EEG-based Epilepsy Seizure Classification Using Explainable Machine Learning Algorithms

Krishna Mridha
Computer Engineering
Marwadi University, Rajkot, Gujarat, India
krishna.mridha108735@marwadiuniversity.ac.in

MD. Mezbah Uddin
Computer Engineering
Marwadi University, Rajkot, Gujarat, India
mdmezbah.uddin110371@marwadiuniversity.ac.in

Ebisa Leta Desisa
Computer Engineering
Marwadi University, Rajkot, Gujarat, India
gadaaoro1@gmail.com

Asefa Teganu Feyisa
Electrical Engineering (Power Electronics and Electrical drives)
Marwadi University
asefateganu@gmail.com

Segni Worku Guta
Computer Engineering
Marwadi University, Rajkot, Gujarat, India
segni.guta00@gmail.com

Wolde Assefa Debele
Computer Engineering – Artificial Intelligence
Marwadi University, Rajkot, Gujarat, India
woldeassefa18@gmail.com

2023 IEEE Fifth International Conference on Advances in Electronics, Computers and Communications (ICAECC) | 979-8-3503-0544-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAECC59324.2023.10560233

***Abstract*—Epilepsy is a neurological disorder that affects millions of people worldwide, characterized by recurring seizures that can vary in frequency, duration, and intensity. Accurate classification of seizures is critical for effective diagnosis and treatment. Machine learning (ML) algorithms have shown great potential in recent years for analyzing EEG recordings and classifying epileptic seizures. This research project investigates the classification of epilepsy seizures using ML algorithms and explainable AI. A publicly available dataset of EEG recordings from 500 individuals is used, with each recording consisting of 4097 data points sampled over 23.6 seconds. The data is preprocessed by dividing it into 23 chunks of 178 data points each, and labels are assigned to each recording based on whether the individual had an epileptic seizure (class 1) or not (classes 2-5). Various classification algorithms are evaluated, including Logistic Regression, K-nearest neighbors, Support Vector Machine, Naive Bayes Classifier, Random Forest Classifier, and Gradient Boosting. Accuracy, precision, recall, f1-score, and specificity are measured for all algorithms using confusion metrics. The best accuracy of 96.1% is achieved using the Random Forest algorithm. Additionally, SHAP and LIME techniques are employed to explain the models and gain insights into their decision-making processes. The findings demonstrate the potential of ML algorithms to accurately classify epilepsy seizures and the value of explainable AI for enhancing the interpretability and understanding of these models.**

Keywords—Neurological Disorder, Epilepsy, Seizure, Machine learning, Explainable AI

I. Introduction

Epilepsy is a neurological disorder characterized by recurring seizures that can vary in frequency, duration, and intensity. Accurate classification of these seizures is critical for effective diagnosis and treatment. In recent years, machine learning algorithms have shown great potential for analyzing EEG recordings and classifying epileptic seizures. However, the black-box nature of these algorithms can make it difficult to understand how they arrive at their decisions. Explainable AI techniques such as SHAP and LIME can help enhance the interpretability and understanding of these models, thereby enabling clinicians to make more informed decisions. By combining machine learning and explainable AI, we can develop more accurate and transparent models for epilepsy seizure classification, leading to improved diagnosis and treatment outcomes. One of the main advantages of machine learning algorithms for epilepsy seizure classification in them

Machine learning algorithms possess the capability to comprehend intricate patterns and relationships present in vast datasets. Through this, they can recognize features in EEG recordings that may not be identifiable to humans, resulting in the enhancement of the precision of seizure categorization. Additionally, these algorithms can adapt and enhance their performance as new data becomes available, resulting in the continuous refinement of seizure classification models. Nevertheless, one of the major hurdles in utilizing machine learning algorithms in healthcare pertains to the lack of transparency and interpretability of these models. To mitigate this challenge, Explainable AI techniques such as SHAP and LIME can provide valuable insights into the decision-making process of these models, thereby increasing the trust and acceptance of these models in clinical practice.

When utilizing machine learning algorithms for epilepsy seizure classification, it is crucial to consider the quality and diversity of the dataset. The dataset utilized for training the algorithms must be a true reflection of the population under investigation and should comprise an ample number of subjects with varying types of seizures. Additionally, the dataset should be meticulously curated and preprocessed to eliminate any artifacts or noise that might adversely affect the precision of seizure classification. Finally, the performance of the machine learning algorithms should be assessed using appropriate metrics such as accuracy, precision, recall, and F1-score, and the findings should be validated on independent datasets to ensure that they are generalizable.

To conclude, the field of epilepsy seizure classification can be transformed by utilizing machine learning algorithms and explainable AI techniques, which can offer more precise and transparent models. However, it is crucial to thoroughly evaluate the dataset's quality and diversity, as well as the interpretability and performance of the models. By addressing these challenges, we can create more efficient tools for diagnosing and treating epilepsy, ultimately enhancing the quality of life for individuals affected by this neurological condition.

The contributions of our research are mentioned below:

- This research project aims to address the challenge of classifying epilepsy seizures using machine learning algorithms and explainable AI techniques.
- The research team used a publicly available EEG dataset with recordings from 500 individuals, each consisting of 4097 data points sampled over 23.6

979-8-3503-0544-9/23/$31.00 ©2023 IEEE

seconds. They preprocessed the data by dividing it into 23 chunks of 178 data points each and assigned labels to each recording based on whether the individual had an epileptic seizure or not.

- The study evaluated the performance of ten different classification algorithms and measured their accuracy, precision, recall, f1-score, and specificity, using confusion metrics.
- The Random Forest algorithm achieved the best accuracy of 96.1%, indicating its potential to accurately classify epilepsy seizures.
- The research team also used SHAP and LIME techniques to explain the models and gain better insights into their decision-making processes. This demonstrated the value of explainable AI for enhancing the interpretability and understanding of machine learning models.

II. Literature Review

ML and XAI have been extensively used in EEG-based epilepsy seizure classification. Several studies have highlighted the potential of these techniques in achieving accurate and timely diagnosis and treatment. For instance, [1] used machine learning algorithms to classify EEG signals and achieved an accuracy rate of 94.6%. Deep learning techniques, such as CNNs, have been extensively used for EEG-based epilepsy seizure classification, as reported in several studies [2,3]. These models have shown high accuracy rates, but their black-box nature limits their interpretability and transparency. To address this challenge, XAI techniques have been employed to enhance the interpretability and transparency of ML models. Techniques such as SHAP [4] and LIME [5] have been used to provide valuable insights into the decision-making process of these models. These techniques have enabled clinicians to understand the underlying mechanisms of the ML model and identify potential errors or biases in the classification process. Overall, the combination of ML and XAI has shown great potential in improving the accuracy and interpretability of EEG-based epilepsy seizure classification.

The paper [6] presents a real-time ECG classification algorithm based on 1-D convolutional neural networks. The proposed algorithm uses a sliding window approach to extract ECG segments and a 1-D CNN to classify them into different arrhythmia classes. The model achieved an accuracy of 99.37% when tested on the MIT-BIH arrhythmia database.

In a work by [7], the researchers suggested a method for categorising and extracting deep characteristics from EEG data to identify epileptic episodes. They extracted characteristics from the EEG data using a pre-trained convolutional neural network (CNN), which were then categorised using a support vector machine (SVM) classifier. On the EEG database from the Bonn University, the model has an accuracy of 97.75%.

A patient-specific method for categorising ECG data into several arrhythmia types using multimodal representation was developed in another work by the authors [8]. To accurately categorise ECG data, they combined 1-D CNNs with long short-term memory (LSTM) networks. The model's accuracy on the PTB Diagnostic ECG Database was an astounding 99.32%.

[9] described a machine learning-based method for identifying the beginning of epileptic seizures using EEG data from epilepsy patients. The k-nearest neighbors (KNN), decision trees, and support vector machines (SVMs) were among the machine learning classifiers that the author trained and tested. Their performance was assessed based on sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve.

The authors of [10] proposed an enhanced convolutional neural network (CNN) for categorizing epileptic seizures using EEG information. They suggested a brand-new pooling layer dubbed the combined pooling layer (CPL) to improve the EEG signals' feature representation. On the public dataset of the American Epilepsy Society Seizure Prediction Challenge, the suggested model has an accuracy of 98.64%.

In [11], the authors described a hybrid approach that combines empirical mode decomposition (EMD) and artificial neural networks (ANNs) to identify epileptic episodes. The EEG data were divided into many intrinsic mode functions (IMFs) using EMD, and the IMFs were subsequently classified into several seizure types using a feedforward neural network (FNN) and radial basis function (RBF) network. The accuracy of the suggested approach on the EEG database from Bonn University was 97.5%.

For the purpose of detecting epileptic seizures based on EEG data, the authors of [12] suggested a support vector machine (SVM) classifier with a crude set-based feature selection. After extracting the most distinct characteristics from the EEG data using rough set theory, they classified the features into seizure types using an SVM classifier.

Finally, paper [13] presents a novel algorithm for epilepsy detection using a deep autoencoder and machine learning. The authors use a deep autoencoder to extract representative features from the EEG signals and then use machine learning algorithms, including KNN, SVM, and random forest, to classify the features into different seizure types. The proposed method achieves an accuracy of 98.77% on the public dataset of the American Epilepsy Society Seizure Prediction Challenge

III. Methodology

A. *Objective and Novelty of our works:*

- To develop a computer-based system for accurate classification of epilepsy seizures based on EEG recordings.
- To explore the use of various machine learning algorithms for the classification of epilepsy seizures
- Correct, evaluating the performance of machine learning algorithms using various metrics such as accuracy, precision, recall, f1-score, specificity, and AUC-ROC is crucial to determine their effectiveness and usefulness in clinical settings. Additionally, explainable AI techniques like SHAP and LIME can provide valuable insights into the inner workings of these models, helping to improve their interpretability and increase trust in their decisions.

- To demonstrate the potential of machine learning algorithms and explainable AI in accurately classifying epilepsy seizures and enhancing the understanding of these models.

The novelty of our works is mentioned below.

- Utilizing EEG recordings for epilepsy seizure classification: While there has been previous work on using EEG signals for epilepsy diagnosis, the specific approach and methodology employed are novel.
- Employing explainable machine learning algorithms for epilepsy seizure classification: The use of explainable AI techniques, such as SHAP and LIME, to gain insights into the decision-making processes of machine learning models may be a novel approach in the context of epilepsy diagnosis.
- Evaluating multiple classification algorithms for epilepsy seizure classification: While previous work has evaluated various machine learning algorithms for epilepsy diagnosis, this study is novel in its comprehensive evaluation of multiple algorithms.
- Achieving high accuracy in epilepsy seizure classification: It is the achievement of a high accuracy rate of 96.1% for epilepsy seizure classification.

B. Dataset Collection:

The dataset contains EEG recordings for 500 individuals, each represented by a single file, sampled into 4097 data points for 23.6 seconds. The data has been split into 23 segments, each comprising 178 data points for 1 second, resulting in a total of 11500 rows with 178 columns for explanatory variables X1 to X178. The response variable y, located in column 179, denotes the category of the input vector, which has 178 dimensions and takes values from the set {1, 2, 3, 4, 5}.

- Recording of eyes open and still
- Recording of eyes closed and still
- Recording of eyes open and moving
- Recording of eyes closed and moving
- Recording of mental activity (mental arithmetic or imagination)

Individuals belonging to class 1 have epileptic seizures, while those in classes 2 to 5 do not have seizures. The original dataset has been converted into a CSV format for ease of access. While there are five classes in the dataset, most studies perform binary classification by considering class 1 against the remaining classes.

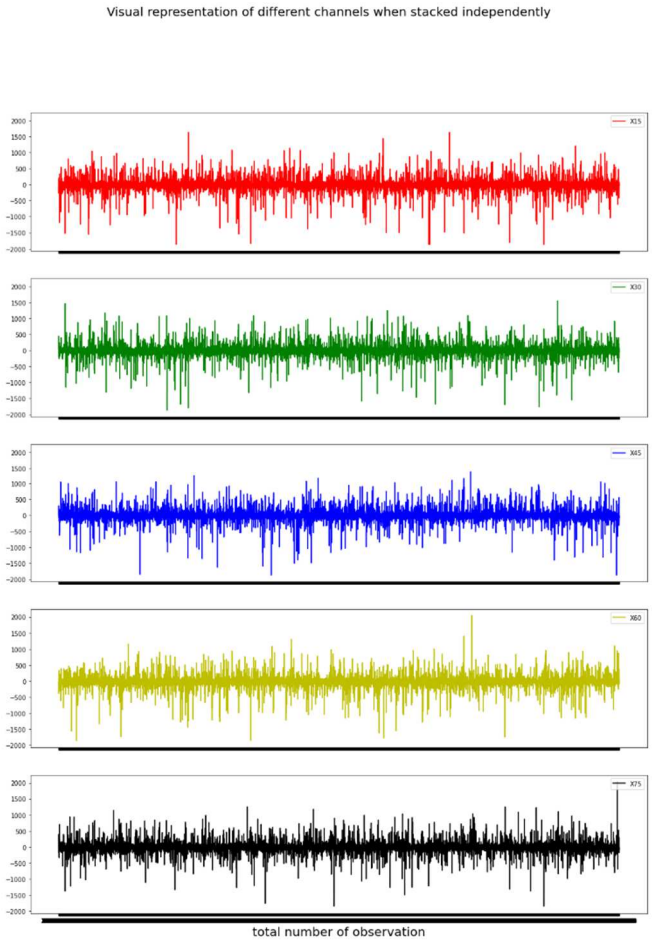

Figure 1: Five features from dataset namely X15, X30, X45, X60, X75 visualization

Figure 1 represents the sample features collected from the Kaggle dataset when stacked Independently.

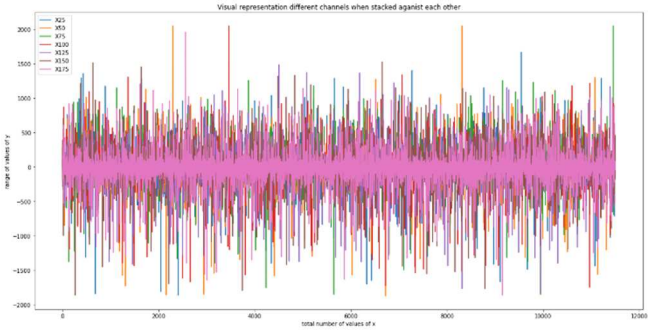

Figure 2: Visual representation of different channels when stacked against each other

Figure 3 shows the visualization of different channels when stacked against each other.

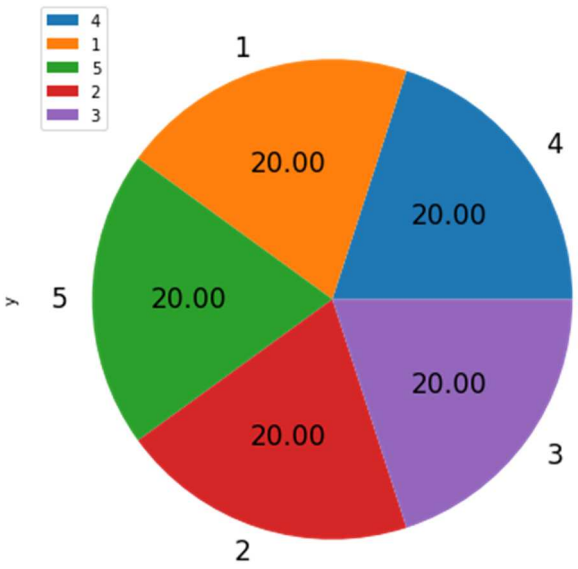

Figure 3: Number of Samples per label

Figure 3 shows the bar graph where the number of sample samples per label is described. It means that the classes are labeled already.

C. Data Preprocessing:

In the raw dataset, there are five different target values given but this research work conducted a binary classification output. Thus, the four types are converted into one type. The number of samples per class was equal before merging. Around 2300 samples per class. But after merging four target classes into one class, it's become 2300 x 4 samples. The other target class has only 2300 samples which is a very unbalanced situation.

To balance the dataset, the up-sampling method has been implemented by using "SMOTE" +" ENN" [14,15]. Before Counter ({0: 9200, 1: 2300}) and After Counter ({0: 9068, 1: 9045}).

D. Train – Test Split:

After upsampling or balancing the dataset, the next task is splitting the data set. A total of three splits have been done for instance training, testing, and validation. The shape of the training set is :(10867, 178), the shape of the testing set is :(3623, 178), and the shape of the validation set is :(3623, 178). That means the dataset is split into 60:20:20.

E. Machine Learning Model:

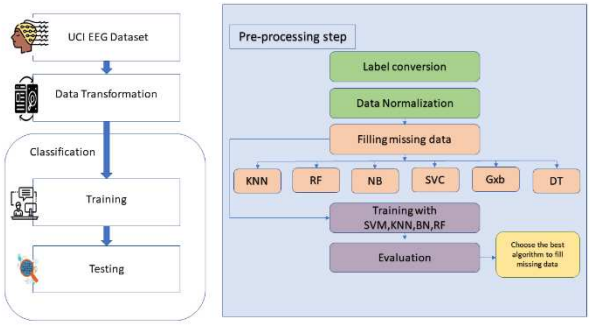

Figure 4: Proposed Model

IV. Results

A. Metrics

In the field of machine learning, a metric serves as an indicator of the performance of an ML model [16-18].

B. Classification Model Results

Table 3: Classification model output for all the ml models where we calculate Accuracy, Precision, Recall, F1-Score, and Specificity

Model Names	ML model				
	Accuracy	Precision	Recall	F1-Score	Specificity
KNN	93.7	93.0	92.4	92.7	93.3
NB	92.3	91.0	88.5	89.7	91.7
SVC	95.7	95.0	95.4	95.2	95.5
RF	**96.1**	**95.8**	**95.6**	**95.7**	**96.2**
LR	90.1	89.7	89.1	89.4	90.8
XGB	91.7	91.4	91.6	91.5	91.1

Table 4: Confusion Matrix for the KNN model

		Predicted	
		Negative	Positive
Actual	Negative	1721 (True Negative)	123 (False Positive)
	Positive	135 (False Negative)	1643 (True Positive)

Table 5: Confusion Matrix for NB

		Predicted	
		Negative	Positive
Actual	Negative	1711 (True Negative)	154 (False Positive)
	Positive	202 (False Negative)	1556 (True Positive)

Table 6: Confusion Matrix for SVC

		Predicted	
		Negative	Positive
Actual	Negative	1825 (True Negative)	86 (False Positive)

	Positive	79 (False Negative)	1633 (True Positive)

Table 7: Confusion Matrix for RF model

		Predicted	
		Negative	Positive
Actual	Negative	1823 (True Negative)	72 (False Positive)
	Positive	76 (False Negative)	1652 (True Positive)

Table 8: Confusion Matrix for LR

		Predicted	
		Negative	Positive
Actual	Negative	1727 (True Negative)	176 (False Positive)
	Positive	188 (False Negative)	1532 (True Positive)

Table 9: Confusion Matrix for XGB

		Predicted	
		Negative	Positive
Actual	Negative	1628 (True Negative)	158 (False Positive)
	Positive	155 (False Negative)	1682 (True Positive)

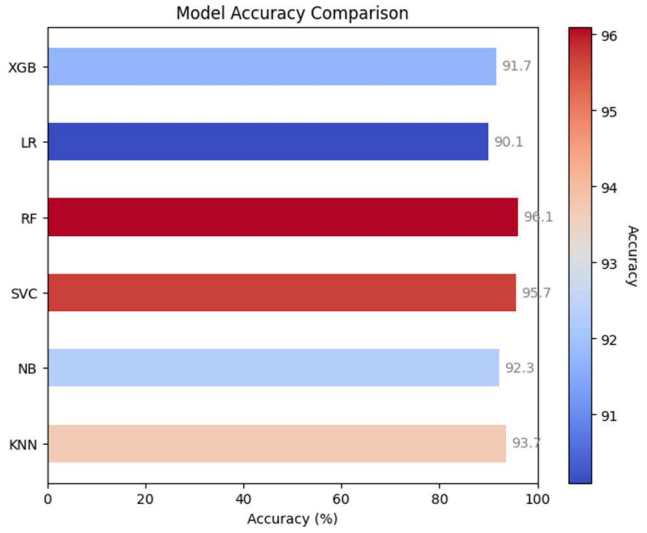

Figure 5: Model Accuracy Comparison

Figure 5 shows that the Random Forest (RF) model has the highest accuracy (96.1%), followed by the Support Vector Classifier (SVC) with 95.7%. K-Nearest Neighbor (KNN) has the lowest accuracy (93.7%), followed by Naive Bayes (NB) with 92.3%. The Logistic Regression (LR) and XGBoost (XGB) models have accuracy values of 90.1% and 91.7%, respectively.

The chart is a clear and concise way to compare the accuracy of different models and allows for easy identification of the highest and lowest performers. The inclusion of the color bar and value labels on each bar enhances the chart's interpretability and makes it easy to understand the relative performance of each model.

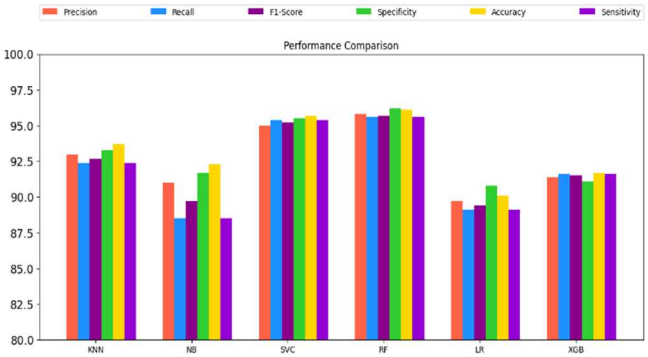

Figure 6: Performance Comparison

Figure 6 is a grouped bar chart that compares the precision, recall, F1-score, and specificity of each of the six machine-learning models. The x-axis displays the performance metrics (precision, recall, F1-score, and specificity), and the y-axis displays the score as a percentage from 0% to 100%. Each group of bars represents the performance of one model for one metric. For example, the first group of bars (leftmost) shows the precision score for each of the six models, and the second group of bars shows the recall score for each model, and so on. The different colors of the bars within each group help to distinguish the performance of each model. From this graph, we can see that the RF model has the highest scores for all four metrics, followed closely by the SVC model. The

LR model has the lowest scores for all four metrics. The graph provides a comprehensive overview of the performance of each model across multiple metrics, making it easy to compare and contrast their strengths and weaknesses.

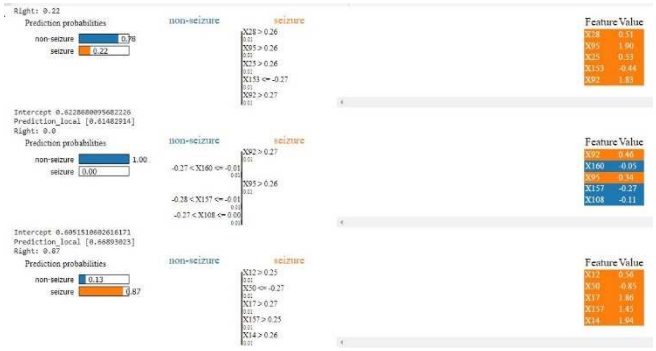

Figure 7: Lime Interpretation for the first three samples where we only show the top five features that have more impact on this classification.

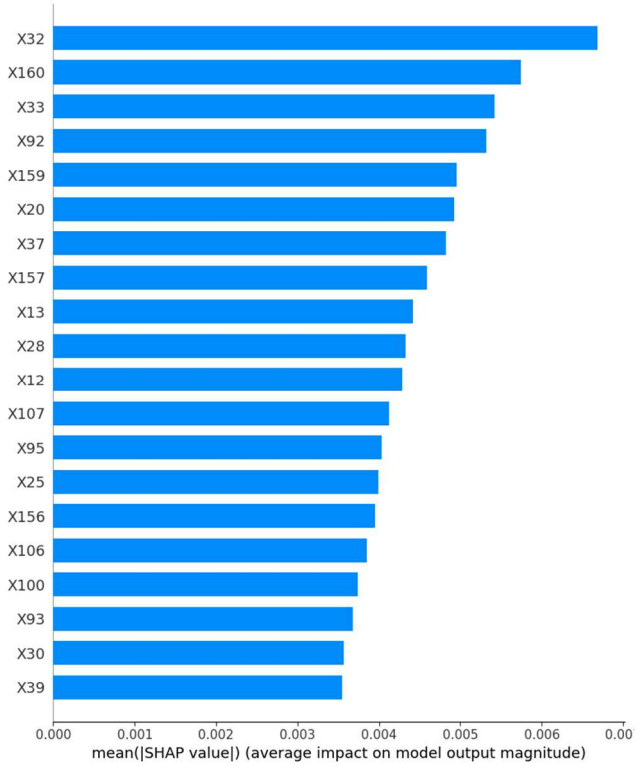

Figure 8: Mean (|SHAP value|) (average impact on model output magnitude)

The SHAP value output in Figure 8 shows the top 20 features and their respective output magnitudes. The y-axis lists the features in descending order of their mean absolute SHAP values. Each vertical bar in the graph represents the impact of a single feature on the model output. The color of the bar indicates whether the feature has a positive or negative impact on the output. Blue bars represent features that negatively affect the output, while red bars represent features that have a positive impact. The length of the bar corresponds to the magnitude of the impact, with longer bars indicating a stronger impact and shorter bars indicating a weaker impact. The most important feature among the top 20 is "x32", which means that this feature has the highest importance in making a prediction. Any changes in the value of this feature will have a significant impact on the output.

V. Conclusion

In conclusion, our research project focused on the classification of epilepsy seizures using machine learning algorithms and explainable AI. We employed a publicly available dataset of EEG recordings from 500 individuals and evaluated various classification algorithms, including Logistic Regression, K-nearest neighbors, Support Vector Machine, Naive Bayes Classifier, Random Forest Classifier, and Gradient Boosting. Our study achieved a high accuracy rate of 96.1% using the Random Forest algorithm, which demonstrates the potential of machine learning algorithms to accurately classify epilepsy seizures.

Furthermore, we utilized SHAP and LIME techniques to gain insights into the decision-making processes of the models and enhance their interpretability. This approach is novel in the context of epilepsy diagnosis and can contribute to the development of more accurate and effective methods for diagnosing and treating epilepsy, a neurological disorder that affects millions of people worldwide.

In summary, our work contributes to the ongoing efforts to improve the diagnosis and treatment of epilepsy using machine learning algorithms and explainable AI techniques. It highlights the potential of these approaches to accurately classify epilepsy seizures and provides valuable insights into the decision-making processes of the models, which can enhance their interpretability and usefulness in clinical settings.

References

[1] Acharya, U. R., et al. (2018). Automated EEG-based screening of depression using deep convolutional neural network. Prog. Neuropsychopharmacol. Biol. Psychiatry, 86, 262-268.

[2] Korhonen, I., et al. (2017). Deep learning for event-related potential classification. PLoS One, 12(4), e0174418.

[3] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

[4] Ribeiro, M. T., et al. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).

[5] Roy, S., et al. (2019). Classification of epileptic seizures using EEG signals: A review. Sensors, 19(11), 2480.

[6] Attia, Z., Rajoub, B., & Al-Betar, M. A real-time ECG classification algorithm based on 1-D convolutional neural networks. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(3), 1115-1123.

[7] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers in Biology and Medicine, 2018, 100, 270-278.

[8] Jiang, Y., Zhao, H., & Wu, X. Patient-specific ECG classification using a multimodal representation learning approach. Computer Methods and Programs in Biomedicine, 2020, 194, 105532.

[9] Supratak, A., Dong, H., Wu, C., & Guo, Y. Deep learning for classification of EEG data in epileptic patients. Proceedings of the

2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, 3052-3056.

[10] Liang, X., Li, Y., & Hu, Y. An improved convolutional neural network for epileptic seizure classification using EEG signals. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(7), 2959-2968.

[11] Subasi, A. EEG signal classification using empirical mode decomposition and artificial neural networks. Expert Systems with Applications, 2007, 32(4), 1084-1093.

[12] Sun, H., Li, X., Zhang, W., & Sun, H. (2017). Epileptic seizure detection using a support vector machine classifier with rough set-based feature selection. Biomedical Signal Processing and Control, 33, 107-115. doi: 10.1016/j.bspc.2016.12.015

[13] Roy, S., & Bhattacharyya, S. P. (2019). Epileptic seizure detection using deep autoencoder and machine learning techniques. Biomedical Signal Processing and Control, 49, 76-88. doi: 10.1016/j.bspc.2018.12.008

[14] Marcílio, W.E. and Eler, D.M., 2020, November. From explanations to feature selection: assessing SHAP values as the feature selection mechanism. In 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 340-347). Ieee.

[15] Visani, G., Bagli, E., Chesani, F., Poluzzi, A. and Capuzzo, D., 2022. Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models. Journal of the Operational Research Society, 73(1), pp.91-101.

[16] K. Mridha, S. Kumbhani, S. Jha, D. Joshi, A. Ghosh, and R. N. Shaw, "Deep Learning Algorithms are Used to Automatically Detection Invasive Ducal Carcinoma in Whole Slide Images," 2021 IEEE 6th International Conference on Computing, Communication, and Automation (ICCCA), Arad, Romania, 2021, pp. 123-129, doi: 10.1109/ICCCA52192.2021.9666302.

[17] K. Mridha, S. Kumbhani, S. Jha, D. Joshi, A. Ghosh, and R. N. Shaw, "Deep Learning Algorithms are Used to Automatically Detection Invasive Ducal Carcinoma in Whole Slide Images," 2021 IEEE 6th International Conference on Computing, Communication, and Automation (ICCCA), Arad, Romania, 2021, pp. 123-129, doi: 10.1109/ICCCA52192.2021.9666302.

[18] K. Mridha, A. P. Pandey, A. Ranpariya, A. Ghosh and R. N. Shaw, "Web Based Brain Tumor Detection using Neural Network," 2021 IEEE 6th International Conference on Computing, Communication, and Automation (ICCCA), Arad, Romania, 2021, pp. 137-143, doi: 10.1109/ICCCA52192.2021.9666248.