Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

EARLY PREDICTION OF EPILEPTIC SEIZURE USING EEG SPECTRAL FEATURES AND MACHINE LEARNING APPROACHES

FAHIMA HOSSAIN*

Lecturer, Department of Computer Science and Engineering, Hamdard University Bangladesh, Gazaria, Munshiganj, Bangladesh. *Corresponding Author Email: fahima@hamdarduniversity.edu.bd

MD. ABUL ALA WALID

Department of Computer Science and Engineering, Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. Email: abulalawalid@gmail.com

ROKEYA AKTER

Department of Computer Science and Engineering, Hamdard University Bangladesh, Gazaria, Munshiganj, Bangladesh. Email: al.mowtushi@gmail.com

S. M. SAKLAIN GALIB

Department of Biomedical Engineering, Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh. Email: shaklain29@gmail.com

MIR MOHAMMAD AZAD

Professor and Dean, Department of Computer Science and Engineering, Hamdard University Bangladesh, Gazaria, Munshiganj, Bangladesh. Email: csdrazad@ hamdarduniversity.edu.bd

Abstract

This paper proposes a system for predicting epileptic seizures from EEG signals using Machine Learning approaches in order to prevent seizures through medication. Electrocorticography (ECoG) and electroencephalography (EEG) media are frequently used to detect these brain impulses. These signals generate a large amount of data and are complicated, noisy, non-linear, and non-stationary. Therefore, identifying seizures and learning about the brain's functions is a difficult undertaking. Without sacrificing performance, machine learning classifiers can classify EEG data, detect seizures, and highlight pertinent, meaningful patterns. In this study, the epileptic seizure dataset was classified using a variety of classifiers. Support vector machines performed better than Naive Bayes, K-Nearest Neighbors, Random Forest classifier, Logistic Regression, Bagging classifier, AdaBoost classifier, Gradient Boosting classifier, Stochastic Gradient Descent (SGD) classifier, Multi-layer Perceptron (MLP) classifier, XGBoost classifier, and Decision Tree classifier, as demonstrated. In this study, we employed the CHBMIT dataset of scalp EEG signals and tested our suggested methodology on the dataset's 22 participants. With superior performance and higher prediction accuracy, our suggested seizure prediction approach is able to reach 95.88% accuracy, 86.91% recall, 1% precision, and 1% sensitivity.

Index Terms: EEG signals, epileptic seizure, prevalence, scaling, machine learning algorithms

1. INTRODUCTION

Epilepsy, a neurological disorder of the brain, is characterized by recurrent seizures and affects approximately 50 million individuals worldwide [1]. The condition is marked by abnormal electrical activity in the cerebral cortex, resulting in excessive neuronal discharges that affect the entire body [2]. Patients with epilepsy may experience sudden and unforeseeable seizures, which can leave them defenseless and at risk of harm from suffocation, drowning, falls, or car accidents [3][4]. Unfortunately, in over 30% of cases, available medical or surgical treatments may not be able to control the occurrence of

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

subsequent seizures [5]. Early detection and medication can prevent attacks, allowing for adequate preparation before the onset of a seizure [6].

Frequent seizures in patients with epilepsy can cause various symptoms, including amnesia, mild depression, and persistent headaches. These seizures can also result in abnormal body movements and even death. Among those affected, approximately 70% are adults and 30% are children. To achieve the goals of personalized medicine, there is a need to automate the detection of epilepsy by identifying abnormal EEG patterns.

Various screening methods, such as MRI, EEG, MEG, and PET, have been developed to identify epileptic seizures. EEG signals are particularly popular due to their affordability, portability, and characteristic frequency patterns. However, EEG data requires careful examination by a neurologist or epileptologist, making the diagnosis of epilepsy time-consuming and laborious. To address this challenge, researchers have explored the creation of a computer-based diagnostic system [7]. An automated categorization and detection system would provide objective results, improve treatment, and greatly enhance epilepsy diagnosis, long-term patient monitoring, and therapy [8]. To overcome the limitations of manual seizure identification and potential human error, researchers have investigated novel ways to predict seizures using epileptic EEG and artificial intelligence.

The analysis of EEG signals often involves the use of wavelet transform techniques for time-frequency estimation. Specifically, the discrete wavelet transform (DWT) technique is commonly used for extracting features from EEG data, as it is a traditional time-frequency analysis method similar to the short-time Fourier transforms [9].

Recently, there has been a growing interest in the use of deep learning (DL) and machine learning (ML) techniques for EEG signal analysis [10]. To address the challenge of patient-independent seizure prediction, research is focused on developing deep learning systems that can learn from data collected from multiple subjects [11]. Machine learning-based approaches have also been developed for detecting abnormal patterns in EEG data during seizures, facilitated by the advancements in IoT-based data collection [12]. In particular, the combination of PCA with neural networks has been proposed for seizure detection.

Various methods have been reported for the categorization of EEG signals, including wavelet transform, PCA, ICA, and linear discriminant analysis (LDA) using support vector machines (SVM) [13]. Directly classifying EEG data patterns by feeding sampled waveforms into classifiers often yields poor performance due to the curse of dimensionality or sparsely dispersed data over a high-dimensional feature space, which leads to a significant decline in classifier accuracy [14].

To address this issue, this paper employs preprocessing techniques such as finding and removing missing and duplicate values, checking the prevalence of the target class, and performing feature scaling using StandardScaler. The dataset is then split into training, testing, and validation sets. Thirteen different classification algorithms are utilized to build the system, including Gaussian Naïve Bayes, Bernoulli Naïve Bayes, Random Forest, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Bagging Classifier, AdaBoost Classifier, Gradient Boosting Classifier, Stochastic Gradient Descent (SGD)

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

Classifier, Multi-layer Perceptron (MLP) Classifier, XGBoost Classifier, and Decision Tree Classifier, to increase overall accuracy.

The rest of the paper is organized as follows: Section 2 provides an overview of related work in seizure prediction. Section 3 describes the dataset used, proposed methodology. Section 3 provides comparison of results our proposed epileptic seizure prediction methods. Section 4 contains the conclusion of the research work and the future work.

2. RELATED WORKS

SangukRyu et al. [1], developed a new hybrid model called DenseNetLSTM that incorporates both DenseNet and LSTM techniques to predict patient-specific epileptic episodes using scalp EEG data. The DenseNet approach, which improves computational efficiency and network information flow beyond current CNN methods, was used in combination with LSTM to learn long-term temporal EEG data properties. However, further testing with additional EEG data is necessary to validate its performance.

Amir Hussein et al. [2], developed an automatic method for accurately detecting pre-ictal seizure states from raw EEG signals. The method is robust against noise and involves filtering the EEG data to remove power line noise, normalizing using z-score, and segmenting the signals using a sliding window. The model employs convolution and recurrent layers, including Gated Recurrent Units (GRUs), to capture temporal dynamics and long-term dependencies in the time series.

John Thomas et al. [13], used the International 10-20 electrode system to collect data for an EEG classification system that distinguishes between EEGs with and without IEDs at Massachusetts General Hospital. The system consists of a pre-processing module, EEG-level classification, and waveform-level classification using CNN and SVM. They identified the best attributes using p-values and a two-sample t-test. The EEG-level classification SVM uses an 8-dimensional feature vector with a threshold of 0.4 for recognizing EEGs with IEDs. They aim to improve the system by implementing artifact rejection and customizing the preprocessing module for other montages and EEG recording devices.

In their study, Syed Muhammad Usman et al. [14], applied a strategy to scalp EEG signals from 22 participants in the CHBMIT dataset. To tackle class imbalance, they employed generative adversarial networks to generate preictal samples during preprocessing. The automated feature extraction used a three-layer convolutional neural network, and classification of preictal and interictal states was done with long short-term memory units. Their method is adaptable to different datasets and can be used on real-time EEG signal recordings for seizure prediction. Combining deep learning and machine learning techniques may increase the average anticipation time in the future.

Yuan Zhang et al. [15], used scalp EEG recordings from the CHB-MIT dataset, which includes data from 23 individuals with medically untreatable focal epilepsy. Their novel approach to seizure prediction involved using convolutional neural networks and common spatial patterns (CSP). To address trial imbalance, artificial pre-ictal EEG signals were created by combining segmented pre-ictal signals. Features were extracted using CSP

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

and wavelet packet decomposition in both the time and frequency domains, which shortened training time and increased overall accuracy. A shallow CNN was then applied to distinguish between pre-ictal and inter-ictal states.

3. PROPOSED METHODOLOGY

This section provides a detailed description about the proposed methodology of this model. The proposed model is illustrated in figure 1 and it consists of five main parts.

A. Data Collection: In the paper, we utilized the CHB-MIT dataset, which consists of scalp EEG recordings obtained from 23 pediatric patients at Children's Hospital Boston. This dataset is publicly available and can be accessed through PhysioNet.org with open access.

B. Data Preprocessing: We used an original dataset consisting of 500 subjects, each with a 23.6-second recording of brain activity. The dataset is divided into five folders, with each folder containing 100 ".edf" files representing a single subject. Each ".edf" file contains 4097 preictal and ictal data points, with each data point representing the value of the EEG recording at a specific point in time. This yields a total of 11,500 rows, where each row represents a one-second chunk with 178 data points and a label y1-y5. The response variable (y) and explanatory variables (X1 to X178) are contained in the last column. The dataset is then converted to ".csv" format for further processing.

Table 1: Classes in the target attribute and their meaning.

Class	Meaning
1	recording of seizure activity
2	the EEG signal are taken from the area where the tumor was located
3	the EEG signal are taken from the healthy brain area
4	the EEG signal was taken when the patient had their eyes closed
5	the EEG signal was taken when the patient had their eyes open

Table 1 depicts the target attribute classes and their meaning in the dataset. All subjects in classes 2, 3, 4, and 5 did not experience an epileptic seizure. Only class 1 subjects have epileptic seizures. Data points are loaded and preprocessed as data frames using Python's pandas library.

To calculate the prevalence rate, we divide the overall percentage of positive samples by the percentage of patients in our dataset experiencing seizures. In this case, the prevalence rate is 20%. This rate helps balance classes and evaluate our model using the "lift" metric.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

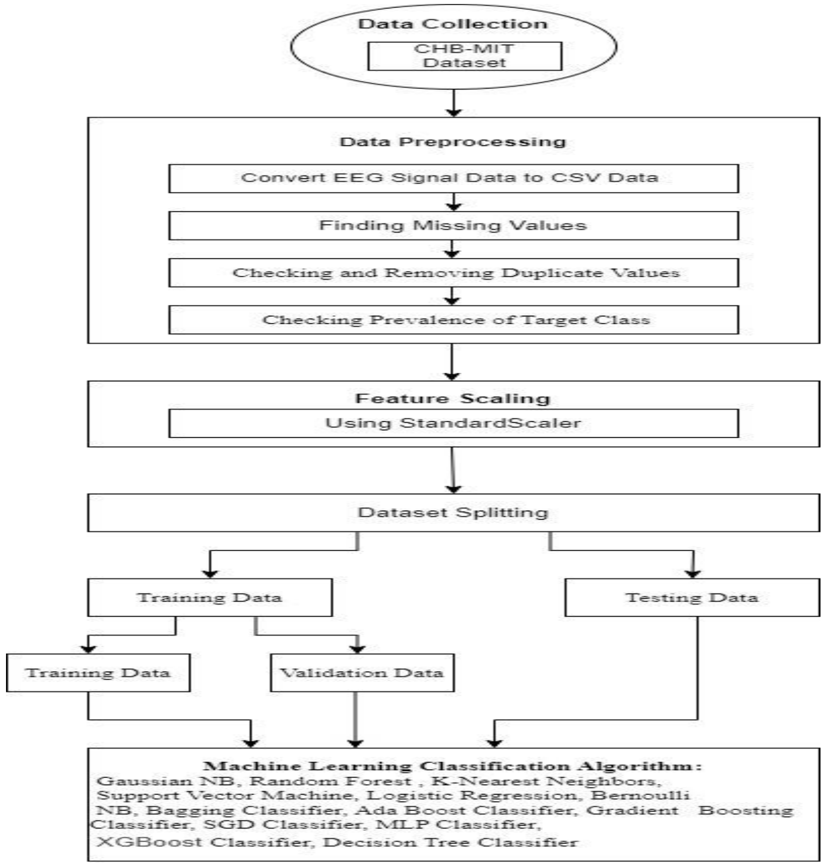

Fig. 1: Proposed Architecture.

C. Feature Scaling: Scaling our variables is essential for proper functioning of our models. To achieve this, we utilized the Standard Scaler from the sklearn library to scale our variables.

D. Splitting Data and Building Training/Validation/Test Sets: Our dataset consists entirely of numerical values of EEG readings, so no feature engineering is required before feeding it into our machine learning model. It's important to keep predictor and response variables separate, and split the dataset into training, validation, and testing sets. The training set typically ranges from 50% to 90% of the core dataset, depending on sample size, while validation and testing sets are usually the same size. To avoid order bias, we first shuffle our dataset before using a 70/15/15 split for training, validation, and testing sets. We then balance the dataset to avoid misclassification of samples as belonging to the majority class, and save each set as a .csv file.

E. Classification and Prediction: The system was constructed with 13 classification algorithms, such as Gaussian/Bernoulli Naive Bayes, Random Forest, K-Nearest Neighbors, and others. Multiple classifiers were used to evaluate the model's predictive capabilities since there is no one-size-fits-all approach.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

4. RESULT ANALYSES

This machine learning model's effectiveness is evaluated through a performance matrix that includes criteria like TP, FP, TN, and FN. Using these metrics, the effectiveness of different feature selection techniques and the classifier has been evaluated.

True positive (TP): The ANN recognizes a seizure segment that the expert identified as a seizure.

True negative (TN): Both the expert and the ANN concur that the EEG pattern does not indicate a seizure.

False positive (FP): the identification of a seizure segment that the expert misdiagnosed as a non-seizure.

False negatives (FN): occur when an expert-identified seizure segment is missed by the ANN.

So, Confusion matrix $=\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$

The following parameters are typically used to assess the performance of classifiers based on the aforementioned metrics:

Area Under the ROC Curve (AUC): AUC measures the complete two-dimensional region beneath the entire ROC curve from (0,0) to (1,1).

Average Accuracy, AA: The percentage of how accurately the model is approximated is defined as accuracy.

Accuracy = ((TP + TN)/(TP + TN + FP + FN)) x 100%

Recall, r: The total amount of data points that are actually retrieved is referred to as recall.

Recall = TP/(T P + F N) x 100%

Precision (p): Precision is the ratio of the number of relevant data points to the total number of relevant and irrelevant ones.

Precision = TP/(T P + F P) x 100%

Specificity (SPE): Specificity is the indicator of how well a classifier can identify non-seizure activity.

Specificity = TN/(T N+ F P) x 100%

Figure 4 demonstrates a comparison of AUC values for various classifiers. The result clearly indicates that Random Forest, Gradient Boosting, and XGboost achieved a comparatively higher AUC value for both the training and validation datasets. The AUC value of a bagging classifier, MLP, AdaBoost, and Decision Tree is higher in the training dataset than in the validation dataset.

Various machine learning techniques are used to evaluate the performance of the model. The most effective classifiers are the Support Vector Machine and Gaussian Naive Bayes. According to figure 5, the classification accuracy of the Support Vector Machine

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

is 95.88%, that of the Gaussian Naive Bayes is 95.30%, and that of the SGDClassifier with the poorest performance is 79.07%.

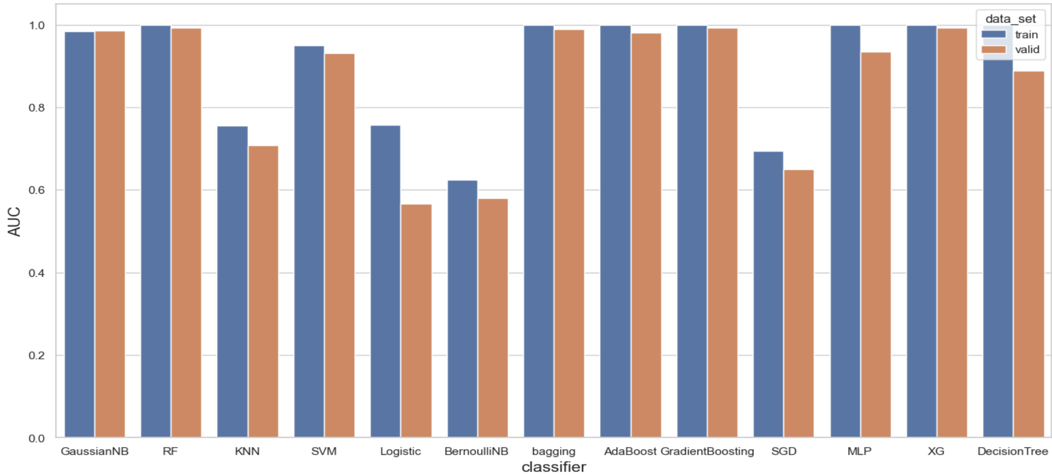

Fig.2: AUC for all Classification Algorithms.

Figures 6, 7, and 8 compare the precision, recall, and specificity of various classifiers. The recall values of Gaussian nave Bayes and Support Vector Machine are higher than those of the other classifiers, while AdaBoost has a recall value of nearly zero. Random Forest, Logistic Regression, and AdaBoost have lower precision and specificity values than other classifiers. Table 2 summarizes all of the results shown in figures 5, 6, 7, and 8. Figure 9 compares accuracy and specificity to other existing systems that is represented in a bar chart. The model's accuracy and specificity are higher than those of the other systems in our proposed work.

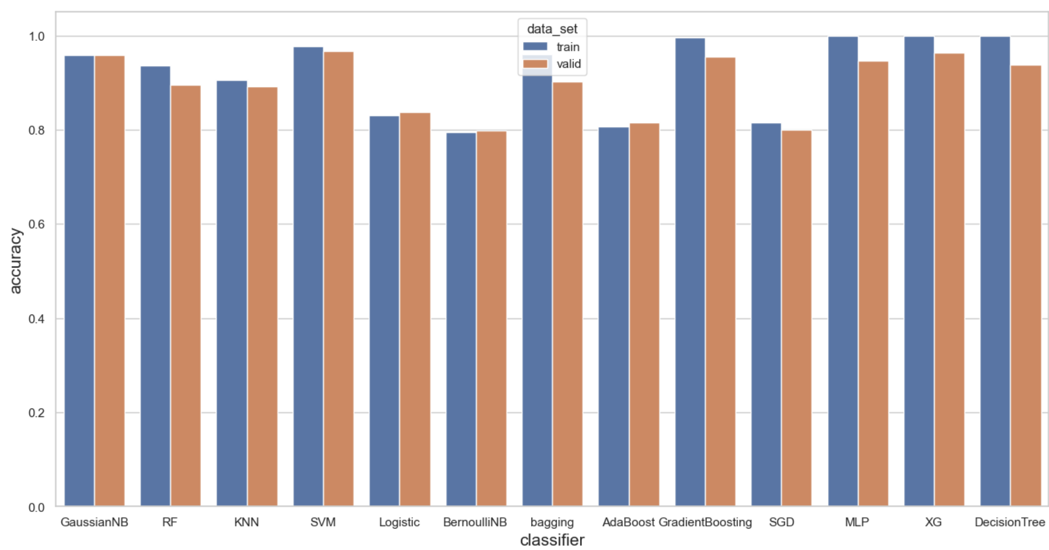

Fig.3: Accuracy for all Classification Algorithms.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

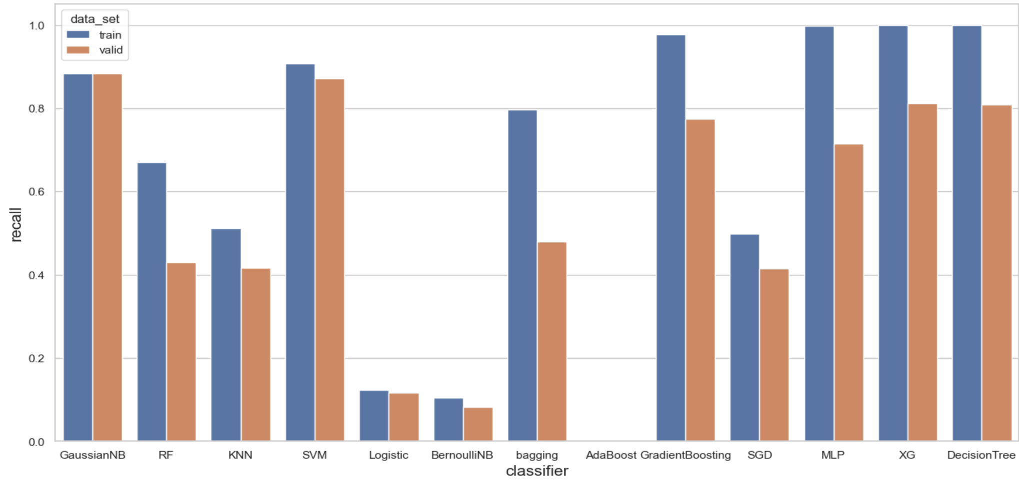

Fig.4: Recall for all Classification Algorithms.

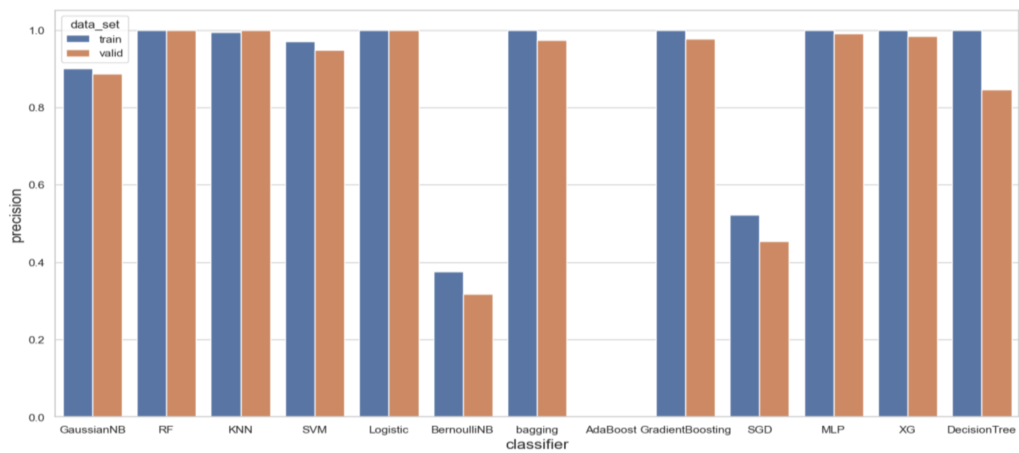

Fig.5: Precision for all Classification Algorithms.

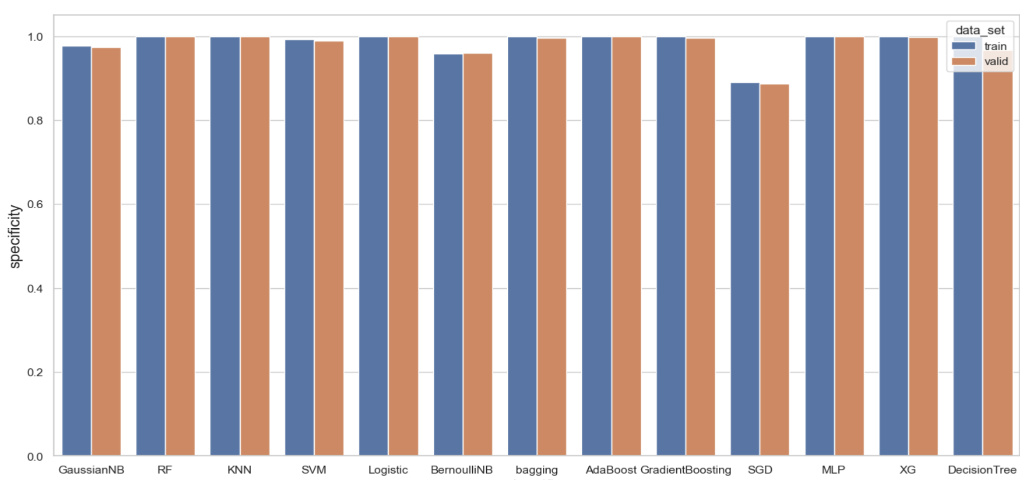

Fig.6: Specificity for all Classification Algorithms.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

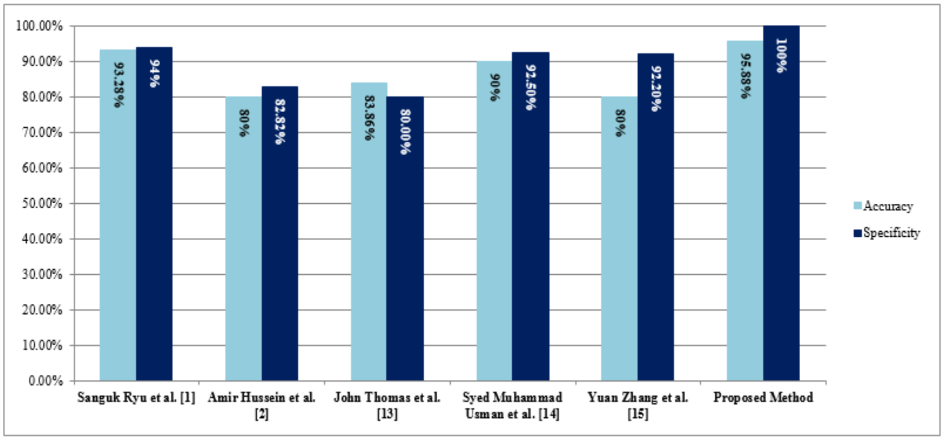

Fig. 9: Comparison of Accuracy and Specificity with other Existing Systems.

5. CONCLUSIONS

This study aims to accurately and specifically detect epileptic seizures using supervised and statistical machine learning algorithms. The model achieved a 95.88% accuracy, 86.91% recall, 1% precision, and 1% sensitivity.

Future development trends for EEG-based epilepsy detection methods include seizure forecasting and localization, which can improve patient quality of life and diagnosis speed and reduce costs for physicians. As machine learning advances, new methods will be applied to feature extraction and the Hybrid Ensemble Classification Algorithm of epileptic EEG signals.

References

1. Ryu, S., & Joe, I. (2021). A hybrid DenseNet-LSTM model for epileptic seizure prediction. Applied Sciences, 11(16), 7661. https://doi.org/10.3390/app11167661
2. Ahmad, I., Wang, X., Zhu, M., Wang, C., Pi, Y., Khan, J. A., Khan, S., Samuel, O. W., Chen, S., & Li, G. (2022). EEG-based epileptic seizure detection via machine/Deep learning approaches: A systematic review. Computational Intelligence and Neuroscience, 2022, 1–20. https://doi.org/10.1155/2022/6486570
3. Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., Guo, H., & Xiang, J. (2018). Epileptic seizure detection based on EEG signals and CNN. Frontiers in Neuroinformatics, 12. https://doi.org/10.3389/fninf.2018.00095
4. Almustafa, K. M. (2020). Classification of epileptic seizure dataset using different machine learning algorithms. Informatics in Medicine Unlocked, 21, 100444. https://doi.org/10.1016/j.imu.2020.100444
5. Muhammad Usman, S., Khalid, S. and Aslam, M.H. (2020) “Epileptic seizures prediction using Deep Learning Techniques,” IEEE Access, 8, pp. 39998–40007. Available at: https://doi.org/10.1109/access.2020.2976866.
6. Usman, S. M., Usman, M., & Fong, S. (2017). Epileptic seizures prediction using machine learning methods. Computational and Mathematical Methods in Medicine, 2017, 1–10. https://doi.org/10.1155/2017/9074759
7. A Shoeibi, M Khodatars, N Ghassemi, M Jafari, P Moridian, R Alizadehsani, M Panahiazar, F

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 05-2023
DOI 10.17605/OSF.IO/GQFVH

Khozeimeh, A Zare, H Hosseini-Nejad, A Khosravi, AF Atiya, D Aminshahidi, S Hussain, M Rouhani, S Nahavandi, and UR Acharya. 2021. “Epileptic seizures detection using deep learning techniques: A review.” International Journal of Environmental Research and Public Health, 18, 11.

8. Gajic D, Djurovic Z, Gennaro SD, Gustafsson F (2014) Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. Biomed EngAppl Basis Commun 26(2):1450021
9. Wang, L., Xue, W., Li, Y., Luo, M., Huang, J., Cui, W., & Huang, C. (2017). Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis. Entropy, 19(6), 222. https://doi.org/10.3390/e19060222
10. Sun, B., Lv, J.-J., Rui, L.-G., Yang, Y.-X., Chen, Y.-G., Ma, C., & Gao, Z.-K. (2021). Seizure prediction in scalp EEG based channel attention dual-input convolutional neural network. Physica A: Statistical Mechanics and Its Applications, 584, 126376. https://doi.org/10.1016/j.physa.2021.126376
11. Dissanayake, T., Fernando, T., Denman, S., Sridharan, S., &Fookes, C. (2021). Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals. IEEE Sensors Journal, 21(7), 9377–9388. https://doi.org/10.1109/jsen.2021.3057076
12. Asif, U., Roy, S., Tang, J., &Harrer, S. (2020). SeizureNet: Multi-spectral deep feature learning for seizure type classification. Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-Oncology, 77–87. https://doi.org/10.1007/978-3-030-66843-3_8
13. Jaiswal, A. K., & Banka, H. (2017). Epileptic seizure detection in EEG signal using machine learning techniques. Australasian Physical & Engineering Sciences in Medicine, 41(1), 81–94. https://doi.org/10.1007/s13246-017-0610-y
14. Aarabi, A., Wallois, F., and Grebe, R., Automated neonatal seizures detection: A multistage classification system through feature selection based on relevance and redundancy analysis. Clin. Neurophysiol. 117:328–340, 2006.
15. Y. Zhang, Y. Guo, P. Yang, W. Chen and B. Lo, "Epilepsy Seizure Prediction on EEG Using Common Spatial Pattern and Convolutional Neural Network", IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 465-474, 2020. Available: 10.1109/jbhi.2019.2933046.