ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul589

Machine Learning-Based Epileptic Seizure Detection using EEG Data in Zimbabwe

PT Chitenhe[1]; E M Hukuimwe[2]; M Gondo[3]

[1,2,3]Data Science & Analytics & Harare Institute of Technology, Zimbabwe School of Information Science and Technology Harare, Zimbabwe

Publication Date: 2025/07/15

Abstract: Epilepsy is a critical neurological disorder, especially prevalent in developing nations like Zimbabwe, where healthcare infrastructure and diagnostic resources remain limited. Electroencephalograms (EEGs) provide essential clinical insight into seizure activity but require expert interpretation, often leading to delays in diagnosis. This research presents a comparative study of three machine learning models—Random Forest (RF), K-Nearest Neighbours (KNN), and Convolutional Neural Networks (CNN)—for the automated detection of epileptic seizures using a Zimbabwe-specific EEG dataset. The dataset comprises 2,216 EEG segments, each with 667 extracted features. The study investigates each model's effectiveness using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Among all models, RF achieved the highest classification performance with an accuracy of 85.78%, suggesting strong potential for integration into clinical decision-support tools to aid early and reliable epilepsy diagnosis in under-resourced settings.

Keywords: *Epilepsy, EEG, Machine Learning, Random Forest, Convolutional Neural Networks, Seizure Detection, Zimbabwe.*

How to Cite: P T Chitenhe; E Hukuimwe; M Gondo (2025), Machine Learning-Based Epileptic Seizure Detection using EEG Data in Zimbabwe. *International Journal of Innovative Science and Research Technology*, 10(7), 659-665. https://doi.org/10.38124/ijisrt/25jul589

I. INTRODUCTION

Epilepsy is a chronic neurological disorder characterized by recurrent seizures, affecting over 50 million individuals worldwide. Because of the limited access to qualified neurologists and diagnostic tools, the burden is especially high in Sub-Saharan Africa. Significant limitations in Zimbabwe's healthcare system frequently cause important epilepsy diagnoses to be delayed. Although EEGs are frequently used to monitor seizures, manual interpretation requires a lot of time and expertise.

Modern developments in machine learning (ML) and artificial intelligence (AI) provide revolutionary answers to these problems. Faster and more reliable diagnosis is made possible by the training of machine learning algorithms to identify seizure patterns from EEG signals. Using an EEG dataset from Zimbabwe, this study assesses the use of three different machine learning models: CNNs, Random Forest, and K-Nearest Neighbours. The objective is to evaluate their potential for application in resource-constrained environments

II. LITERATURE REVIEW

This study expands on earlier research showing how well machine learning analyses EEG data to detect seizures. The World Health Organization (2019) estimates that over 50 million people worldwide suffer from epilepsy, the majority of whom reside in low- and middle-income nations where access to resources frequently delays diagnosis and treatment. These difficulties are particularly noticeable in Zimbabwe because there aren't enough neurologists or EEG equipment (Mavhunga et al., 2019).

Manually interpreting EEG signals is a time-consuming and subjectively variable process that is used in traditional seizure detection. Shoeb (2009) highlighted that even for skilled professionals, consistent manual interpretation of EEG data is challenging due to its complexity. Machine learning (ML) algorithms that automatically classify EEG signals based on learned patterns have become more and more popular among researchers as a solution to this problem.

In EEG analysis, feature engineering is essential. Power bands that represent brain activity at different frequencies and have been connected to different neurological states include delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz). Acharya et al. (2018), for example, showed that changes in the delta and theta bands are frequently seen during epileptic seizures. Furthermore, brief spikes and shifts in EEG signals—which are indicative of seizures—can be effectively captured by Local Stationary Wavelet Transform (LSWT) features (Subasi, 2007).

Seizures have been used to test a number of machine learning algorithms. Random Forest (RF) classifiers are commended for their capacity to handle high-dimensional

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul589

feature spaces and their resilience to noise. They lessen overfitting by combining predictions from several decision trees (Khan et al., 2021). Despite being straightforward and easy to understand, K-Nearest Neighbours (KNN) frequently performs poorly on intricate EEG datasets because it is sensitive to unimportant features (Abiyev & Abizade, 2016).

Recently, Convolutional Neural Networks (CNNs) have become the cutting edge of EEG classification. CNNs are capable of learning spatial-temporal features without the need for manual feature extraction, as demonstrated by Goyal et al. (2020), who trained CNNs directly on raw EEG signals and obtained accuracies above 90%. Their research supports the notion that, particularly in large and rich datasets, deep learning models can perform better than conventional machine learning techniques.

Nonetheless, a large portion of this research comes from settings with abundant resources. Empirical research utilizing machine learning techniques on EEG data from African populations is conspicuously lacking. Zimbabwe's diagnostic gap was brought to light by Mavhunga et al. (2019), who pointed out that the majority of rural areas do not have access to EEG services. By using machine learning techniques on locally sourced EEG data and assessing how well RF, KNN, and CNN models perform in this setting, this project fills that gap.

The application of features like power bands and wavelet coefficients, a focus on seizure detection, and the use of machine learning algorithms with EEG data were among the inclusion criteria for the cited studies. Excluded studies were those that did not deal with the classification of epileptic seizures or that did not use machine learning.

In conclusion, the literature highlights the importance of automated EEG analysis in minimizing diagnostic delays and supports the use of ML for seizure detection. By confirming these methods on EEG data from Zimbabwe, this study advances the field and shows that machine learning can be successfully modified for healthcare systems with limited resources.

III. METHODOLOGY

➢ *Dataset Description*

The dataset used in this study comprises 2,216 EEG samples recorded from patients in Zimbabwe. Each sample is represented by 667 features including statistical, time-domain, and frequency-domain attributes. The target variable is binary: epileptic (1) or non-epileptic (0).

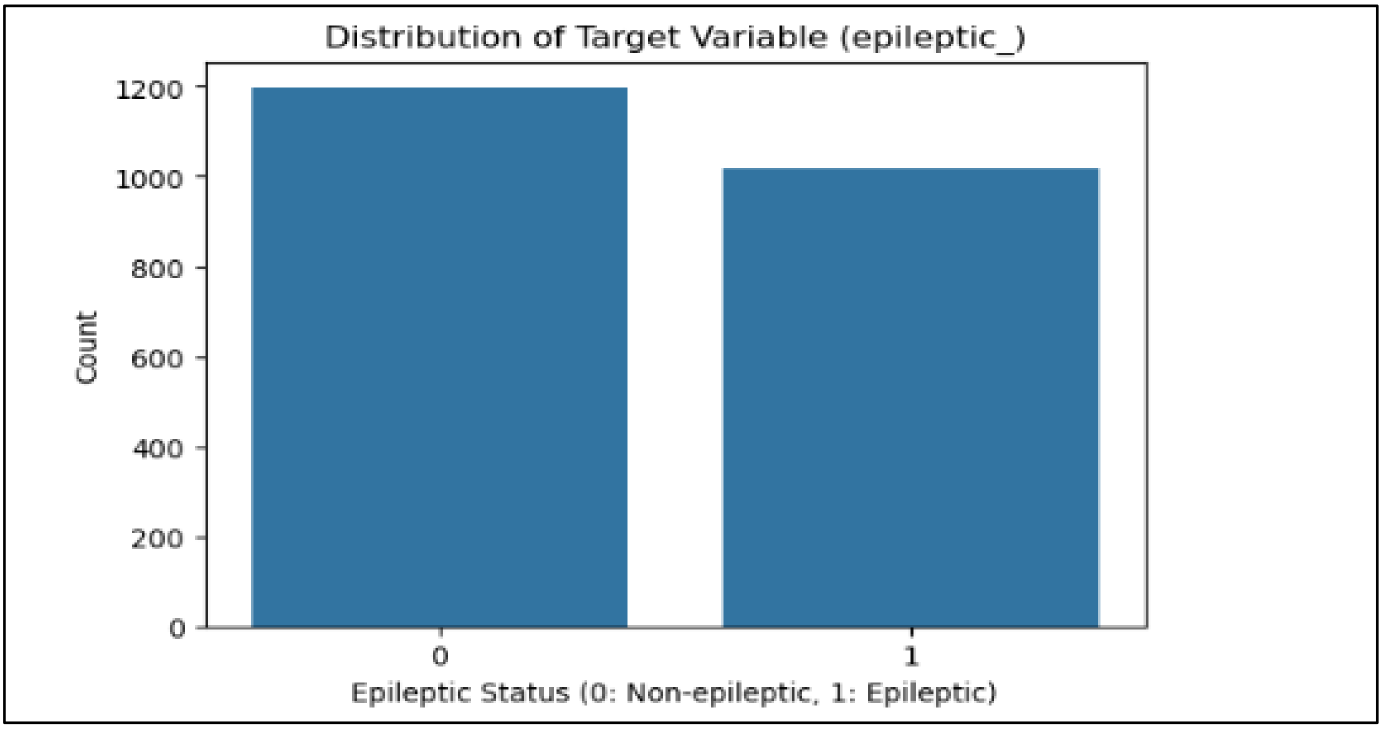

Fig 1 Distribution of Target Variable (Epileptic)

➢ *Feature Engineering :*

Feature extraction and selection are crucial steps for model performance. The dataset includes:

- Power spectral density features: delta, theta, alpha, beta, gamma bands
- Localized Slope Wavelet Transform (LSWT) coefficients: top 50 features capturing time-frequency characteristics
- Statistical measures: mean, standard deviation, absolute mean, skewness, and kurtosis

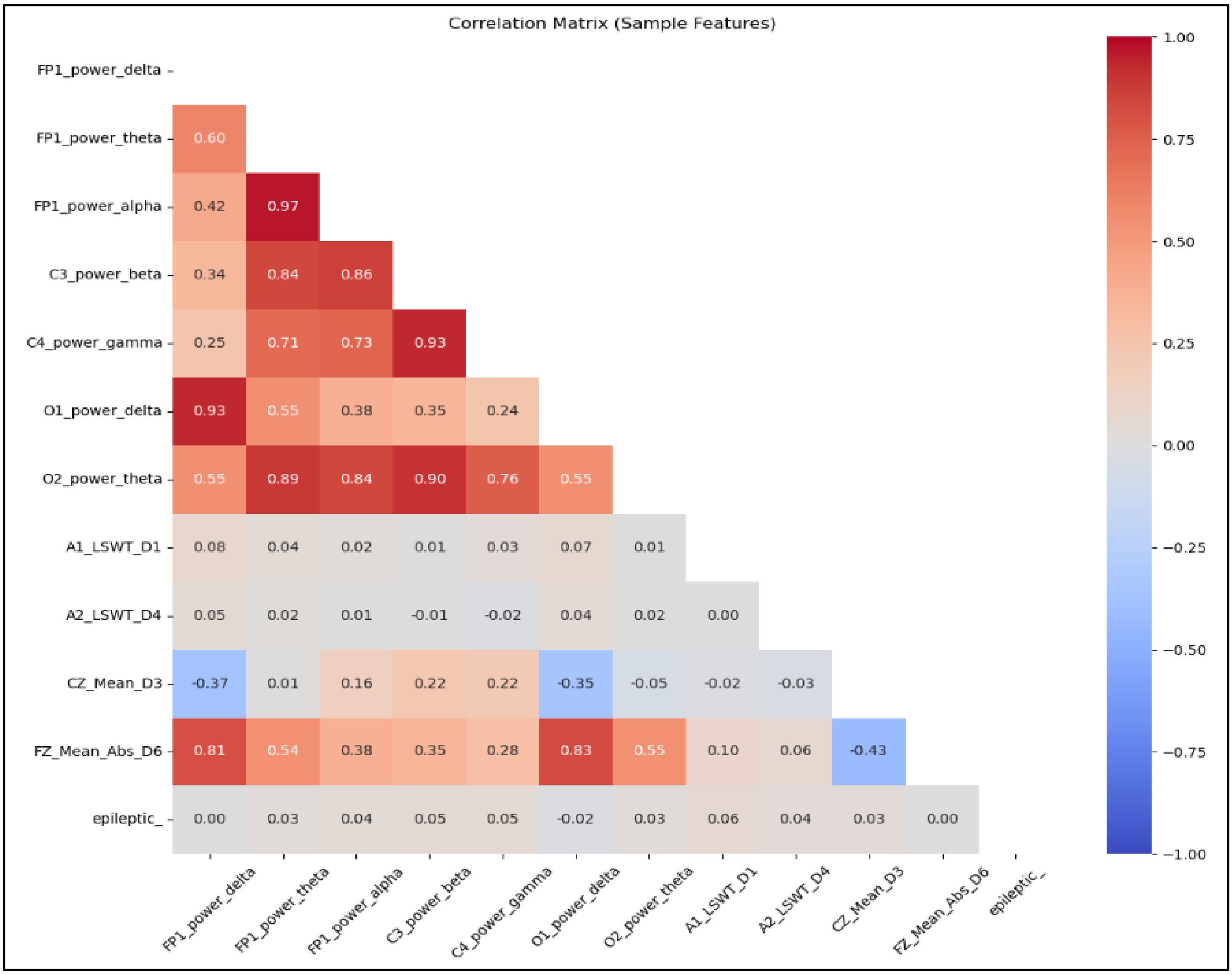

Fig 2 Correlation Matrix (Sample Features)

Dimensionality reduction was achieved using a combination of expert selection and correlation-based filtering, resulting in 185 optimized features.

➢ *Model Development:*

- *Random Forest (RF):*

Configured with 50 decision trees, max depth of 10, and minimum sample split of 5. RF provides feature importance ranking and handles multicollinearity.

- *K-Nearest Neighbours (KNN):*

Used with default k=5 and Euclidean distance metric. No hyperparameter tuning due to its baseline role.

- *Convolutional Neural Network (CNN):*

Implemented as a 1D CNN with multiple Conv1D and MaxPooling layers, ReLU and tanh activations, and dropout regularization. The architecture was trained over 20 epochs.

➢ *Evaluation Strategy A stratified 5-fold cross-validation method ensured balanced training and validation sets. The following evaluation metrics were calculated:*

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC (Receiver Operating Characteristic - Area Under Curve)

IV. RESULTS AND DISCUSSION

➢ *Random Forest:*

- Accuracy: 85.78%
- Precision: 86.86%
- Recall: 81.67%
- F1-score: 84.08%
- ROC-AUC: 0.89

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul589

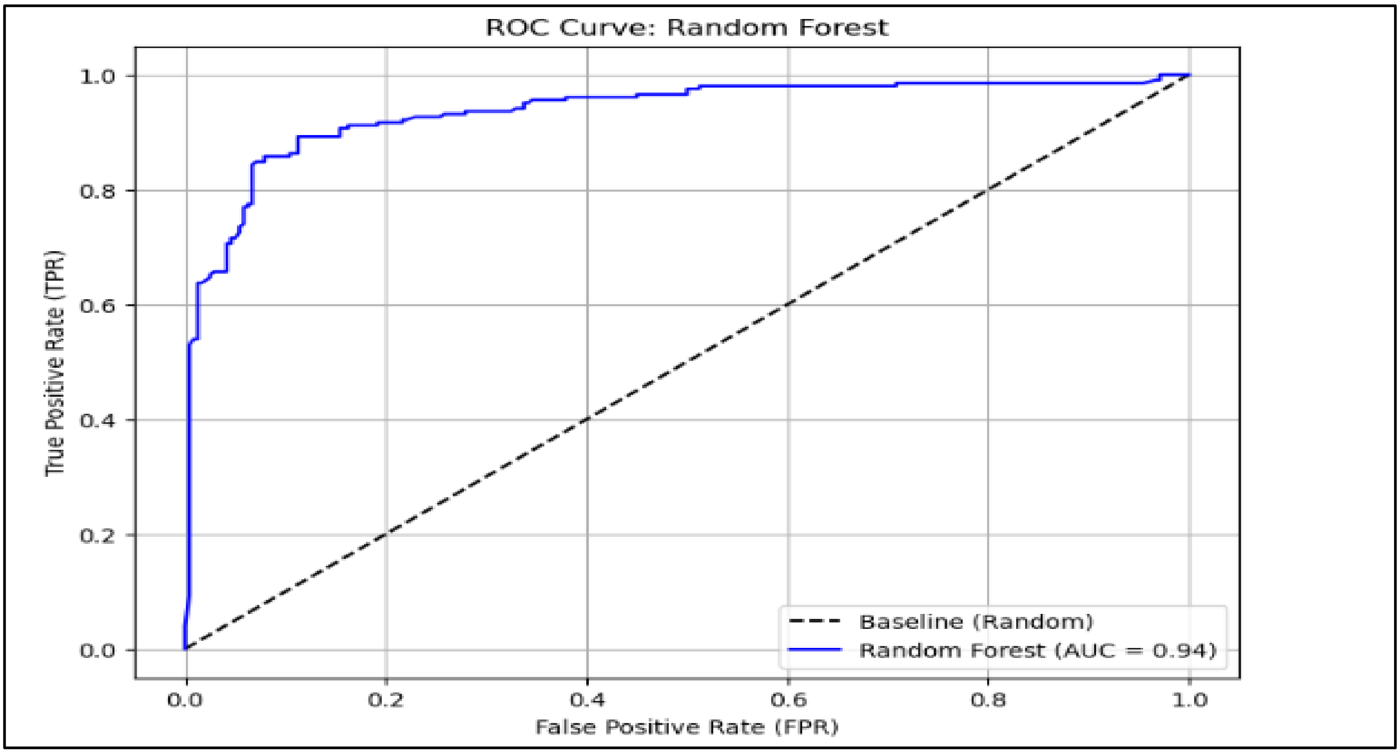

Fig 3 ROC Curve: Random Forest

➢ *CNN:*

- Accuracy: 83.93%
- Precision: 88.93%
- Recall: 74.31%
- F1-score: 80.85%
- ROC-AUC: 0.84

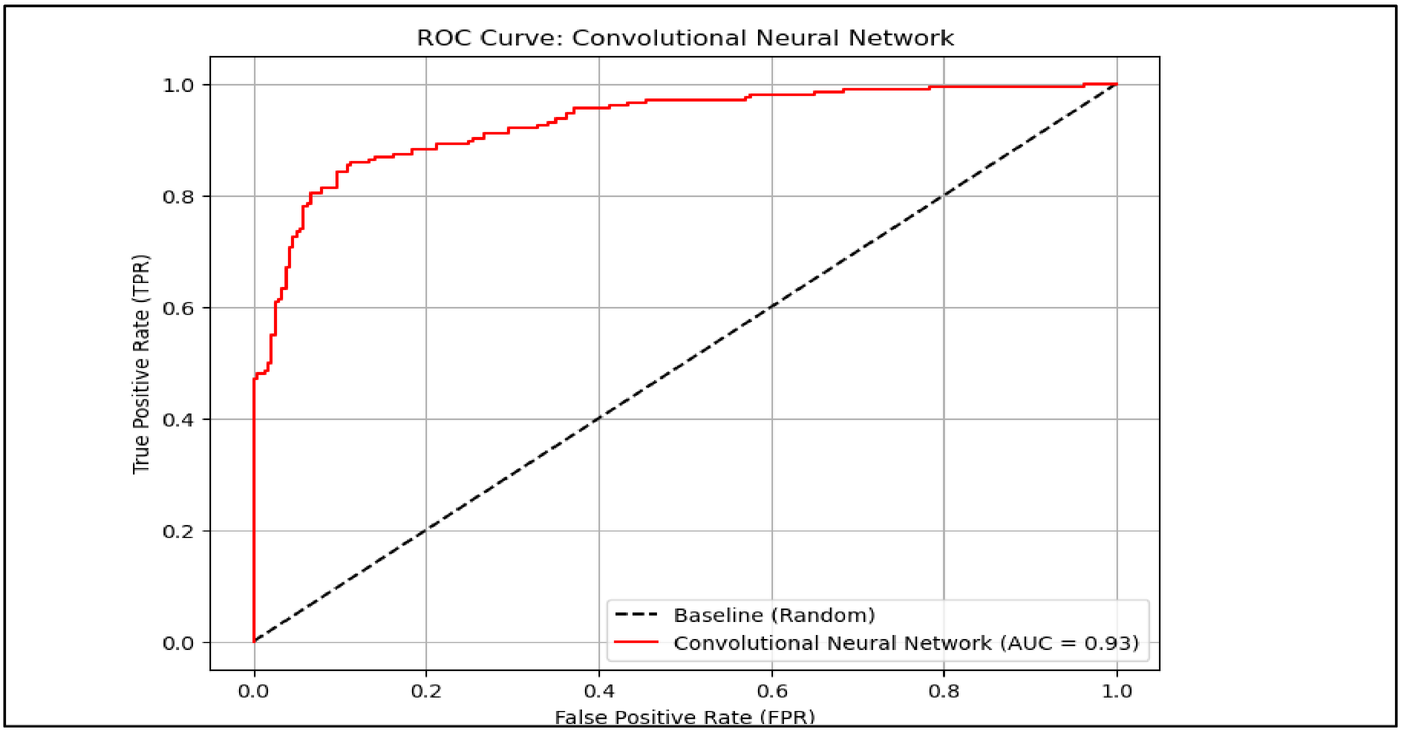

Fig 4 ROC Curve: Convolutional Neural Network

➢ *KNN:*

- Accuracy: 64.21%
- Precision: 60.37%
- Recall: 64.90%
- F1-score: 62.70%
- ROC-AUC: 0.62

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul589

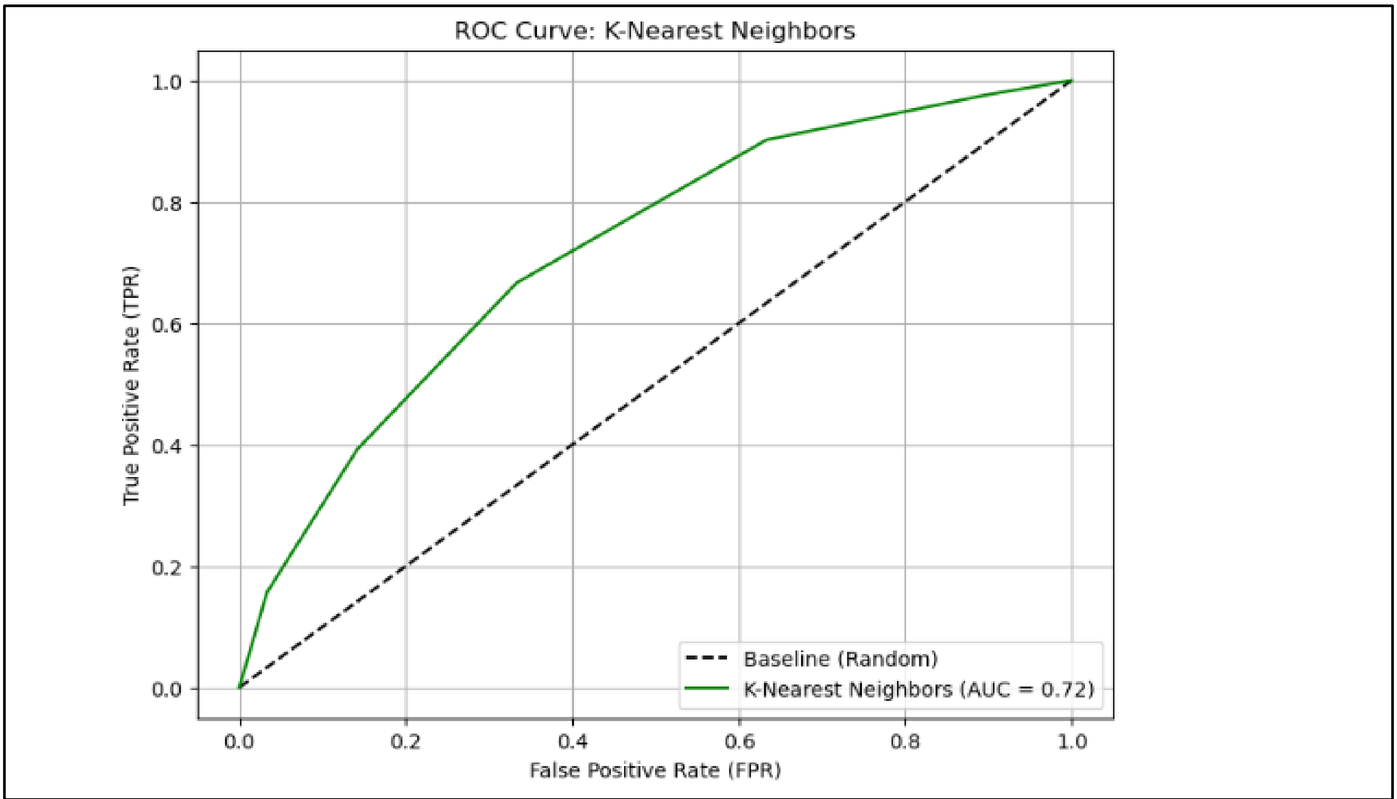

Fig 5 ROC Curve: K-Nearest Neighbors

➢ *Model Comparison*:

The Random Forest model demonstrated superior overall performance with high recall and balanced precision, indicating strong generalization. Convolutional Neural Network achieved higher precision but suffered from a lower recall rate, highlighting potential under-detection of seizures. K Nearest Neighbours yielded the weakest performance and suffered from high variance.

1	MODEL	METRIC	SCORE
2	Random Forest	Accuracy	0.8631
3	Random Forest	Precision	0.8734
4	Random Forest	Recall	0.8235
5	Random Forest	F1-Score	0.8408
6	KNN	Accuracy	0.6421
7	KNN	Precision	0.6037
8	KNN	Recall	0.649
9	KNN	F1-Score	0.6254
10	CNN	Accuracy	0.8393
11	CNN	Precision	0.8893
12	CNN	Recall	0.7431
13	CNN	F1-Score	0.8094

Fig 6 Model Comparison

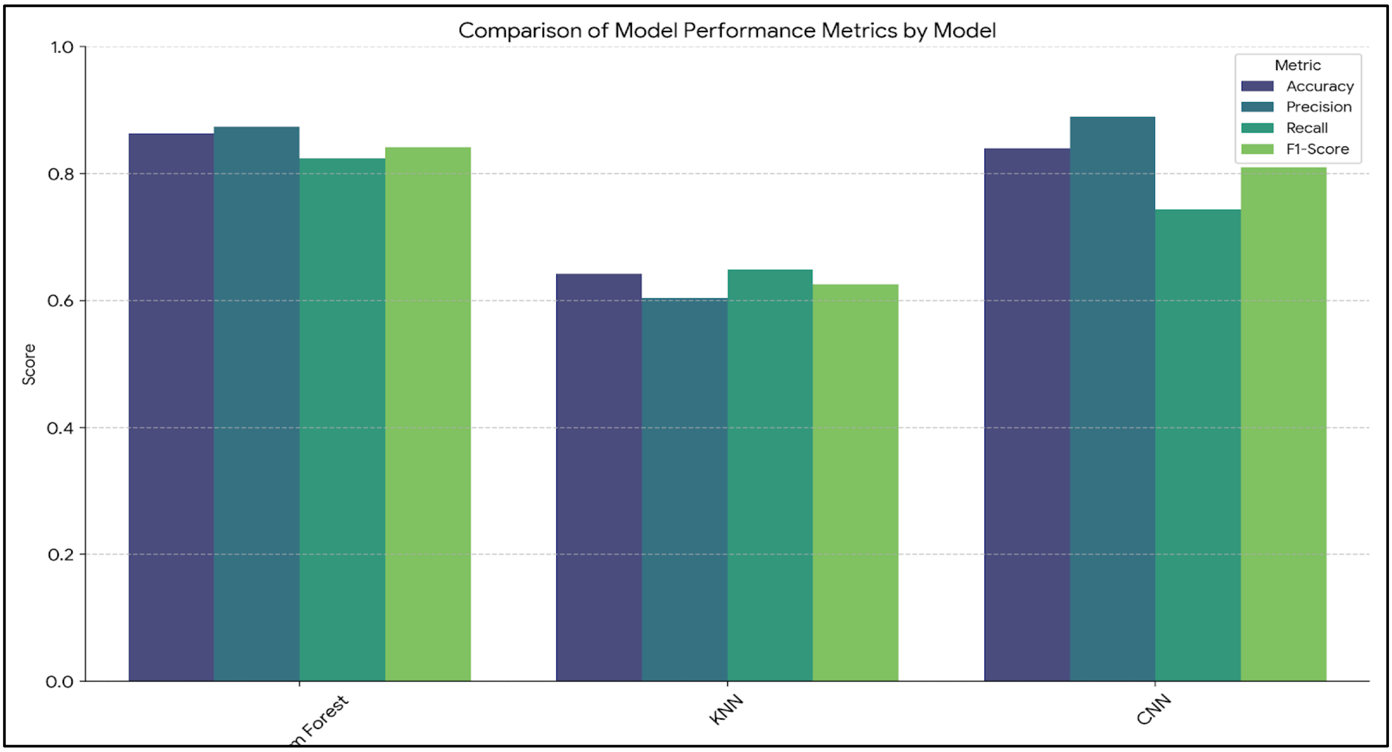

Fig 7 Comparison of Model Performance Metrics by Model

V. CONTRIBUTIONS

- Developed a machine learning pipeline tailored to Zimbabwe-specific EEG data
- Demonstrated superior performance of RF in seizure classification tasks
- Highlighted the potential of ML in addressing diagnostic resource gaps in low-income healthcare systems
- Provided empirical evidence for model generalizability and interpretability

VI. CONCLUSION

This study validates the efficacy of machine learning models in automating epileptic seizure detection using EEG signals collected in Zimbabwe. The Random Forest model outperformed both KNN and CNN in most metrics, making it suitable for clinical deployment due to its robustness and explainability. Future work will explore model deployment on edge devices, real-time EEG monitoring, and integration with electronic health records. Additionally, incorporating explainable AI frameworks could improve clinician trust and model transparency.

REFERENCES

[1]. Acharya, U. R., et al. (2018). [6] A deep convolutional neural network for automated diagnosis of epilepsy using EEG signals. [6][7] Future Generation Computer Systems, 86, 415-422.

[2]. Duan, R., et al. (2019). Epileptic seizure detection using CNN-based spectral-spatial feature representation of EEG signals. IEEE Access, 7, 77969-77977.

[3]. Goyal, M., et al. (2020). Application of machine learning in epileptic seizure detection. Journal of Healthcare Engineering, 2020, 8886297.

[4]. Khan, Y. U., et al. (2021). Automated seizure detection in EEG signals using ensemble learning. Biomedical Signal Processing and Control, 69, 102922.

[5]. Li, X., & Chen, W. (2021). Epileptic seizure detection based on EEG signals and CNN. [7][8] Frontiers in Neuroinformatics, 12, 95.

[6]. Truong, N. D., et al. (2021). A deep learning framework for epileptic seizure detection using EEG signals. [7][9] IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 687-696.

[7]. Ullah, I., et al. (2018). Automated epileptic seizure detection using উন্নত features of EEG signals. Biocybernetics and Biomedical Engineering, 38(4), 734-744.

[8]. Duta, I. (2025). [5] A cautiously optimistic future for affordable epilepsy diagnosis. Oxford Martin School. https://www.oxfordmartin.ox.ac.uk/blog/a-cautiously-optimistic-future-for-affordable-epilepsy-diagnosis

[9]. Ifakara Health Institute. [10] (2025). [3][5] Leveraging AI in epilepsy diagnosis. https://www.ihi.or.tz/blog/news/digital-health-leveraging-ai-in-epilepsy-diagnosis/

[10]. Kabir, M.S.A., Roy, T., Rahimi, M.J. (2022). Robust Classifier Selection for Cost-Effective EEG Seizure Detection in Remote Areas. 2022 2nd International Conference on Computing and Communication Systems (I3CS). IEEE.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul589

[11]. The IC² Institute. (n.d.). AI-Based Seizure Diagnosis for Low-Resource Areas. https://ic2.utexas.edu/research/ai-based-seizure-diagnosis-for-low-resource-areas/
[12]. Federated Machine Learning for Epileptic Seizure Detection using EEG. (2024). [11] International Journal of Advanced Computer Science and Applications, 15(4). [11]
[13]. Jia, K., et al. (2022). Epileptic seizure detection with deep EEG features by convolutional neural network and shallow classifiers. [6] Frontiers in Neuroscience, 16, 840141.
[14]. Nahzat, F., & Yağanoğlu, M. (2024). [12] Epileptic seizure detection using machine learning on electroencephalogram data. [7][13] Journal of Emerging Investigators, 24(028).
[15]. Roy, S., et al. (2019). Detection Analysis of Epileptic EEG Using a Novel Random Forest Model Combined With Grid Search Optimization. [14][15] Frontiers in Human Neuroscience, 13, 52.
[16]. Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016 - PMC - PubMed Central.
[17]. Evaluating the generalisability of region-naïve machine learning algorithms for the identification of epilepsy in low-resource settings - PubMed Central.