

Optimizing Epileptic Seizure Recognition with Machine Learning Algorithms

Sanagavarupu Sunitha¹, Dr. Umadevi Ramamoorthy²

¹Research Scholar, SSCS, CMR University, Bengaluru, Karnataka, India

Emails: sanagavarupu.sunitha@cmr.edu.in

ORCID ID- 0000000311672112

²Associate Professor, SSCS, CMR University, Bengaluru, Karnataka, India

Emails: umadevi.r@cmr.edu.in

ARTICLE INFO

Received: 10 Oct 2024

Revised: 06 Dec 2024

Accepted: 20 Dec 2024

ABSTRACT

According to the WHO, Epilepsy is a significant public health issue and increases every year from 1% to 2% in all age groups. It is one of the oldest recognized neurological disorders. Early detection and proper medication reduce the risk to the person. EEG is one of the methods to identify epilepsy, The continuous monitoring of EEG signals recognizes seizures. These occur in the partial or total body or several parts of a brain for a person and it causes unconsciousness. A person who suffers from any one of the following health problems: high fever, sleepless nights, anxiety, and stress might cause epileptic seizures. Surviving with epilepsy is stressful and limited to employability. This work aims to build the best model using Machine Learning algorithms with high performance and accuracy values, by using the conventional Machine Learning algorithms like K- Nearest Neighbor, Support Vector Classifiers, Support Vector Regression, Lasso, Ridge, Decision Tree, Gradient Boost, eXtreme Gradient Boosting, Light Gradient Boosting Machine, Categorical Boosting, and Linear Regression. Optuna used for tuning hyperparameters in the ML models to improve the performance of a model and to obtain best results. The Kaggle data set used for training and validation purposes. In these models, according to classifiers models, all the "gradient boost" classifiers produce accuracy, precision, recall and F1-score with 1.0 values. In Regression models, the best model is "Linear Regression" with R2 score as 1.0 to detect epileptic seizures.

Keywords: EEG, Epilepsy, Seizures, ML Classifiers models, Regression models, Boosting Algorithms

INTRODUCTION

Epilepsy is a common disorder of brain signals. Around 0.625% of the people in the world effected by epilepsy and it causes seizures. Seizures occur through genetic problems or due to health issues or food habits. If seizures cannot identify within a stipulated time, he or she might be the last consciousness and may go through paralysis and severe seizures cause death also. So many researchers are working on this area to identify epileptic seizures of your patient or person.

EEG is one of the methods for detecting seizures by recording the electrical signals of the brain. Those signals information converted as EEG dataset to develop an accurate model by using machine learning algorithms and various techniques to analyse easy data to detect seizures.

ML Classifiers and Statistical features to analyse normal brain activity and scissor activity. These algorithms process substantial amounts of EEG datasets to recognise this.

Many people have gone through this domain to reach validation accuracy and performance metrics with perfect values. Those models have pitfalls. Gradient Boost algorithms show best results to identify epileptic seizure detection through classifiers and Linear regression through regressors for the Kaggle EEG dataset.

RELATED WORK

Wide research has done in the field of Epileptic Seizures detection but required levels of accuracy has never accomplished.

In this paper, K-Nearest Neighbor (KNN) and Fuzzy Rough Nearest Neighbor (FRNN) give the highest classification accuracy scores, with improved sensitivity and specificity percentages on the Bonn and Children's Hospital of Boston-Massachusetts Institute of Technology (CHB-MIT) datasets [1]. A deep learning novel model called Bi-LSTM with 98% of performance metrics worked on Children's Hospital of Boston-Massachusetts Institute of Technology (CHB-MIT) datasets from physio net [2].

In this paper, authors studied epilepsy research related work from 2005 to 2013 by using R programme language for Biblio Metrix analysis & SCIMAT and Vosviewer for data analysis and visualization [3]. DNN algorithms which are better and how CNN and sparse autoencoders are giving best scores to predict cancer [4]. A deep learning model 1D-CNN is a model to improve the performance of the classifier on a given data set from upf.edu ntsa [5]. This paper used to evaluate egg assemblages correctly by using taphonomic, sedimentological and geochemical analysis [6].

In this paper, Authors used EMD (empirical mode) with feature extraction to identify seizure detection on the CHB – MIT database [7]. In this paper, Authors explain how to convert age signals into images, and they proposed an alternative architecture on pre-trained data set and easily identified epilepsy [8]. This paper proposed a system with feature extraction and classification of the Taylor Fourier rhythm specific models and the filter bank for processing and classification SVM used with accuracy 94.88 authors worked on database Born University [9].

This paper, study detection and analysis of lifestyle-based diseases in Nelly phases of life how it applied on lifestyle of a person with the help of surveys [10]. The Authors developed the multimodal fusion approach with deep learning-based features to increase the accuracy of epilepsy diagnosis [11]. Authors developed CNN and Gaussian filters used to sort the skin cancer with accuracy [12]. This paper describes two models, one is ensemble and second one is Choquet fuzzy integral with deep neural network in the system [13].

In this paper Authors explain how classification of Elite scissor data set using different machine learning algorithms as well as PCA feature reduction technique [14]. Authors presented a review on eclectic scissors prediction by using machine learning and deep learning approaches [15]. A novel model for epilepsy detection for binary and multi class classifiers based on multitaper spectral features developed by authors [16]. Authors proposed 1CNN,2CNN,3CNN & 4CNN: 1CNN and 2CNN to identify epileptic seizures automatically, 3CNN and 4CNN predicted with the accuracy value of 95% [17].

Author proposed a CNN and transfer learning based eight class-type for EEG based Multiclass seizure [[18]. In this paper, Authors developed a multi featured learning model for epilepsy classification supervised by a highly robust heterogenous deep symbol [19]. In this paper, authors explained students' mentality using Generative AI tools [20]. The Author describes how ML and DL algorithms or techniques used on different data sets to produce a predictive model [21]. In this paper authors proposed discrete wavelet transform based singular value decomposition for K nearest neighbour classified technique to identify epileptic seizure [22].

In this paper, Authors used ensemble empirical mode decomposition and least squares support vector nation for classification of epileptic seizures [23]. Authors explain the machine learning methods to predict epileptic seizure with sensitivity [24]. In this paper, a novel independent RNN approach for classification of seizures and non-seizures [25]. In this study CNN proposed to identify epileptic seizures based on EEG signals [26]. Machine Learning has applied on the previous seizure activity values but also on other features like eye state and brain tumours to detect seizures in more efficient manner [27].

PROPOSED METHODS

The Kaggle dataset is pre-processed, divided into Trained and Tested data sets. Apply ML algorithms on trained dataset and next apply on tested dataset. The best model evaluated with different parameters like accuracy, precision, recall and F1-score, MSE, R² Score and RMSE.

11495	X22.V1.114	-22	-22	-23	-26	-36	-1	-18	-37	-47	-48	2
							...						
11496	X19.V1.354	-47	-11	28	77	141	27	48	77	117	170	1
							...						
11497	X8.V1.28	14	6	-13	-16	10	-67	-30	-2	-1	-8	5
							...						
11498	X10.V1.932	-40	-25	-9	-12	-2	116	86	68	59	55	3
							...						
11499	X16.V1.210	29	41	57	72	74	5	4	-2	2	20	4
							...						

[11500 rows x 180 columns]

Data Preprocessing

while recording EEG Signals of a patient, those divided into five classes.

- 5 - eyes open,
- 4 - eyes closed,
- 3 - Yes, they identify the tumour region from the healthy brain area
- 2 - EEG data points area about the location of the tumour
- 1 - Recording of seizure activity

Here 5,4,3,2 classes are not for identifying epileptic seizures. Only class 1 required to detect seizures. By using binary encoding divide all the classes as 0 and 1.

ML model

The following machine learning algorithms applied on the dataset to get the best model. Among these KNN, RF, SVC, DT, SVR, LR, Linear Regression, Lasso, Ridge, GB, XGBoost, LightGBM, CatBoost and LR, Random Forest and Support vector Classifier selected and XGBoost algorithm selected with Optuna.

RESULTS & DISCUSSIONS

This model divides the dataset into 80% as trained and 20% as test data. It is suitable for trained datasets and no overfitting on unseen test data. Hyperparameters used to optimize the model performance and control the complexity and regularization. Feature importances used to help which features prominently work on model prediction. There is no overfitting.

CLASSIFICATION MODELS

The performance of various classification models used in predicting Epileptic Seizures has evaluated based on key metrics such as F1-score, Accuracy, Precision and Recall. These metrics provide a comprehensive understanding of how well each model fits the data and its predictive accuracy. The results are as follows in table 2 and shown in figure 2.

Table 2: Classification models performance metrics

Model Name	Datase t	Accurac y	Precisio n	Recall	F1-Score	TN	F P	FN	TP
Linear Regression	Trained	1	1	1	1	7365	0	0	1835
	Tested	1	1	1	1	1835	0	0	465
Random	Trained	1	1	1	1	7365	0	0	1835

Forest	Tested	0.98	0.97	0.95	0.96	1823	12	23	442
Support Vector	Trained	1	1	1	1	7365	0	0	1835
	Tested	1	0.98	1	0.99	1827	8	1	464
Decision Tree	Trained	1	1	1	1	7365	0	0	1835
	Tested	0.98	0.94	0.94	0.94	1809	26	28	437
k - Nearest Neighbor	Trained	0.95	1	0.74	0.85	7361	4	476	1359
	Tested	0.93	0.99	0.68	0.8	1833	2	151	314
Gradient Boost	Trained	1	1	1	1	7365	0	0	1835
	Tested	1	1	1	1	1835	0	0	465
XGradient Boost	Trained	1	1	1	1	7365	0	0	1835
	Tested	1	1	1	1	1835	0	0	465
Light Gradient	Trained	1	1	1	1	7365	0	0	1835
	Tested	1	1	1	1	1835	0	0	465
CatBoost	Trained	1	1	1	1	7365	0	0	1835
	Tested	1	1	1	1	1835	0	0	465

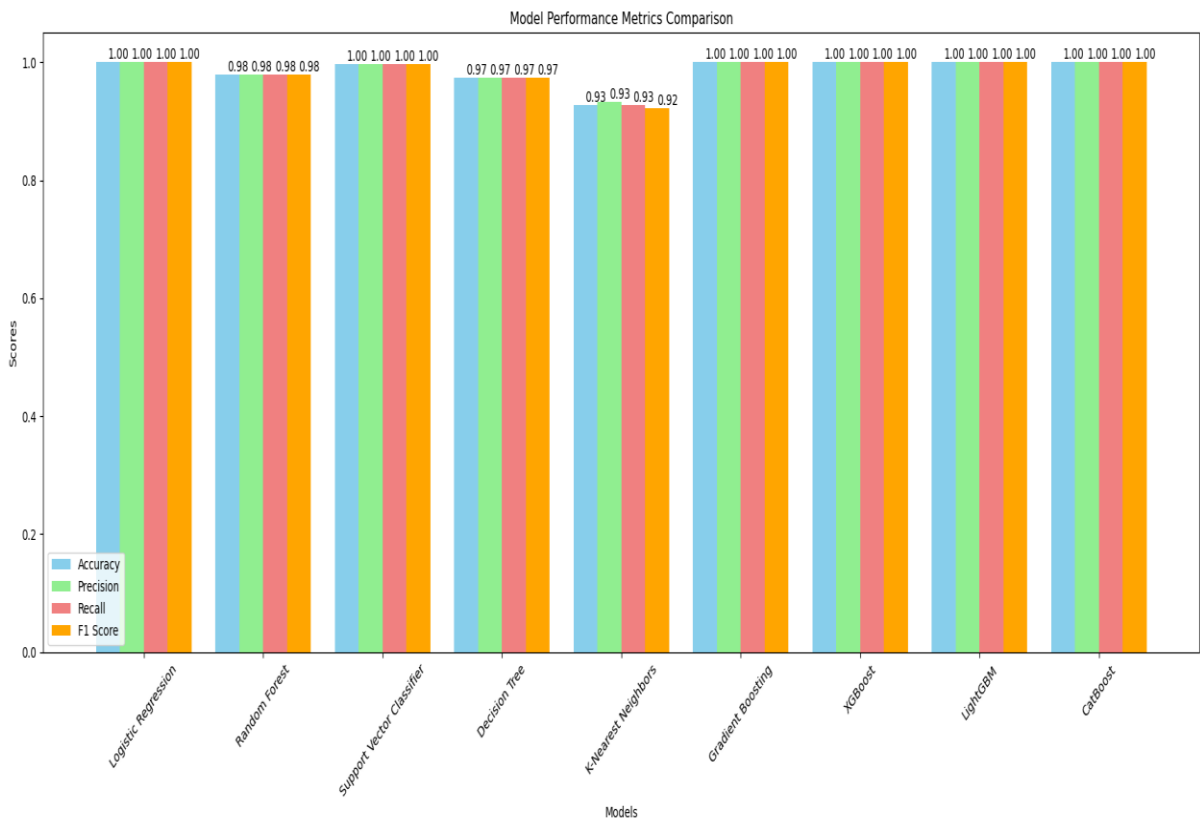


Fig. 2 Comparative chart of Classification models performance metrics

All the classification models like LR, SVC, RF, DT, KNN, GB, XGBoost, LightGBM and CatBoost perform well on trained datasets with all the metric values as 1 (One) except KNN.

LR and Gradient Boost all models performed very well on the test dataset also with perfect scores as 1 for all the metric values. The confusion matrices as shown in Fig 3, Fig 4a, Fig 4b, Fig4c and Fig 4d.

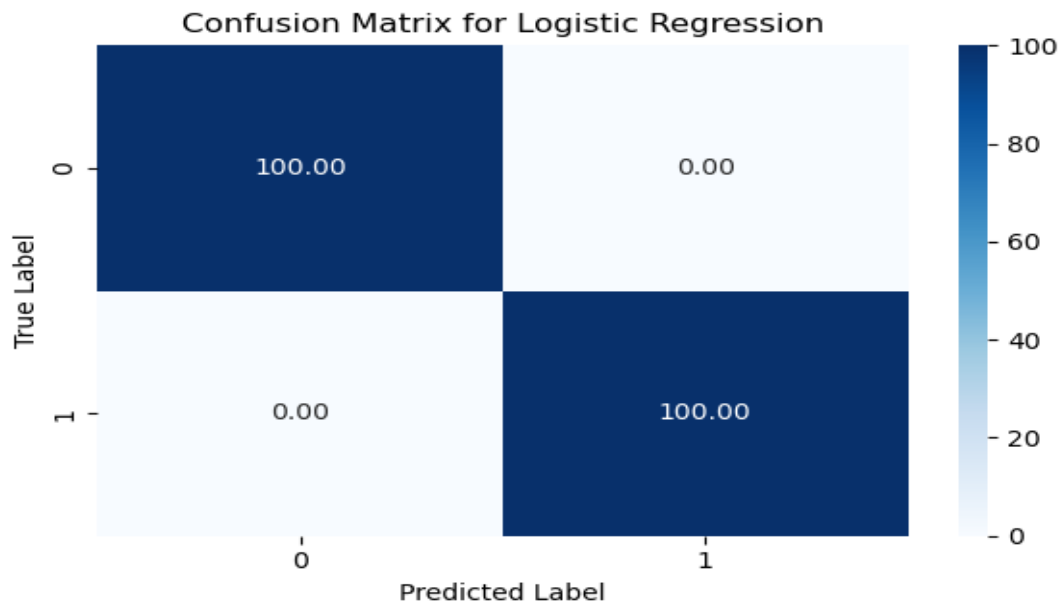


Fig 3. Logistic regression confusion matrix

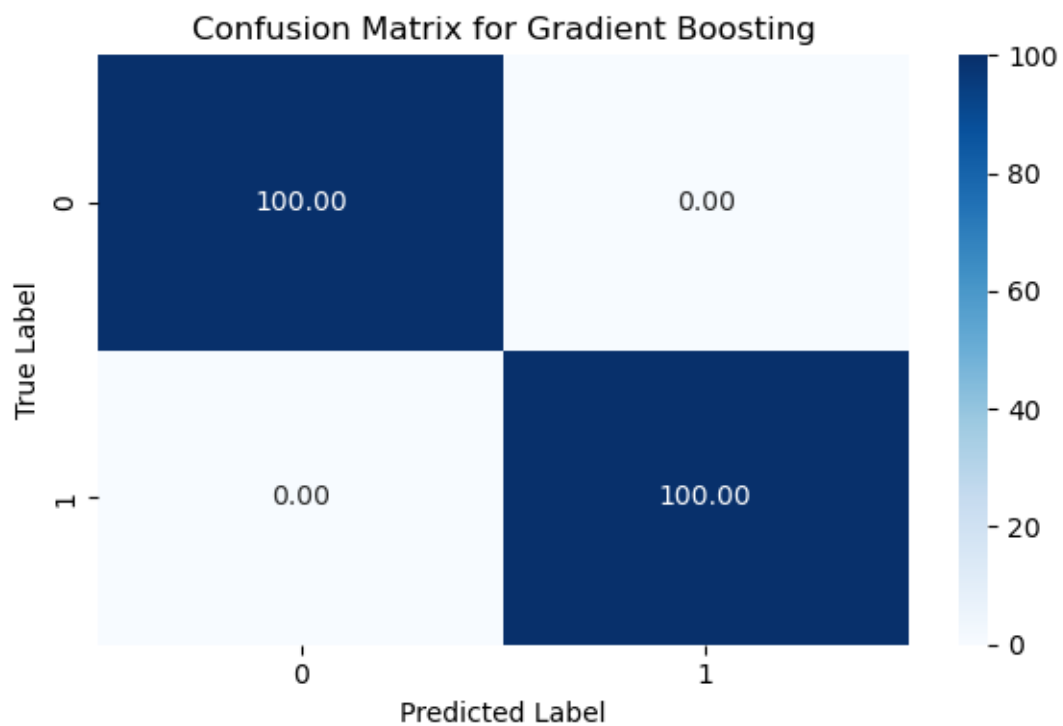


Fig 4a. Gradient Boosting confusion matrix

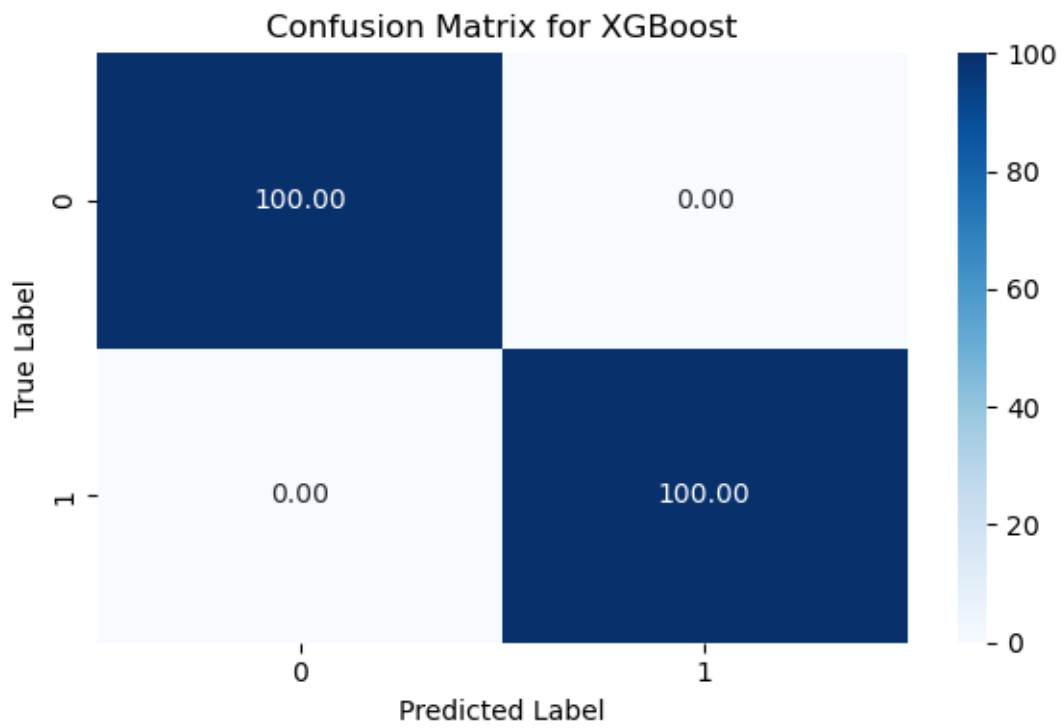


Fig 4b. XGradient Boost confusion matrix.

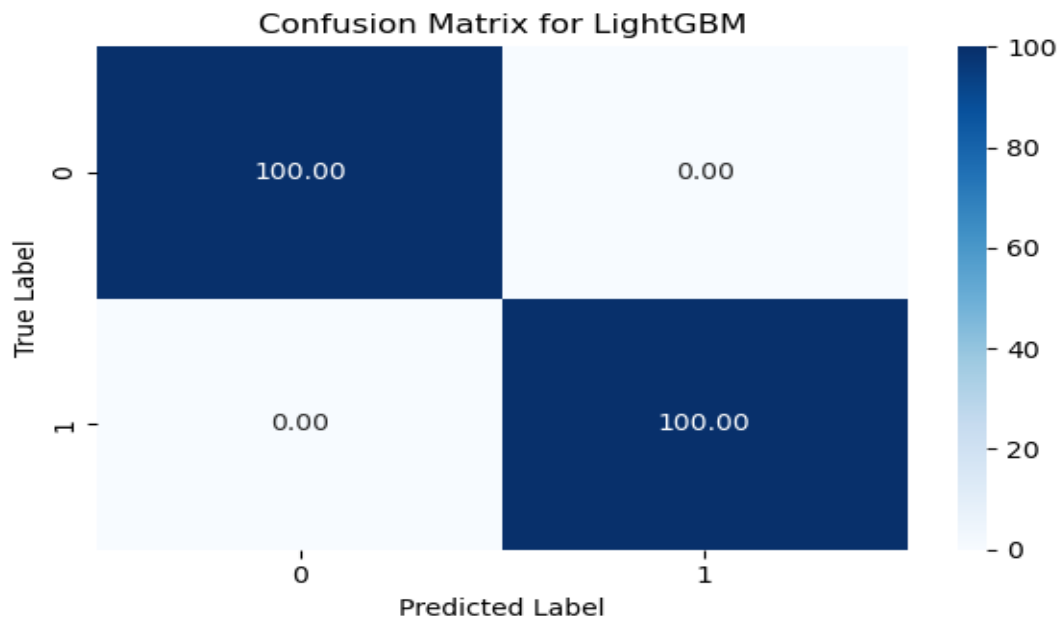


Fig 4c. LightGBM confusion matrix

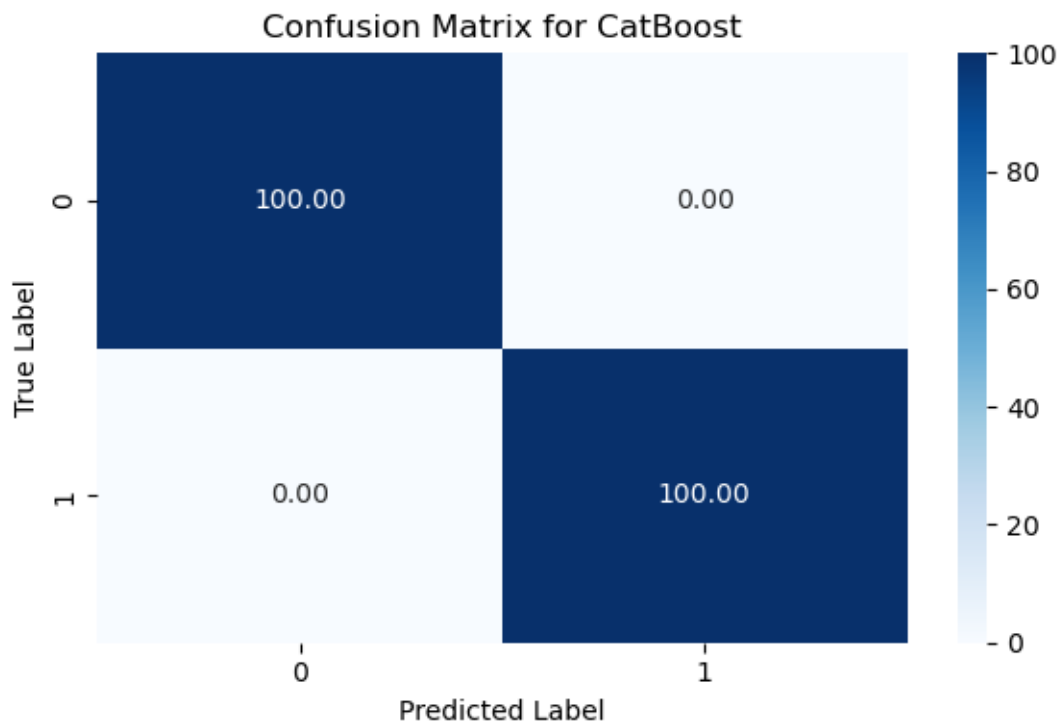


Fig 4d. CatBoost confusion matrix

RF model has a slight drop in all metric values depending on the FP and FN values shown in Fig 5.

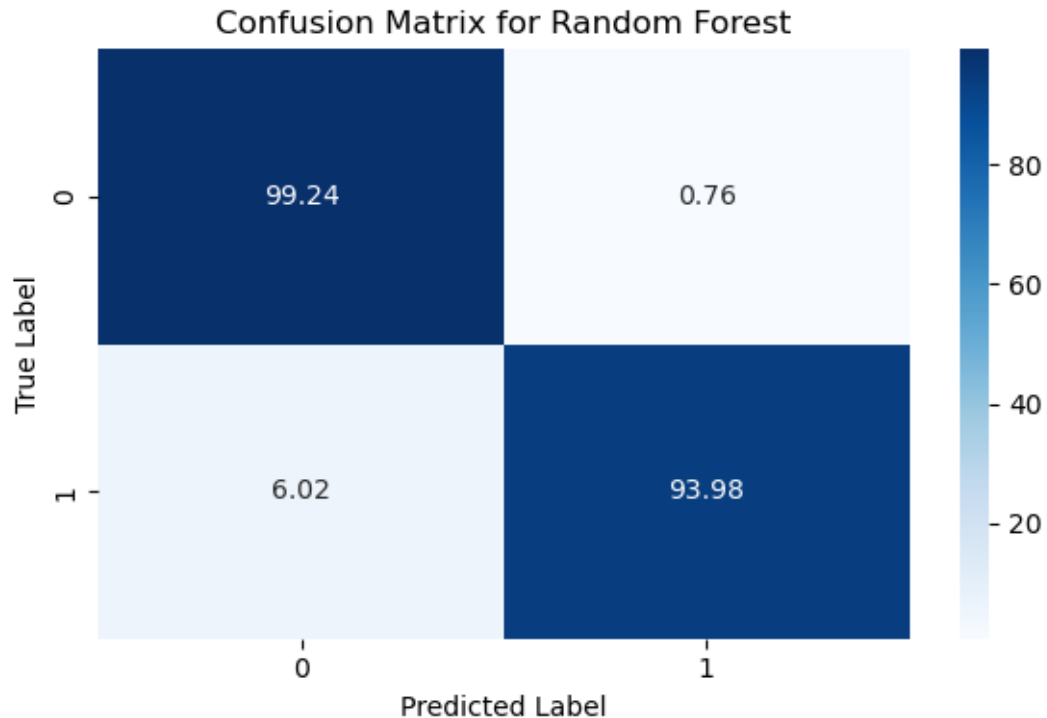


Fig 5. Random Forest confusion matrix

SVC performs excellently but there is a slight drop in precision shown in Fig. 6.

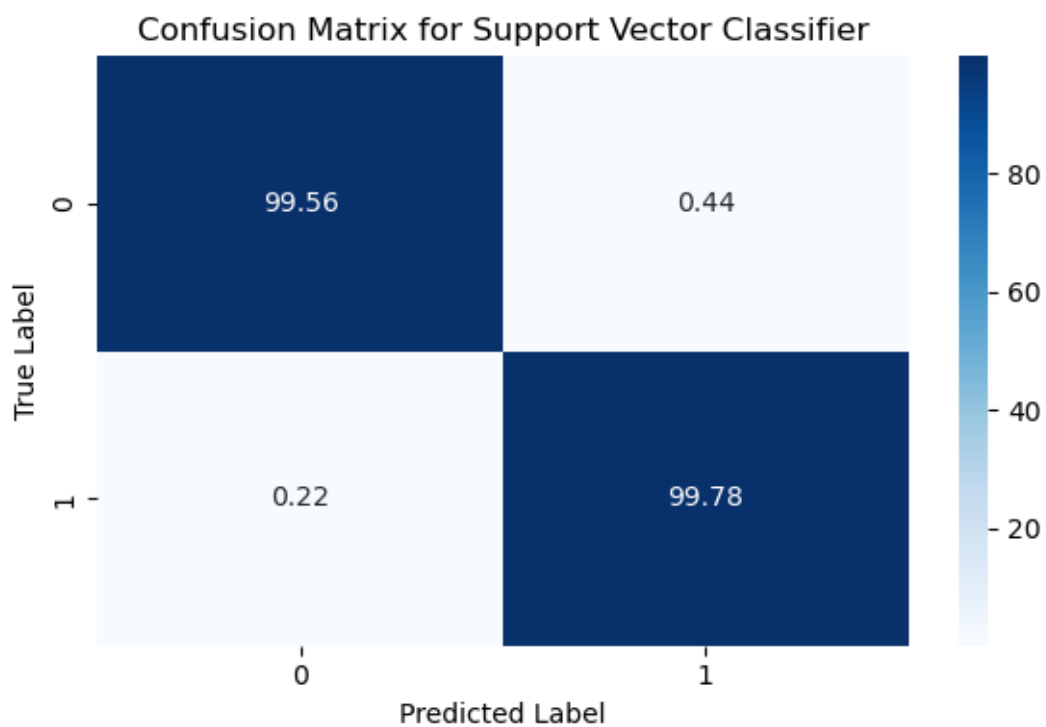


Fig 6. Support Vector confusion matrix.

The DT model performs well on the test dataset not as compared to other models, but it works in Fig 7.

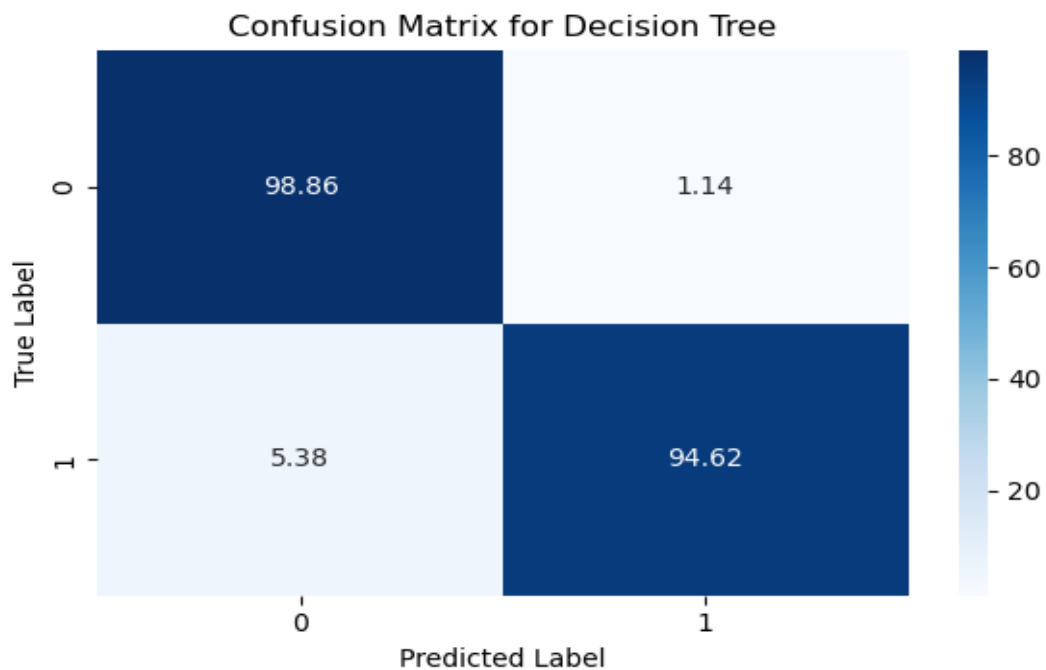


Fig 7. Decision Tree confusion matrix

KNN model indicates poor performer model as shown in Fig 8.

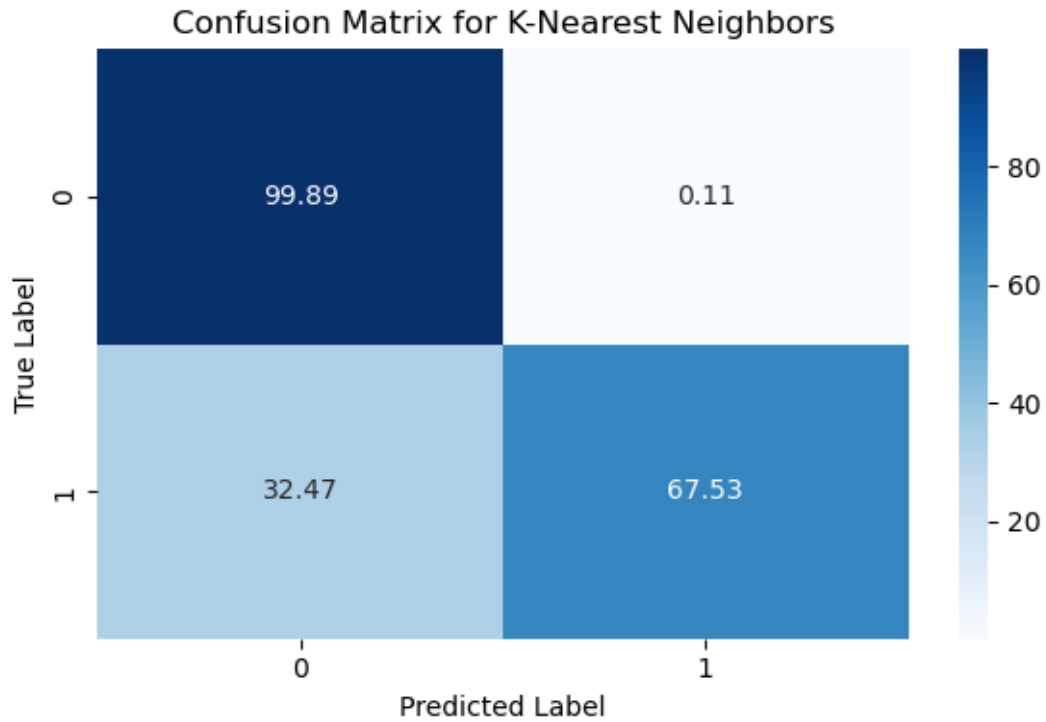


Fig 8. K-Nearest Neighbors confusion matrix

REGRESSION MODELS

The performance of various regression models used in predicting Epileptic Seizures. Evaluation based on key metrics such as the R² score, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics provide a comprehensive understanding of how well each model fits the data and its predictive accuracy. The results are as follows in table 3 and shown in figure 3.

Table 3: Regression model performance metrics

MODEL	R SQUARE	MEAN SQUARE ERROR	ROOT MEAN SQUARE ERROR
Linear	1.0	7.863388758529942e-31	8.86757506792581e-16
Ridge	0.9999996414328779	5.799132199180844e-08	0.00024081387416801474
Lasso	-0.00011572299386974905	0.16174944480262124	0.40218086081093074
ElasticNet	-0.00011572299386974905	0.16174944480262124	0.40218086081093074
Decision Tree	0.8387012987012987	0.02608695652173913	0.15314016588874438

Random Forest	0.943084531286895	0.009204980237154152	0.09493580106635574
Gradient Boosting	0.9222801524906472	0.012569687582903088	0.112124802333707
Support Vector	0.9464246186008695	0.008664785481480884	0.09308482949160343
K-Nearest Neighbors	0.6568736717827627	0.055494071146245064	0.23557196160417

Linear Regression has the highest R^2 score, indicating it predicts accurate detection of epilepsy. It describes adaptability in the data. The low MSE and RMSE values show minimal error in predictions. Linear Regression is the most effective model among those evaluated.

Ridge model performs well with an R^2 score of 0.99, suits to identify epileptic seizure detection on the dataset. The MSE and RMSE values are less indicates model predictions are close to the actual values.

Lasso and Elastic Net models got the same MSE and R^2 with negative values. These models not fit for this dataset to detect Epileptic seizures.

Decision Tree Regression has a slope R^2 score indicating that the model fits up to 83.8% of the data. The RMSE and MSE values indicate few errors in prediction. The overall performance of this model is reasonable.

Random Forest Regression shows better performance with an R^2 score of 94.4% of the variability. The MSE and RMSE values are highly effective to predict accuracy.

Gradient Boosting Regression has a high R^2 score of 92.2%, indicating that high variance in the prediction. The low MSE and RMSE values indicate good predictive performance. [3]

SVR also performs well, with a high R^2 score of 94.64% explains a massive portion of the variance in prediction of epileptic seizures. However, its MSE and RMSE are low measures accurate and predictable.

KNN Regression model with a R^2 score of 65.7% of the variability on the dataset. The MSE and RMSE values are high indicating model predictions are less proportionate to the actual values. [2]

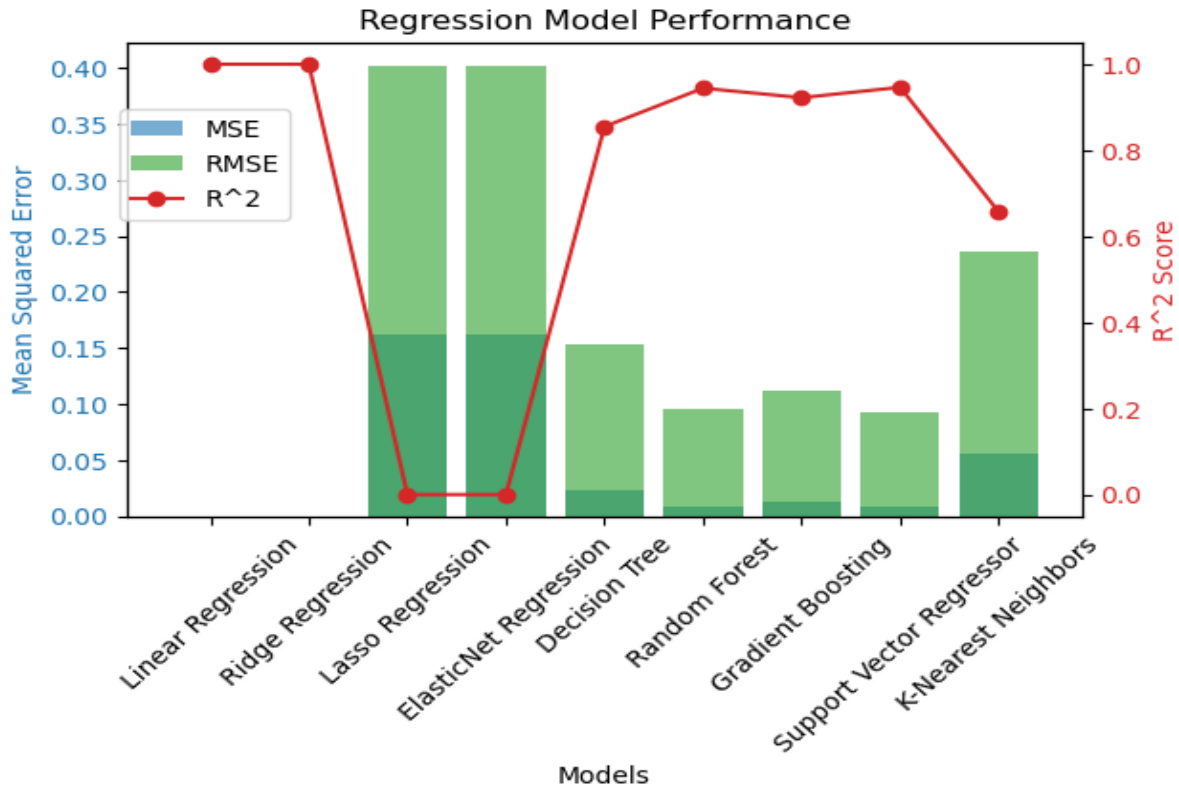


Fig. 9 Comparative chart of Regression models performance metrics

Best hyperparameters

Create the Optuna study and optimize the objective function to get best hyper parameters.

```
{'max_depth': 8, 'learning_rate': 0.04963543801829856, 'n_estimators': 457, 'subsample': 0.8621557245253744, 'colsample_bytree': 0.942678490847755, 'gamma': 0.01384987069049426, 'min_child_weight': 3, 'reg_alpha': 0.00461845318465644, 'reg_lambda': 0.00989699528229048}[1]
```

Feature Importances in the sorted Order

Train XGBoost model to get feature importances and it tells you in which way the given features used to predict the objective. The results in Table 4 and graphical representation in Fig

Table 4: Feature Importances

Feature	Value
X25	0.038312
X159	0.03824
X28	0.035563
X157	0.032701
X17	0.02719
X117	0.000158
X150	0.00015
X87	0.000108
X118	0.000102
X149	0.000075

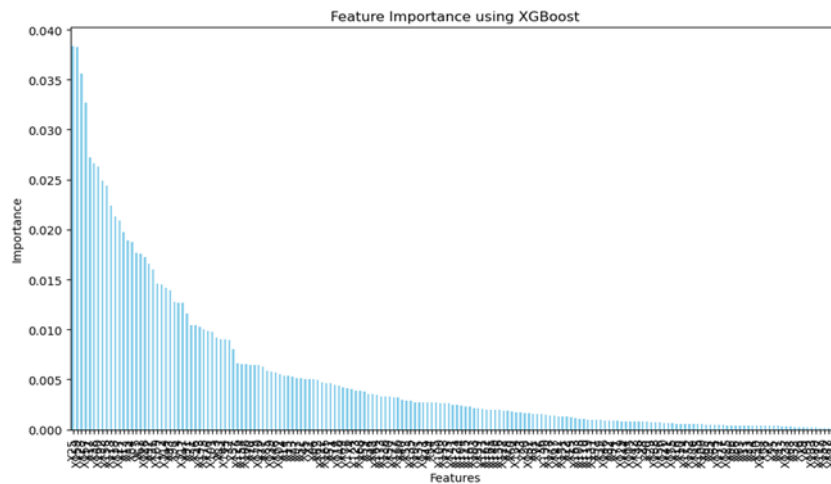


Fig 10: Feature Importance using XGBoost

CONCLUSIONS

All the models went through the dataset of the five hundred persons under healthy and unhealthy categories. These models divided the healthy person's brain EEG signal and seizure patient brain EEG signal, the system classifies the signal data with the XGBoost, CatBoost and LightGBM a validation accuracy of 100% as shown in Table 2, The common machine learning algorithms logistic regression, SVM and KNN achieve good accuracy but not work fine in classification. In conclusion, all the 3 proposed models work better as compared to other models used in this study. According to regression models, Linear regression, Ridge Regression and SVR with high R^2 score 1.0 and low MSE and RMSE. Lasso and Elastic Regression with higher errors and less R^2 score. The remaining algorithms are moderate to detect epileptic seizures on the taken dataset.

These algorithms must apply on the other EEG datasets which are available through the net like Oasis datasets, open Neuro, CHB-MIT, TUH EEG Corpus to find out the performance metrics of the datasets.

Availability of Data and Materials

Data is available in Kaggle.

REFERENCES

- [1]. Aaysha, Z., Qureshi, M., Afzaal, M., Qureshi, M., & Fayaz, M. (2021). Machine learning-based EEG signals classification model for epileptic seizure detection. *Multimed Tools Appl.*, 80. <https://doi.org/10.1007/s11042-021-10597-6>
- [2]. Abdelhameed, A., & Bayoumi, M. (2021). A deep learning approach for automatic seizure detection in children with epilepsy. *Front Comput Neurosci*, 15. <https://doi.org/10.3389/fncom.2021.650050>
- [3]. Manisha Sharma, Satyajit Anand, & Rajeev Pourush (2023). Landscape of epilepsy research: Analysis and future trajectory. <https://doi.org/10.1016/j.inat.2023.101879>
- [4]. Chakrapani, I. S., Tyagi, N., Tyagi, S., Kunekar, P., Padmaja, D. L., & Pant, K. (2022). Applications of Deep Learning (DL) techniques in detecting breast cancer and malignant cells. In *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*. <https://doi.org/10.1109/IC3I56241.2022.10072720>
- [5]. Chirasani, S. K. R., & Manikandan, S. (2022). A deep neural network for the classification of epileptic seizures using hierarchical attention mechanisms. *Appl Soft Comput*, 26. <https://doi.org/10.1007/s00500-022-07122-8>
- [6]. Ezquerro, L., Coimbra, R., Bauluz, B., Núñez-Lahuerta, C., Román-Berdiel, T., & Moreno-Azanza, M. (2024). Large dinosaur egg accumulations and their significance for understanding nesting behaviour. *Geoscience Frontiers*, 15(5). <https://doi.org/10.1016/j.gsf.2024.101872>
- [7]. George, F. (2022). Epileptic seizure prediction using EEG images. In *2020 International Conference on Communication and Signal Processing (ICCSP)*.

- [8]. Jiang, Y., Yao, L. u., & Yang, L. (2022). An epileptic seizure prediction model based on a time-wise attention simulation module and a pretrained ResNet. *Methods*, 202. <https://doi.org/10.1016/j.ymeth.2021.07.006>
- [9]. Kumar Boddu, R. S., Chakravarthi, D. S., Venkateswararao, N., Chakravarthy, D. S. K., Devarajan, A., & Kunekar, P. R. (2021). The effects of artificial intelligence and medical technology on the life of humans. *J Pharm Res Int*, 33. <https://doi.org/10.9734/jpri/2021/v33i50A33378>
- [10]. Kunekar, P. R., Gupta, M., & Agarwal, B. (2019). Detection and analysis of lifestyle based diseases in the early phase of life: a survey. In A. Somani, S. Ramakrishna, A. Chaudhary, C. Choudhary, & B. Agarwal (Eds.), *Emerging technologies in computer engineering: microservices in big data analytics. ICETCE 2019. Communications in Computer and Information Science*. Springer. https://doi.org/10.1007/978-981-13-8300-7_6
- [11]. Kunekar, P. R., Gupta, M., & Agarwal, B. (2020). Deep learning with multi modal ensemble fusion for epilepsy diagnosis. In *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*. <https://doi.org/10.1109/ICETCE48199.2020.9091742>
- [12]. Kunekar, P., Vaishnav, R., Kalaivani, E., Gangodkar, D., Kaur, C., & Dhanraj, J. A. (2022). Applications of machine learning techniques in detecting skin cancer. In *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*. <https://doi.org/10.1109/IC3I56241.2022.10072834>
- [13]. Ludwig, S. A. (2020). Multi-label classification for epileptic seizure recognition: deep neural network ensemble versus Choquet fuzzy integral fusion. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. <https://doi.org/10.1109/SSCI47803.2020.9308541>
- [14]. Nahzat, S., & Yaganoglu, M. (2021). Classification of epileptic seizure dataset using different machine learning algorithms and PCA feature reduction technique. *J Invest Eng Technol*, 4.
- [15]. Natu, M., Bachute, M., Gite, S., Kotecha, K., & Vidyarthi, A. (2022). Review on epileptic seizure prediction: machine learning and deep learning approaches. *Comput Math Methods Med*, 2022. <https://doi.org/10.1155/2022/7751263>
- [16]. Oliva, J. T., & Rosa, J. L. G. (2021). Binary and multiclass classifiers based on multitaper spectral features for epilepsy detection. *Biomed Signal Process Control*, 66. <https://doi.org/10.1016/j.bspc.2021.102469>
- [17]. Ouichka, O., Echioui, A., & Hamam, H. (2022). Deep Learning Models for Predicting Epileptic Seizures Using iEEG Signals. *Electronics*, 11. <https://doi.org/10.3390/electronics11040605>
- [18]. Raghu, S., Sriraam, N., Temel, Y., Rao, S. v, & Kubben, P. L. (2020). EEG based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Netw*, 124. <https://doi.org/10.1016/j.neunet.2020.01.017>
- [19]. Raibag, M. A. A., Franklin, J. v, & Sarkar, R. (2022). Multi-feature learning model for epilepsy classification supervised by a highly robust heterogeneous deep ensemble. *Turk J Comput Math Educ*, 13.
- [20]. Shahzad, M. F., Xu, S., Lim, W. M., Yang, X., & Khan, Q. R. (2024). Artificial intelligence and social media on academic performance and mental well-being: Student perceptions of positive impact in the age of smart learning. *Heliyon*, 10(8). <https://doi.org/10.1016/j.heliyon.2024.e29523>
- [21]. Shoeibi, A., Khodatars, M., Ghassemi, N., Jafari, M., Moridian, P., & Alizadehsani, R. (2021). Epileptic Seizures Detection Using Deep Learning Techniques: A Review. *Int J Environ Res Public Health*, 18. <https://doi.org/10.3390/ijerph18115780>
- [22]. Singh, N., & Dehuri, S. (2020). Multiclass classification of EEG signal for epilepsy detection using DWT based SVD and fuzzy kNN classifier. *Intel Decis Technol*, 14. <https://doi.org/10.3233/IDT-190043>
- [23]. Torse, D. A., & Khanai, R. (2022). Classification of epileptic seizures using ensemble empirical mode decomposition and least squares support vector machines. In *2021 International Conference on Computer Communication and Informatics (ICCCI)*.
- [24]. Usman, S. M., Usman, M., & Fong, S. (2017). Epileptic seizures prediction using machine learning methods. *Comput Math Methods Med*, 2017. <https://doi.org/10.1155/2017/9074759>
- [25]. Yao, X., Cheng, Q., & Zhang, G. Q. (2019). A novel independent RNN approach to classification of seizures against non-seizures.
- [26]. 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. *Front Neuroinform*, 12. <https://doi.org/10.3389/fninf.2018.00095>
- [27]. <https://www.kaggle.com/datasets/harunshimanto/epileptic-seizure-recognition/data>

Acknowledgements

Corresponding author would like to thank her Guide Dr. Umadevi Ramamoorthy, SSCS, CMR University and family members and friends.

Funding

No funding for this work.

Author Information

Sanagavarapu Sunitha: Conceptualization, Investigation, Methodology, Writing and Editing.

Dr. Umadevi Ramamoorthy: Supervision.

Ethics Declarations

No competing interests and no conflicts.

Additional Information

This Scopus is unbiased regarding jurisdictional claims in published maps and institutional affiliations.

Rights & Permissions

Free access.