SPERTL: Epileptic **S**eizure **P**rediction using **E**EG with **R**esNets and **T**ransfer **L**earning

Umair Mohammad and Fahad Saeed
Knight Foundation School of Computing and Information Sciences, Florida International University
Miami, FL 33199, USA
Emails: {umohamma,fsaeed}@fiu.edu

2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-8791-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/BHI56158.2022.9926767

Abstract—**Epilepsy is a chronic condition that causes repeat unprovoked seizures and many epileptics either develop resistance to medications and/or are not suitable candidates for surgical solutions. Hence, these recurring unpredictable seizures can have a severely negative impact on quality of life including an elevated risk of injury, social stigmatization, inability to take part in essential activities such as driving and possibly reduced access to healthcare. A predictive system that informs patients and caregivers about a potential upcoming seizure ahead of time is not only desirable but an urgent necessity. In this paper, we contribute by designing and developing patient-specific epileptic seizure (ES) prediction models using only electroencephalography (EEG) data with residual neural networks (ResNets) and transfer learning (TL) - (SPERTL). We train our proposed model on EEG data from 20 patients with a seizure prediction horizon (SPH) of 5 minutes and use the validation data to plot precision-recall curves for selecting the best thresholds. Testing on unseen data shows our model outperforms the state-of-the-art methods by achieving the highest average sensitivity of 88.1%, specificity of 92.3%, and accuracy of 92.3%. Our results also demonstrate the proposed model is less susceptible to false positives while maintaining a high positive prediction rate.**

Index Terms—**residual neural networks, CNN, ResNet, Electroencephalography, EEG, epileptic seizure prediction**

I. Introduction

Epilepsy is a neurological disorder that causes repeat unprovoked seizures and affects over 3.4 million people in the USA and over 65 million worldwide. An epileptic seizure (ES) is defined as two or more unprovoked recurring seizures with a gap greater than 1 day [1], [2]. During a seizure, increased electrical activity of the brain results in symptoms including but not limited to shaking, convulsions, and short-term memory loss [1]. Seizures may include one or more of the following phases: aura, preictal (before a seizure), ictal (main seizure activity) and interictal (period between two seizures) [3]. Repeat seizures can be debilitating because they interfere with daily activities, may lead to social stigma, and can potentially result in serious injury or death. Although treatable with medications or surgical resection, epilepsy can become drug-resistant and many patients can be excluded from surgery due to comorbidities and other risk factors, risk of loss of cognitive functions, failure to localize epileptogenic zone or even a low socioeconomic status.

Existing commercial solutions based on wearables are only capable of detection, mostly based on movement sensors which are unreliable, exclude some types of seizures and have a high drop-rate due to their unreliability. Alternative seizure control methods based on implants such as responsive neuro-stimulation (RNS) [4] and vagus nerve stimulation (VNS) [5] have some of the same obstacles as surgical solutions. Therefore, wearable solutions that can reliably predict seizures ahead of time are urgently needed so that patients/caregivers can take the necessary precautionary measures. Consequently, the focus has shifted to seizure prediction using electroencephalography (EEG) data which directly measures electrical activity of the brain. Although the protocol for diagnosing epilepsy in clinical settings is well-established, ES prediction with EEG for use in ambulatory settings remains a significant challenge; especially the design of accurate predictive models.

In this work, we design and develop an ML model for patient-specific ES prediction based on the residual neural network (ResNet) and transfer learning (TL). Our proposed model SPERTL performs epileptic **S**eizure **P**rediction using only **E**EG data by applying **R**esNets with **TL**. ResNet is a form of convolutional neural network (CNN) that features additional residual layers with skip connections. At the time, it provided the highest performance for ImageNet classification when trained on 1.28 million images with $\sim$2 million parameters [6] and has recently found use in time-series classification [7]. TL works by re-training a successfully pre-trained model for one problem on another problem with either a different task or a different domain with the same task. Re-use of pre-trained models with TL is of high research and commercial interest due to the amount of resources (time, money, energy, etc.) required for training these models in the first place [8].

A. Motivation and Contributions

Initially, techniques such as the recurrent neural network (RNN), gated recurrent unit (GRU) and long short-term memory (LSTM) were popular for time-series classification but, 1d-CNNs have made a comeback [7]. Adding residual layers further enhances CNN performance as a recent benchmark [7] shows that ResNets provide the best performance for time-series classification. Overall, ResNets provide superior performance, are better able to prevent overfitting compared to CNNs, and can deal with class imbalance. All of these properties are desirable for the ES prediction task with severe class-imbalance (e.g. less than 2% of the EEG recordings in the used dataset contain seizure activity). To design the ResNet, we can either start from scratch or adopt well-

TABLE I
COMPARISON OF EXISTING MODELS IN THE LITERATURE FOR ES PREDICTION.

Ref.	Feature Extraction	End-to-end?	Classifier	Class-imbalance	Sensitivity
[9]	FT + Genetic Algorithm	No	Logistic Regression	Considered	61.7%
[10]	Several Handcrafted	No	Ensemble (SVM, CNN, LSTM)	GANs	96.3%
[11]	Short-time FT	No	LSTM	GANs	93.0%
[12]	AE+CNN	Yes	LSTM	Equal preictal and interictal	99.6%
[13], [14]	CWT conversion to 2D images	Yes	Semi-Dilated CNN +FC	Equal preictal and interictal	99.7%
[15]	Neural arch. search + CNN	Yes	FC	Overlapping preictal segments	99.8%
[16]	CNN with 1D & 2D pooling (each lead a dimension)	Yes	FC	Overlapping preictal segments	98.8%
[17]	CNNs	Yes	FC	Considered	81.9%
[18]	CNNs	Yes	FC	Considered	68.8%
[19]	CNNs	Yes	Bi-LSTM	Considered	76.6%

known architectures (e.g. ResNet101). However, we employ TL because it speeds up the model development process, can achieve a higher accuracy compared to a completely new architecture, and enhance the accuracy when there is less data by transferring the knowledge from models trained on larger datasets [20]. For example, the adopted model was trained on a dataset of $>$ 2.5 million patients [21] whereas for patient-specific ES prediction, data is available from only 1 patient at a time.

Specifically, the contributions of this paper are as follows:

- It represents the first work to use a ResNet for ES prediction from raw EEG data with TL.
- Does not make an assumption of a large preictal period within which a seizure should be predicted. Rather, a stricter approach is used as explained in Section III.
- The complete data is used for seizure prediction which better reflects the real-world scenario where the occurrence of a seizure is expected to be an extremely rare event. Several previous works use only parts of the EEG recordings to eliminate the class imbalance.
- Achieves a higher accuracy compared to the state-of-the-art given these constraints.

II. Related Work

With the advent of machine learning (ML), researchers have leveraged large amounts of data collected from long-term EEG recordings, such as the Children Hospital Boston-Massachusetts Institute of technology (CHB-MIT) dataset [22], to develop predictive models. Early ES prediction models involved feature extraction with time-frequency analysis e.g., continuous/discrete wavelet transform (CWT/DWT) or a version of the Fast Fourier Transform (FFT) followed by traditional techniques such as the support vector machine (SVM), random forest (RF), naïve Bayes (NB) or the neural network (NN) [2]. Unfortunately, these techniques reached a performance ceiling. The introduction of deep learning (DL) has revived interest in seizure prediction because of its ability to provide a superior performance and do end-to-end prediction from raw EEG data which is desirable for deployment on low-cost hardware. DL techniques include the deep NN (DNN), CNNs, LSTMs and the generative adversarial network (GAN) among others.

Though DL application to raw data is desirable, some works [9]–[11] apply DL techniques after manual feature extraction. In contrast, the works of [12]–[16] are end-to-end and achieve a sensitivity $>98\%$, but they assume balanced classes whereas seizures are very rare events making the problem severely imbalanced. For example, [12]–[14] select data samples such that there is an equal preictal and ictal duration whereas [15], [16] artificially increase the amount of preictal data by using overlapping segments while keeping the interictal segments non-overlapping; both are not reflective of the real-world. The works of [17]–[19] take into account the class imbalance and design simple 1d-CNNs for ES prediction. Therefore, they suffer from low performance with average sensitivities of 81.9%, 68.8% and 76.6%, respectively. Furthermore, the state-of-the-art work by [19] which uses a Bi-LSTM for classification, shows models tested after eliminating imbalance perform poorly when tested with the complete data. For example, the performance of the model in [12] comes down to below 70% from a near-perfect score. Table I summarizes the details of all discussed models.

As observable, existing techniques for ES prediction from raw EEG data mostly focus on simple 1-D CNNs or LSTM networks for feature extraction and prediction whereas our work is the first to introduce ResNet. Moreover, many existing works do not include all of the data for training and testing or use techniques to eliminate class imbalance which is not reflective of the real-world. In contrast, the works that consider the whole dataset do not achieve high performance metrics. Our work is the first to introduce ResNets for ES prediction with a higher performance on class-imbalanced data.

III. Methods

A. Data

CHB-MIT dataset [22]: The CHB-MIT dataset comprises approximately 24-hour wearable EEG scalp recordings from 22 patients split into files of approximately 1-hour recordings each, with 1 patient providing 2 sets for a total of 23 recordings. The data was collected from 23 electrodes placed using the international 10-20 system sampled at 256Hz with a 16-bit resolution. A total of 198 seizure events were recorded and they are clearly marked in the annotation files which states the number of seizures in every recording as well as the respective start and end times for each seizure. The data is split into smaller segments of $T_{seg} = 20$s to be used as raw input to the model. Recall that we propose an end-to-end DL model without any manual feature extraction. Two important concepts

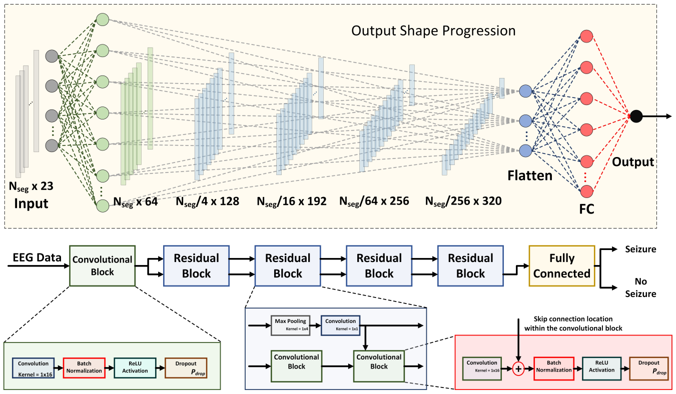

Fig. 1. An illustration of the model. The top part shows how the output shape changes over successive residual layers. The center part shows a block overview of the model. The bottom-left describes the basic convolutional block whereas the bottom-center part describes the overall residual block. The position of the skip connection within the second convolutional block of the residual block is shown on the bottom-right.

in ES prediction include the seizure prediction horizon (SPH), which defines how far ahead the seizure should be predicted, and the seizure occur period (SOP), which is the window of time within which the seizure should occur.

With respect to our data, let us further define the start duration of the seizure as marked by experts as t_{start} and the end by t_{end}. Then, any segment that contains even one reading falling within the window of the seizure duration is considered an ictal segment. An additional variable defining the desired prediction time T_{pred} represents the duration before start of a seizure such that the segment that contains the first reading $t_{start} - T_{pred}$ will be associated with the preictal label. All subsequent segments before t_{start} will be marked as preictal and all remaining segments interictal. In this way, the prediction problem can be reformulated as a simple classification task. The choices for T_{seg} and T_{pred} will dictate the accuracy and will be further dictated by how far ahead the user wants to be warned of an upcoming seizure. In this work, we set T_{pred} or the SPH to be 5 minutes. We do not evaluate our model based on its capability to predict a seizure within a given SOP. Rather, we use a more stringent evaluation and purely quantify its performance based on the capability to correctly predict all preictal segments. More details about the evaluation metrics are provided after the model description; direct comparison regarding the capability to predict seizures within a given SOP will be left for future work.

B. Proposed Model

As described, our proposed model SPERTL applies TL to the ResNet architecture developed by [21]. TL is characterized by a source task S and a target task T. Each task has a domain denoted by D_S and D_T comprising the feature space (X_S and X_T) and associated probability distributions P(X_S) and P(X_T). The tasks are defined by the labels Y_S and Y_T and the predictive functions are denoted by P_S and P_T. Because our tasks are the same (binary classification of a disorder), it is a heterogeneous TL problem since $X_T \neq X_s$. However, the feature space is only different in the sense that the raw EEG data with 23 leads will have a different shape than the ECG data with 12 leads. In contrast, the overall strategy is justified because the final feature space after the convolutional layers is similar. Recall that these layers are considered automatic feature extractors. We modify the input layer and re-train the residual layers to do domain adaptation whereas the fully connected (FC) layers perform the function of P_T which is re-trained using P_S as a starting point.

The overall model accepts raw EEG data as input into a convolutional block, which is followed by 4 residual blocks for feature extraction. The output of the last block is flattened, and an FC layer is introduced for classification. The convolutional block comprises a convolution filter, batch normalization, rectified linear unit (ReLU) activation, followed by dropout. Each residual block features 2 convolutional blocks where the second convolutional block accepts the output of the previous convolutional block within the residual block, and also the output of the skip connection from the previous block after max pooling and 1x1 convolution. The dropout probability in the first convolutional block is zero but is set to 0.8 for all other convolutional blocks inside the residual blocks. The convolutional filter length starts with 64, increasing by 64 in each subsequent residual block. Our implementation is adapted to start with channel length similar to the input size which may range from 1024 to 5120 depending upon the selection of segment duration and is sub-sampled by 4 after every residual block. The output of the last residual block is flattened followed by the FC layers and sigmoid activation for classification. Figure 1 describes the model.

C. Evaluation Metrics

Consider a seizure has occurred, a prediction of a seizure represents a true positive (TP) whereas a prediction of no seizure is a false negative (FN). In contrast, if a seizure has not occurred, a prediction of a seizure is a false positive (FP) and a true negative (TN) otherwise. Although the existence of FN's reduces the value of seizure prediction/detection technologies, FP's are as dangerous because they create anxiety, cause stress, and can lead to the discontinuation of usage over time. Let us define the following metrics:

$$
\begin{aligned}
Sensitivity &= \frac{TP}{TP+FN} \\
Specificity &= \frac{TN}{TN+FP} \\
Accuracy &= \frac{TP+TN}{TP+TN+FP+FN}
\end{aligned}
\tag{1}
$$

The sensitivity is also known as the TP rate which (TPR) which is the complement of the FN rate (FNR) such that ($Sensitivity = TPR = 1 - FNR$). This sensitivity will capture the ability of the classifier to reduce FN's. Because specificity is the TN rate which is the complement of the FP rate ($Specificity = TNR - 1 - FPR$), it will capture the ability to reduce FP's while correctly identifying TN's. Both

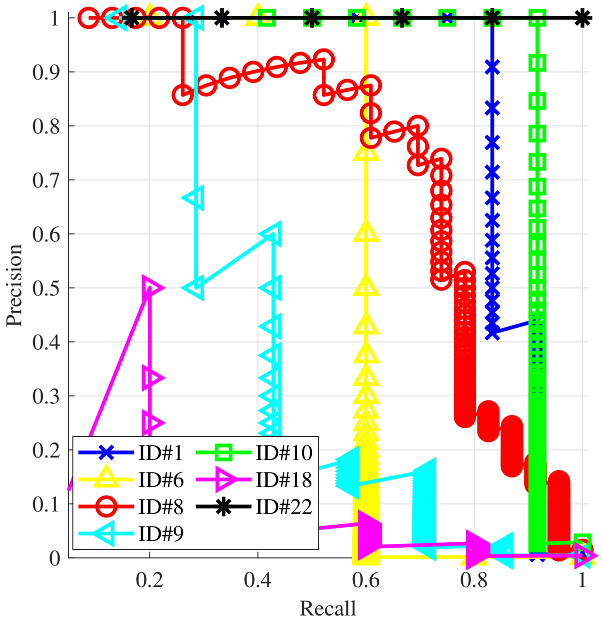

Fig. 2. The PRC curve drawn for several patients from their respective validation dataset. The threshold was selected to maximize the F-1 score.

metrics are important because the accuracy may not capture the true performance due to the severe class imbalance.

Recall that the classifier outputs a probability threshold and for balanced binary class problems, the threshold for a positive label is a probability of greater than 0.5. On the other hand, this strategy fails for class-imbalanced problems. During validation, the precision-recall curve (PRC) will be used to select an initial threshold based on the highest F1-score ($\frac{2 \times Precision \times Recall}{(Precision+Recall)}$). The threshold will be tuned to ensure the maximum possible accuracy with lowest numbers of FNs/FPs and the final accuracy will be evaluated on the unseen test data. For the sake of stringent evaluation, we report the results from the model directly where classification as a pre-ictal segment indicates a seizure and interictal indicates no seizure.

IV. Results and Discussion

Figures 2 illustrates the PRC curve for a set of patients chosen from the complete 20 patients for illustrative purposes. Using validation data, we generate a set of probabilities that are output by the model, and each point on the plot corresponds to a pair of either (Recall, Precision) or (FPR, TPR) achieved by setting that probability as a threshold for seizure vs non-seizure prediction. The ideal model should be able to achieve a perfect score for both, the precision and the recall, which is rarely possible and hence, we select the threshold that maximizes the F-1 score initially and then tune it. From Figure 2, it is clear that SPERTL only achieves a perfect model for patient 22 during validation and this also translated into 100% test accuracy.

Table II shows the achieved sensitivity (sen.), specificity (spe.), and accuracy (acc.) for 20 patients in the CHB-MIT dataset by SPERTL. The patient ID's highlighted in bold font are for patients which were used for ES prediction in [19]. The achieved average sensitivity is 91.20% and an average specificity of 93.80% is achieved for predicting a seizure 5 minutes ahead of time. Furthermore, a higher specificity indicates that the model was able to reduce the likelihood of generating FP's. However, this comes at a cost of a few FN's as it resulted in a lower sensitivity. For patients 5, 11 and 22, the designed model was able to differentiate between the pre-ictal and interictal segments perfectly even during the testing phase. For several patients, the model was able to achieve a perfect score for either sen. (ID's: 3, 14, 17, 18, 19, 21) or for spe. (ID's: 1, 3, 4 and 10). For the remaining patients, the model did achieve scores of greater than 90% for at least one metric except for patient ID 7. Further, the case of patients 6 and 18 are outliers because despite performing very well for one metric, the other metric had a low score of 60%.

Comparing our work to the state-of-the-art [19], it can be seen that our proposed ResNet model provided a higher accuracy for all patients except for patient 6 where our achieved accuracy of 60.0% was lower compared to 77.4%. For patients 1 and 2, our model had a lower sensitivity of 83.3% and 60.0% compared to 88.4% and 82.7%, respectively, and a lower specificity of 60.0% compared to 66.5% for patient 18. All of these results are summarized in Table III. Specificity measures the ability of the model to identify true negatives while minimizing the false positives. For patient 18, our model actually has a sensitivity of 1 which means that all seizures are predicted (none is missed), i.e. there are 0 false negatives. In contrast, both works show a low specificity which indicates a high false positive rate for this particular patient. This may be because there are very few seizures to begin with, the preictal duration is too low, and lastly, the preictal profile makes it very hard to predict an upcoming seizure.

TABLE II
Testing results for seizure prediction with SPERTL

ID	Sen.	Spe.	Acc.
1	83.3%	100.0%	99.9%
3	100.0%	100.0%	100.0%
4	90.0%	100.0%	100.0%
5	100.0%	100.0%	100.0%
6	60.0%	97.6%	97.5%
7	87.5%	72.2%	72.2%
8	95.7%	92.5%	92.6%
9	85.7%	95.8%	95.8%
10	91.7%	100.0%	100.0%
11	100.0%	100.0%	100.0%
12	92.5%	74.5%	74.8%
14	100.0%	98.6%	98.6%
15	96.2%	89.1%	89.2%
16	50.0%	96.1%	96.0%
17	100.0%	99.9%	99.9%
18	100.0%	60.0%	60.0%
19	100.0%	99.9%	99.9%
21	100.0%	99.8%	99.8%
22	100.0%	100.0%	100.0%
23	90.9%	99.9%	99.8%
Avg.	**91.2%**	**93.8%**	**93.8%**

TABLE III
SPERTL SEN., SPE. AND ACC. COMPARED TO STATE-OF-THE-ART [19]

	Wang et al. [19]			SPERTL		
ID	Sen.	Spe.	Acc.	Sen.	Spe.	Acc.
1	**88.4%**	91.2%	89.8%	83.3%	**100.0%**	**99.9%**
6	**82.7%**	66.2%	74.4%	60.0%	**97.6%**	**97.5%**
8	78.5%	83.3%	80.9%	**95.7%**	**92.5%**	**92.6%**
9	81.4%	71.9%	76.7%	**85.7%**	**95.8%**	**95.8%**
10	75.8%	71.7%	73.8%	**91.7%**	**100.0%**	**100.0%**
18	88.4%	**66.5%**	**77.4%**	**100.0%**	60.0%	60.0%
22	90.7%	63.3%	77.0%	**100.0%**	**100.0%**	**100.0%**
Avg	82.7%	72.4%	77.6%	**88.1%**	**92.3%**	**92.3%**

The higher accuracy is the result of higher specificity achieved by SPERTL, which also implies an ability to improve the TNR which reduces the FPR. In [19], because only seizures with certain properties are chosen (at least one hour of preictal time and 4 hours of interictal time), the class imbalance is lower. Despite using the complete dataset, the ability of SPERTL to have a lower number of FP's is desirable. One reason for this maybe the ability of ResNets to deal with class-imbalance. For example, a simple 34-layer ResNet without stacked layers can reduce the top-1 error rate by 3.50% compared to plain CNNs for ImageNet classification. Compared to such tasks, medical event detection/prediction is even more imbalanced which highlights ResNet superiority.

V. CONCLUSION

In this work, we successfully developed a model called SPERTL which used TL techniques to train a ResNet for early ES prediction from EEG data. Our proposed model was trained and tested on a set of 20 patients from the CHB-MIT dataset for patient-specific 5-minute-ahead seizure prediction. The experiments have shown SPERTL has a superior capability to differentiate between preictal and interictal segments and it outperformed the state-of-the-art method in terms of sensitivity by 5.4%, specificity by 19.9%, and accuracy by 14.7%. These are encouraging preliminary results, and we are working towards improving the seizure prediction accuracy within a given SOP which will result in lower false alarm rates. Other future goals include developing energy-efficient federated models for real-time deployment on hardware.

REFERENCES

[1] Cure Epilepsy, "What are the phases of seizures? – CURE Epilepsy," 2020. [Online]. Available: https://www.cureepilepsy.org/for-patients/what-are-the-phases-of-seizures-of-epilepsy/

[2] K. Rasheed, A. Qayyum, J. Qadir, S. Sivathamboo, P. Kwan, L. Kuhlmann, T. O'Brien, and A. Razi, "Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review," pp. 139–155, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9139257

[3] A. Saeed Malik and H. Ullah Amin, *Designing EEG Experiments for Studying the Brain: Design Code and Example Datasets*, 1st ed. Elsevier, 2017. [Online]. Available: https://www.elsevier.com/books/designing-eeg-experiments-for-studying-the-brain/malik/978-0-12-811140-6

[4] B. Razavi, V. R. Rao, C. Lin, K. A. Bujarski, S. E. Patra, D. E. Burdette, E. B. Geller, M. M. Brown, E. A. Johnson, C. Drees, E. F. Chang, J. E. Greenwood, C. N. Heck, B. C. Jobst, R. P. Gwinn, N. M. Warner, and C. H. Halpern, "Real-world experience with direct brain-responsive neurostimulation for focal onset seizures," *Epilepsia*, vol. 61, no. 8, pp. 1749–1757, aug 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/epi.16593

[5] B. Fisher, J. A. DesMarteau, E. H. Koontz, S. J. Wilks, and S. E. Melamed, "Responsive Vagus Nerve Stimulation for Drug Resistant Epilepsy: A Review of New Features and Practical Guidance for Advanced Practice Providers," *Frontiers in Neurology*, vol. 11, p. 1863, jan 2021.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv e-prints*, pp. 1–12, 2015. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[7] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, jul 2019. [Online]. Available: https://doi.org/10.1007/s10618-019-00619-1

[8] Seldon, "Transfer Learning for Machine Learning," 2021. [Online]. Available: https://www.seldon.io/transfer-learning

[9] M. F. Pinto, A. Leal, F. Lopes, A. Dourado, P. Martins, and C. A. Teixeira, "A personalized and evolutionary algorithm for interpretable EEG epilepsy seizure prediction," *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–12, feb 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-82828-7

[10] S. Muhammad Usman, S. Khalid, and S. Bashir, "A deep learning based ensemble learning method for epileptic seizure prediction," *Computers in Biology and Medicine*, vol. 136, p. 104710, sep 2021.

[11] S. M. Usman, S. Khalid, and Z. Bashir, "Epileptic seizure prediction using scalp electroencephalogram signals," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 211–220, jan 2021.

[12] 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, oct 2019.

[13] R. Hussein, S. Lee, R. Ward, and M. J. McKeown, "Epileptic seizure prediction: A semi-dilated convolutional neural network architecture," in *Proceedings - International Conference on Pattern Recognition*. Institute of Electrical and Electronics Engineers Inc., 2020, pp. 5436–5443.

[14] ——, "Semi-dilated convolutional neural networks for epileptic seizure prediction," *Neural Networks*, vol. 139, pp. 212–222, jul 2021.

[15] S. Zhao, J. Yang, and M. Sawan, "Energy-Efficient Neural Network for Epileptic Seizure Prediction," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 401–411, jan 2022.

[16] Y. Xu, J. Yang, S. Zhao, H. Wu, and M. Sawan, "An End-to-End Deep Learning Approach for Epileptic Seizure Prediction," *Proceedings - 2020 IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2020*, pp. 266–270, aug 2020.

[17] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, pp. 104–111, sep 2018.

[18] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance1, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, jul 2018. [Online]. Available: https://iopscience.iop.org/article/10.1088/1741-2552/aace8c https://iopscience.iop.org/article/10.1088/1741-2552/aace8c/meta

[19] Z. Wang, J. Yang, H. Wu, J. Zhu, and M. Sawan, "Power efficient refined seizure prediction algorithm based on an enhanced benchmarking," *Scientific Reports*, vol. 11, no. 1, p. 23498, dec 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-02798-8

[20] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, dec 2016. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6

[21] A. H. Ribeiro, M. H. Ribeiro, G. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, M. Wagner, T. B. Schön, and A. L. P. Ribeiro, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Communications*, vol. 11, no. 1, pp. 1–9, dec 2020. [Online]. Available: https://doi.org/10.1038/s41467-020-15432-4

[22] A. H. Shoeb and J. Guttag, "Application of Machine Learning To Epileptic Seizure Detection," in *ICML*, jan 2010, pp. 975–98. [Online]. Available: https://icml.cc/Conferences/2010/papers/493.pdf