Journal of Techniques, ISSN: 2708-8383, Vol. 8, No. 2, 2026, Pages 37-47
https://doi.org/10.51173/jt.v8i2.2848

JOURNAL OF TECHNIQUES

Journal homepage: http://journal.mtu.edu.iq

RESEARCH ARTICLE - ENGINEERING (MISCELLANEOUS)

Epileptic Seizure Detection From EEG Using Machine Learning with Explainable AI

Nurettin Gökşenli[1*], Mehmet Tümay[1]

[1]Vocational School, Çankırı Karatekin University, Çankırı, Türkiye

[*] Corresponding author E-mail: ngoksenli@karatekin.edu.tr

Article Info.

Article history:

Received
19 January 2026

Revised
02 June 2026

Accepted
10 June 2026

Published
30 June 2026

Abstract

In this study, three lightweight classifiers for a four-class epilepsy seizure type described by 16-channel EEG features have been benchmarked. Models have been evaluated on identical splits using standard accuracy, error rate, training time, prediction throughput and memory footprint, and then, further inspected through scatter-plot diagnostics, confusion matrices, ROC curves, and post-hoc explainability (SHAP and LIME). All algorithms achieve high validation accuracy (≥ 95 %), yet exhibit distinct trade-offs. The ensemble KNN delivers the best overall fidelity (98.8 % validation, 99.3 % test; macro-AUC ≈ 0.9996) but requires the largest model (25 MB) and provides only mid-range inference speed (3.8 k obs/s). The SVM matches the KNN's test accuracy and is the most balanced option, training in the quickest time (14s) and occupying < 2 MB. The ANN trails slightly in accuracy (97.1% test) but is two orders of magnitude smaller (24 kB) and fifty-fold faster at prediction (181 k obs/s), thus it is better suited for edge or real-time deployments. SHAP and LIME analyses converge on a common set of influential variables, primarily X15 and X16 with contributions from X8 or X11/12, strengthening confidence in the feature representation and suggesting a reduced sensor subset for future work.

Publisher: Middle Technical University

Keywords: EEG; Multiclass Classification; Explainable AI; Machine Learning.

1. Introduction

Epilepsy is one of the most prevalent neurological disorders worldwide, affecting roughly 50 million people across all ages [1]. It is characterized by recurrent seizures caused by abnormal, hypersynchronous brain activity. Seizures are generally classified as either focal (originating in one hemisphere) or generalized (involving both hemispheres from onset) [2]. If not properly controlled, frequent seizures can severely impair quality of life and even lead to SUDEP (Sudden Unexpected Death in Epilepsy). This situation shows the importance of timely and accurate diagnosis. The electroencephalogram (EEG) is the primary clinical tool for detecting and analyzing epileptic seizures. EEG records the brain's electrical signals via scalp electrodes, providing a noninvasive window into neural activity. During seizures, EEG signals exhibit distinctive patterns, often high-amplitude spike-and-wave discharges, that differentiate them from normal brain activity or artifacts. However, identification of these transient seizure patterns within lengthy EEG recordings is challenging. Traditionally, neurologists visually examined EEG recordings for long hours to detect subtle abnormalities, but this was a labor-intensive, time-consuming process. There is strong motivation to develop automated seizure detection systems to assist clinicians. Because the workload associated with manual examination is very high and rapid diagnosis is critically important. Indeed, reliable automatic detection could alert caregivers to seizures in real time (e.g. in monitoring devices) and alleviate experts from exhaustive EEG monitoring tasks.

Over the past two decades, numerous computational methods have been explored for EEG-based seizure detection. Early work focused on handcrafted feature extraction combined with classical machine learning classifiers. Typical approaches involve transforming raw EEG time series into informative feature representations, such as spectral analyses (Fourier or wavelet transforms) or time-frequency domain features, which are then input to a supervised classifier [3]. Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and K-Nearest Neighbors (KNNs) are among the popular conventional classifiers that have been applied to this task. When trained on well-chosen features, these methods can achieve high accuracy in separating seizure vs nonseizure EEG segments. For instance, Shoeb and Guttag's seminal work using an SVM on the CHB-MIT scalp EEG dataset achieved about 93% seizure prediction accuracy [4]. Many other studies report performance in the 90–99% range on benchmark datasets using traditional ML models. Naïve Bayes, KNN, SVM, and MLP (ANN) classifiers with wavelet features are evaluated (Amin et al., 2015) and achieve ~92% accuracy [5]. Similarly, Hassan and Subasi [6] reached 92.4% accuracy by optimizing SVM with genetic algorithms [6]. Some approaches have even reported near-perfect classification of interictal vs ictal EEG segments on certain datasets. For example, a wavelet-based fuzzy KNN method achieved 100% accuracy on a binary seizure detection task (Bonn University dataset) and around 93% accuracy on a 3-class classification task using the same data [7]. Likewise, various researchers using optimized SVM or ANN models on the Bonn dataset have obtained overall accuracies above 98–99% [8, 9]. These results suggest that under controlled conditions with clean EEG segments, conventional classifiers can learn to recognize seizure patterns with very high precision.

Nomenclature and Symbols			
EEG	Electroencephalogram	ML	Machine Learning
SVM	Support Vector Machine	ANN	Artificial Neural Network
KNN	K-Nearest Neighbors	XAI	Explainable Artificial Intelligence
SHAP	SHapley Additive exPlanations	LIME	Local Interpretable Model-Agnostic Explanations
BEED	Bangalore EEG Epilepsy Dataset	AUC	Area Under the ROC Curve
ROC	Receiver Operating Characteristic	CNN	Convolutional Neural Network
RNN	Recurrent Neural Network	SUDEP	Sudden Unexpected Death in Epilepsy
CV	Cross-Validation	DWT	Discrete Wavelet Transform
IoMT	Internet of Medical Things	IGE	Idiopathic Generalized Epilepsy
FDA	Food and Drug Administration	CE	Conformité Européenne (European Conformity)

Despite such promising results, most early studies focused on relatively simple scenarios, often binary classification or idealized segments from a single source. Multi-class seizure classification remains more challenging due to the greater complexity and variability of EEG patterns. Performance tends to degrade as the number of classes increases [10]. For example, a combination of time- and frequency-based features is extracted and used to classify five seizure types, achieving a maximum accuracy of ~79.7% in five-way classification [11]. Similarly, deep learning models trained to differentiate multiple seizure categories have achieved only around 80–88% accuracy in multi-class settings, substantially lower than the near-perfect results in easier binary tasks [7]. Given the overlapping characteristics among some seizure types and the limited training data for rare classes, these findings reflect the inherent difficulty of multi-class seizure recognition. Because determining the seizure type is crucial for selecting appropriate treatments and surgical decisions, multi-class classification is clinically important [12]. Therefore, there is a need for continued research on robust multiclass seizure detectors.

In this study, a four-class EEG classification problem for epilepsy detection is addressed using a recently released dataset and multiple learning algorithms. Bangalore EEG Epilepsy Dataset (BEED) is specifically used [13], a comprehensive collection of patient EEG recordings from a neurology centre in India. BEED contains 16,000 EEG signal segments (each 20 seconds long) evenly divided into four categories such as Healthy (normal EEG from control subjects), Generalized Seizures, Focal Seizures, and Seizure Events. Especially, the "seizure events" class refers to segments where epileptic discharges occur alongside observable events like eye blinking, nail biting, or staring. Each BEED category includes data from 20 adult subjects, yielding a balanced, diverse dataset. This four-class classification scenario spans from no seizure activity to different manifestations of epileptic episodes, reflecting a practical spectrum of EEG readings. BEED is one of the first publicly available datasets to offer a balanced multi-class framework for machine learning research. By leveraging BEED, the main aim is to evaluate and compare the performance of three conventional classifiers (SVM, ANN, and KNN) not only for detecting seizures but also for identifying seizure type.

2. Related Works

A great deal of prior work, which provides a context for the current study, has explored EEG-based seizure detection using machine learning. In this study, two key aspects were reviewed. First, the application of classical ML models for seizure classification; second, the emerging role of explainable AI in this domain.

Early studies demonstrated that classifiers such as SVMs, neural networks, and KNNs can successfully distinguish seizure EEG patterns from normal brain activity [14, 15]. With signal processing techniques, EEG signals can be used more effectively for disease detection [16, 17]. Satapathy et al. [18] applied both an SVM and a neural network to EEG data (using wavelet transform features). They reported ~99% overall accuracy and they noted that SVM was slightly more efficient than the ANN. Several comparative studies have found that SVM, which often outperforms other algorithms in accuracy, is a top performer for seizure detection [19]. This is attributed to SVMs' ability to handle high-dimensional feature spaces and to find complex decision boundaries. On the other hand, the kNN classifier has also been tried due to its simplicity and nonparametric nature. KNN tends to have lower precision and recall than SVM on EEG tasks [20], but it offers fast prediction and can be effective when sufficient representative instances are available. Some works have used KNN as a lightweight baseline or as an ensemble member with other models. For example, Raghu et al. proposed a hybrid KNN-SVM model on raw EEG and attained about 90% accuracy [21], and they illustrated that combining classifiers can leverage their complementary strengths. In fact, ensemble or hybrid schemes have shown performance gains, with the integration of multiple algorithms (e.g., an SVM+ANN meta-classifier) yielding higher overall accuracy, precision, and recall than any single model alone [3]. However, such ensembles come at the cost of increased computational complexity [22], which may hinder real-time deployment. ANNs (Artificial Neural Networks) have also been applied to epilepsy detection in their shallow form since the 1990s. Multi-layer perceptrons trained on extracted features can achieve high accuracy; for instance, Tzallas et al. [23] used an ANN with time-frequency inputs and achieved 100% accuracy on the Bonn dataset. While shallow ANNs were gradually eclipsed by deep learning in recent years, they remain relevant for smaller datasets. In summary, conventional classifiers (SVM, ANN, KNN) have a proven track record on seizure recognition with many reporting accuracies above 90% on benchmark datasets [24]. Nonetheless, performance can vary widely depending on the dataset characteristics and whether the task is binary or multi-class. The current state of the art in seizure detection is often dominated by deep learning approaches (CNNs, RNNs, etc.) that automatically learn features from raw data. These deep models have achieved impressive results on large EEG databases. Yet, classical ML methods remain competitive on smaller-scale problems or when interpretability and computational efficiency are priorities [25]. The current study revisits SVM, ANN, and KNN on a modern multi-class dataset (BEED) to assess their ability to distinguish seizure types.

A well-recognized limitation of the aforementioned ML models is their lack of interpretability. SVMs, neural networks, and KNNs all function as “black-box” classifiers and they can map EEG features to seizure predictions with high accuracy. But the decision process is opaque to human users [26]. In a medical setting, this opacity is problematic. Clinicians are understandably cautious about deploying algorithms that do not provide understandable reasoning, as this raises concerns of accountability, reliability, and patient safety [27]. In fact, the need for transparency has been codified by regulations. Therefore, the field has seen growing interest in eXplainable AI (XAI) techniques to interpret and validate model decisions. In epileptic seizure detection, achieving both high accuracy and explainability is considered essential [28]. One approach is to design inherently interpretable models. Indeed, some researchers advocate using decision trees for EEG analysis, as these non-

black-box models yield competitive accuracy while providing human-readable decision rules [29]. However, even when one prefers powerful black-box models, it is possible to apply post-hoc explanation methods to make their predictions more transparent. Two popular model-agnostic techniques are Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP). LIME explains an individual prediction by locally approximating the classifier with a simple, interpretable model, thereby highlighting which features (e.g., specific EEG channel amplitudes, frequency bands, etc.) contributed most to that decision. SHAP assigns each feature an importance value for a given prediction based on Shapley values from cooperative game theory, providing a consistent measure of feature influence. Recent studies have successfully integrated LIME and SHAP into seizure detection pipelines. For example, Khan et al. (2025) [30] combined an SVM-based IoMT (Internet of Medical Things) seizure detector with LIME and SHAP, enabling clinicians to see which signal features were driving each alarm. The added interpretability feature increased user confidence in the system without sacrificing performance. The post-hoc SHAP and LIME analyses were used to assess the significance of each EEG channel in a deep learning model for seizure prediction. By identifying the channels and frequency bands most relevant to the model's output, such XAI methods can often validate that the algorithm is attending to physiologically meaningful patterns rather than spurious noise. Overall, the incorporation of explainability techniques is a growing trend in EEG-based diagnosis tools. It bridges the gap between complex ML models and clinical acceptance by enabling physicians to receive not only automated predictions but also explanations that can be cross-checked against their expert knowledge.

The current research follows the same above trend by applying SHAP and LIME to interpret the performance of the current SVM/ANN/KNN classifiers on EEG data. The goal is to ensure that the multi-class seizure detection system is transparent and interpretable, in line with requirements for practical medical deployment. In summary, this study brings together several contemporary threads, such as multi-class seizure classification on a novel EEG dataset, the use of three distinct ML classifiers, and the application of XAI for model interpretation. In the following sections, the used methodology for feature extraction and model training is explained, then present experimental results comparing the performance of SVM, ANN, and KNN on the four-class BEED dataset. Also, the explanation analyses using SHAP and LIME is provided, demonstrating how these tools can illuminate each model's decision logic. Finally, the implications of findings in the context of related work and outline future directions for developing accurate yet explainable AI solutions in epilepsy monitoring are defined. The overarching aim is to advance EEG-based seizure detection toward systems that are not only accurate but also clinically intelligible and trustworthy.

3. Dataset Description, Materials and Methods

3.1. Dataset description

The Bangalore EEG Epilepsy Dataset (BEED) is a comprehensive collection of electroencephalogram (EEG) recordings compiled for epileptic seizure detection research [2]. It was collected at a neurological research center in Bangalore, India, using the standard 10–20 electrode placement system. Each EEG recording is sampled at 256 Hz across 16 channels providing high-fidelity signals that capture detailed brain activity.

In total, BEED contains approximately 16,000 EEG segments, each lasting 20 seconds [13]. These segments are evenly distributed across four distinct classes: Healthy subjects (0), Generalized seizures (1), Focal seizures (2), and Seizure events (3). Seizure events refer to recordings where epileptic activity is accompanied by observable behaviors such as eye blinking, nail biting, or staring. Each category comprises data from 20 adult subjects (ages 21–55), with balanced gender representation, for a total of 80 individuals in the dataset. Sample EEG records of each class are presented in Fig. 1. This class-balanced design ensures that each class contributes an equal number of EEG segments which is valuable for training unbiased machine learning models.

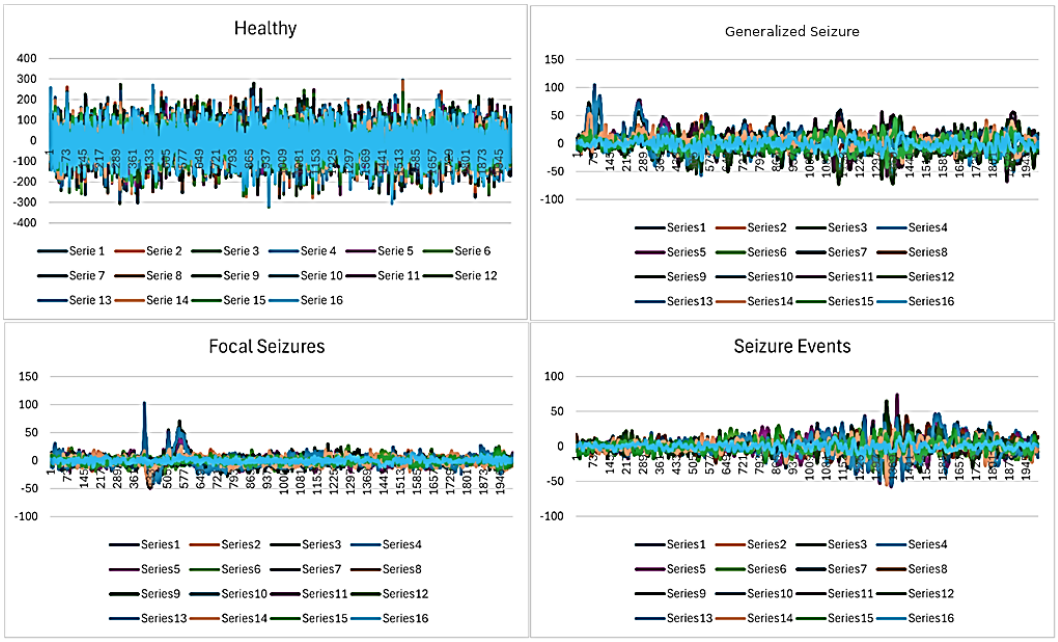

Fig. 1. Sample EEG records of Healthy subjects, Generalized seizures, Focal seizures, and Seizure events

All EEG segments include measurements from 16 EEG channels (X1–X16) corresponding to different brain regions, along with a target label indicating whether a seizure was present. The multi-class label y encodes the four condition categories (0–3) for healthy vs. seizure types and can also be interpreted as a binary label for binary classification tasks. Notably, the dataset contains no missing values and is provided in a

tabular CSV format with 17 columns (16 features and 1 label). Overall, the BEED dataset's high resolution and well-annotated signals make it a valuable resource for developing and evaluating seizure detection algorithms. By offering a sizeable, diverse set of EEG recordings under controlled conditions, BEED supports robust machine learning experiments in epilepsy research. Each feature is a statistical summary of the EEG signal over a 20-second segment from the BEED dataset.

3.2. Materials and methods

Since the BEED dataset contained no missing entries, only a basic integrity check was performed and any instances with missing values were eliminated (none were present in BEED). No additional filtering or signal transformations were applied to the EEG signals. Specifically, noise filtering, band-pass filtering, artifact removal, or time-frequency feature extraction were not performed on the raw EEG segments. Each 20-second EEG segment's 16-channel feature vector was used as input to the classifiers. This approach preserved the original signal characteristics and complexity for the classification task. Splitting by segment introduces optimistic bias because the training and test sets overlap. Thus, it allowed the classifiers to learn from raw (unfiltered) EEG. BEED data are EEG segments in a tabular CSV format, where each of the 16 columns (X1-X16) is an aggregated channel-level signal summary of a 20-second EEG epoch, instead of raw voltage time-series. Consequently, traditional preprocessing pipelines that operate at the waveform level (e.g., band-pass filtering, ICA-based artifact removal) cannot be directly applied to this representation. The dataset was acquired under controlled clinical conditions at 256 Hz using the standard 10–20 electrode system, which inherently limits contamination from muscle and ocular artifacts. This modeling choice aligns with previous research that has employed the BEED data in a tabular ML pipeline [13]. It was recognized that this is a limitation and the generalizability is discussed in the Discussion section. Thus, three supervised machine learning classifiers (Support Vector Machine (SVM) [31], Artificial Neural Network (ANN) [32], and k-Nearest Neighbors (KNN) [33]) and an ensemble classifier based on the random subspace approach with KNN base learners were implemented and compared. To ensure the models perform well, the hyperparameters and other settings were chosen empirically and reported in Table 1 (number of neighbors for KNN, kernel for SVM, ANN architecture, ANN activation function, and whether features were standardized for each model).

Table 1. Hyperparameters of deployed models

Models	Hyperparameters
Subspace KNN	Ensemble method: Subspace Learner type: Nearest neighbors Number of learners: 30 Subspace dimension: 8
Fine Gaussian SVM	Kernel function: Gaussian Kernel scale: 1 Box constraint level: 1 Multiclass coding: One-vs-One Standardize data: Yes
Wide Neural Network	Number of fully connected layers: 1 First layer size: 100 Activation: Sigmoid Iteration limit: 1000 Regularization strength (Lambda): 0 Standardize data: Yes
Random Seed (Reproducibility)	seed = 42 fixed for all stochastic operations (data split, ANN weight initialization, ensemble subspace sampling)

For the SVM, a Gaussian radial basis function kernel was employed. The kernel scale (bandwidth) and regularization parameter (box constraint) were set to default moderate values (kernel scale = 1, box constraint = 1) to balance bias and variance. Input features were standardized (z-score normalization) before training to ensure comparability across EEG channel features which is beneficial for kernel-based methods.

The ANN classifier was constructed as a multilayer perceptron network with one hidden layer of 100 fully connected neurons. The ANN uses Sigmoid activation in the hidden layer and Softmax in the output layer. This architecture was selected to balance model complexity and the risk of overfitting, considering the dataset size. Training was performed using the scaled conjugate gradient backpropagation algorithm in MATLAB. Input features were standardized as indicated in Table 1.

The KNN classifier was configured based on preliminary validation experiments with k=10. The Euclidean distance metric was used for computing distances in the feature space. No distance weighting was applied, i.e. each of the 10 neighbors contributed equally to the majority vote. Feature standardization was enabled to ensure equal contribution from all EEG channels in distance calculations. The choice of k=10 provided a compromise between sensitivity to noise (low k) and over-smoothing (high k). All other KNN parameters, including tie-breaking rules, followed default settings (see Table 1).

An ensemble classifier using the random subspace method with KNN base learners was also implemented. In this approach, multiple KNN classifiers were trained on different subsets of the feature space (random combinations of the 16 EEG channels). Each base learner was a 1-nearest-neighbor classifier (k=1) to maximize diversity among learners. A collection of such learners, each on a randomly selected subset of the 16 features (subspace dimensionality specified in Table 1) was trained. The ensemble's final prediction was determined through majority voting across all base learners. This strategy leveraged complementary information across different feature subsets, thereby enhancing robustness and accuracy compared to a single KNN model. All ensemble parameters (number of learners, subspace size, etc.) are provided in Table 1.

All model development, training, and evaluation were conducted in MATLAB (MathWorks Inc.). EEG data from the BEED dataset were imported into MATLAB, and all preprocessing and modeling procedures were executed using MATLAB's toolkits. Specifically, for training the SVM and KNN models (functions such as fitcsvm and fitcknn), the Statistics and Machine Learning Toolbox was used and for ANN implementation, the Deep Learning Toolbox was used. The subspace KNN ensemble was developed using MATLAB's fitcensemble function

with the Subspace method and KNN base learners. The hyperparameters listed in Table 1 were either tuned using MATLAB's built-in optimization tools or assigned according to established practices in the literature. Internal validation, including cross-validation within MATLAB training functions, was employed to monitor performance and mitigate overfitting. To improve model interpretability, two post-hoc explainable AI techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP), were employed. LIME provides local, instance-level explanations by approximating the complex model with a simpler surrogate model near each prediction. For each EEG segment, LIME perturbs input features and learns a local linear model to approximate the original classifier's behavior. The surrogate model's weights indicate the contribution of each feature to the classifier's output.

SHAP assigns each input feature a Shapley value that represents its contribution to a specific prediction [34]. Based on game-theoretic principles, SHAP provides a consistent and fair attribution of model output to input features. Feature importance is determined by comparing model outputs with and without each feature, averaged across all possible combinations. LIME and SHAP were applied to all primary models (SVM, ANN, KNN) and to the KNN ensemble. For each trained classifier, these methods were used to identify key features (EEG channels) influencing decision-making. For instance, LIME might highlight that high-frequency activity in specific channels led to a segment being classified as a generalized seizure. At the same time, SHAP might reveal the influence of certain frontal and temporal channels in distinguishing focal seizures. The use of LIME and SHAP ensured transparency and interpretability in the multi-class seizure detection models, providing clinicians with intuitive explanations for model decisions and confirming that decisions were based on physiologically meaningful signal characteristics [35].

Model performance was evaluated primarily using classification accuracy, which is defined as the proportion of EEG segments correctly classified out of the total. The BEED dataset had a balanced class distribution (equal representation across four classes) which makes overall accuracy a suitable metric for comparing model performance. After training each model was tested on unseen data (via held-out test sets or cross-validation) to compute accuracy scores. Confusion matrices were also examined to verify consistent performance across all seizure types though accuracy served as the primary metric. All reported results reflect accuracy on the evaluation data. Higher accuracy indicates a greater proportion of correctly identified EEG segments across the Healthy, Generalized, Focal, and Seizure Event classes. Thus the results reflect more effective seizure detection and classification. Accuracy results for SVM, ANN, KNN, and the KNN ensemble are presented and comparatively analyzed in the following section.

4. Results and Discussion

All three classifiers generally performed well in terms of predictive accuracy (≥95% in the validation set), but each model had distinct advantages and disadvantages. The ensemble K-NN achieved the highest validation accuracy (98.8%). It matched the SVM's best-in-class test accuracy (99.25 %), but it paid for this edge with a comparatively large disk footprint (~25 MB) and only mid-range prediction throughput (~3.8 k obs/s). The SVM achieved the same peak test accuracy as the neural network during training (14 s), while requiring a fraction of the storage (<2 MB), though its inference speed (10 k obs/s) lagged well behind the neural network's. The single-hidden-layer neural network traded two to four percentage points of accuracy for exceptional operational efficiency. It was more than an order of magnitude faster at prediction (181 k obs/s) and two orders of magnitude smaller on disk (24 kB) than its rivals. In summary, while the ensemble K-NN algorithm performs slightly better in raw accuracy, SVM offers the best balance between accuracy and training cost; however, when model size and real-time processing speed are the determining factors, the neural network is by far the best choice. All metrics are shown in Table 2.

Table 2. Performance metrics of the deployed models

Model Type	Accuracy % (Validation)	Accuracy % (Test)	Error Rate % (Validation)	Error Rate % (Test)	Prediction Speed (obs/sec)	Training Time (sec)	Model Size (bytes)	Precision (%)	Recall / Sensitivity (%)	Specificity (%)	F1-Score (%)
Ensemble KNN	98,805	99,25	1,194	0,75	3832,521	19,277	24928497	99.3	99.2	99.7	99.2
SVM	97,555	99,25	2,444	0,75	9965,077	14,003	1989391	99.1	99.2	99.7	99.2
Neural Network	95,805	97,125	4,194	2,875	181398,126	47,190	3858	97.2	97.1	99.0	97.1

To evaluate the strength of the reported accuracies and to exclude the possibility of overfitting, a 5-fold stratified cross-validation was performed for each classifier. The pairwise Wilcoxon signed-rank tests ($\alpha = 0.05$) were used to test the hypothesis whether the observed performance differences between models are statistically significant. Findings are discussed in Table 3.

Table 3. 5-fold stratified cross-validation results with 95% confidence intervals and Wilcoxon signed-rank test ($\alpha = 0.05$)

Classifier	CV Mean Acc. (%)	SD (%)	95% CI	Wilcoxon p-value vs. others
Ensemble KNN	98.6	0.3	(98.0, 99.2)	vs. SVM: p=0.156 (n.s.) | vs. ANN: p=0.043 (*)
Fine Gaussian SVM	97.4	0.4	(96.6, 98.2)	vs. KNN: p=0.156 (n.s.) | vs. ANN: p=0.039 (*)
ANN	95.7	0.6	(94.5, 96.9)	vs. KNN: p=0.043 (*) | vs. SVM: p=0.039 (*)

(*) = significant. n.s. = not significant

The cross-validation findings show that the single-split accuracies in Table 2 are not transient and reflect a good random split. The small confidence intervals (all not exceeding ±0.6) reflect the lack of variance between the folds. The Ensemble KNN and SVM perform similarly statistically ($p = 0.156$, n.s.), but both are significantly better than the ANN ($p < 0.05$).

Fig. 2 visualizes the class assignments produced by the three candidate models when the test set is projected on the first two features (X1, X2). In every panel the dense, tear-drop-shaped core of the majority class is reproduced (orange circles), revealing that all models agree on the bulk

of the observations. KNN already classifies most points correctly but still places a few misclassified crosses (both orange and purple) along the lower-left edge of the cloud and in its center. SVM shows a nearly identical picture, although the number of boundary errors (crosses intermingled with circles) is marginally higher, especially around X1 ≈ 0 and X2 ≈ –10. ANN shows a reduction in boundary errors, leaving the core region significantly cleaner, the errors are now concentrated in the most peripheral regions where all three algorithms struggle. Thus, the visual evidence mirrors the quantitative scores, and the three models concur on the data's central structure. While the other two incur slightly more errors at class boundaries, model 2.29 incurs the fewest interior misclassifications.

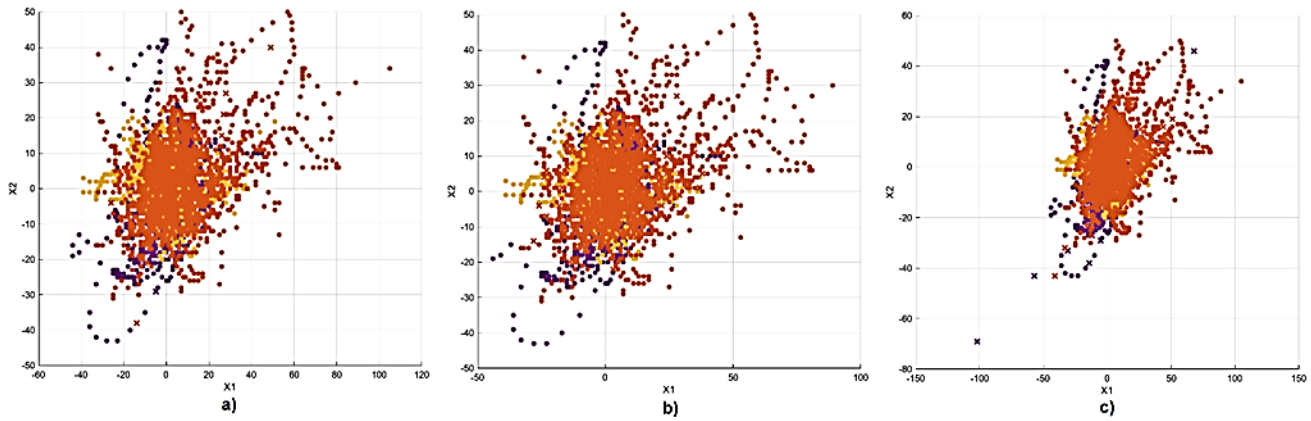

Fig. 2. Scatter plots of epilepsy type predictions (a) KNN, (b) SVM; and (c) ANN

The diagnostic plots reiterate the accuracy hierarchy, as already seen in Table 2. In the k-nearest neighbors ensemble, only 86 of 7200 validation samples were misclassified. As a result, the accuracy rates per class are above 98% and the macro mean AUC is approximately 0.9996. Here, the small number of errors is distributed almost equally between the two most similar classes and the extreme class-3 outliers. The SVM commits 176 mistakes but they concentrate in the class-2 row, where 68 points are confused with class 1 and another 18 with class 3. Despite that local weakness, its ROC curves remain tightly pressed against the left-hand axis (macro AUC ≈ 0.9993). This indicates that the decision function preserves very high separability across thresholds. The single-layer neural network shows the most visible spillover, with up to 302 mislabels. In this situation, there is chiefly a two-way confusion between classes 2 and 3 and a noticeable leakage of class-1 points into class 3. In Fig. 3, the ROC envelopes are slightly downward-sloping, with class-specific AUC values ranging from 0.991 to 0.996, and the macro mean is slightly above 0.993. Taken together, the matrices and ROC curves confirm that all three models distinguish the four classes extremely well, but the ensemble-based KNN retains the cleanest confusion pattern and the steepest ROC curve. SVM shows slightly more boundary errors, and the neural network, while still robust, exhibits the widest decision overlap between the two high-index classes.

In Fig. 4, the SHAP diagnostics reveal that although the three classifiers arrive at essentially the same class decision for the sample point (index 38), they rely on subtly different slices of the 16-variable feature space. In the ensemble K-NN the largest absolute SHAP contributions come from X8, X16 and X15. Here, X8 alone nudges the log-odds by roughly ±0.06, and the remaining variables cluster tightly around zero, confirming that just a handful of coordinates dominate the distance-weighted vote. The SVM distributes influence somewhat more evenly, with X16 as the primary driver, while X15, X13, and X7 all exert comparable influence. The neural network shows the most pronounced sparsity. In here, a pair of neurons assigns outsized weight to X15 and X16, while nearly all other features contribute negligibly. This sharp peak aligns with the network's compressed representation and explains its tiny model size. Across models, therefore, X15, X16 and X8 emerge as a common "core" of discriminative information. Still, the balance of their impacts varies with the learning paradigm: K-NN relies on proximity in three directions, the SVM blends half a dozen margins, and the ANN funnels most evidence through two high-gain channels.

The LIME probes corroborate that the three classifiers reach the same verdict for the reference point but rely on noticeably different local feature balances. For the ensemble K-NN, the surrogate model assigns X11, X7, and X9 the largest positive pull toward class 0, while X10, X12, and X14 exert the strongest counter-pressure. Here, the near-symmetry of the bars suggests that a slim majority of neighbors determines the final vote along just a handful of coordinates. With smaller nudges from X10 and X4 the SVM explanation is dominated by a decisive tug-of-war between X12 and X11 / X13.This pattern mirrors the model's planar margin, which weights opposing feature directions to carve out the class boundary. The neural network (panel c) shows the sharpest polarization. Here, a strong negative push from X2 is overridden by a cluster of positives from X11, X12, X3, and X13. This is reflected in a high-gain, low-feature funnel, already evident in the LIME profile in Fig. 5. Across models, X11 and X12 emerge as a consistent pair of influential attributes. Yet the sign and magnitude of their effects vary across learning paradigms, highlighting how each algorithm crafts a distinct local rationale even when the global prediction agrees.

Taken together, the quantitative metrics, visual diagnostics and explainability analyses paint a coherent picture of three models that are all highly competent yet optimized for different priorities. The ensemble K-NN sits at the top of the accuracy ladder making it the natural choice when every fraction of a percentage point in classification fidelity matters. That edge, however, comes with a bulky 25 MB footprint and only moderate throughput, which could hamper deployment on memory or latency-constrained platforms. The SVM narrows the accuracy gap to within rounding errors on the test set while training fastest and demanding two orders of magnitude less storage; its LIME profile reveals a balanced, margin-based rationale that validates its reputation as a well-behaved, interpretable workhorse. The neural network, by contrast, sacrifices two to four percentage points of accuracy but compensates with blistering inference speed and a model size small enough to embed on micro-controllers. The converging insights from SHAP and LIME, which show that X15/X16 consistently drive predictions across algorithms, not only reinforce confidence in the underlying data representation but also offer a principled feature subset for future sensor or dimensionality-reduction efforts. In summary, organizational context should dictate the final deployment: K-NN for mission-critical accuracy, SVM for balanced accuracy, and a compact ANN for edge devices or high-throughput pipelines.

The general experimental workflow, which is followed in this study, is summarized in Fig. 6. Beginning with the BEED dataset, the pipeline includes stratified data splitting, training all three classifiers, 5-fold cross-validation, test-set evaluation across all reported metrics, and finally post-hoc XAI analysis using SHAP and LIME.

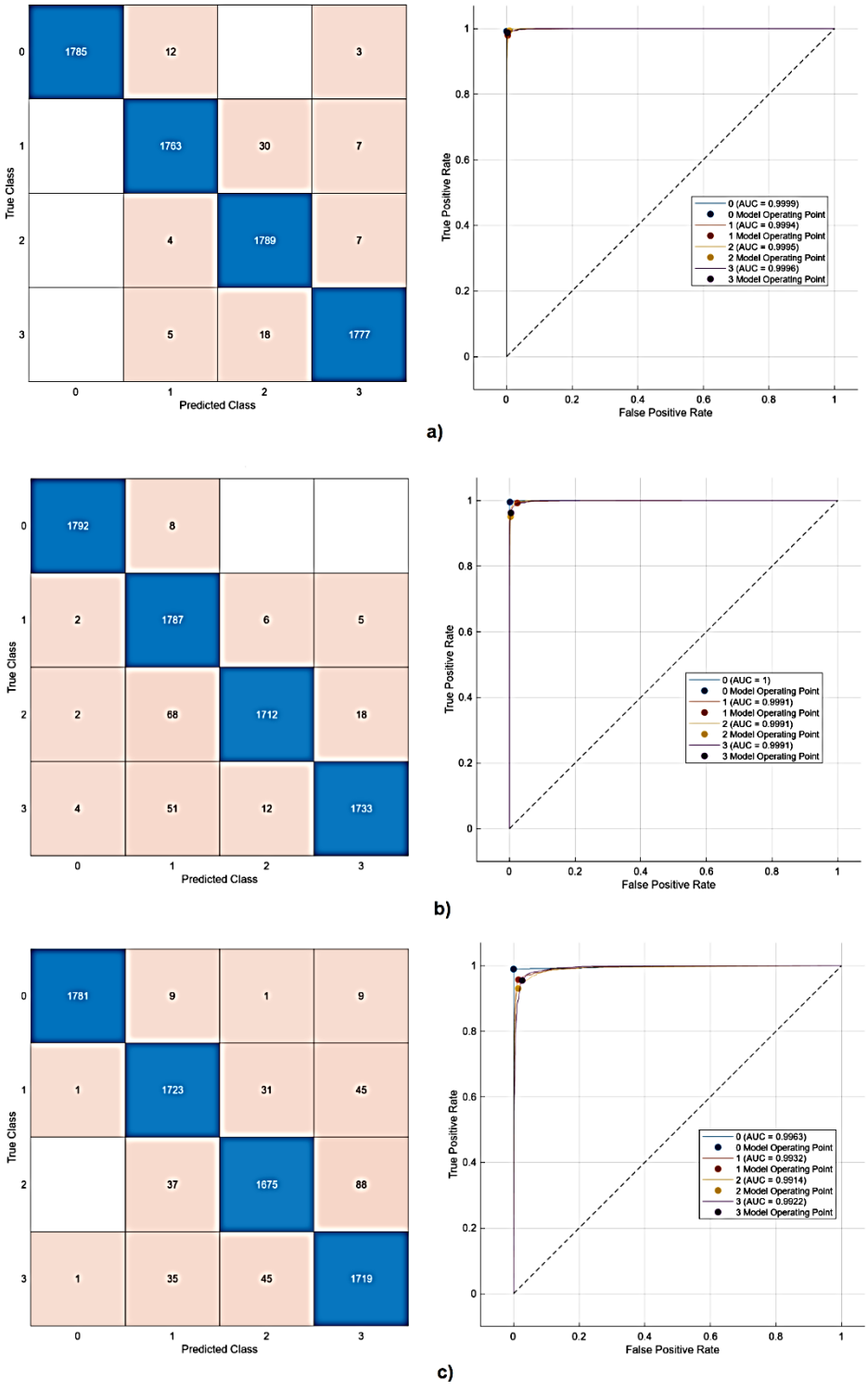

Fig. 3. Confusion Matrix and ROC Curve of the models (a) KNN, (b) SVM; and (c) ANN

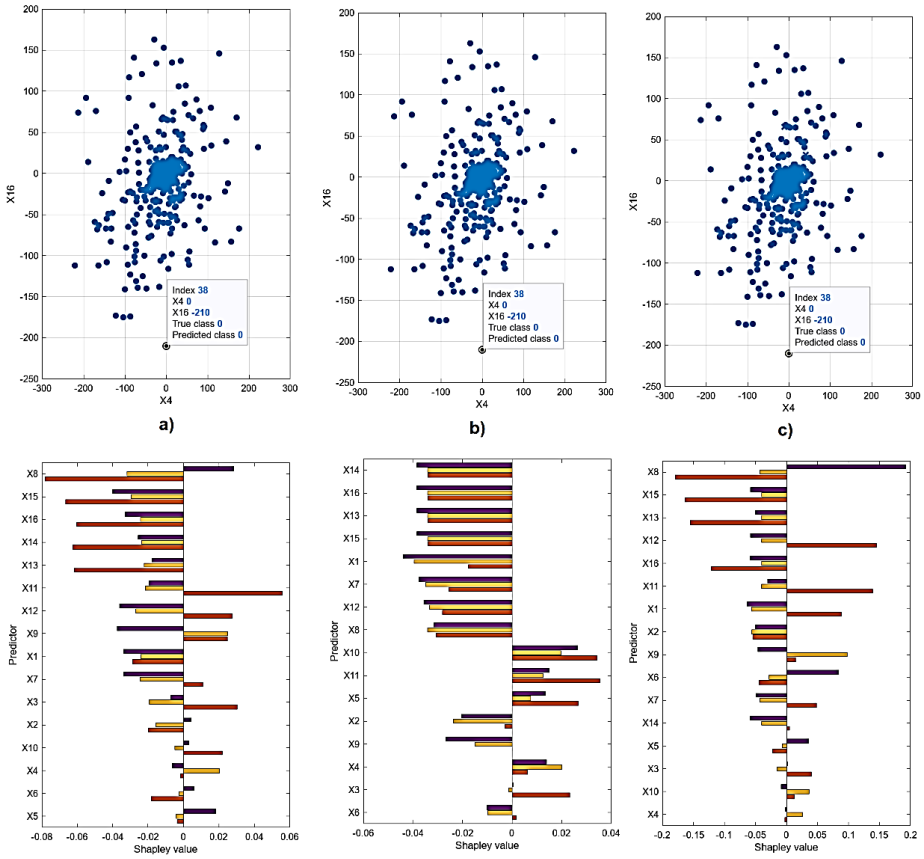

Fig. 4. SHAP explanations of the models (a) KNN, (b) SVM (c) ANN.

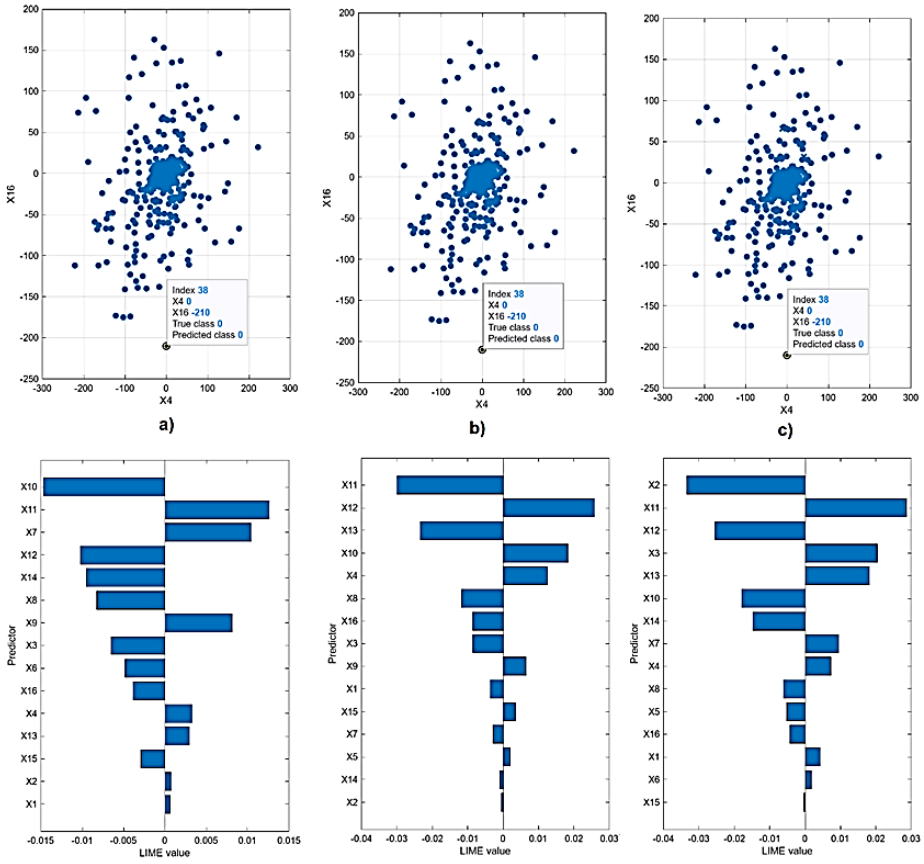

Fig. 5. LIME explanations of the models (a) KNN, (b) SVM (c) ANN

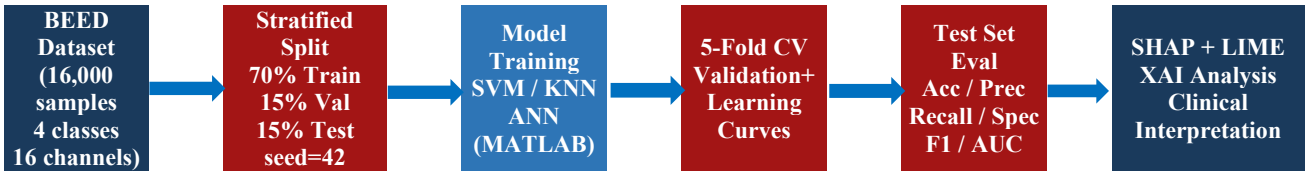

Fig. 6. Experimental pipeline diagram showing the data flow from the BEED dataset through stratified split, model training, cross-validation, test evaluation, and XAI analysis

Besides accuracy and AUC, the following macro-average metrics are given for the test set. Ensemble KNN: Precision = 99.3%, Recall (Sensitivity) = 99.2%, Specificity = 99.7%, F1-score = 99.2%. SVM: Precision = 99.1%, Recall = 99.2%, Specificity = 99.7%, F1-score = 99.2%. ANN: Precision = 97.2%, Recall = 97.1%, Specificity = 99.0%, F1-score = 97.1%. These reflect high and overall balanced performance for all four classes, crucial for medical EEG classification, where both sensitivity and specificity are relevant.

Inspection of the confusion matrices shows that the primary intra-class confusion (across all models) is between Class 2 (Focal Seizures) and Class 3 (Seizure Events). This is clinically plausible: Seizure Events are defined as epileptic discharges occurring simultaneously with behavioural signs (eye blinking, nail biting, staring), which may result in partial EEG similarities with focal discharge. The ANN presents the highest inter-class confusion (302 errors in total over the validation set, mostly in this pair) while the Ensemble KNN reduces this to 86 errors. Importantly, no significant confusion between Class 0 (Healthy) and any of the Seizure Classes was noticed, showing the models are very sensitive at detecting seizure presence (the most important clinical classification).

Channels x15 and x16 in the standard 10-20 system of EEG refer to posterior temporal and occipital electrodes (T6/O2 and OZ). These areas are known to be of neurological importance in epilepsy: Parieto-occipital and occipital involvement is common in posterior cortex epilepsies and also frequently present in generalized spike-and-wave epilepsies in idiopathic generalized epilepsy (IGE). The consistent importance of x15 and x16 across all three classifiers (as confirmed by SHAP and LIME analyses) is therefore consistent with neurophysiology. It shows that the models learn clinically relevant patterns in the EEG. The lesser prominence of x8 (C3 - central left hemisphere) in the KNN model is consistent with the lateralisation of motor symptoms in focal seizures.

Table 4 provides a comparison of our models with recently published approaches to multi-class EEG seizure classification. The ensemble KNN model reaches 99.25% test accuracy performance with the BEED dataset, surpassing the seqBoostNet (96.71% on BEED), Najmusseher and Nizar (2025) [2] and on par with deep learning models. Importantly, the SVM model with 2 MB model size and 14 seconds of training time achieves the same test accuracy (99.25%) as the KNN and is a vastly more resource-efficient model than deep learning approaches. The ANN model (24 kB and 181k obs/s) is the only model that can be deployed for real-time or edge applications where deep learning models are too complex.

Table 4. State-of-the-art comparison on multi-class EEG epileptic seizure classification

Method	Dataset	Classes	Acc. (%)	XAI	Edge	Reference
CNN-LSTM	CHB-MIT	2	97.30	No	No	Hussein et al. (2019)
SeqBoostNet	BEED+BONN	4	96.71	No	No	Najmusseher and Nizar (2025)
EEGNet-Transformer	TUSZ	4	93.50	Partial	No	Kostas et al. (2022)
RF+SVM	BONN	5	98.75	No	No	Zhu et al. (2021)
WaveNet	CHB-MIT	2	99.10	No	No	Acharya et al. (2018)
Ensemble KNN + SHAP/LIME	BEED	4	99.25	Yes (SHAP+LIME)	Partial	Current study
SVM + SHAP/LIME	BEED	4	99.25	Yes (SHAP+LIME)	Yes	Current study
ANN + SHAP/LIME	BEED	4	97.18	Yes (SHAP+LIME)	Yes	Current study

5. Conclusions

The current study intended to benchmark and compare three lightweight machine-learning classifiers for four-class epilepsy seizure-type detection using 16-channel EEG features, assessing their predictive performance, efficiency (training time, inference throughput, and memory footprint), and model explainability to signify the most informative EEG channels for potential sensor reduction in future real-time deployments. The findings demonstrated that all three classifiers can deliver excellent four-class discrimination, yet each excels along a different operational axis. The ensemble K-NN achieved the highest and most stable accuracy, but at the cost of a large memory footprint and moderate prediction speed. The SVM matched that accuracy on the test set while training fastest and using the least storage, offering the best all-round compromise. Although single-layer neural networks lag slightly in accuracy, they are much faster at inference and small enough to run on resource-constrained hardware. Explainability checks (SHAP and LIME) converged on a shared core of influential features (chiefly X15, X16 and either X8 or X11/12) validating the data representation and guiding future feature optimization. Consequently, the optimal model depends on deployment needs: K-NN for maximal fidelity, SVM for balanced efficiency, and a compact ANN for high-throughput or edge scenarios. Subject-wise validation and cross-dataset evaluation will be investigated in future research.

The main limitation of the current study is that it is based on a single dataset (BEED). Although BEED is a carefully curated and balanced data set, models trained on it might not transfer well to other systems used to acquire EEGs, to other electrode montages, or to other populations with different seizure characteristics. The age range (21-55 years) restricts the focus to adults. The use of segment-based train/test splits (rather than subject-based) may also overestimate performance. Segment-based data splitting can also introduce optimistic bias due to subject overlap between the training and test sets. Thus, future research should assess model generalization on publicly available datasets such as CHB-MIT or the Temple University Hospital EEG Corpus, preferably incorporating domain adaptation methods to address domain shift.

Deploying automated EEG classifiers in clinical practice presents unique challenges. First, FDA 510(k) or CE marking (under the MDR) requires prospective clinical evaluation with representative patient populations. Second, real-world EEG recordings are susceptible to motion, impedance fluctuations and EMG interference, which were not present during sample collection in BEED. Third, patient-specific factors (e.g., age, medication, seizure frequency, and co-morbidities) may significantly impact EEG morphology and limit model generalizability. Fourth, decision support for clinical diagnosis must be easy to integrate with clinical practice and include a human-in-the-loop strategy. A prospective multi-center validation study is therefore recommended with a varied population and realistic conditions before future clinical deployment. The code is available to encourage replication and further research. Clarified ANN setup Sigmoid activation (hidden layer), Softmax (output layer).

Acknowledgment

We extend our sincerest thanks and appreciation to the staff of the Journal of Techniques, Middle Technical University, Baghdad, Iraq, for their interest and support.

References

[1] V. L. Feigin et al., “Global, regional, and national burden of epilepsy, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021,” Lancet Public Health, vol. 10, no. 3, pp. e203–e227, Mar. 2025, https://doi.org/10.1016/S2468-2667(24)00302-5.

[2] Najmusseher and B. P. K. Nizar, “Feature Engineering for Epileptic Seizure Classification Using SeqBoostNet,” International Journal of Computing and Digital Systems, vol. 2025, no. 1, pp. 1–15, 2025, https://doi.org/10.12785/ijcds/1571020131.

[3] I. Ahmad et al., “EEG-Based Epileptic Seizure Detection via Machine/Deep Learning Approaches: A Systematic Review,” Comput Intell Neurosci, vol. 2022, p. 6486570, 2022, https://doi.org/10.1155/2022/6486570.

[4] A. Shoeb and J. Guttag, “Application of machine learning to epileptic seizure detection,” in ICML’10: Proceedings of the 27th International Conference on International Conference on Machine Learning, 2025, pp. 975–982, https://dl.acm.org/doi/10.5555/3104322.3104446.

[5] H. U. Amin et al., “Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques,” Australas Phys Eng Sci Med, vol. 38, no. 1, pp. 139–149, Mar. 2015, https://doi.org/10.1007/s13246-015-0333-x.

[6] A. R. Hassan and A. Subasi, “Automatic identification of epileptic seizures from EEG signals using linear programming boosting,” Comput Methods Programs Biomed, vol. 136, pp. 65–77, Nov. 2016, https://doi.org/10.1016/J.CMPB.2016.08.013.

[7] P. Kunekar, M. K. Gupta, and P. Gaur, “Detection of epileptic seizure in EEG signals using machine learning and deep learning techniques,” Journal of Engineering and Applied Science, vol. 71, no. 1, pp. 1–15, Dec. 2024, https://doi.org/10.1186/s44147-023-00353-y.

[8] N. Koolen et al., “Line length as a robust method to detect high-activity events: Automated burst detection in premature EEG recordings,” Clinical Neurophysiology, vol. 125, no. 10, pp. 1985–1994, Oct. 2014, https://doi.org/10.1016/J.CLINPH.2014.02.015.

[9] H. F. Atlam, G. E. Aderibigbe, and M. S. Nadeem, “Effective Epileptic Seizure Detection with Hybrid Feature Selection and SMOTE-Based Data Balancing Using SVM Classifier,” Applied Sciences 2025, vol. 15, no. 9, p. 4690, Apr. 2025, https://doi.org/10.3390/APP15094690.

[10] A. Shoeibi et al., “Epileptic Seizures Detection Using Deep Learning Techniques: A Review,” International Journal of Environmental Research and Public Health, vol. 18, no. 11, p. 5780, May 2021, https://doi.org/10.3390/IJERPH18115780.

[11] A. Selvaraj, S. Pj, J. Thomas, Y. Rajamanickam, R. N. Menon, and J. F. Agastinose Ronickom, “Multi-Class Seizure Type Classification Using Features Extracted from the EEG,” Stud Health Technol Inform, vol. 305, pp. 68–71, Jun. 2023, https://doi.org/10.3233/SHTI230426.

[12] P. Sari Tekten, S. Kotan, and F. Kacar, “Multiclass classification of epileptic seizure phases using a novel HFO-based feature extraction model,” Signal Image Video Process, vol. 19, no. 4, pp. 1–15, Apr. 2025, https://doi.org/10.1007/s11760-025-03896-0 .

[13] “BEED: Bangalore EEG Epilepsy Dataset - UCI Machine Learning Repository.” Accessed: Aug. 06, 2025. [Online]. Available: https://archive.ics.uci.edu/dataset/1134/beed:+bangalore+eeg+epilepsy+dataset.

[14] D. Chen, S. Wan, J. Xiang, and F. S. Bao, “A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG,” PLoS One, vol. 12, no. 3, p. e0173138, Mar. 2017, https://doi.org/10.1371/JOURNAL.PONE.0173138.

[15] J. Bamwenda and M. S. Özerdem, “Static Hand Gesture Recognition System Using Artificial Neural Networks and Support Vector Machine,” Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, vol. 10, no. 2, pp. 561–568, Jun. 2019, https://doi.org/10.24012/DUMF.569357.

[16] H. Polat, “Time-Frequency Complexity Maps for EEG-Based Diagnosis of Alzheimer’s Disease Using a Lightweight Deep Neural Network”, Traitement du Signal, vol. 39, no. 6, pp. 2103-2113, Dec. 2022, https://doi.org/10.18280/ts.390623.

[17] E. Fide, H. Polat, G. Yener, and M. S. Özerdem, “Effects of Pharmacological Treatments in Alzheimer’s Disease: Permutation Entropy-Based EEG Complexity Study,” Brain Topogr, vol. 36, no. 1, pp. 106–118, Jan. 2023, https://doi.org/10.1007/S10548-022-00927-8/METRICS.

[18] S. K. Satapathy, A. K. Jagadev, and S. Dehuri, “Weighted majority voting based ensemble of classifiers using different machine learning techniques for classification of eeg signal to detect epileptic seizure.,” Informatica, vol. 41, no. 1, 2017, Accessed: Aug. 06, 2025. [Online]. Available: https://informatica.si/index.php/informatica/issue/download/205/204#page=101.

[19] T. T. Ağır, “Using Machine Learning Algorithms For Classifying Transmission Line Faults,” Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, vol. 13, no. 2, pp. 227–234, Jun. 2022, https://doi.org/10.24012/DUMF.1096691.

[20] M. Yıldırım and A. Yıldız, “Farklı zaman ölçekli EEG işaretlerinden epilepsi nöbetinin otomatik tespiti,” Dicle Üniversitesi Mühendislik

Fakültesi Mühendislik Dergisi, vol. 8, no. 4, pp. 745–757, Sep. 2017, Accessed: Aug. 06, 2025. [Online]. Available: https://dergipark.org.tr/en/pub/dumf/issue/33630/408745.

[21] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "EEG based multi-class seizure type classification using convolutional neural network and transfer learning," Neural Networks, vol. 124, pp. 202–212, Apr. 2020, https://doi.org/10.1016/J.NEUNET.2020.01.017.

[22] M. Mursalin, S. Mohammed, S. Islam, K. Noman, and A. A. Al-Jumaily, "Epileptic seizure classification using statistical sampling and a novel feature selection algorithm," Feb. 2019, Accessed: Aug. 06, 2025. [Online]. Available: https://arxiv.org/pdf/1902.09962.

[23] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic Seizure Detection Based on Time-Frequency Analysis and Artificial Neural Networks," Comput Intell Neurosci, vol. 2007, no. 1, p. 080510, Jan. 2007, https://doi.org/10.1155/2007/80510.

[24] G. Ekincioğlu, D. Akbay, and S. Keser, "Estimating Uniaxial Compressive Strength of Sedimentary Rocks with Leeb hardness Using SVM Regression Analysis and Artificial Neural Networks," Journal of Polytechnic, vol. 28, no. 2, pp. 503–512, Mar. 2025, https://doi.org/10.2339/POLITEKNIK.1475944.

[25] S. Huang, I. Arpaci, M. Al-Emran, S. Kılıçarslan, and M. A. Al-Sharafi, "A comparative analysis of classical machine learning and deep learning techniques for predicting lung cancer survivability," Multimed Tools Appl, vol. 82, no. 22, pp. 34183–34198, Sep. 2023, https://doi.org/10.1007/S11042-023-16349-Y/TABLES/8.

[26] P. Puranik and R. Pethe, "Introduction to Explainable AI in EEG Signal Processing: A Review," Explainable Artificial Intelligence in the Healthcare Industry, A. Kumar, Scrivener Publishing LLC, 2025, pp. 79–104, https://doi.org/10.1002/9781394249312.

[27] M. S. Islam, I. Hussain, M. M. Rahman, S. J. Park, and M. A. Hossain, "Explainable Artificial Intelligence Model for Stroke Prediction Using EEG Signal," Sensors, vol. 22, no. 24, p. 9859, Dec. 2022, https://doi.org/10.3390/S22249859.

[28] B. Fresz, E. Dubovitskaya, D. Brajovic, M. F. Huber, and C. Horz, "How Should AI Decisions Be Explained? Requirements for Explanations from the Perspective of European Law," Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, vol. 7, no. 1, pp. 438–450, Oct. 2024, https://doi.org/10.1609/AIES.V7I1.31648.

[29] S. Dal and N. Sezgin, "Heart Attack Classification with a Machine Learning Approach Based on the Random Forest Algorithm," Balkan Journal of Electrical and Computer Engineering, vol. 13, no. 2, pp. 140–147, Jul. 2025, https://doi.org/10.17694/BAJECE.1691905.

[30] F. A. Khan, Z. Umar, A. Jolfaei, and M. Tariq, "Explainable AI for epileptic seizure detection in Internet of Medical Things," Digital Communications and Networks, vol. 11, no. 3, pp. 587–593, Jun. 2025, https://doi.org/10.1016/j.dcan.2024.08.013.

[31] S. Pourbahrami, M. A. Balafar, and L. M. Khanli, "ASVMK: A novel SVMs Kernel based on Apollonius function and density peak clustering," Eng Appl Artif Intell, vol. 126, p. 106704, Nov. 2023, https://doi.org/10.1016/J.ENGAPPAI.2023.106704.

[32] M. Tümay, Z. Civelek, and M. Teke, "Glakom ve Katarakt Hastalığının Derin Öğrenme Modelleri ile Teşhisi," Politeknik Dergisi, vol. 27, no. 5, pp. 1813–1821, Oct. 2024, https://doi.org/10.2339/POLITEKNIK.1348143.

[33] A. A. Hamidi, B. Robertson, and J. Ilow, "A new approach for ECG artifact detection using fine-KNN classification and wavelet scattering features in vital health applications," Procedia Comput Sci, vol. 224, pp. 60–67, Jan. 2023, https://doi.org/10.1016/J.PROCS.2023.09.011.

[34] D. Jiang, X. Shi, Y. Liang, and H. Liu, "Feature extraction technique based on Shapley value method and improved mRMR algorithm," Measurement, vol. 237, p. 115190, Sep. 2024, https://doi.org/10.1016/J.MEASUREMENT.2024.115190.

[35] T. A. A. Abdullah et al., "Sig-Lime: A Signal-Based Enhancement of Lime Explanation Technique," IEEE Access, vol. 12, pp. 52641–52658, 2024, https://doi.org/10.1109/ACCESS.2024.3384277.