Przemysław WISZNIEWSKI[1], Marcin KOŁODZIEJ[2], Andrzej MAJKOWSKI[3], Andrzej RYSZ[4]

Warsaw University of Technology (1,2,3), 1 Wojskowy Szpital Kliniczny z Polikliniką SPZOZ w Lublinie Filia w Ełku (4)
ORCID: 1.0000-0002-2336-841X; 2.0000-0003-2856-7298; 3.0000-0002-6557-836X; 4.0000-0002-8272-7275

doi:10.15199/48.2023.02.07

Detection of epileptic seizures with the use of convolutional neural networks

***Abstract**. The purpose of the article is to investigate whether the implementation of a CNN consisting of several layers will allow the effective detection of epileptic seizures. For the research, a publicly available database registered for 4 dogs and 8 people was used. The 1-second iEEG recordings were marked by a neurophysiologist as interictal, early seizure, and seizure. A CNN was trained for each patient individually. Coefficients such as precision, AUC, sensitivity, and specificity were calculated, and the results were compared with the best algorithms published in one of the contests on the Kaggle platform. The average accuracy for the recognition of seizures using CNN is 0.921, the sensitivity is 0.850, and the specificity is 0.927. For early seizures these values are 0.825,* 0.782, and 0.828, *respectively.*

***Streszczenie.** Celem artykułu było zbadanie czy zastosowanie sieci CNN, składającej się z kilku warstw umożliwi skuteczną detekcję napadów epileptycznych. Na użytek badań zastosowano ogólnodostępną bazę danych zarejestrowaną dla 4 psów oraz 8 ludzi. Jednosekundowe zapisy sygnału iEEG zostały oznaczone przez neurofizjologa jako: międzynapadowe, wczesnonapadowe oraz napadowe. Zaproponowano strukturę sieci CNN, a następnie wytrenowano ją dla każdego pacjenta indywidualnie. Zostały wyliczone współczynniki takie jak: trafność, AUC, czułość, specyficzność. Następnie wyniki zostały porównane do osiągniętych w najlepszych algorytmach opublikowanych w konkursie na platformie Kaggle. Średnia skuteczność rozpoznawania napadów z wykorzystaniem sieci CNN wynosi 0.921, czułość 0.850, a specyficzność 0.927. Dla okresów wczesnonapadowych wartości te wynoszą odpowiednio 0.825, 0.782 i 0.828. (**Wykrywanie napadów padaczkowych z wykorzystaniem konwolucyjnych sieci neuronowych**)*

Keywords: seizure detection, convolutional neural network, iEEG, classification, feature extraction.
Słowa kluczowe: detekcja napadów epileptycznych, konwolucyjne sieci neuronowe, iEEG, klasyfikacja, ekstrakcja cech.

Introduction

One of the significant problems of today's neurology is the diagnosis and treatment of epilepsy [1], [2]. It is estimated that 1-2% of people worldwide suffer from epilepsy [3]. In recent years, we can observe an increased interest in the use of EEG/iEEG signals to recognize and detect the onset of a seizure. Using automatic seizure detection methods would enable easier diagnosis and monitoring of a patient's condition [4]. We are currently observing the rapid development of devices for a non-invasive and invasive recording of brain activity [5]. Developing an effective algorithm for the detection and/or prediction of seizures is one of the important challenges [6].

Standard and dedicated EEG/iEEG signal analysis combined with machine learning are commonly used to recognize seizures [7], [8]. The most popular methods are correlation analysis, spectral analysis, and wavelet analysis [9], [10]. Techniques such as LDA, SVM, MLP, and newer ones such as LSTM networks, CNN, and autoencoders [11]–[13] are used for classification. To validate algorithms, unique databases were created dedicated to recognizing the onset of seizures [14]. The authors of many studies report the achievement of high accuracy and recognition rates and increased algorithms sensitivity.

The purpose of the article is to demonstrate the usefulness of CNN in detecting early seizure and seizure. The iEEG database registered for one of the contests on the Kaggle platform was used for network training and evaluation of the results [15]. The results were then discussed and compared with other detection algorithms. Also, the aspects of practical application of the proposed solution were considered.

Materials

For effective training of seizure detection algorithms, particularly with the implementation of deep networks, it is necessary to have representative data with sufficient examples. This study used the database made available to develop algorithms for detecting epileptic seizures based on electrocorticographic signal (iEEG) [16]. The authors of the database shared it in the hope that the algorithms developed will be implemented in devices that stimulate the brain in the area of epileptic seizures. The database contains the signal collected during the examination of people and dogs. The developed algorithms should give the lowest possible level of false alarms. The challenge was published on the platform kaggle.com [17]. Data should be classified into interictal, early seizure, and seizure.

Data were recorded for 4 dogs and 8 people suffering from epilepsy. iEEG signals were recorded with a sampling frequency of 500 Hz to 5 kHz. A low-pass and anti-aliasing filter was used during signal acquisition. Incorrect recordings were deleted. The data was described and categorized by two neurophysiologists. The marked segments covered the entire seizure from the earliest stage until the end of the seizure. The data was divided into a training set and a test set. Only epileptic seizures preceded by a seizure-free period of at least 4 hours were considered. The interictal periods were chosen to occur at least one hour before or after the seizure. When dividing the data, the inequality of the occurrence of epileptic seizures was preserved in relation to the periods between seizures. The training data was sorted chronologically, while the test data was devoid of chronology. In each segment, the probability of a seizure or early seizure (the first 15 seconds of the seizure) was determined. During the training and testing of the network, the original data division was maintained as in the "UPenn and Mayo Clinic's Seizure Detection Challenge". Detailed information on the database can be found in the article [18].

CNN for seizure detection

In the classical convolutional network, the first layers act as feature extractors and consist of filters used repeatedly during a single epoch. Successive convolutional layers generate more complex features that cover larger areas of the processed data [19]. A single convolution layer often consists of convolution, activation (for example ReLU), and pooling operations. The pooling operation calculates a local value group's maximum or mean. Behind the last convolutional layer, a network is used, which acts as a classifier. A fully connected (FC) network, such as a multilayer perceptron (MLP), can be used for classification. In the case of multidimensional and multichannel data, the convolution can be performed along any dimension or several dimensions, for each channel separately or for all channels simultaneously. The convolution result can also be a multichannel, single channel, or a combination of convolution for selected channels to give a multichannel result.

In our experiment, a 2D convolutional network was used. The network consisted of three convolutional layers with filters of the same sizes 16x16. The number of filters for subsequent layers of the network was 8, 16, and 32. A fully connected layer with 256 neurons was used at the end of the network. Figure 1 shows a block diagram of the applied network. During network training, the hyperparameters listed in Table 1 were tested. Hyperparameters such as the number of epochs, the learning coefficient, the optimizer, and the scheduler were selected. The influence of the hyperparameters on the convolutional and fully connected layers was tested. All hyperparameters were chosen manually. The hyperparameters for which the best results were obtained were retained. Default methods of initializing the weights of individual layers were used. Due to the different number of signal channels recorded for each patient and the different locations of epilepsy foci, the networks were individually trained for each patient.

Table 1. Range of tested hyperparameters

Parameter name	Range / possible values
Number of epochs	10 - 1000
Learning rate	1e-7 – 1e-2
Optimizer	Adam, SGD
Scheduler	None, ReduceLROnPlateau, CyclicLR
Number of convolutional layers	1 – 4
Number of filters in a layer	4 – 2048
Filter size	3x3 – 32x32
Filter step	(1,1), (2,2)
Double layer filtration	yes, no
Normalization within the training set	yes, no
Dimension reduction	yes, no
Size of reduction kernel (MaxPooling)	2x2
Reduction step (MaxPooling)	2x2
Number of fully connected network layers	1, 2
Number of neurons in a fully connected network	32 – 2048
Dropout in various places on the network	0 – 0,8

Python programming language was used to implement the solution. Python is characterized by dynamic typing and automatic memory management. A deep network implementation library, PyTorch, was used to test various network architectures. The Pytorch library enables the use of GPUs with CUDA technology to accelerate calculations on tensors. The Pytorch Lightning library was used to organize the code. The Tensorboard tool was implemented to diagnose the network, allowing real-time visualization of the training process. The weights of the trained models are saved for each patient, allowing the models to be used later.

Results and discussion

The Area Under the Curve (AUC) measures how well a model can distinguish between positive and negative classes. The AUC is calculated as the area under the ROC curve. ROC curves have the desired property in the case of unbalanced data sets - they are insensitive to changes in the distribution of classes [20].

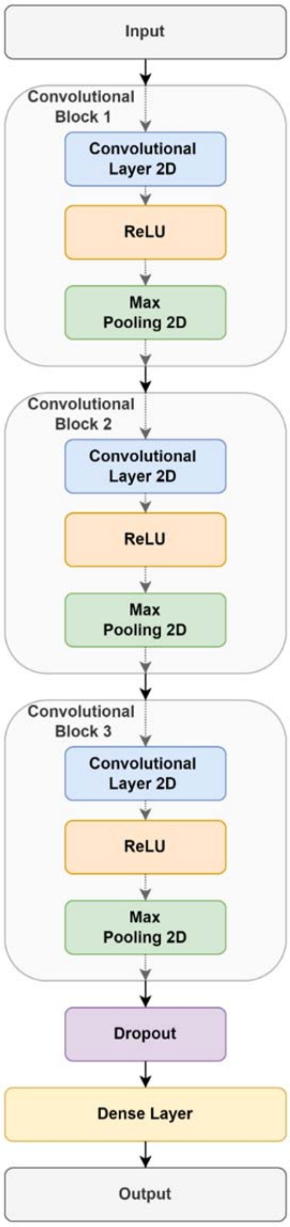

Figure 1. Block diagram of the applied CNN

Table 2. Applied CNN hyperparameters

Parameter name	Value
Number of epochs	100
Learning rate	1e-4
Optimizer	Adam
Scheduler	None
Number of convolutional layers	3
Number of filters in a layer	8, 16, 32
Filter size	16x16, 16x16, 16x16
Filter step	(1,1), (1,1), (1,1)
Double layer filtration	no, no, no
Normalization within the training set	no, no, no
Padding with zeros	(0,0), (0,0), (0,0)
Dropout after convolution	0, 0, 0, 0.1
Reduction	yes, yes, yes
Reduction size	(2,2), (2,2), (2,2)
Filter step	(2,2), (2,2), (2,2)
Number of fully connected network layers	1
Number of neurons in a fully connected network	256
Dropout after fully connected layers	0, 0

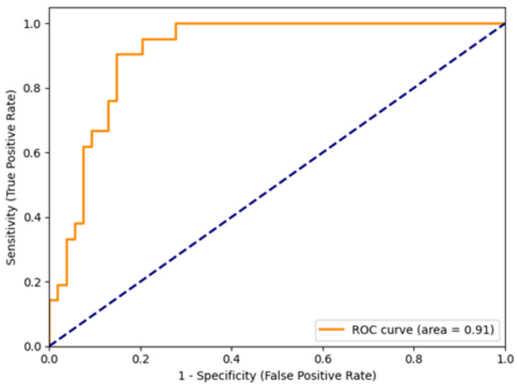

Figure 2. ROC curve (the ROC curve for the random classifier is marked with a dashed line)

If the ratio of positive to negative cases in the test set changes, the ROC curve remains the same. However, the AUC is only one measure used to evaluate seizure detection systems and does not give a complete picture of the system's performance. Figure 2 shows an example of the ROC curve calculated for patient 1 for the task of distinguishing between seizure and non-seizure signals. The area under the curve (AUC) for seizure detection is 0.918, which is a relatively good result. Accuracy in the context of binary classification is the proportion of correct predictions (both true positive and true negative) to the total number of cases examined [21]. Accuracy can be misleading when used with unbalanced datasets in which the size of individual classes differs [22]. Sensitivity, on the other hand, measures how well the model can predict positive classes [23]. Sensitivity expresses the percentage of positive cases classified as positive by the model. The higher the sensitivity, the better the model at predicting positive classes. Sensitivity itself does not form the basis for an informed judgment of the results, as the results may contain many false positives that are not considered [24]. Specificity is a measure of how well the model can predict negative classes. Specificity can be interpreted as the probability of identifying negative cases overall negative cases. The higher the specificity, the better the model in predicting negative classes. As with the sensitivity, it should be noted that the specificity calculation does not consider false negative cases. The F1 score is the harmonic mean of the classifier's precision and sensitivity. The F1 works well for unbalanced data sets [25]. At the same time, note that this measure is sensitive to a change in the set's class size ratio. Therefore, the F1 scores for sets with different ratios of the number of positive and negative classes should not be compared directly [21], [26].

We have performed an in-depth analysis to verify our CNN solution and its applicability in clinical practice. For this purpose, measures were calculated, such as AUC, recognition accuracy, F1 score, sensitivity, and specificity. A summary of the results for the CNN solution was presented in Table 3. The *seizure* suffix means that the results were calculated for no-seizure as the negative class and early seizure or seizure as the positive class. The *early* suffix means that results were calculated for seizure or no seizure as the negative class and early seizure as the positive class. *AUC total* is calculated as mean of *AUC seizure* and *AUC early*. The results presented in Table 3 show that the measures obtained are satisfactory. However, for Patient 3 and Patient 4 the results are worse. Therefore, it seems that the developed CNN algorithm is not suitable for every user, or the recorded iEEG signals for Patient 3 and Patient 4 do not give similar iEEG signal patterns for seizure, early seizure, and no seizure periods. It is also worth noting that the mean accuracy values obtained for dogs are better than for humans. An important indicator that describes the solution is sensitivity and precision. In Table 4, we can observe satisfactory sensitivity and specificity values, but worse results were also obtained for Patient 3 and Patient 4.

Table 3. AUC, accuracy and F1-score obtained for the test set

	AUC seizure	**AUC early**	**AUC total**	**ACC seizure**	**ACC early**	**F1 seizure**	**F1 early**
Dog_1	0.987	0.966	0.976	0.987	0.949	0.879	0.424
Dog_2	0.947	0.752	0.849	0.935	0.598	0.563	0.040
Dog_3	0.983	0.920	0.951	0.950	0.888	0.772	0.362
Dog_4	0.993	0.974	0.983	0.976	0.965	0.770	0.537
Patient_1	0.967	0.898	0.932	0.956	0.865	0.758	0.320
Patient_2	0.993	0.965	0.979	0.984	0.943	0.874	0.356
Patient_3	0.755	0.576	0.666	0.746	0.678	0.340	0.072
Patient_4	0.559	0.559	0.559	0.702	0.702	0.221	0.221
Patient_5	0.923	0.855	0.889	0.924	0.918	0.548	0.260
Patient_6	0.973	0.912	0.942	0.932	0.886	0.666	0.241
Patient_7	1.000	0.916	0.958	0.999	0.909	0.994	0.219
Patient_8	0.934	0.624	0.779	0.963	0.600	0.802	0.049
Mean	0.918	0.826	0.872	0.921	0.825	0.682	0.258

Table 4. Sensitivity and specificity obtained for the test set

	Sensitivity seizure	**Sensitivity early**	**Specificity seizure**	**Specificity early**
Dog_1	0.956	0.938	0.988	0.949
Dog_2	0.829	0.781	0.940	0.596
Dog_3	0.940	0.888	0.951	0.888
Dog_4	0.976	0.953	0.976	0.965
Patient_1	0.865	0.813	0.964	0.868
Patient_2	0.969	0.953	0.984	0.943
Patient_3	0.656	0.500	0.755	0.683
Patient_4	0.460	0.460	0.726	0.726
Patient_5	0.821	0.672	0.930	0.923
Patient_6	0.918	0.844	0.933	0.887
Patient_7	0.992	0.958	1.000	0.908
Patient_8	0.811	0.625	0.978	0.600
Mean	0.850	0.782	0.927	0.828

From Tables 4 and 5 it can be concluded that in each case the results for early seizures are clearly worse than for seizures in general. Worse results for early seizures indicate difficulties in classifying these types of cases. For Patient 4 only, there is no difference in the results. This is due to the fact that there were only early seizures in both the training set and the test set for that patient.

A slightly more detailed look at the results can be obtained by analyzing the confusion matrices for the test set. Table 5 presents the confusion matrix for Patient 1 and Table 6 for Patient 3.

Table 5. Confusion matrix for the test set for Patient1

		Predicted	
		interictal	seizure
True	interictal	1819	68
	seizure	22	141

Table 6. Confusion matrix for the test set for Patient3

		Predicted	
		interictal	seizure
True	interictal	871	282
	seizure	44	84

The best confusion matrix is the one for which the diagonal values have the highest values. In the case of Patient 1, we can observe 22 cases classified as interictal and evaluated by experts as a seizure. Additionally, 68 cases were assessed as seizures, but experts evaluated them as interictal. For Patient 3, we can observe 44 cases classified as interictal but assessed by experts as seizure. Moreover, as many as 282 cases were classified by the system as seizures and were considered by experts as interictal.

We compared our results with the best algorithms published in one of the contests on the Kaggle platform concerning the detection of epileptic seizures. Many researches participated in this competition. The methods used by the winners are described below. The first algorithm is based on classification using three sets of features. The first set of features were FFT magnitudes in the low-frequency range of 1 to 47 Hz. The spectrum was normalized in each frequency range. The second set of features were correlation coefficients (between EEG channels) and their eigenvalues in the frequency domain. The third set of features were correlation coefficients (between EEG channels) and their eigenvalues in the frequency domain. The classifier, a random forest consisting of 3,000 trees, was trained on such a set of features. In the second algorithm, the data was preprocessed in an automated filter selection step. The filter combinations were selected from a bank of 10 partially overlapping bandpass filters covering the entire frequency range from 5 Hz to 200 Hz. Three combinations were chosen that performed the best in the cross-validation test. After filtering, the covariance matrices of the EEG data were calculated and normalized for each 1-second fragment of the signal. The classification was carried out using a set of 100 multilayer perceptrons, each consisting of two hidden layers with 200 and 100 neurons in each layer. In the third algorithm, the signal was initially re-sampled to 100 Hz. The algorithm used features calculated for each channel separately and global features. Channel-specific features included maximum amplitude, mean amplitude, absolute deviation, and variance, as well as features derived from a fast Fourier transform of the signal, such as maximum power, mean power, variance, and frequencies at which maximum power occurred. Global features included time domain features such as maximum amplitude, mean amplitude, maximum absolute deviation, maximum value, mean value, and covariance between channel signals. Global features were also obtained using a fast Fourier transform of the signal - maximum power, average power, maximum variance between channels, and maximum, average, and variance of frequency at which the maximum power occurred. The first and second derivatives for the features mentioned above were also calculated. The classification was made by averaging the results of 1000 decision trees using the Extremely Randomized Trees method implemented in Python with the ExtraTreesClassifier from the scikit-learn library.

In the contests on the Kaggle platform, only AUC (Area Under the Curve) was used as an algorithm quality assessment method. The comparison of the three winning algorithms and our CNN algorithm is presented in Table 7. The *Final score* is the arithmetic mean of the *AUC total* for all dogs and patients. It should be noted that the application of expert knowledge on signal processing and analysis resulted in better AUC results than the use of CNN. The size of filters in individual layers of the CNN does not allow for the identification of interdependencies between all registered channels.

Table 7. Comparison of the three winning algorithms and the proposed CNN solution

	Final score
Algorithm 1	0.975
Algorithm 2	0.968
Algorithm 3	0.962
CNN	0.872

When analyzing the obtained results, we must not forget that the classification took place after only one second of the recorded iEEG signal. One second may not be sufficient in some applications to effectively classify and recognize the phase of a seizure. Including longer fragments of signals in the target algorithm in the authors' assessment would allow for much better results.

The solutions developed require training the classifier for a specific patient. In practice, this means recording many hours of the iEEG signal and manually marking selected fragments as interictal, early seizure, and seizure by a neurophysiologist. Obstacles to creating a fully automatic and universal detection system for each patient are different epilepsy focus for each patient and different morphologies of the iEEG signal.

Conclusion

The use of CNN enables an effective detection of epileptic seizures. The results obtained indicate that CNN independently searches for iEEG signal features that are useful for classification. The average seizure recognition accuracy using CNN is 0.921, the sensitivity is 0.850, and the specificity is 0.927. The network must be trained individually for each patient. It is related to the acquisition of iEEG signals, the location of the seizure foci, and the morphology of seizures. It should also be noted that the results are not satisfactory for all patients. In order to reliably assess the usefulness of the proposed solution, it is necessary to conduct research for a larger number of people in the future.

***Authors**: mgr inż. Przemysław Wiszniewski, dr hab. inż. Marcin Kołodziej, dr hab. inż. Andrzej Majkowski, Warsaw University of Technology, pl. Politechniki 1, 00-661 Warszawa, marcin.kolodziej@.pw.edu.pl, andrzej.majkowski@pw.edu.pl. przemyslaw.wiszniewski.dokt@pw.edu.pl, dr n. med Andrzej Rysz, 1 Wojskowy Szpital Kliniczny z Polikliniką SPZOZ w Lublinie Filia w Ełku, Oddział Neurochirurgii, ul. Kościuszki 30,19-300 Ełk.*

REFERENCES

[1] H. Oguni, "Diagnosis and treatment of epilepsy," *Epilepsia*, vol. 45 Suppl 8, pp. 13–16, 2004, doi: 10.1111/j.0013-9580.2004.458003.x.

[2] E. Magiorkinis, A. Diamantis, K. Sidiropoulou, and C. Panteliadis, "Highights in the History of Epilepsy: The Last 200 Years," *Epilepsy Res. Treat.*, vol. 2014, pp. 1–13, Aug. 2014, doi: 10.1155/2014/582039.

[3] L. J. S. Knutsen and M. Williams, "6.11 - Epilepsy," in *Comprehensive Medicinal Chemistry II*, J. B. Taylor and D. J. Triggle, Eds. Oxford: Elsevier, 2007, pp. 279–296. doi: 10.1016/B0-08-045044-X/00171-1.

[4] A. Ulate-Campos, F. Coughlin, M. Gaínza-Lein, I. S. Fernández, P. L. Pearl, and T. Loddenkemper, "Automated seizure detection systems and their effectiveness for each type of seizure," *Seizure*, vol. 40, pp. 88–101, Aug. 2016, doi: 10.1016/j.seizure.2016.06.008.

[5] B. Byrom, M. McCarthy, P. Schueler, and W. Muehlhausen, "Brain Monitoring Devices in Neuroscience Clinical Research: The Potential of Remote Monitoring Using Sensors, Wearables, and Mobile Devices," *Clin. Pharmacol. Ther.*, vol. 104, no. 1, pp. 59–71, Jul. 2018, doi: 10.1002/cpt.1077.

[6] W. J. Bosl, A. Leviton, and T. Loddenkemper, "Prediction of Seizure Recurrence. A Note of Caution," *Front. Neurol.*, vol. 12, p. 675728, 2021, doi: 10.3389/fneur.2021.675728.

[7] M. Natu, M. Bachute, S. Gite, K. Kotecha, and A. Vidyarthi, "Review on Epileptic Seizure Prediction: Machine Learning and Deep Learning Approaches," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–17, Jan. 2022, doi: 10.1155/2022/7751263.

[8] M. Kołodziej, A. Majkowski, and R. Rak, "A new method of feature extraction from EEG signal for brain-computer interface design," *Przeglkad Elektrotechniczny*, Jan. 2010.

[9] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz, "Assessing seizure dynamics by analysing the correlation

structure of multichannel intracranial EEG," *Brain*, vol. 130, no. 1, pp. 65–77, Nov. 2006, doi: 10.1093/brain/awl304.
[10] M. Borowska and N. Białobłocka, "Multiresolution Analysis of EEG Signals," *Stud. Log. Gramm. Rhetor.*, vol. 47, Dec. 2016, doi: 10.1515/slgr-2016-0044.
[11] R. Atangana, D. Tchiotsop, G. Kenne, and L. Djoufack, "EEG Signal Classification using LDA and MLP Classifier," vol. 9, pp. 14–32, Feb. 2020, doi: 10.5121/hiij.2020.9102.
[12] K. Singh and J. Malhotra, "Two-layer LSTM network-based prediction of epileptic seizures using EEG spectral features," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2405–2418, Jun. 2022, doi: 10.1007/s40747-021-00627-z.
[13] M. Zhou *et al.*, "Epileptic Seizure Detection Based on EEG Signals and CNN," *Front. Neuroinformatics*, vol. 12, p. 95, Dec. 2018, doi: 10.3389/fninf.2018.00095.
[14] V. Patel, J. Tailor, and A. Ganatra, "Essentials of Predicting Epileptic Seizures Based on EEG Using Machine Learning: A Review," *Open Biomed. Eng. J.*, vol. 15, no. 1, Oct. 2021, doi: 10.2174/1874120702115010090.
[15] A. Temko, A. Sarkar, and G. Lightbody, "Detection of seizures in intracranial EEG: UPenn and Mayo Clinic's Seizure Detection Challenge," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2015, pp. 6582–6585, 2015, doi: 10.1109/EMBC.2015.7319901.
[16] S. Kovac, V. N. Vakharia, C. Scott, and B. Diehl, "Invasive epilepsy surgery evaluation," *Seizure*, vol. 44, pp. 125–136, Jan. 2017, doi: 10.1016/j.seizure.2016.10.016.
[17] "UPenn and Mayo Clinic's Seizure Detection Challenge." https://kaggle.com/competitions/seizure-detection (accessed Aug. 16, 2022).
[18] S. N. Baldassano *et al.*, "Crowdsourcing seizure detection: algorithm development and validation on human implanted device recordings," *Brain J. Neurol.*, vol. 140, no. 6, pp. 1680–1691, Jun. 2017, doi: 10.1093/brain/awx098.
[19] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," *Procedia Comput. Sci.*, vol. 132, pp. 679–688, Jan. 2018, doi: 10.1016/j.procs.2018.05.069.
[20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.
[21] A. Tharwat, "Classification assessment methods," *Appl. Comput. Inform.*, vol. 17, no. 1, pp. 168–192, Jan. 2020, doi: 10.1016/j.aci.2018.08.003.
[22] Y. Ma and H. He, *Imbalanced Learning: Foundations, Algorithms, and Applications*. John Wiley & Sons, 2013.
[23] R. Parikh, A. Mathai, S. Parikh, G. Chandra Sekhar, and R. Thomas, "Understanding and using sensitivity, specificity and predictive values," *Indian J. Ophthalmol.*, vol. 56, no. 1, pp. 45–50, 2008.
[24] R. Trevethan, "Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice," *Front. Public Health*, vol. 5, p. 307, 2017, doi: 10.3389/fpubh.2017.00307.
[25] W. Siblini, J. Fréry, L. He-Guelton, F. Oblé, and Y.-Q. Wang, "Master Your Metrics with Calibration," in *Advances in Intelligent Data Analysis XVIII*, Cham, 2020, pp. 457–469. doi: 10.1007/978-3-030-44584-3_36.
[26] C. K. I. Williams, "The Effect of Class Imbalance on Precision-Recall Curves," *Neural Comput.*, vol. 33, no. 4, pp. 853–857, Apr. 2021, doi: 10.1162/neco_a_01362.