. Mula and F. Monaco, "Ictal and peri-ictal psychopathology," *Behavioural Neurology*, vol. 24, pp. 21–25, 2011.

[5] https://www.epilepsy.com/learn/treating-seizures-and-epilepsy/devices/responsive-neurostimulation-rns.

[6] https://www.epihunter.com/professionals.

[7] https://neuroeventlabs.com/for-patients.

[8] https://www.epitel.com/.

[9] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 804–813, 2019.

[10] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, and O. Kavehei, "A generalised seizure prediction with convolutional neural networks for intracranial and scalp electroencephalogram data analysis," *CoRR*, vol. abs/1707.01976, 2017.

[11] K. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in Biology and Medicine*, vol. 99, pp. 24 – 37, 2018.

[12] D. Thara, B. PremaSudha, and F. Xiong, "Epileptic seizure detection and prediction using stacked bidirectional long short term memory," *Pattern Recognition Letters*, vol. 128, pp. 529–535, 2019.

[13] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 465–474, 2019.

[14] R. Hussein, M. O. Ahmed, R. Ward, Z. J. Wang, L. Kuhlmann, and Y. Guo, "Human intracranial EEG quantitative analysis and automatic feature learning for epileptic seizure prediction," *CoRR*, vol. abs/1904.03603, 2019.

[15] K. Rasheed, J. Qadir, T. J. O'Brien, L. Kuhlmann, and A. Razi, "A generative model to synthesize EEG data for epileptic seizure prediction," *arXiv preprint arXiv:2012.00430*, 2020.

[16] R. Hussein, S. Lee, R. Ward, and M. J. McKeown, "Epileptic seizure prediction: A semi-dilated convolutional neural network architecture," *arXiv preprint arXiv:2007.11716*, 2020.

[17] F. Ali, S. El-Sappagh, S. R. Islam, D. Kwak, A. Ali, M. Imran, and K.-S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208 – 222, 2020.

[18] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Special issue on learning from imbalanced data sets," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 1–6, 2004.

[19] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.

[20] A. R. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari, and F. Kolahdouzan, "Segmentation of dermoscopy images using wavelet networks," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1134–1141, 2012.

[21] S. Gibson, J. W. Judy, and D. Markovic, "Technology-aware algorithm design for neural spike detection, feature extraction, and dimensionality reduction," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 5, pp. 469–478, 2010.

[22] F. Bao, X. Liu, and C. Zhang, "PyEEG: An open source python module for EEG/MEG feature extraction," in *Computational Intelligence and Neuroscience*, 2011.

[23] R. Rahman, S. Maleki Varnosfaderani, O. Makke, N. J. Sarhan, E. Asano, A. Luat, and M. Alhawari, "Comprehensive analysis of EEG datasets for epileptic seizure prediction," *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, Daegu, Korea, 2021.

[24] L. Kuhlmann, P. Karoly, D. R. Freestone, B. H. Brinkmann, A. Temko, A. Barachant, F. Li, J. Titericz, Gilberto, B. W. Lang, D. Lavery, K. Roman, D. Broadhead, S. Dobson, G. Jones, Q. Tang, I. Ivanenko, O. Panichev, T. Proix, M. Náhlík, D. B. Grunberg, C. Reuben, G. Worrell, B. Litt, D. T. J. Liley, D. B. Grayden, and M. J. Cook, "Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG," *Brain*, vol. 141, pp. 2619–2630, 08 2018.

[25] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018.

[26] O. Stojanović, L. Kuhlmann, and G. Pipa, "Predicting epileptic seizures using nonnegative matrix factorization," *PloS one*, vol. 15, no. 2, p. e0228025, 2020.