Comparison of Handcrafted Features and Autoencoders for Epileptic Seizure Detection

Fatemeh Akrami, Fynn Aurand, Dominik T. Brockmann, Lena Faske, Laura Tiemann
University of Osnabrueck, Neurodynamics SoSe 2022
Osnabrueck, Germany
{fakrami, faurand, dobrockmann, lfaske, lautiemann}@uni-osnabrueck.de

Abstract— **More than 50 million people worldwide suffer from epilepsy and thus from recurrent and uncontrollable seizures. Here, Machine Learning (ML) approaches may help to detect seizures based on EEG recordings. In recent years, the interest in automated classification grew bigger and consequently, the methods used to detect seizures are numerous. There exist many ML techniques, for example end-to-end approaches that are built upon deep neuronal networks (DNNs) and feature-based approaches that rely on expert knowledge and handcrafted feature extraction. In this study, we compared which of these two approaches performs best with regards to accuracy, needed time and effort. Therefore, we developed an autoencoder and a handcrafted feature pipeline and compared them given the same dataset (UBonn) and classifier. Our results indicate that the handcrafted feature approach performs better with regards to the accuracy as well as to the needed time and effort.**

Index Terms—**epilepsy, seizure detection, EEG, autoencoder, deep neural networks, feature extraction, handcrafted features, principal component analysis, discrete wavelet transform**

I. Introduction

Epilepsy is one of the most common neurological disorders and seriously affects the quality of life of epilepsy patients. Therefore, finding accurate diagnostic methods is a crucial issue [15]. ML methods have the potential to automate clinical diagnostics based on EEG data. An automatic seizure detection system consists of many steps: First of all, EEG data needs to be acquired and preprocessed. However, the two main stages of such a system are: feature extraction and classification [20]. During the feature extraction step, the most representative or characterizing features of the EEG data need to be extracted and selected [2]. This step directly affects the precision and sophistication of the classifier, as the classification system is built and trained upon these extracted features for the discrimination of different EEG signal types [1], [2], [20]. Strategies for extracting features can be categorized into (1) feature-based (handcrafted features), and (2) end-to-end approaches (automatic feature learning from signals) [5], [14].

In this study, we want to compare these two approaches. Therefore, we have chosen to implement two simple pipelines that are representative of one of the approaches. As a feature-based approach, we are extracting statistical features on wavelet sub-bands and further reduce the dimensionality of the feature subspace using Principal Component Analysis (PCA), a statistical procedure to explore distinctive features using orthogonal transformation. As an end-to-end approach, we are applying a stacked sparse autoencoder that automatically learns the important features of the EEG pattern. To compare the two approaches, both systems were trained on the same dataset and their balanced accuracy was tested for 2, 4, 8, 16 and 32 features.

Hypothesis: We expect that the end-to-end approach will have a higher accuracy detecting epileptic seizures based on EEG signals than the feature-based approach. Further, we expect that the needed time and effort will be less for the end-to-end approach as well.

II. Related Work

The design of handcrafted features often involves a significant amount of expert knowledge to derive the characteristic epileptic patterns in the EEG signals. Additionally, it isn't generally ensured that these manually extracted features are optimal. Therefore, extremely good efforts have been exerted to enhance automatic seizures detection systems [2]. It has been proven that automatically learned features have been more robust than handcrafted features and that they achieve better detection performance [20]. However, while there are many papers that evaluated the performance of either handcrafted features or deep neural networks (DNNs), there are only a few papers that directly compare those approaches.

In one of those by Shoeibi et al. [15], they compared the performance of handcrafted features and DNNs in detecting epileptic seizures on the UBonn dataset. They extracted three different types of handcrafted features: time domain (e.g. mean, variance, skewness and mode), frequency domain (e.g. power spectrum of frequency sub-bands) and non-linear features (Lyapunov exponents and entropy-based ones). All in all, 50 features were extracted and evaluated by using the Fisher feature scoring algorithm. For the classification based on handcrafted features, they used a k-nearest-neighbor (KNN) algorithm and a support vector machine (SVM). For the convolutional autoencoder (CNN-AE), a softmax classifier was applied. Additionally, the authors introduced a hybrid design that was based on AE-features and handcrafted features. The results show that handcrafted features as well as autoencoders have a good performance and are well suited to detect seizures. However, their hybrid design performed best. Another paper by Gemein et al. [5] showed, like Shoeibi et al., that feature based approaches can achieve accuracies on the same level as state-of-the-art DNNs.

In a paper of Lin et al. [8], a stacked sparse autoencoder (SSAE) with softmax classifier was proposed for the automatic detection of epileptic EEG. They first preprocessed the data from the UBonn dataset by dividing every raw EEG signal into 4 segments before normalizing the data. After that, they used 3-hidden-layer SSAE networks for feature extraction with a Sigmoid activation. The hidden layers had 300, 250 and 150 units. For the classification, a softmax classifier was used and trained for 150 epochs. Their proposed framework achieved an average classification accuracy of 96%.

Wang et al. [18] proposed an automatic epilepsy diagnosis framework based on the combination of multi-domain feature extraction and nonlinear analysis of EEG signals. They first preprocessed the data from the UBonn dataset using the fourth-order Daubechies (Db4) wavelet threshold denoising. After that, they extracted multiple features in the time, frequency and time-frequency domains and nonlinear features on five clinically relevant frequency sub-bands. To eliminate irrelevant or redundant features, they used PCA for dimensionality reduction and then ranked the features using ANOVA. Using different classifiers, they were able to identify the epileptic seizures from EEG signals with an average accuracy of 99.25%.

III. Methods

Both pipelines, as seen in Figure 2, were implemented in Python relying heavily on the following modules: `pywt`, `scipy.signal`, `sklearn.preprocessing` and `tensorflow`. The exact implementation can be found in the respective GitHub repository.[1]

A. Dataset

In this study, we used the UBonn dataset from the University of Bonn [3]. It consists of five sets denoted as Z,N,O,F and S, each containing 100 single-channel EEG signals of 23.6 s duration with a sampling rate of 173.61 Hz. The data in set Z and O are EEG signals from 5 healthy volunteers. Set N and F are the EEG signals from 5 epileptic patients during seizure-free intervals. Only set S contains signals during seizure activity. For our classification task we created a non-seizure group by joining the sets Z,N,O and F, while set S makes up the seizure group.The recordings of each group were then randomly assigned to the training (70%), testing (15%) and validation set (15%).

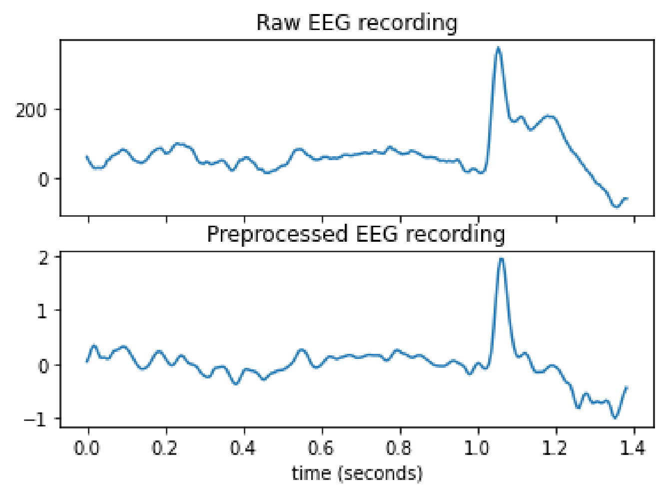

Fig. 1. Example of a raw vs. preprocessed EEG recording (in the autoencoder pipeline).

[1]https://github.com/dbrockmann/eeg-classification

B. Preprocessing

1) Autoencoder Pipeline: At first, we applied a fifth-order Butterworth bandpass filter [13], [16] with a 0.5 Hz and a 40 Hz cutoff frequency to eliminate the extra signal noise and unwanted artifacts. In the next preprocessing step, we normalized the data [15] to a range between 0 and 1 using min-max scaling because we use a sigmoid activation function later on, which is sensitive for this range [12]. In order to provide enough training examples for the autoencoder and the classifier, we separated each EEG signal into 16 epochs of 256 samples by utilizing the non-overlapping sliding window technique [12].

2) Handcrafted Feature Pipeline: As a first step in the handcrafted feature pipeline, we standardized the data by removing the mean and scaling to unit variance. After that, we windowed the data like described in III-B1. Then we applied Daubechies (Db4) wavelet threshold denoising for noise reduction as it has showcased the best performance compared to other methods [9]. The key idea of wavelet-based denoising is to apply discrete wavelet transform (DWT) on a single-channel EEG signal, remove the wavelet coefficients that fall below some statistical noise threshold and reconstruct the signal back using the inverse DWT. In our implementation we used a decomposition level of 5 and the so-called universal threshold $\lambda = \sigma\sqrt{2 \log N}$, where λ is the wavelet threshold, σ is the standard deviation of the noise and N is the length of the sample signals. [18]

C. Feature Extraction

1) Autoencoder Pipeline: Autoencoders can extract features by reducing the dimensionality of the data. An autoencoder consists of an encoder that reduces the input to a latent representation and a decoder that restores the original input from this latent representation. By using the input as the target output, autoencoders are an unsupervised learning algorithm [10].

We used a stacked sparse autoencoder (SSAE) to learn the high level representations from the preprocessed data. Stacked autoencoders can learn low-level characteristics from more abstract data by using deep networks with multiple hidden layers. Additionally, an activation penalty is added to obtain sparse representations [8].

The autoencoder that we used consists of 6 fully-connected layers with 128, 64, 32, 64, 128 and 256 nodes. The first 3 layers make up the encoder, with the third layer being the bottleneck layer representing the extracted features. Apart from the last layer, we used the exponential linear unit (ELU) as the activation function. ELUs can alleviate the vanishing gradient problem while increasing learning speed by pushing mean unit activations closer to zero [4]. The use of ELU seems to achieve a lower reconstruction error and is shown to perform better than alternatives in tasks of dimension reduction using stacked autoencoders [11]. We used Sigmoid as the activation for the last layer to reconstruct the input in the range between 0 and 1. The weights are initialized according to [6]. This initialization is optimized for networks using the rectified linear unit (ReLU) and is used in the paper proposing ELU [4].

The model is trained using the Adam optimizer [7] with a learning rate of 0.001, which seems to achieve the best results.

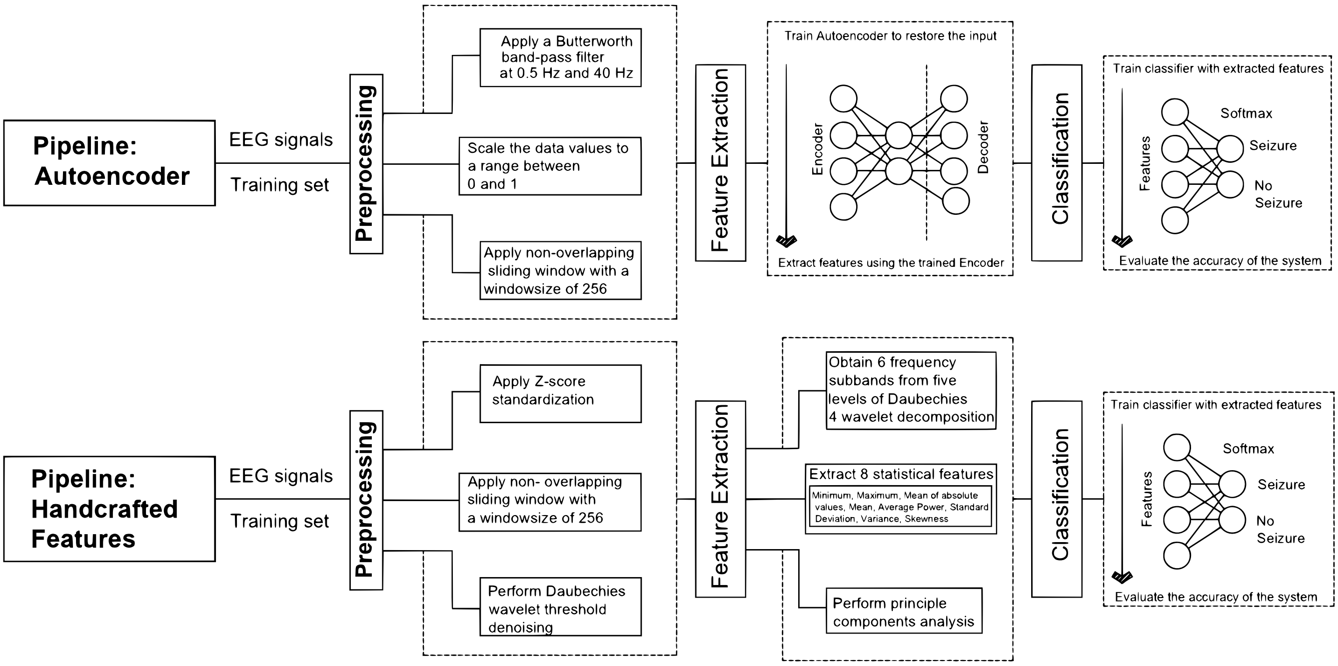

Fig. 2. Autoencoder Pipeline vs Handcrafted Feature Pipeline.

We used the mean squared error as the loss and evaluation metric. To achieve sparse representations, an L1 penalty is added to the cost function for the activation of the bottleneck layer with a coefficient of 0.001.

2) Handcrafted Feature Pipeline: Systems that use wavelet domain feature extraction methods have been shown to produce superior signal processing results for EEG seizure detection, as they preserve time and frequency components in the signals [18]. Specifically, DWT has been widely used.

In practice, the DWT is always implemented as a filter-bank. It can be seen as a cascade of high-pass and low-pass filters. Using the group of filters, the signal can be decomposed into several frequency sub-bands. At each decomposition stage a high and low-pass filter is applied that provide the detail and approximation coefficients. Thus, it is important to choose the right number of decomposition levels and a proper wavelet function [17]. In this study, we applied a fifth-level DWT to break the signal into 6 clinically relevant sub-bands. The signal sub-bands and with their related frequencies are tabulated in Table I. As the mother wavelet we chose to use fourth-order Daubechies (Db4) due to its orthogonal properties [18].

After decomposing the signal, we calculated 8 statistical features from the wavelet coefficients of each sub-band to further decrease the dimensionality of the extracted feature vectors. The statistical features included are the *minimum, maximum, mean, mean of absolute values, average power, standard deviation, variance and skewness* of the coefficients. They are also commonly extracted in other similar systems. In total, 48 features were extracted. However, these features may still contain redundant or irrelevant features, that could negatively affect the accuracy of the classification process. Thus, we additionally applied PCA, a widely used algorithm for feature dimension reduction in EEG classification problems.

TABLE I
SIGNAL SUB-BANDS AND THEIR RESPECTIVE FREQUENCIES

Decomposed Segment	Frequency
D1	43.4 - 86.8 Hz
D2	21.7 - 43.4 Hz
D3	10.8 - 21.7 Hz
D4	5.4 - 10.8 Hz
D5	2.7 - 5.4 Hz
A5	0 - 2.7 Hz

D. Classification

The classifier model used for both pipelines consists of 3 fully connected layers with 32, 16 and 2 nodes. We used the rectified linear unit (ReLU) as the activation function for the first two layers. A probability vector for the two classes is obtained by a softmax activation in the last layer.

We used the Adam optimizer to train the model with a learning rate of 0.0001 and the sparse categorical crossentropy as loss. Accuracy of the model is computed by selecting the class with an output value greater than 0.5 and taking the percentage of correct predictions. Additionally, the balanced accuracy is calculated by weighting using the ratio of class labels in the dataset. The dataset we used contains seizure and non-seizure samples at a ratio of 1:4, which results in weighting non-seizure samples with a higher coefficient when calculating the balanced accuracy.

IV. RESULTS

We trained the autoencoder using different numbers of features for 40 epochs in order to avoid overfitting. The classifier was trained for 400 epochs in the autoencoder pipeline and for 150 epochs in the handcrafted pipeline. We used a validation set for fine-tuning the hyperparameters of the models that is not used for training. The results are obtained by evaluating the performance on the completely unseen test set. It shows that the classifier trained with handcrafted

TABLE II
RESULTS

	autoencoder			*handcrafted*	
features	**reconstruction loss**	**accuracy**	**balanced accuracy**	**accuracy**	**balanced accuracy**
2	0.008553295	0.825	0.56718767	0.9575	0.9265624
4	0.008035725	0.90833336	0.77552086	0.96	0.9359375
8	0.0070872433	0.9433333	0.8723958	0.9633333	0.9364583
16	0.005512166	0.95916665	0.9135417	0.9658333	0.9380208
32	0.0034239881	0.96	0.9171875	0.975	0.953125

features achieves over 95% accuracy with only 2 features while the autoencoder pipeline only reaches above 95% with 16 or more features. The best accuracy for both pipelines was achieved using 32 features with an accuracy of 96% and 97.5% (91.72% and 95.31% balanced accuracy) for the autoencoder and handcrafted pipeline, respectively.

In Figure 3, one can see the balanced accuracy of the models on the y-axis depending on the number of features on the x-axis. While the autoencoder performs poorly with less than 8 features, the handcrafted approach already starts with a balanced accuracy of 92.66% and overall performs better than the autoencoder.
The results can be seen in more detail in Table II. Here, reconstruction loss, accuracy and balanced accuracy are displayed for the specific numbers of features.

V. Discussion

The obtained results do not match our hypothesis: Using features extracted by an autoencoder to train the classifier does not result in an overall higher accuracy than using handcrafted features with PCA for the feature extraction. As other studies show, autoencoder based feature extraction for EEG data can perform worse than statistical methods, especially for a low number of extracted features [19]. We see multiple possible improvements to our methods that are mostly due to time and computational power constraints.

The dataset that we used contains more non-seizure samples, which results in an unbalanced number of samples in each class. Balancing the dataset for the training process and not only for evaluation may increase overall accuracy.

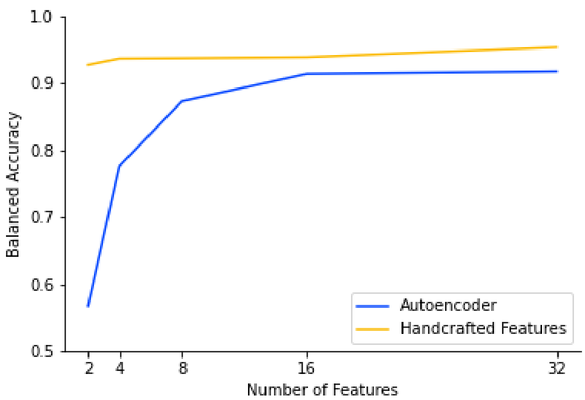

Fig. 3. Balanced Accuracy of Autoencoder and Handcrafted Features given a specific number of features.

Training on unbalanced datasets can lead to a bias towards the class that occurs more often. By using balanced accuracy, we are able to show the performance of the system regardless of the unbalanced data. In addition to that, calculating precision, recall and the F1-score could improve evaluation. Furthermore, the achieved accuracy is obtained using the same number of epochs for each different number of features. This number is optimized for 16 features to avoid overfitting. Instead of manually setting the number of epochs, early stopping could improve training for different numbers of features. We restrained ourselves by using rather simple models due to time and computational power constraints. As presented in section II, other papers use more complex models that in their design may be better suited for this specific task. Evaluation could be improved by using cross fold validation, which is especially beneficial if only limited data is available, but it requires more time.

In addition to comparing the resulting accuracy of the two pipelines, the time and effort has to be considered. Using handcrafted features requires extensive expert knowledge on the research topic and knowledge about the specific dataset that is used. Autoencoders on the other hand, can extract features without knowledge from the specific research area, are more dataset independent and therefore allow a more general usage. But optimizing the hyperparameters of such a system can be challenging and requires additional data.

VI. Conclusion

Our obtained results show that handcrafted feature extraction with PCA performs well at classifying epileptic seizures using EEG data with an accuracy of greater than 97% and a balanced accuracy of above 95%. The autoencoder model we used is able to achieve 96% accuracy but only above 91% balanced accuracy. Autoencoders can extract features without requiring expert knowledge. By taking the required time for hyperparameter optimization into account, our results indicate that handcrafted features perform better even when considering the needed time and effort for this specific task. Autoencoders can be a great alternative, if expert knowledge is missing or a more general approach is needed. Further research could look more intensively at the difference in time and computational effort between the two approaches. This could include investigating whether the complexity of feature engineering in the feature-based approach is simply shifted to network engineering in the end-to-end approach.

References

[1] Ahmed Abdelhameed and Magdy Bayoumi. A deep learning approach for automatic seizure detection in children with epilepsy. *Frontiers in Computational Neuroscience*, 15:29, 2021.

[2] Ahmed M Abdelhameed, Hisham G Daoud, and Magdy Bayoumi. Epileptic seizure detection using deep convolutional autoencoder. In *2018 IEEE International Workshop on Signal Processing Systems (SiPS)*, pages 223–228. IEEE, 2018.

[3] Ralph G. Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E*, 64:061907, Nov 2001.

[4] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus), 2015.

[5] Lukas A.W. Gemein, Robin T. Schirrmeister, Patryk Chrabaszcz, Daniel Wilson, Joschka Boedecker, Andreas Schulze-Bonhage, Frank Hutter, and Tonio Ball. Machine-learning-based diagnostics of eeg pathology. *NeuroImage*, 220:117021, 2020.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015.

[7] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[8] Qin Lin, Shu-qun Ye, Xiu-mei Huang, Si-you Li, Mei-zhen Zhang, Yun Xue, and Wen-Sheng Chen. Classification of epileptic eeg signals with stacked sparse autoencoder based on deep learning. In De-Shuang Huang, Kyungsook Han, and Abir Hussain, editors, *Intelligent Computing Methodologies*, pages 802–810, Cham, 2016. Springer International Publishing.

[9] John R Martin and SL Swapna. A machine learning framework for epileptic seizure detection by analyzing eeg signals. *International Journal of Computing and Digital Systems*, pages 1383–1391, 2021.

[10] Qinxue Meng, Daniel Catchpoole, David Skillicom, and Paul J. Kennedy. Relational autoencoder for feature extraction. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 364–371, 2017.

[11] Nirmalajyothi Narisetty, Gangadhara Rao Kancherla, Basaveswararao Bobba, and K Swathi. Investigative study of the effect of various activation functions with stacked autoencoder for dimension reduction of nids using svm. *International Journal of Advanced Computer Science and Applications*, 12(5), 2021.

[12] Yang Qiu, Weidong Zhou, Nana Yu, and Peidong Du. Denoising sparse autoencoder-based ictal eeg classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(9):1717–1726, 2018.

[13] Marzieh Savadkoohi, Timothy Oladunni, and Lara A. Thompson. A machine learning approach to epileptic seizure prediction using electroencephalogram (eeg) signal. *Biocybernetics and Biomedical Engineering*, 2020.

[14] Afshin Shoeibi, Navid Ghassemi, Roohallah Alizadehsani, Modjtaba Rouhani, Hossein Hosseini-Nejad, Abbas Khosravi, Maryam Panahiazar, and Saeid Nahavandi. A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in eeg signals. *Expert Systems with Applications*, 163:113788, 2021.

[15] Afshin Shoeibi, Navid Ghassemi, Marjane Khodatars, Mahboobeh Jafari, Parisa Moridian, Roohallah Alizadehsani, Ali Khadem, Yinan Kong, Assef Zare, Juan Manuel Górriz, Javier Ramírez, Maryam Panahiazar, Abbas Khosravi, and Saeid Nahavandi. Applications of epileptic seizures detection in neuroimaging modalities using deep learning techniques: Methods, challenges, and future works. *arXiv: Learning*, 2021.

[16] Athar Shoka, Mohamed Dessouky, Ahmed el sherbeny, and Ayman El-Sayed. Literature review on eeg preprocessing, feature extraction, and classifications techniques. *Menoufia Journal of Electronic Engineering Research*, 28:292–299, 12 2019.

[17] Abdulhamit Subasi, Jasmin Kevric, and M Abdullah Canbaz. Epileptic seizure detection using hybrid machine learning methods. *Neural Computing and Applications*, 31(1):317–325, 2019.

[18] Lina Wang, Weining Xue, Yang Li, Meilin Luo, Jie Huang, Weigang Cui, and Chao Huang. Automatic epileptic seizure detection in eeg signals using multi-domain feature extraction and nonlinear analysis. *Entropy*, 19(6):222, 2017.

[19] Tingxi Wen and Zhongnan Zhang. Deep convolution neural network and autoencoders-based unsupervised feature learning of eeg signals. *IEEE Access*, 6:25399–25410, 2018.

[20] Ye Yuan, Guangxu Xun, Kebin Jia, and Aidong Zhang. A multi-view deep learning framework for eeg seizure detection. *IEEE journal of biomedical and health informatics*, 23(1):83–94, 2018.