EEG-based epileptic seizure detection with a bidirectional long short-term memory deep learning model

Lyudmila D. Egorova[1,2*]

[1]Siberian Federal University, 79, Svobodny pr., 660041, Krasnoyarsk, Russia
[2]Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarskiy Rabochiy pr., 660037, Krasnoyarsk, Russia

Abstract. This paper presents a method for detecting epileptic seizures based on electroencephalogram (EEG) analysis using a deep learning model based on Bidirectional Long Short-Term Memory (BiLSTM). The proposed model architecture allows taking into account temporal dependencies and nonlinear dynamics of EEG signals, which makes it effective for recognizing patterns associated with epileptic seizures. The model uses frequency, dynamic, fractal, correlation and statistical characteristics of the EEG signal as informative features. The study includes the stages of data preprocessing, feature extraction and neural network training. To improve the accuracy of the model, data normalization and regularization methods were used. The experimental results obtained on the publicly available TUH EEG dataset demonstrate high performance of the model in detecting epileptic activity: Sensitivity 96.2, Specificity 99.8, F1-score 0.77, AUC 0.98.

1 Introduction

Epilepsy is a common neurological disorder characterized by involuntary seizure activity. This disease can lead to a variety of social restrictions, including restrictions in professional activities, which significantly reduces the quality of life of patients. Finding effective methods for diagnosing and treating epilepsy is a major task in modern medicine. Therefore, the development of algorithms and models of machine learning and artificial intelligence that can improve the quality of diagnosis and treatment of this disease is an urgent task [1].

Despite significant advances in the development of algorithms for classification tasks based on neural networks, including the task of diagnosing epilepsy from an EEG signal, the de facto standard today is still manual analysis of electroencephalogram data by a neurologist. This is due to a number of problems, not least of which are the difficulty of interpreting the classification result issued by a neural network, the high degree of individuality of the EEG, the low signal-to-noise ratio, and the high level of false alarms of existing algorithms, forcing doctors to review too much data [2-4].

* Corresponding author: egorova_ld@rambler.ru

Deep learning methods are often a "black box". It is difficult to explain why a neural network made a particular decision. Therefore, researchers are currently making great efforts to find methods that allow interpreting the results of neural network models [5].

Developing efficient methods for interpreting the results produced by models could improve the clinical acceptance of neural network methods in medical practice.

Examples of software for automated seizure detection approved for clinical use already exist. The FDA-approved Persyst software suite has comparable performance to experienced senior EEG technologists, approaching the level of human EEG seizure detection performance [6-7]. There are also other commercial tools such as Encevis (EpiScan) and Besa [2].

However, existing methods still do not provide all the requirements necessary for their clinical use. A model that realizes seizure detection automation for patients of different genders and ages, taking into account individual characteristics, while having high sensitivity and low false alarm rate is still an urgent need for epilepsy clinics [2].

2 Methodology

2.1 Dataset description

In our experiment, we used the open dataset TUH EEG, Temple University, Philadelphia, USA [8]. This dataset is the world's largest set of EEG recordings designed to support research on the problem of automatic detection of epileptic seizures. The dataset contains several corpuses, each of which is devoted to specific aspects of this problem, such as, for example, identifying artifacts in EEG recordings or detecting a specific group of epileptic events. In our work, we used the TUH EEG Seizure Corpus (TUSZ). We considered the task of automatic detection of epileptic seizures as a problem of binary classification (pathology/norm) of supervised learning. For our experiments, we used a truncated dataset formed in such a way that it preserved the balance of classes present in the original dataset.

2.2 Calculation of informative features

There are different approaches for detecting epileptic events. The choice of a suitable classification model has a decisive impact on the quality of the result obtained by the algorithm. Recent studies show that deep learning methods on raw EEG signals outperform traditional machine learning methods [3,9]. The advantage of deep learning models is their ability to automatically extract and learn informative features from the input data.

Convolutional neural networks (CNNs) are often used to analyze EEG data. Recently, long short-term memory (LSTM) networks have proven themselves to be effective in dealing with sequential data, such as EEG signal data. They cope well with raw time series, as they are able to capture temporal dependencies. At the same time, the data fed to the neural network can be either raw time series or pre-processed using various methods, such as the sliding window method for computing informative features.

The choice of the method for generating input data for a neural network depends on the specific task and the characteristics of the data. In [2], independent component analysis (ICA) was used to clean the recording from artifacts. In [1], EEG signals were pre-processed using empirical mode decomposition followed by bandpass filtering to remove noise, and then automated features were extracted from the signal using a three-layer convolutional neural network. In [4], multispectral informative features were pre-calculated from raw EEG, and then these prepared data were fed to the input of the neural network. In [10], informative features were calculated and then principal component analysis (PCA) was used to reduce the dimensionality of the input data. We chose an approach in which informative features are calculated using a sliding window method on raw EEG data. This approach improves the performance of the neural network model by providing it with additional features that characterize the nature of the data. Since EEG data have complex time dependencies, calculating additional features provides the model with additional information for generalization.

To calculate informative features, the original raw EEG recording was divided into windows for each channel independently. We selected the window width based on the experiments conducted in the previous work related to the calculation of fractal characteristics of the signal. Although shorter windows are found in the works of other authors, they used raw EEG data [9,11-12].

The calculation of informative features was performed using the sliding overlapping window method with a window width of 4 s and a shift of 25% of the window width. For each window, 26 informative features were calculated, which can be divided into the following categories:

- Frequency characteristics: spectral power of the signal in the delta, theta, alpha, beta and gamma frequency ranges. These parameters represent information about the level of brain activity in different frequency ranges.
- Dynamic characteristics: sample entropy, maximal Lyapunov exponent, correlation dimension. These indicators help to assess the complexity and randomness of time series.
- Fractal characteristics: Hurst exponent, detrended fluctuation analysis (DFA), Higuchi fractal dimension. Fractal analysis provides additional information about the structure and dynamics of EEG signals.
- Correlation characteristics: autocorrelation. This parameter describes the degree of linear relationship between the values of a time series at different time intervals, which allows one to evaluate the temporal structure of the signal.
- Statistical characteristics: maximum, minimum, moments: mean (1st raw moment M1), variance (2nd central moment M2), skewness (normalized 3rd central moment M3), kurtosis (normalized 4th central moment M4); median, standard deviation, the first (Q1) and the third quartile (Q3), interquartile range limits (IQR): the lower Q1−1.5 IQR and the upper Q3 + 1.5 IQR; 5th and 95th percentiles. Statistical analysis provides information about the distribution of signal values, which gives insight into the overall behavior of the system.

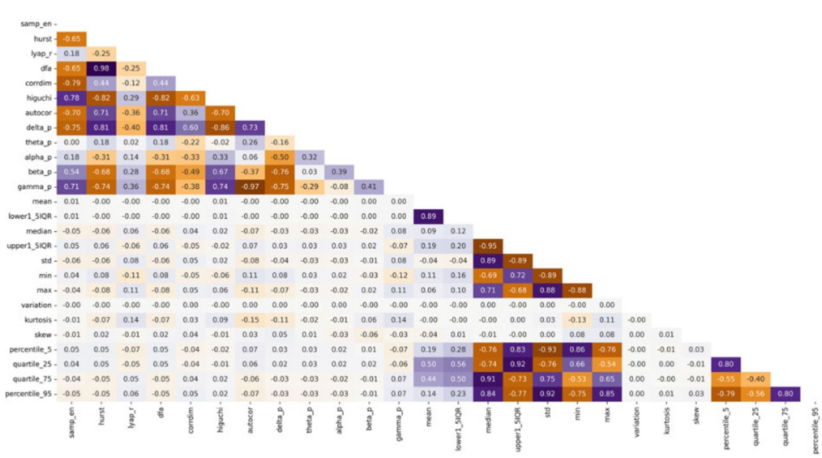

Fig. 1 Heatmap of correlations between informative features.

Figure 1 shows a heatmap representing the correlations between different information features.

The correlation heat map is a powerful tool for visualizing the relationships between features. It is used in conjunction with feature selection and model testing methods to select the optimal set of features to find a balance between the best performance and model complexity.

2.3 Preliminary data preparation

2.3.1 Splitting into sequences

When working with time series for classification problems, especially when using LSTM networks, it is important to properly prepare the data. The choice of sequence length is especially important. Splitting into sequences of fixed length is performed both when using raw EEG records and when using the approach with preliminary calculation of informative features. The choice of sequence length is an important parameter, the optimal value of which depends on the nature of the data and the task. The choice of this parameter can be influenced by the presence of short-term or long-term dependencies in the data. When choosing the sequence length, we used an empirical approach, testing different values and then choosing the best value, which turned out to be seq_length = 16.

2.3.2 Data resampling

The dataset we used to train and test the model is highly imbalanced. The imbalance of this dataset stems from the nature of the data, as epileptic events in EEG recordings are quite rare, accounting for about 1% of the total number of samples. Many machine learning algorithms do not work well with imbalanced data.

There are various methods to solve the problem of class imbalance. One of the widely used approaches to solve this problem is the method of resampling the training set. Another approach is to assign different weights to each of the classes.

There are two main types of resampling: majority undersampling methods, which reduce the majority class to the size of the minority class, and oversampling methods, which enlarge the minority class to the size of the majority class. For example, [3] used undersampling, while [1] enlarged the minority class using synthetic segments generated by generative adversarial networks. The choice of an appropriate data resampling method has a significant impact on the classification result. For each specific dataset, it is advisable to determine the most suitable method through additional experiments. We chose the oversampling method "Oversample using Adaptive Synthetic (ADASYN)", which creates synthetic samples for the smaller class based on the existing samples, thereby increasing the amount of data for the smaller class [13].

In addition to rebalancing, other data pre-preparation procedures were performed, such as cleaning data from missing values, scaling and normalization.

2.4 Proposed model

We used a Bidirectional Long Short-Term Memory (BiLSTM) neural network with three layers. BiLSTM processes sequences in both directions, which allows the model to better capture complex temporal dependencies in the data and improves the model's performance.

To reduce overfitting between BiLSTM layers, DROPOUT layers were used. DROPOUT is a regularization method that randomly sets some neurons to 0 during training. This technique prevents the model from overfitting to the training data.

The ADAM optimizer was used as the optimizer. The hyperbolic tangent (tanh) was used as the activation function.

The final (output) fully connected layer of DENSE uses a sigmoid activation function, which transforms the outputs of the biLSTM layer into probabilities of classes 0 or 1. The the binary cross-entropy was used as the loss function. This loss function is designed specifically for binary classification problems [14] and works with the probabilities that the model produces.

The graphical diagram of the neural network model of binary classification is presented in Figure 2.

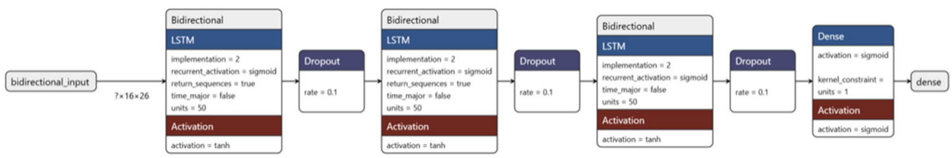

Fig. 2 Proposed binary classification model based on BiLSTM

2.4.1 Model training

After data preprocessing and resampling, the model was trained on the resampled data, which was already a balanced sample.

The presence of model overfitting was monitored using the training graph. The training graph, shown in Figure 3, shows the change in the sensitivity metric (recall) on the training and validation sets by epoch. This graph allows to get an idea of how the model learns and improves over time.

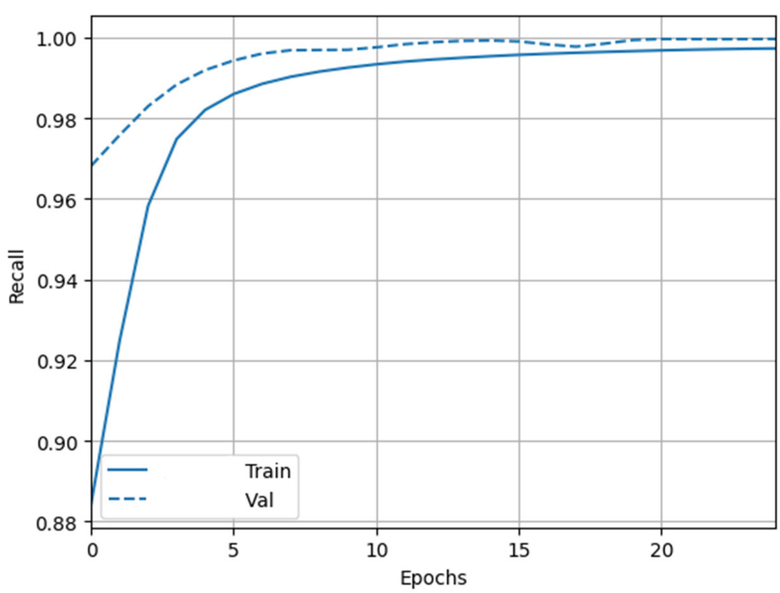

Fig. 3 Graph of model training on training and validation samples.

3 Results

Table 1 presents the results of computational experiments. As the main indicators for assessing the effectiveness of the model the metrics Sensitivity, Specificity, F1-Score and AUC-ROC was used. The choice of these metrics was dictated by the strong imbalance of the data set. Using these metrics allows to take into account the imbalance of classes.

The Sensitivity metric (Recall, True Positive Rate (TPR)) evaluates the proportion of correctly predicted positive events among all real positive events and shows how correctly the model identifies positive samples. Sensitivity is calculated using the formula:

$$Sensitivity = \frac{TP}{(TP+FN)} \tag{1}$$

Specificity shows how often the classifier does not correctly classify objects into the positive class. Specificity is determined by the formula:

$$Specificity = \frac{TN}{(FP+TN)} \tag{2}$$

Precision shows what proportion of positive class predictions turned out to be correct and is calculated using the formula:

$$Precision = \frac{TP}{(FP+TP)} \tag{3}$$

The False Positive Rate (FPR) is described by the formula:

$$FPR = \frac{FP}{(FP+TN)} \tag{4}$$

F1-Score conveys the balance of Sensitivity and Precision, being their harmonic mean. F1-Score is calculated using the formula:

$$F1 = \frac{2*Precision*Sensitivity}{Precision+Sensitivity}, \tag{5}$$

where:
TP = True positives;
TN = True negatives;
FP = False positives;
FN = False negatives.

Specificity and F1-Score metrics are very important for assessing the clinical applicability of the method, since their value is affected by the number of false positive results. The clinician needs to review the events that the algorithm considers pathological. An excessive number of false positive results, even with a high level of Sensitivity, makes the method unsuitable for clinical use, since, given the class imbalance, in this case, clinicians have to review too many events labeled by the model as pathology, but in fact are not [15].

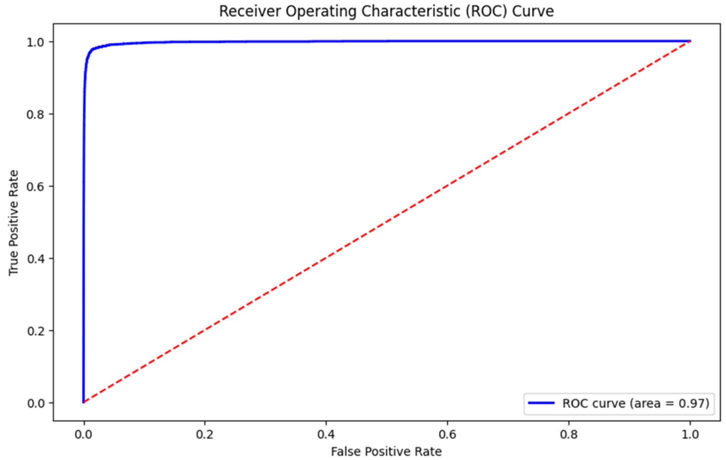

Fig. 4. ROC-curve and AUC for the proposed model.

ROC curve (Receiver Operating Characteristic) and AUC (Area Under the Curve) are important tools for assessing the quality of binary classification models. ROC curve shows how the TPR changes depending on the FPR at different classification thresholds. AUC evaluates the overall ability of the model to distinguish between classes, where a value of 1.0 indicates a perfect model and 0.5 indicates a random model.

Figure 4 shows the ROC curve and AUC of the model, allowing to judge how well the model distinguishes between classes.

The values of these metrics were obtained on test data that was not involved in the training process.

Table 1. Comparison of the performance of the proposed model with models of other authors tested on the TUH EEG dataset

Author	Classifier	Task	Specifi-city	Sensiti-vity	FPR	F1	AUC
X. Zhang at al. [16]	CNN+Attention	M	97.4	88.1	-	-	0.95
D. M. Shama et al. [17]	Transformer+ BiLSTM	B	89.0	67.9	-	-	0.901
M. Golmohammadi at al. [18]	Hybrid HMM/DL	M, B	-	95.11	-	-	-
Y. Ma et al. [19]	Transformer	M	-	-	-	-	0.921
S. Tang et al. [20]	GNN	M	-	-	-	0.749	0.875
Saab K. at al. [3]	CNN	-	-	-	-	-	0.88
Y. Yang at al. [2]	ConvLSTM	B	-	-	-	-	0.84
Asif at al. [4]	CNN	M	**-**	**-**	-	**0.94**	**-**
Ours	biLSTM	B	**99.8**	**96.2**	0.002	0.85	**0.979**

4 Conclusion

In this paper, a binary classification model of EEG signals based on the bidirectional LSTM neural network is proposed. As input data, the neural network receives informative features calculated by the sliding overlapping window method. Informative features include frequency, fractal, nonlinear and statistical characteristics of the EEG signal. Comparison of the obtained results with existing methods tested on the same TUH EEG dataset shows that the proposed model demonstrates high values of Specificity 99.8, Sensitivity 96.2, F1-score 0.85 and AUC 0.98. The proposed model also has low level of FPR 0.002.

The work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2022-1121).

References

1. M. Hosseinzadeh, P. Khoshvaght, S. Sadeghi, P. Asghari, P. Varzeghani, A. Varzeghani, M. Mohammadi, H. Mohammadi, J. Lansky, S. Lee, *A Model for Epileptic Seizure Diagnosis Using the Combination of Ensemble Learning and Deep Learning*, IEEE Access. pp. 1-1. (2024), doi: 10.1109/ACCESS.2024.3457018

2. Y. Yang, N. D. Truong, C. Maher, A. Nikpour, O. Kavehei, *Continental generalization of a human-in-the-loop AI system for clinical seizure recognition*, Expert Systems with Applications, 207:118083, (2022)
3. K. Saab, J. Dunnmon, C. Ré, D. Rubin, C. Lee-Messer, *Weak supervision as an efficient approach for automated seizure detection in electroencephalography*, NPJ digital medicine, **3**, 59 (2020). doi: 10.1038/s41746-020-0264-0
4. U. Asif, S. Roy, J. Tang, S. Harrer, *SeizureNet: A Deep Convolutional Neural Network for Accurate Seizure Type Classification and Seizure Detection*. arXiv Prepr. arXiv1903.03232 (2019).
5. Y. Sun et al., *Continuous Seizure Detection Based on Transformer and Long-Term iEEG,* in IEEE Journal of Biomedical and Health Informatics, **26(11)**, 5418-5427 (2022). doi: 10.1109/JBHI.2022.3199206.
6. M. L. Scheuer, S. B. Wilson, A. Antony, G. Ghearing, A. Urban, A. I. Bagi ́c, *Seizure detection: interreader agreement and detection algorithm assessments using a large dataset*, J. Clin. Neurophysiol., **38,** 439–47 (2021).
7. C. N. Joshi, K. E. Chapman, J. J. Bear, S. B. Wilson, D. J. Walleigh, M. L. Scheuer, *Semiautomated spike detection software persyst 13 is noninferior to human readers when calculating the spike-wave index in electrical status epilepticus in sleep,* J. Clin. Neurophysiol., **35,** 370–4 (2018)
8. I. Obeid, J. Picone, *The temple university hospital EEG data corpus,* Frontiers in Neuroscience, **10**, 00196 (2016).
9. C. da Silva Lourenço, M. C. Tjepkema-Cloostermans, M.J. van Putten, *Machine learning for detection of interictal epileptiform discharges,* Clin. Neurophysiol., **132** 1433–43 (2021).
10. A. Omar, T. Abd El-Hafeez, *Optimizing epileptic seizure recognition performance with feature scaling and dropout layers*, Neural Computing and Applications **36,** 2835–2852 (2023). doi: 10.1007/s00521-023-09204-6.
11. Y. Hao, H. M. Khoo, N. von Ellenrieder, N. Zazubovits, J. Gotman, *DeepIED: an epileptic discharge detector for EEG-fMRI based on deep learning,* NeuroImage, **17**, 962–75 (2018).
12. S. Clarke, *Computer-assisted EEG diagnostic review for idiopathic generalized epilepsy Epilepsy,* Behav., **121,** 106556 (2019)
13. H. He, B. Yang, A. G. Edwardo, L. Shutao, *ADASYN: Adaptive synthetic sampling approach for imbalanced learning,* In IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322-1328, (2008).
14. I. J. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (Cambridge, MA, USA: MIT Press, 2016).
15. D. Nhu, M. Janmohamed, A. Antonic-Baker, et al., *Deep learning for automated epileptiform discharge detection from scalp EEG: A systematic review*, J Neural Eng. **19(5)** 051002 (2022), doi: 10.1088/1741-2552/ac9644
16. X. Zhang, L. Yao, M. Dong, Z. Liu, Y. Zhang, Y. Li, *Adversarial representation learning for robust patient-independent epileptic seizure detection,* IEEE Journal of Biomedical and Health Informatics, **24**, 2852–2859 (2020)
17. D. Moodasarige Shama, J. Jing, A. Venkataraman, *Deep- SOZ: A Robust Deep Model for Joint Temporal and Spatial Seizure Onset Localization from Multichannel EEG Data*, **10**, 184–194 (2023).

18. M. Golmohammadi, A. H. Harati Nejad Torbati, S. Lopez de Diego, I. Obeid, J. Picone, *Automatic analysis of EEGS using big data and hybrid deep learning architectures,* Frontiers in Human Neuroscience, **13**, 00076 (2019).
19. Y. Ma, C. Liu, M. S. Ma, Y. Yang, N. D. Truong, K. Kothur, A. Nikpour, O. Kavehei, *Tsd: Transformers for seizure detection,* bioRxiv, (2023).
20. S. Tang, J. A. Dunnmon, K. Saab, X. Zhang, Q. Huang, F. Dubost, D. Rubin, C. Lee-Messer, *Self-supervised graph neural networks for improved electroencephalographic seizure analysis,* in International Conference on Learning Representations (2021).