

Interpretable EEG-Based Epileptic Seizure Detection Using XGBoost Optimized by an Enhanced Bird Swarm Algorithm

Md Daud Ibrahim¹, Rana Ahmed²,
Anup Chandra Sagor³, Md Belayet Hossain Babla^{4*}

^{1,4}School of Electronics and Information Engineering,
Nanjing University of Information Science and Technology, Nanjing, China

²Electronics and Communication Engineering Discipline,
Khulna University, Khulna, Bangladesh

³School of Computer Science and Technology,
Nanjing University of Information Science and Technology, Nanjing, China

E-mail: ibrahimmdaud@nuist.edu.cn; imranaahmed0101@gmail.com;
202353080061@nuist.edu.cn; belayet@nuist.edu.cn

*Corresponding author details: Md Belayet Hossain Babla; belayet@nuist.edu.cn

ABSTRACT

Epilepsy affects approximately 50 million people worldwide, and because unexpected seizures can cause serious injuries, efficient automated seizure detection is a clinical priority. This study analyzes electroencephalogram (EEG) signals from the Bangalore EEG Epilepsy Dataset (BEED) using an interpretable machine learning framework that differentiates seizure from non-seizure states. Extreme Gradient Boosting (XGBoost) was used as the classifier and optimized with an Enhanced Bird Swarm Algorithm (EBSA) incorporating a dynamic inertia weight, fitness-weighted vigilance, and Lévy flight-based exploration for feature selection and hyperparameter tuning. The dataset was split into 70% training and 30% testing subsets, and stratified 5-fold cross-validation was used to assess reliability. SHapley Additive exPlanations (SHAP), partial dependence plots (PDPs), and a global surrogate decision tree were used to enhance transparency and clinical interpretability. The baseline XGBoost model achieved 99.71% accuracy, 99.61% precision, 100.00% recall, and a 99.81% F1-score. With EBSA-based feature selection, the optimized model used only 11 features and achieved 99.54% accuracy, 99.39% precision, 100.00% recall, and a 99.70% F1-score. The optimization converged within approximately 60 epochs, and 5-fold cross-validation yielded a mean accuracy of 99.74% and a mean recall of 100.00%. SHAP analysis identified X15, X8, and X6 as the most influential features, PDP analysis revealed positive relationships between X1 and X4 and seizure prediction and non-monotonic effects for X3 and X8, and the surrogate tree identified X8 as the primary decision-boundary feature. The framework can be extended to multi-class seizure classification and validated on additional EEG datasets such as Bonn and CHB-MIT.

Keywords: epilepsy; electroencephalography (EEG); XGBoost; feature selection; seizure classification; Bangalore EEG Epilepsy Dataset (BEED); Bird Swarm Algorithm (BSA); SHAP; partial dependence plots; model interpretability

INTRODUCTION

Epilepsy is a chronic neurological disorder characterized by recurrent, unprovoked seizures that significantly disrupt normal brain function. These episodes are primarily attributable to underlying factors such as head injuries, severe central nervous system infections, congenital brain abnormalities, and brain tumors. Without timely diagnosis and appropriate therapeutic intervention, the condition can lead to severe adverse outcomes, including irreversible cognitive impairment, long-term disability, or fatal consequences.

Epilepsy remains one of the most prevalent neurological conditions worldwide, affecting approximately 50 million people [1] and causing an estimated 140,000 deaths in 2021 [2]; people with epilepsy also face a risk of premature death up to three times higher than that of the general population [1]. Although the exact etiology frequently remains unidentified (classified as idiopathic), recognized contributors include stroke, genetic disorders, and cranial trauma [1,3].

In many cases, patients experience unpredictable seizures that cause physical injuries from falls and secondary trauma, particularly if the condition is not detected or managed early [1]. The clinical presentation of these seizures ranges from brief lapses in attention to severe episodes involving complete loss of consciousness, compromised motor control, and, in rare instances, sudden death [1,4]. The global burden of epilepsy spans both urban and rural populations, with incidence rates varying across regions. In resource-limited countries, the prevalence is notably higher because of restricted access to healthcare services and diagnostic tools. In Bangladesh, a national household survey estimated the prevalence of epilepsy at 8.4 per 1,000 individuals, implying that more than one million people are living with the disorder [5]. A major challenge in managing this population is that many individuals do not exhibit consistent clinical symptoms, which makes early seizure detection exceptionally difficult before severe injuries occur [6].

To address these diagnostic challenges, modern neuroimaging and computational tools are widely used. The Bangalore EEG Epilepsy Dataset (BEED), which contains electroencephalogram (EEG) recordings from individuals presenting various seizure types, plays a critical role in modeling seizure patterns and improving automated detection frameworks [7,8]. The BEED dataset categorizes seizures into specific manifestations, encompassing generalized seizures, focal seizures, and distinctive behavioral indicators such as involuntary eye blinking or nail biting [7]. Given the severe clinical risks associated with unmanaged episodes, developing a robust, high-performance computational model for automated early seizure detection is paramount [8,9]. To this end, this study proposes a machine learning framework that leverages Extreme Gradient Boosting (XGBoost) to achieve high accuracy in automated classification. Furthermore, the Bird Swarm Algorithm (BSA) is employed for intelligent feature selection, which optimizes the model by eliminating redundant features and enhancing its overall interpretability. This integrated framework aims to identify epileptic events with high accuracy and processing speed, thereby assisting clinicians in real-time diagnosis and improving long-term patient management.

A substantial body of literature has explored the BEED dataset for EEG-based epileptic seizure detection and classification. For instance, Najmusseher et al. [8] investigated multi-domain EEG features using both the BEED and Bonn datasets, in which time-domain, frequency-domain, and time-frequency-domain features were extracted and classified using supervised learning.

Subsequently, Najmusseher and Banu [10] proposed SeqBoostNet, an architecture that combines uniform manifold approximation and projection (UMAP)-based temporal representations with fast Fourier transform (FFT)-based spectral features within a stacking-based classification structure.

Najmusseher et al. [11] also introduced an adaptive fast Fourier transform with non-uniform sampling (AIFFT-NS) paired with an improved deep belief network (IDBN) for robust seizure classification. More recently, Obaido and Esenogho [12] evaluated a tabular foundation model, TabPFN, to benchmark EEG-based seizure classification on the BEED dataset. Although these pioneering studies achieved promising results, most prior work has focused predominantly on feature extraction and engineering, deep learning, or standard stacking-based classification models. Crucially, limited research has addressed optimizer-based feature selection specifically tailored to the BEED dataset. To bridge this research gap, this study introduces an optimized XGBoost-based seizure detection framework in which feature selection is performed by the BSA. The proposed approach explicitly aims to isolate the most relevant features from the BEED feature pool, reduce computational overhead, maximize model interpretability, and maintain high classification performance.

METHODOLOGY

This study follows the computational pipeline illustrated in Figure 1. The analysis begins with the BEED dataset, which undergoes data preprocessing to improve signal quality. Subsequently, feature selection is applied to identify the most informative attributes, after which an XGBoost classifier is trained on the selected features. Finally, model performance is evaluated using standard classification metrics to ensure both predictive accuracy and interpretability.

First, the raw EEG data from the BEED dataset undergo thorough preprocessing, including data cleaning, noise removal, and signal normalization, to ensure consistency and readiness for analysis. The dataset is then divided into two parts: 70% of the data are used for training the model, and the remaining 30% are used for testing.

The model is subsequently optimized using the BSA for feature selection. The BSA identifies the most relevant features, reducing unnecessary complexity while improving the performance and interpretability of the model. For the classification task, the XGBoost algorithm is trained on the selected features from the BEED dataset. XGBoost is a robust machine learning model known for its high accuracy in classification tasks.

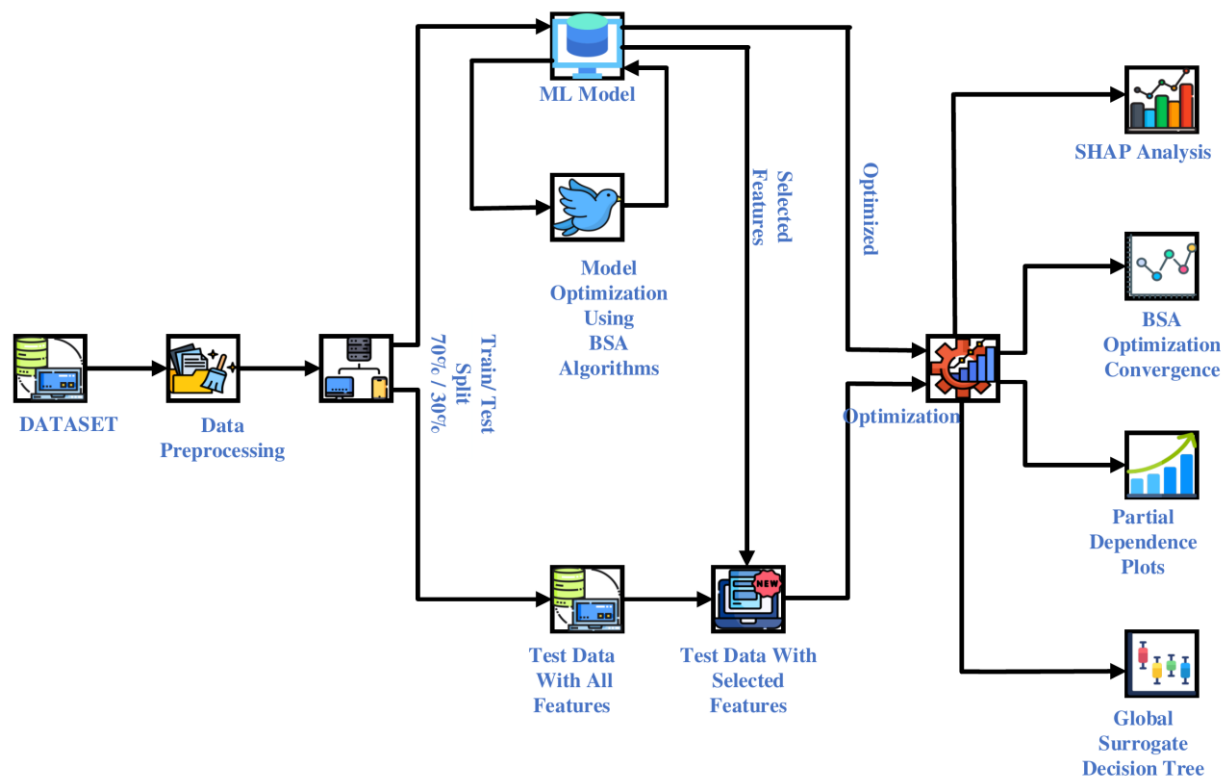


FIGURE 1: Proposed Computational Research Framework.

In this research, model performance is evaluated using several techniques. SHAP analysis is used to assess the importance of each feature in the model's decision-making process, providing transparency. In addition, the convergence of the BSA optimization is analyzed to ensure effective optimization. PDPs illustrate how key EEG channel features influence the model's predictions of seizure and non-seizure states. Finally, a global surrogate decision tree is used to approximate the decisions made by the optimized XGBoost model, providing a simplified and interpretable view of the model's logic. This methodology is designed to ensure that the epilepsy detection system not only performs efficiently but also remains interpretable, supporting its suitability for real-world clinical use.

Bangalore EEG Epilepsy Dataset (BEED)

The BEED is a publicly available dataset designed specifically for research on epilepsy detection and seizure classification. Collected at a neurological research center in Bangalore, India, the dataset includes EEG recordings from healthy subjects and from patients with various types of epileptic seizures, and it is increasingly used for training and testing machine learning models aimed at automatic seizure detection [7].

The EEG signals were recorded from 16 channels at a sampling rate of 256 Hz, providing detailed information about the electrical activity of the brain [7]. Each EEG signal is annotated with the presence of seizures or normal activity, making the dataset well suited for supervised learning tasks. The electrode placement used for recording the EEG signals follows the international 10–20 system [13], as shown in Figure 2.

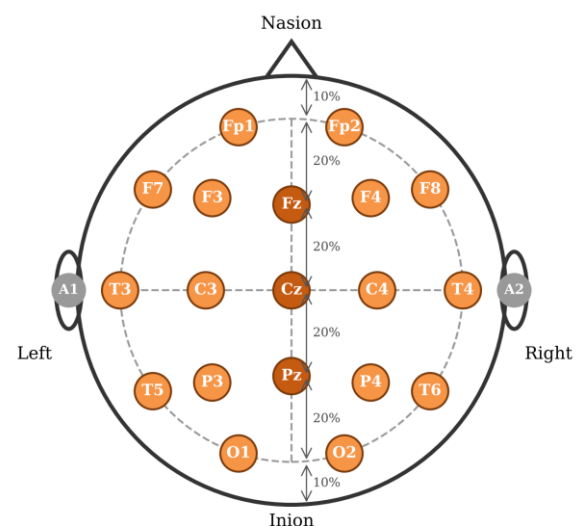


FIGURE 2: Electrode Positions of the International 10–20 System (Schematic Top View) [13].

Dataset Composition

The BEED dataset comprises EEG recordings from 80 adult participants (20 per class), each contributing 100 segments, yielding a total of 8,000 samples (2,000 per class). The dataset contains 16 numerical input features (X1–X16), each corresponding to one EEG channel [7]. The target variable (y) consists of four classes: No Seizure (Class 0), representing healthy subjects with normal brain activity; Seizure (Class 1), representing generalized tonic-clonic seizures; Seizure (Class 2), representing focal seizures limited to specific brain regions; and Seizure (Class 3), representing seizure-like events such as eye blinking or nail biting. Standard preprocessing techniques, including normalization, filtering, and artifact removal, were applied to ensure data quality.

Although the four original classes are balanced, merging Classes 1–3 into a single Seizure class in the binary formulation adopted in this study yields a 3:1 ratio (6,000 Seizure versus 2,000 No Seizure samples). This imbalance was addressed through stratified data partitioning and by reporting precision, recall, and F1-score alongside accuracy. The dataset is summarized in Table 1.

TABLE 1: Summary of Dataset Features.

Variable	Role	Type	Missing values
X1	Feature	Integer	No
X2	Feature	Integer	No
X3	Feature	Integer	No
X4	Feature	Integer	No
X5	Feature	Integer	No
X6	Feature	Integer	No
X7	Feature	Integer	No
X8	Feature	Integer	No
X9	Feature	Integer	No
X10	Feature	Integer	No
X11	Feature	Integer	No
X12	Feature	Integer	No
X13	Feature	Integer	No
X14	Feature	Integer	No
X15	Feature	Integer	No
X16	Feature	Integer	No
y	Target	Integer	No

Figure 3 shows the correlation heatmap of the selected features, which illustrates the relationships among the EEG features. The Pearson correlation coefficient is used to measure these relationships, which are categorized into three types:

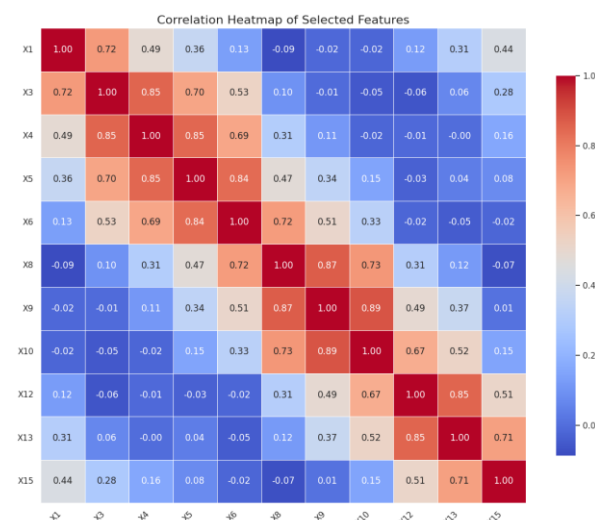


FIGURE 3: Correlation Heatmap of the Selected Features.

Strong positive linear relationship: Both features move in the same direction; as one feature increases, the other also increases. A perfect positive correlation is denoted by a coefficient of +1.

Strong negative linear relationship: Also known as a strong inverse relationship, this occurs when two features move in opposite directions; as one feature increases, the other decreases. A perfect negative correlation is denoted by a coefficient of -1.

No correlation: A change in one feature does not result in a predictable change in another, as indicated by a coefficient near 0.

In this context, let the output label be y , and the input features be $X1, X2, \dots, X16$. For illustration, suppose that high correlations are observed between features and the target y :

$$\begin{aligned} (X1, X2) &\rightarrow y \text{ (correlation: 0.90)} \\ (X1, X3) &\rightarrow y \text{ (correlation: 0.80)} \\ (X1, X16) &\rightarrow y \text{ (correlation: 0.65)} \end{aligned}$$

In this case, features such as $X1, X2$, and $X3$ (with high correlation) are retained for further analysis, whereas $X16$ (with lower correlation) is removed.

Classification Model

In this study, the XGBoost classifier is used for EEG-based epileptic seizure detection on the BEED dataset [7]. XGBoost is a powerful and fast machine learning algorithm built on decision trees [14]. It is an ensemble method that creates a strong final model by combining a series of trees, in which each new tree corrects the errors made by the previous ones.

Several key properties enable XGBoost to perform well in this research, including regularization to prevent overfitting, the handling of missing data, fast histogram-based training, and efficient memory usage. These properties provide high accuracy and a strong ability to handle complex data and non-linear relationships, making XGBoost a strong choice for classifying EEG-based seizures [15,16].

In this research, the original BEED labels were simplified into a binary classification format, in which Class 0 denotes the non-seizure state and Class 1 denotes the seizure state (original Classes 1–3 merged). The XGBoost model was trained using features extracted from the EEG data to separate seizure samples from non-seizure samples. Following the standard XGBoost formulation of Chen and Guestrin [14], the objective function combines a training loss and a regularization penalty, as expressed in Eq. (1):

$$\text{Obj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

Where $l(y_i, \hat{y}_i)$ represents the loss between the actual label y_i and the predicted output $\hat{y}_i$, n is the number of training samples, K is the number of trees, and $\Omega(f_k)$ denotes the regularization term used to control the complexity of the k th decision tree f_k , thereby reducing overfitting and improving the generalization ability of the model [14].

• Regularization

XGBoost includes regularization in its objective function to control model complexity and reduce overfitting [14]. The regularization term penalizes overly complex trees and helps the model generalize better to unseen data. It mainly includes two types of regularization:

1. *L1 regularization (Lasso)*: Encourages sparsity in the leaf weights by penalizing their absolute values, thereby reducing the influence of less informative splits.
2. *L2 regularization (Ridge)*: Improves model stability by preventing large leaf weights.

By balancing the training loss with the regularization term, XGBoost avoids memorizing the training data and improves its performance on unseen EEG samples [14]. The hyperparameters of the XGBoost classifier and their search ranges are listed in Table 2.

TABLE 2: Hyperparameter Search Ranges for the XGBoost Classifier.

Hyperparameter	Type	Range
Number of estimators (n_estimators)	Integer	100–1000
Maximum depth (max_depth)	Integer	3–15
Learning rate (learning_rate)	Float	0.001–0.3
Subsampling ratio (subsample)	Float	0.5–1

Evaluation Metrics

Model performance was assessed using the following confusion-matrix outcomes, with the seizure class (Class 1) treated as the positive class:

- *True positive (TP)*: a positive sample correctly identified as positive.
- *False negative (FN)*: a positive sample incorrectly identified as negative.
- *False positive (FP)*: a negative sample incorrectly identified as positive.
- *True negative (TN)*: a negative sample correctly identified as negative.

Performance was evaluated using the metrics defined in Eqs. (2)–(5):

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\begin{aligned} \text{F1-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= \frac{2TP}{2TP+FP+FN} \end{aligned} \quad (5)$$

Feature Selection Algorithm

Feature selection (FS) is the process of identifying and extracting the most relevant subset of features from an original dataset. In most datasets, not all features contribute equally to model performance; therefore, FS plays a vital role in selecting the most significant features to enhance accuracy. In this research, the metaheuristic BSA was used to identify the most prominent features and hyperparameters for the classifier. The BSA is a swarm intelligence-based optimization technique inspired by the foraging, vigilance, and flight behaviors observed in bird communities; it was first proposed by Meng et al. in 2016 [17]. We propose an enhanced BSA (EBSA) that incorporates a dynamic inertia weight, a fitness-weighted vigilance mechanism, and Lévy flight with self-learning to overcome stagnation in local optima. Unless otherwise stated, “BSA” in the remainder of this paper refers to this enhanced variant. The three modifications are described below.

• Foraging Behavior with Dynamic Inertia Weight

In standard foraging models, birds move on the basis of their current positions. To improve the search process, we introduce an inertia weight (w), which helps the algorithm explore a wider area at the beginning of the search and focus on finding a precise optimal solution towards the end. The position of each bird is updated using Eq. (6):

$$\begin{aligned} x_{i,j}^{t+1} &= w \cdot x_{i,j}^t + C \cdot \text{rand}(0,1) \cdot (p_{i,j} - x_{i,j}^t) \\ &\quad + S \cdot \text{rand}(0,1) \cdot (g_j - x_{i,j}^t) \end{aligned} \quad (6)$$

Where $x_{i,j}^{t+1}$ is the new position of bird i in dimension j , $x_{i,j}^t$ is its current position (scaled by the inertia weight w), $p_{i,j}$ is the personal best position of bird i (cognitive component), g_j is the global best position (social component), C and S are the cognitive and social acceleration coefficients, respectively, and $\text{rand}(0,1)$ denotes a random number uniformly distributed in $[0, 1]$. The value of the inertia weight changes at every iteration according to Eq. (7):

$$w = w_{\max} - (w_{\max} - w_{\min}) \cdot \frac{t}{\text{MaxIter}} \quad (7)$$

Where $w_{\max}$ and $w_{\min}$ are the maximum and minimum inertia weights, t is the current iteration, and MaxIter is the maximum number of iterations.

• Vigilance Behavior Based on a Fitness-Weighted Mean

In this study, we propose an improved vigilance behavior for the BSA to enhance its convergence accuracy. The standard BSA treats all birds in the swarm as equally important during the vigilance behavior. However, birds with better fitness usually have a stronger influence on the group's movement. To implement this, we introduce a fitness-weighted mean (M_w), which is computed in two main steps.

Weighting process: First, a weight (W_i) is assigned to each bird on the basis of its fitness value using Eq. (8):

$$W_i = \frac{f_{\text{worst}} - f(x_i)}{f_{\text{worst}} - f_{\text{best}}} \quad (8)$$

Where $f(x_i)$ is the current fitness of bird i , and f_{best} and f_{worst} are the best (lowest) and worst (highest) fitness values in the swarm, respectively. Because the fitness (1 - F1-score) is minimized, this weighting assigns $W_i = 1$ to the best bird and $W_i = 0$ to the worst, ensuring that birds with better performance are given higher importance.

Position update based on the weighted center: The swarm then moves toward a center (M_w) that is heavily influenced by the most successful birds, as given by Eq. (9):

$$M_w = \frac{\sum (W_i \cdot x_i)}{\sum W_i} \quad (9)$$

The new position of each bird ($x_{i,j}^{t+1}$) is then obtained using Eq. (10):

$$x_{i,j}^{t+1} = x_{i,j}^t + A_1 \cdot (M_w - x_{i,j}^t) \cdot \text{rand}(0,1) + A_2 \cdot (p_{k,j} - x_{i,j}^t) \cdot \text{rand}(-1,1) \quad (10)$$

Where k ($k \neq i$) is a randomly selected bird, $\text{rand}(-1,1)$ is a random number uniformly distributed in $[-1, 1]$, and A_1 and A_2 are the fitness-dependent coefficients of the original BSA [17], defined in Eqs. (11) and (12):

$$A_1 = a_1 \cdot \exp\left(-\frac{\text{pFit}_i}{\text{sumFit} + \varepsilon}\right) \quad (11)$$

$$A_2 = a_2 \cdot \exp\left(\frac{\text{pFit}_i - \text{pFit}_k}{|\text{pFit}_i - \text{pFit}_k| + \varepsilon} \cdot \frac{N \cdot \text{pFit}_k}{\text{sumFit} + \varepsilon}\right) \quad (12)$$

Where a_1 and a_2 are positive constants in $[0, 2]$, pFit_i is the best fitness value of bird i , sumFit is the sum of the best fitness values of the swarm, N is the population size, and ε is a small constant that prevents division by zero [17]. By using this modified logic, the algorithm can prioritize high-quality search areas, which leads to a more stable and faster path to the global optimum.

• Flight Behavior with Lévy Flight and Self-Learning

In this study, the flight phase was enhanced using two mechanisms. First, Lévy flight was incorporated into the producer phase to improve global exploration through occasional long-distance jumps, allowing the algorithm to escape local traps [18], as expressed in Eq. (13):

$$x_{i,j}^{t+1} = x_{i,j}^t \cdot (1 + \text{step} \cdot \text{Lévy}(\lambda)) \quad (13)$$

Where step is the step-size scaling factor and $\text{Lévy}(\lambda)$ is a random step drawn from a heavy-tailed Lévy distribution with stability index λ [18]. This form follows the producer update of the original BSA [17], with the Gaussian step replaced by a Lévy step. Second, the scrounger phase was extended with self-learning. In the original BSA, a scrounger i follows a randomly selected producer k according to Eq. (14) [17]:

$$x_{i,j}^{t+1} = x_{i,j}^t + \text{FL} \cdot \text{rand}(0,1) \cdot (x_{k,j}^t - x_{i,j}^t) \quad (14)$$

Where $\text{FL} \in [0, 2]$ indicates the degree to which the scrounger follows the producer. In the proposed variant, a self-learning term is added to this update so that each bird balances following the leader (x_k) and its own best experience (p_i), maintaining a better balance between social and cognitive intelligence, as given in Eq. (15):

$$x_{i,j}^{t+1} = x_{i,j}^t + \text{FL} \cdot \text{rand}(0,1) \cdot (x_{k,j}^t - x_{i,j}^t) + c_s \cdot \text{rand}(0,1) \cdot (p_{i,j} - x_{i,j}^t) \quad (15)$$

Where c_s is the self-learning coefficient.

SHAP Analysis

In this analysis, the SHapley Additive exPlanations (SHAP) technique was employed to interpret the contribution of individual features to the model's decisions on the dataset [19]. Rooted in cooperative game theory [20], SHAP provides a mathematically rigorous and theoretically grounded framework for distributing the final prediction fairly among the input features, treating each feature as a "player" contributing to a total reward [19]. This choice of explainability method is particularly advantageous in clinical applications, where understanding the reasoning behind model predictions is as critical as the predictions themselves [21]. The horizontal axis of the SHAP summary plot measures the impact of each feature on the model output: points to the right of the center line indicate a positive influence toward predicting Seizure (Class 1), whereas points to the left indicate a tendency toward No Seizure (Class 0) [19]. By integrating SHAP analysis into the evaluation pipeline, this study supports responsible AI development, ensuring that predictive performance is complemented by explainability, a quality increasingly demanded by both the scientific community and regulatory bodies in healthcare AI [21,22].

Convergence Plot

Convergence curves are essential diagnostic tools in optimization studies because they visualize the progression of the objective function value across iterations, thereby providing insight into the search behavior and efficiency of the optimization algorithm [17]. In this study, the best fitness value is defined as (1 - F1-score), which serves as the objective function to be minimized by the BSA. This formulation ensures that minimizing the fitness value directly corresponds to maximizing the classification performance of the model [17]. The convergence curve was monitored over 100 epochs (iterations) to evaluate the stability, exploration capability, and exploitation efficiency of the BSA during the feature selection and hyperparameter optimization process.

Interpretability Analysis Using Partial Dependence Plots (PDPs)

PDPs were employed in this study to gain deeper insight into how individual features influence the model's output [23]. Unlike global feature-importance metrics, PDPs visualize the marginal effect of a specific

feature on the predicted outcome while averaging out the influence of all other features [24]. This makes PDPs particularly valuable in clinical settings, where understanding the directional relationship between a feature and a predicted class is essential for informed decision-making. In this analysis, PDPs were generated separately for Seizure Activity (Class 1) and Normal State (Class 0, i.e., No Seizure) to capture class-specific feature behavior and to validate the logical consistency of the proposed model.

Global Surrogate Model Analysis

To enhance the interpretability of the XGBoost model, a global surrogate model was constructed using a decision tree approximation [24]. A surrogate model is an inherently interpretable model that is trained to approximate the predictions of a complex black-box model, thereby providing human-readable decision rules that reflect the overall behavior of the original model [25]. In this study, the surrogate decision tree was fitted to the predictions generated by the optimized XGBoost model, using the Gini impurity criterion to measure node purity at each split. This approach enables clinicians and domain experts to understand the global decision-making logic of the model through a transparent tree structure without sacrificing the predictive power of the underlying XGBoost classifier.

Reproducibility

To ensure reproducibility, all experiments were conducted in Python 3.10 using XGBoost (v2.0.3), scikit-learn (v1.3.0), SHAP (v0.48.0), and Matplotlib (v3.7.1). A fixed random seed of 42 was used in all experiments to ensure consistent and reproducible results. The BEED dataset used in this study is publicly available from the UCI Machine Learning Repository [7]. The data were divided into training (70%) and testing (30%) subsets and stratified 5-fold cross-validation was used to evaluate model performance. For the BSA optimization, the population size was 30, with 100 iterations for the final optimization run and 20 iterations for each cross-validation fold; the flight behavior was performed every 10 iterations, and the foraging probability was 0.8. Each bird encodes a candidate solution as a vector in $[0, 1]^{20}$, with one element per feature (16) and one element per hyperparameter (4; Table 2). A feature is retained when its element exceeds 0.5, and each hyperparameter element is mapped linearly onto its range in Table 2. The control parameters were $w_{\max} = 0.9$, $w_{\min} = 0.4$, $C = S = 1.5$, $a_1 = a_2 = 1$, FL drawn uniformly from $[0.5, 0.9]$, $c_s = 0.5$, $\text{step} = 0.01$, and $\lambda = 1.5$ [17,18]. The fitness ($1 - \text{F1-score}$) of each candidate was estimated by stratified three-fold cross-validation on the training data. In the final run, the hyperparameters were held at their cross-validation means while the feature subset was optimized. A reference implementation of the complete XGBoost-BSA pipeline (ebsa_xgboost.py) is provided as supplementary material. Details of the BEED dataset and the BSA are provided in the corresponding subsections above.

RESULTS

This section evaluates both the baseline and optimized models. First, the baseline model was evaluated to establish a benchmark. Next, the optimized XGBoost model was analyzed with a focus on hyperparameter tuning and the cost function, followed by its fold-wise performance metrics and the performance obtained with the mean optimized hyperparameters. Finally, SHAP values were calculated to show how each feature affects the output, the BSA convergence curve was examined, PDPs were used to visualize the marginal effects of features on the predicted probabilities of Class 0 and Class 1, and a global surrogate decision tree was applied to approximate the decision logic of the optimized XGBoost model. Together, these three explainable AI (XAI) techniques and the convergence analysis address both predictive accuracy and model transparency.

Table 3 presents the performance of the XGBoost classifier on the BEED dataset with and without feature selection. Without removing any features, the model achieved 99.71% accuracy, 99.61% precision, 100.00% recall, and a 99.81% F1-score. The 100.00% recall is clinically critical because missed seizure events can adversely affect patients; in this setting, the model did not miss any seizure cases, suggesting strong suitability for clinical decision support. Next, the BSA was used to optimize the feature set. The BSA selected 11 features: X1, X3, X4, X5, X6, X8, X9, X10, X12, X13, and X15. The optimized model achieved 99.54% accuracy, 99.39% precision, 100.00% recall, and a 99.70% F1-score. Recall remained at 100.00%, indicating that no seizure samples in the test set were missed. The BSA reduced the number of features by 31.25% (from 16 to 11) with only a marginal decrease in performance. This reduction makes the XGBoost model lighter and potentially faster for hospital use. These findings indicate that the XGBoost classifier, combined with BSA-based feature selection, provides an effective framework for EEG-based seizure detection on the BEED dataset.

TABLE 3: Performance Comparison of the XGBoost Classifier with and without Feature Selection.

Metric	Without feature selection	With feature selection
Accuracy (%)	99.71	99.54
F1-score (%)	99.81	99.70
Precision (%)	99.61	99.39
Recall (%)	100.00	100.00
Number of features	16	11
Features used	All (X1–X16)	X1, X3, X4, X5, X6, X8, X9, X10, X12, X13, X15

Cross-Validation Performance of the Optimized Model

As shown in Figure 4, the BSA successfully minimized the fitness value ($1 - \text{F1-score}$) across all five cross-validation folds. Each plot illustrates the progression of the objective value over the approximately 20 epochs of each per-fold optimization run (the final optimization run, presented later in this section, was monitored over 100 epochs). Folds 1 and 2 exhibit early improvements, with noticeable drops in the fitness value occurring at

approximately epochs 4 and 3, respectively, before stabilizing. Fold 3 demonstrates a later improvement near epoch 18, after which the curve flattens completely. In contrast, Folds 4 and 5 converge immediately from the first epoch, indicating that the BSA identified near-optimal solutions without requiring further exploration. The rapid and stable convergence observed across all folds suggests that the selected features possess strong discriminative properties, enabling the BSA to locate optimal solutions efficiently with minimal iterations.

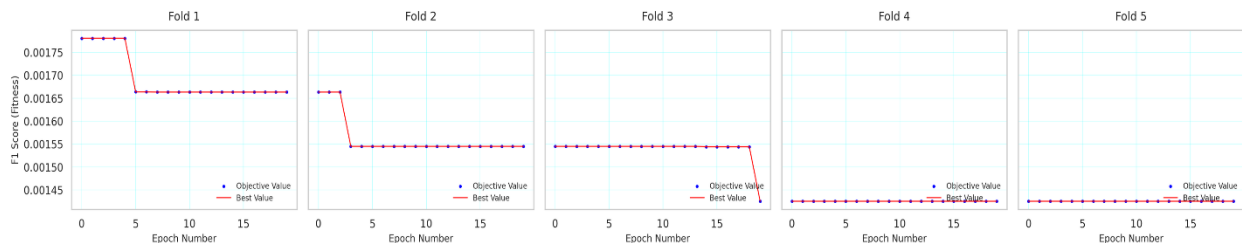


FIGURE 4: Cost Function (Fitness Value) for Each Cross-Validation Fold.

TABLE 4: Model Performance Metrics Per Fold.

Fold	Train				Test			
	Acc (%)	F1 (%)	Pre (%)	Rec (%)	Acc (%)	F1 (%)	Pre (%)	Rec (%)
1	100.00	100.00	100.00	100.00	99.75	99.83	99.67	100.00
2	100.00	100.00	100.00	100.00	99.81	99.88	99.75	100.00
3	100.00	100.00	100.00	100.00	99.75	99.83	99.67	100.00
4	100.00	100.00	100.00	100.00	99.81	99.88	99.75	100.00
5	100.00	100.00	100.00	100.00	99.56	99.71	99.42	100.00
Mean	100.00	100.00	100.00	100.00	99.74	99.83	99.65	100.00

Acc, accuracy; F1, F1-score; Pre, precision; Rec, recall.

The model was evaluated using stratified 5-fold cross-validation, which helps ensure that the results are fair and unbiased. The training and testing results are presented in Table 4. In the training phase, the model achieved 100.00% for all metrics, including accuracy and F1-score, across all five folds, indicating that it fitted the training data well.

In the testing phase, performance varied slightly across folds. Folds 2 and 4 performed best, with 99.81% accuracy, a 99.88% F1-score, 99.75% precision, and 100.00% recall. Folds 1 and 3 also performed well, with 99.75% accuracy, a 99.83% F1-score, 99.67% precision, and 100.00% recall. Fold 5 showed slightly lower performance, with 99.56% accuracy, a 99.71% F1-score, 99.42% precision, and 100.00% recall; nevertheless, it still demonstrated that the model can detect seizures reliably. The consistent performance across all folds suggests that the model does not exhibit bias toward any specific data partition.

On average, the model achieved 99.74% accuracy, a 99.83% F1-score, 99.65% precision, and 100.00% recall. Importantly, the model did not miss any

seizures, which is critical in a medical context. The results were also highly consistent across folds, suggesting that the model generalizes well to unseen data within the BEED dataset. These results demonstrate that the XGBoost-BSA framework is accurate, stable, and reliable for seizure detection, highlighting its potential applicability in real-world clinical settings.

Figure 5 shows the confusion matrix for each fold of the BSA-optimized XGBoost model, where Class 0 denotes No Seizure and Class 1 denotes Seizure. The model showed strong performance: in every fold, it correctly identified all 1,200 Seizure instances as Class 1, with no false negatives. The only errors occurred in the No Seizure class (Class 0). In Folds 1 and 3, the model correctly identified 396 No Seizure instances and misclassified only 4 as Seizure. Folds 2 and 4 performed slightly better, with only 3 errors and 397 correctly classified No Seizure instances. Fold 5 showed the lowest performance among the folds; it misclassified 7 No Seizure instances as Seizure and correctly identified 393, while still classifying all Seizure cases correctly.

Figure 6 shows the receiver operating characteristic (ROC) curve for each fold of the BSA-optimized XGBoost model. The ROC curve shows how well the model separates the two classes by plotting the true positive rate against the false positive rate at different classification thresholds. Folds 1 and 2 achieved areas under the curve (AUCs) of 0.9995 and 0.9992, respectively; Folds 3 and 4 achieved

0.9993 and 0.9994, respectively; and Fold 5 achieved 0.9993. Every fold achieved an AUC above 0.999, indicating that the model distinguishes between seizure and non-seizure EEG samples very well regardless of the data partition. These results suggest that the XGBoost model is highly reliable for seizure detection.

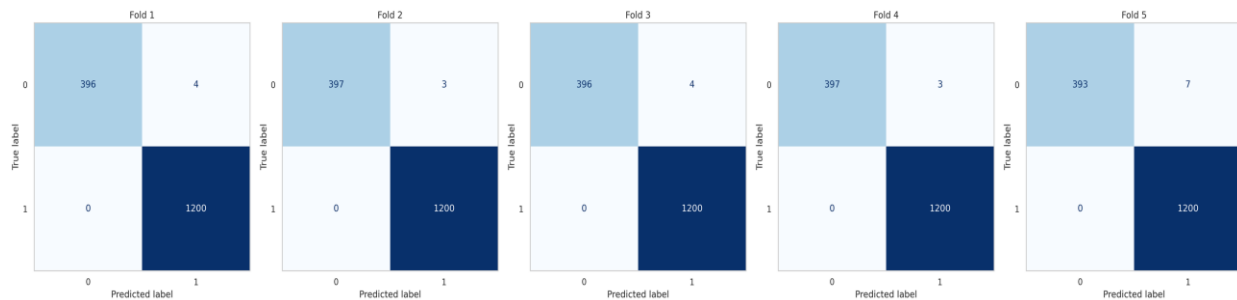


FIGURE 5: Fold-Wise Confusion Matrices of the BSA-Optimized XGBoost Model.

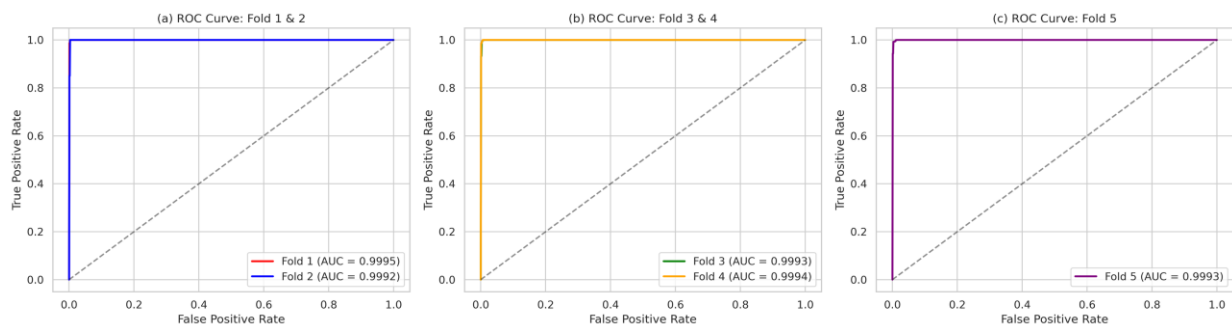


FIGURE 6: Fold-wise ROC Curves: (A) Folds 1 and 2, (B) Folds 3 and 4, and (C) Fold 5.

Performance with Mean Optimized Hyperparameters of XGBoost

The average hyperparameter values obtained from the five cross-validation folds were calculated for the XGBoost model. This approach helps ensure that the chosen parameters are not biased toward a single data split, making the model more reliable. To prevent overfitting and improve predictions, the four hyperparameters listed in Table 2 ($n_{\text{estimators}}$, max_depth , learning_rate , and subsample) were tuned. As shown in Table 5, their average values were 461, 8, 0.2619, and 0.7800, respectively. Using 461 estimators with a learning rate of 0.2619 provides a good balance between training speed and accuracy, while a subsample ratio of 0.7800 helps prevent the model from overfitting the training data. These average hyperparameter values were then used with the BSA to select important features and evaluate the overall performance. This step is intended to make the classification metrics stable, accurate, and easy to reproduce.

TABLE 5: Average Hyperparameter Values.

Hyperparameter	Average value
$n_{\text{estimators}}$	461.0
max_depth	8.0
learning_rate	0.2619
subsample	0.7800

Figure 7 shows the confusion matrix and the ROC curve of the optimized XGBoost model. As illustrated by the confusion matrix, the model correctly classified 589 instances of Class 0 and 1,800 instances of Class 1. The model misclassified 11 instances, all of which were Class 0 samples predicted as Class 1. In contrast, the model classified all Class 1 instances correctly, without any errors. The ROC curve also indicates excellent performance, with an AUC of 0.9997. These results indicate that the selected hyperparameters enable the XGBoost model to distinguish between positive and negative samples with little overlap between the two classes.

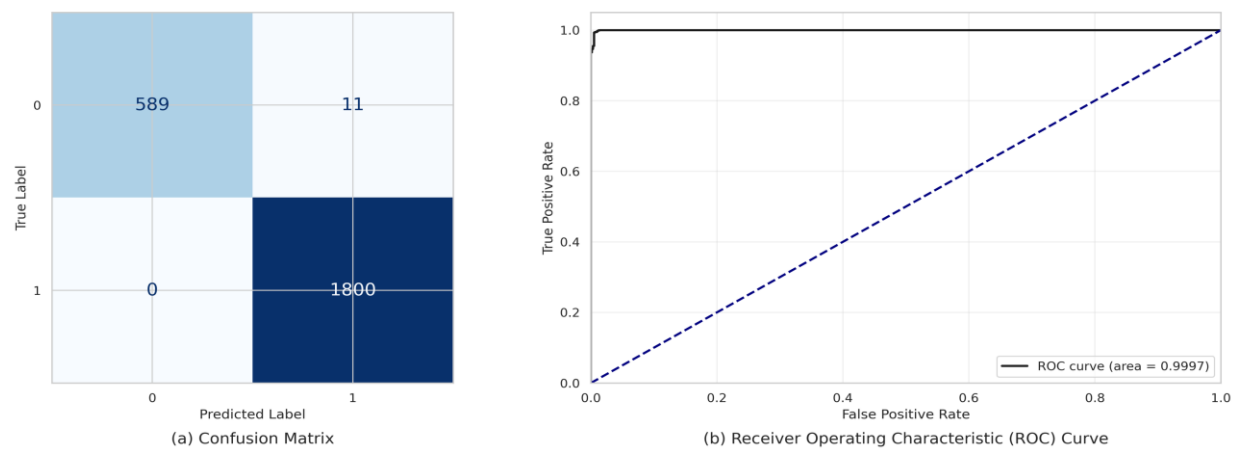


FIGURE 7: (A) Confusion Matrix and (B) ROC Curve of the XGBoost Model with the Mean Optimized Hyperparameters.

SHAP Analysis

Figure 8 illustrates the feature importance ranking and the distribution of SHAP values for the most significant attributes. The plot identifies X15 as the most influential feature in the model's predictions, followed by X8 and X6, as indicated by their broad distributions of SHAP values along the horizontal axis. This directional interpretation allows clinicians and domain experts to assess directly which features drive the model toward a particular diagnosis, thereby bridging the gap between machine learning outputs and clinical decision-making.

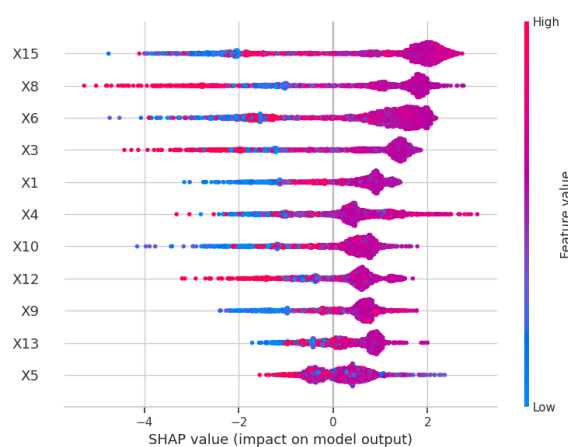


FIGURE 8: Feature Importance Ranking and Distribution Based on Shap Values.

For instance, high values of X15 (represented by red clusters) are predominantly positioned on the right side, suggesting that an increase in this variable strongly drives the model toward Seizure classification. Similarly, X8 exhibits a wide spread of SHAP values in both directions, indicating a strong but mixed influence on the model output depending on the feature value magnitude. In contrast, higher values of X3 tend to push the model toward a No Seizure prediction, reflecting an inverse relationship with the Seizure class. Meanwhile, X5 shows values concentrated near zero, suggesting a negligible contribution to the model's final decision.

Furthermore, the consistency of SHAP value distributions across the feature set suggests that the model has learned stable and meaningful decision boundaries rather than relying on spurious correlations. The identification of X15, X8, and X6 as dominant predictors may also provide valuable clinical insights, potentially guiding future biomarker research and feature selection strategies in seizure detection studies.

Convergence Plot

Figure 9 presents the BSA optimization convergence curve, illustrating the progression of the best fitness value ($1 - F1\text{-score}$) over 100 epochs. The curve shows a clear staircase-like descending pattern, reflecting the stepwise improvement characteristic of the BSA during its search process.

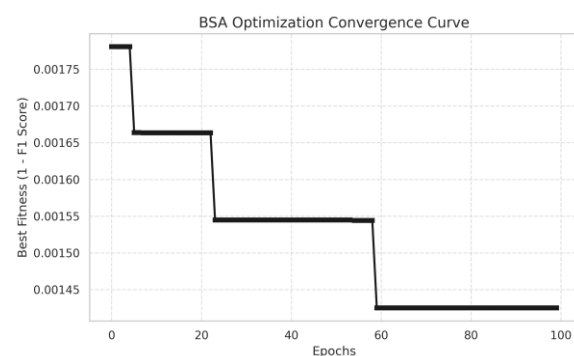


FIGURE 9: Convergence Characteristics of the Proposed BSA-Based Optimization Framework.

Specifically, the fitness value begins at approximately 0.001780 at epoch 0 and undergoes its first significant drop at around epoch 5, settling at approximately 0.001660. A second notable improvement occurs near epoch 22, where the fitness value decreases to approximately 0.001548. The algorithm then stabilizes briefly before achieving another substantial reduction at around epoch 59, reaching a near-optimal fitness value of approximately 0.001430. Beyond epoch 60, the curve remains completely flat, indicating that the algorithm has converged to a stable solution, with no further improvement observed in the remaining iterations.

This stepwise convergence behavior demonstrates that the BSA effectively balances exploration and exploitation during the optimization process. The final fitness value of 0.001430 corresponds to an F1-score of approximately 0.99857, confirming that the BSA successfully identified an optimized feature subset and hyperparameter configuration that yields near-perfect classification performance.

Interpretability Analysis Using Partial Dependence Plots (PDPs)

Figures 10 and 11 present the PDPs of the top features for Normal State (Class 0) and Seizure Activity (Class 1), respectively.

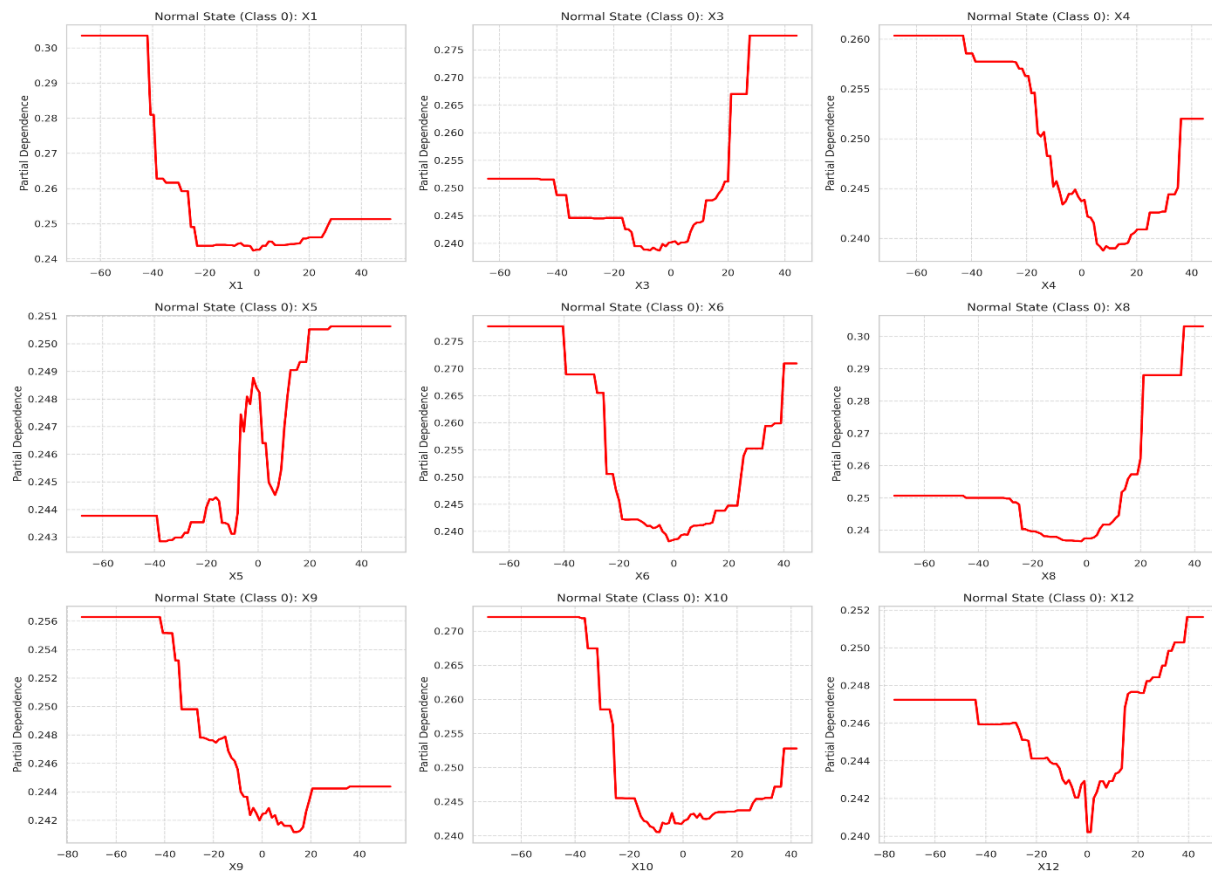


FIGURE 10: PDPs of the Top Features for Normal State (Class 0).

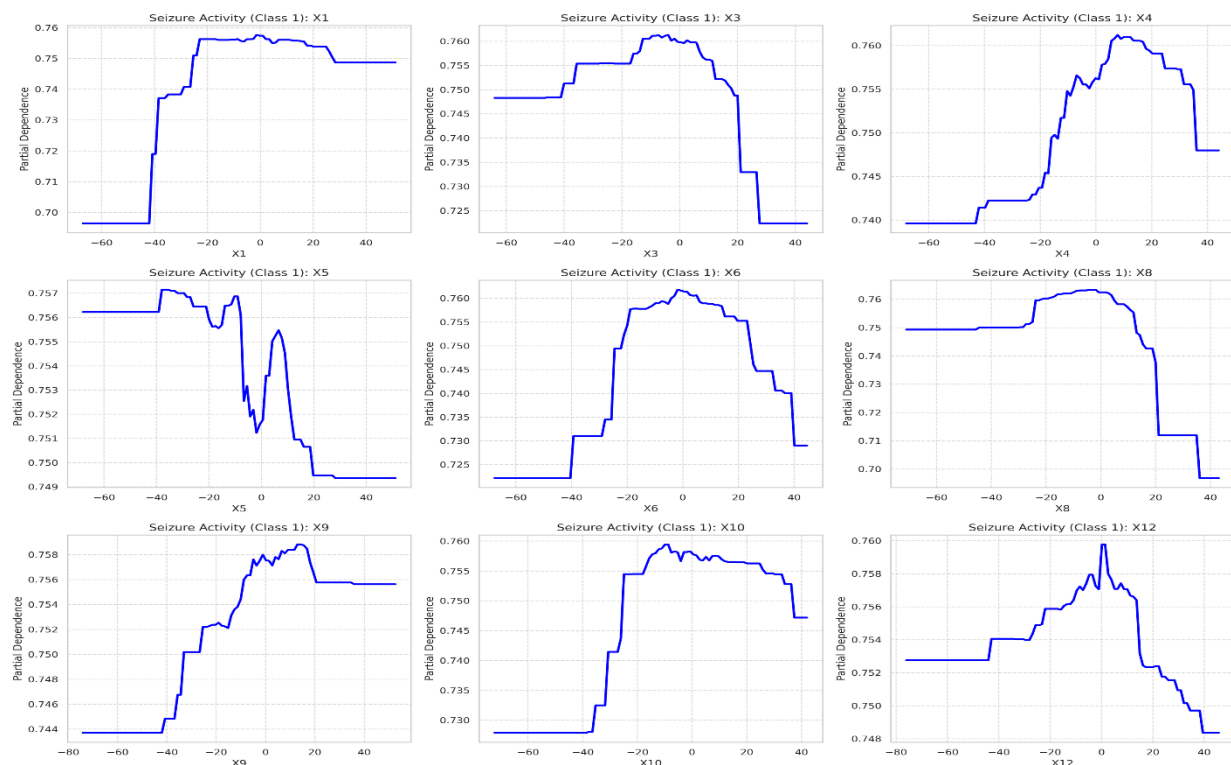


FIGURE 11: PDPs of the Top Features for Seizure Activity (Class 1).

As observed in Figure 11, X1 and X4 exhibit a positive relationship with seizure prediction over most of their ranges: the predicted seizure probability rises as their values increase from the low end of the range toward zero, with only a slight decline at the highest values. X3 shows a non-monotonic pattern, increasing up to values near zero and then dropping sharply at high values, which is consistent with the SHAP result that high X3 values push the model toward No Seizure. X5 exhibits an irregular oscillating pattern near the center, reflecting a complex non-linear relationship with the Seizure class. The variation in partial dependence is modest (approximately 0.03–0.06 in predicted probability), indicating that each feature contributes in combination with others rather than in isolation.

As shown in Figure 10, the PDP patterns for Class 0 exhibit an inverse behavior compared with those for Class 1, as expected for a binary classifier whose two class probabilities sum to one. X8 presents a notably different, U-shaped pattern: the Normal State probability is lowest for values near zero and rises sharply for high values, while the Seizure probability correspondingly decreases at high values. Together with the surrogate tree (Figure 12), in which both very low ($X8 \leq -23.5$) and very high ($X8 > 22.5$) values lead predominantly to Normal State, this indicates that intermediate X8 values are most strongly associated with seizure activity.

The value-dependent patterns identified above are consistent with the SHAP and surrogate-tree results. By employing PDP analysis, this study moves beyond simple global importance metrics to uncover the exact value ranges that drive the decision-making process, providing a transparent and clinically relevant interpretation [24].

Global Surrogate Model Analysis

Figure 12 presents the global surrogate model, represented as a decision tree approximation of the optimized XGBoost model. The tree reveals the key decision boundaries learned by the model and provides a transparent, rule-based explanation of its classification behavior.

Root node (primary decision boundary): The root node splits on $X8 \leq -23.5$, covering 100% of the samples, with a Gini impurity of 0.37 and a class

distribution of [0.245, 0.755], indicating a strong overall bias toward Seizure Activity (Class 1). This establishes X8 as the single most critical feature in the model's global decision-making process.

Left branch ($X8 \leq -23.5$, true; Normal State region): When $X8 \leq -23.5$, the model routes 11.8% of the samples to a predominantly Normal State (Class 0) node with a Gini impurity of 0.094 and a class distribution of [0.951, 0.049]. This branch further splits on $X9 \leq -9.5$; the majority of these samples (10.8%) are classified as Normal State with a very low Gini impurity of 0.03 and a distribution of [0.985, 0.015], confirming high confidence in Normal State prediction for this region. A smaller subset (1.0%) is further split on $X5 \leq -20.0$, ultimately leading to a pure Normal State leaf node (Gini = 0.0).

Right branch ($X8 > -23.5$, false; Seizure Activity region): When $X8 > -23.5$, the model routes 88.2% of the samples toward Seizure Activity (Class 1), with a Gini impurity of 0.257 and a distribution of [0.151, 0.849]. This branch further splits on $X8 \leq 22.5$, where the dominant path covers 80.0% of the samples and splits again on $X15 \leq -34.5$ with a Gini impurity of 0.14. The largest leaf node in this branch covers 77.8% of all samples with a near-pure Seizure classification (Gini = 0.097; distribution = [0.051, 0.949]), confirming that the combination of $-23.5 < X8 \leq 22.5$ and $X15 > -34.5$ is the strongest indicator of Seizure Activity. Conversely, the remaining 8.2% of samples with $X8 > 22.5$ are routed to a predominantly Normal State node (Gini = 0.199; distribution = [0.888, 0.112]), which further splits on $X9 \leq 8.5$.

The surrogate decision tree reveals that X8, X9, X15, and X5 are the most influential features in the global decision logic of the model. The clear separation between the Normal State and Seizure Activity regions, supported by low Gini values at the leaf nodes, indicates that the XGBoost model has learned well-defined and clinically meaningful decision boundaries. This transparent rule-based representation not only validates the model's predictive behavior but also provides actionable insights for clinical interpretation and future biomarker research [24].

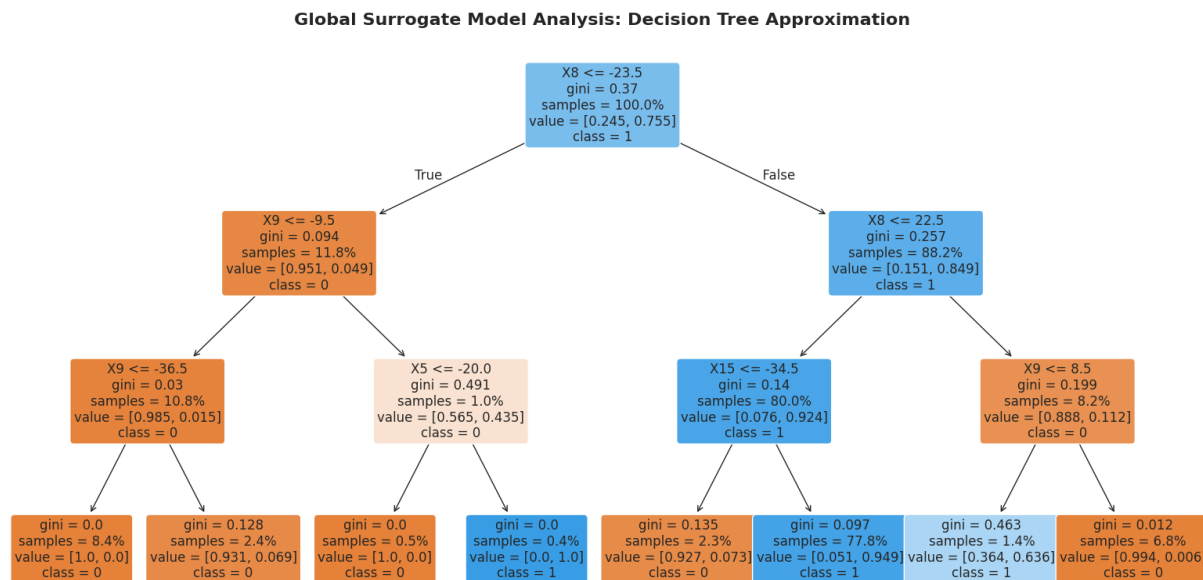


FIGURE 12: Global Surrogate Decision Tree Approximation for Interpreting the Optimized Black-Box Model.

DISCUSSION

Table 6 compares the outcomes of the proposed method with those of previous BEED-based epileptic seizure classification studies. Several studies have explored the BEED dataset for EEG-based epileptic seizure detection and classification. Najmousseher et al. [8] investigated multi-domain EEG features using the BEED and Bonn datasets, in which time-domain, frequency-domain, and time-frequency-domain features were extracted from EEG signals and classified using supervised learning techniques, reporting an accuracy of approximately 99.00% on BEED. Najmousseher and Banu [10] proposed SeqBoostNet by combining UMAP-based temporal representations and FFT-based spectral features with a stacking-based classification model, achieving an average accuracy of 96.71% on the BEED dataset. Another study employed AIFFT-NS and an IDBN for seizure classification, achieving 96.16% accuracy on BEED [11]. More recently, Obaido and Esenogho [12] evaluated a TabPFN-based foundation model for EEG-based seizure classification and reported 99.70% accuracy on the BEED dataset. It should be noted that the present study converts the original BEED labels into a binary non-seizure-versus-seizure classification problem; therefore, comparisons with previous multi-class BEED studies should be interpreted with caution.

Although previous studies achieved promising results, most of them relied mainly on feature extraction, deep learning, stacking-based models, or foundation-model-based classification. In contrast, the proposed study applies an XGBoost classifier optimized with BSA-based feature selection to maintain high classification performance while reducing the number of input features. In this research, the baseline XGBoost model without feature selection achieved an accuracy of 99.71%, an F1-score of 99.81%, a precision of 99.61%, and a recall of 100.00%. After BSA-based feature selection was applied, the model achieved an accuracy of 99.54%, an F1-score of 99.70%, a precision of 99.39%, and a recall of 100.00%.

Although the accuracy decreased slightly after feature selection, the model still maintained excellent classification performance, indicating that the BSA successfully removed less important features without causing any major loss in predictive accuracy. Stratified 5-fold cross-validation was used to assess the reliability of the proposed model. The accuracy for each fold was very high, ranging from 99.56% to 99.81%. On average, the model achieved an accuracy of 99.74%, an F1-score of 99.83%, a precision of 99.65%, and a perfect recall of 100.00%.

For feature selection, the BSA-based method selected only 11 of the original 16 features: X1, X3, X4, X5, X6, X8, X9, X10, X12, X13, and X15. In addition, explainable artificial intelligence techniques were applied to analyze the model's decision-making process. SHAP analysis showed that X15, X8, and X6 were the most influential features for seizure prediction. PDPs further demonstrated how important EEG features influenced the predicted probabilities of the seizure and normal states. Moreover, the global surrogate decision tree provided a simplified rule-based explanation of the optimized XGBoost model. Overall, the proposed XGBoost-BSA framework demonstrates strong potential for automated epilepsy detection using the BEED dataset.

However, this study has some limitations. First, the original multi-class BEED labels were converted into a binary classification problem, so detailed seizure-type classification was not evaluated. Second, the model was validated only on the BEED dataset. Third, data partitioning was stratified by class at the segment level; because each participant contributes multiple segments, segments from the same participant may appear in both the training and test sets, which can yield optimistic performance estimates. In addition, the hyperparameters were derived from cross-validation folds before evaluation on the hold-out set, so a nested, subject-independent validation scheme (e.g., leave-subjects-out) would

provide a stricter estimate of generalization to unseen patients. In future work, the proposed framework can be tested on additional EEG datasets, such as the Bonn and CHB-MIT datasets, while

preserving the original multi-class seizure labels for detailed seizure-type classification and real-time EEG-based seizure detection.

TABLE 6: Comparison of Previous BEED-Based Studies with the Proposed Method.

Reference	Dataset	Method	Results
Najmusseher et al. [8]	BEED and Bonn EEG datasets	Multi-domain features with supervised learning	Accuracy: 99.00% on BEED
Najmusseher and Banu [10]	BEED EEG dataset	UMAP, FFT, and SeqBoostNet	Accuracy: 96.71%
Najmusseher et al. [11]	BEED, Bonn, and CHB-MIT	AIFFT-NS with IDBN	Accuracy: 96.16% on BEED
Obaido and Esenogho [12]	BEED EEG dataset	TabPFN foundation model	Accuracy: 99.70%
Proposed method 1	BEED EEG dataset	XGBoost without feature selection	Accuracy: 99.71% F1-score: 99.81% Precision: 99.61% Recall: 100.00%
Proposed method 2	BEED EEG dataset	XGBoost with BSA feature selection	Accuracy: 99.54% F1-score: 99.70% Precision: 99.39% Recall: 100.00%
Proposed method 3	BEED EEG dataset	5-fold CV XGBoost-BSA	Accuracy: 99.74% F1-score: 99.83% Precision: 99.65% Recall: 100.00%

CONCLUSIONS

This study presented an optimized and interpretable machine learning framework for automated epileptic seizure detection using the BEED dataset. The proposed framework combines an XGBoost classifier with the BSA for feature selection and hyperparameter tuning. The experimental results demonstrate that this optimization approach yields robust classification performance. When evaluated using a stratified 5-fold cross-validation protocol, the model achieved a mean testing accuracy of 99.74%, an F1-score of 99.83%, a precision of 99.65%, and a perfect recall of 100.00%. In addition, the BSA feature selection process successfully streamlined the input space, reducing the feature set from 16 to 11 highly discriminative attributes while maintaining strong predictive performance (99.54% accuracy and 100.00% recall). To address the interpretability limitations common in ensemble architectures, the pipeline integrates three complementary XAI techniques: SHAP analysis, PDPs, and a global surrogate decision tree. These methods successfully mapped the model's underlying decision logic, identifying X15, X8, and X6 as the most influential predictive features. Although the current study focused on binary classification within a single dataset, the proposed framework provides a solid foundation for automated diagnostics; because performance was assessed at the segment level, these estimates should be confirmed with subject-independent validation. Future work will involve validating the system's generalizability on additional clinical EEG datasets, including the Bonn and CHB-MIT repositories, and adapting the architecture to support granular multi-class seizure-type identification in real-time clinical monitoring environments.

REFERENCES

- [1] World Health Organization. (2024). *Epilepsy* [Fact sheet]. <https://www.who.int/news-room/fact-sheets/detail/epilepsy> (accessed 24 April 2026).
- [2] Zhao, J., Chen, Q., Dong, J., Gao, M., Ge, S., & Wang, A. (2026). Global, regional, and national burden of epilepsy, 1990–2021: A Global Burden of Disease study. *Journal of Global Health*, 16, 04066.
- [3] Johns Hopkins Medicine. (2026). *Epilepsy causes*. <https://www.hopkinsmedicine.org/health/conditions-and-diseases/epilepsy/epilepsy-causes> (accessed 24 April 2026).
- [4] Centers for Disease Control and Prevention. (2024). *Health and safety concerns*. <https://www.cdc.gov/epilepsy/health-safety-concerns/index.html> (accessed 24 April 2026).
- [5] Mohammad, Q. D., Saha, N. C., Alam, M. B., Hoque, S. A., Islam, A., Chowdhury, R. N., Hussain, M. E., Chowdhury, Y. S., Hossain, S., Chowdhury, M. A., Rahman, M., Majumder, B. K., Salam, A., Sarker, A., Uddin, M. K., Moniruzzaman, M., Hakim, F., Bhuiyan, R., Anwar, N., & Zaman, M. M. (2020). Prevalence of epilepsy in Bangladesh: Results from a national household survey. *Epilepsia Open*, 5(4), 526–536.
- [6] Pellinen, J., Tafuro, E., Yang, A., Price, D., Friedman, D., Holmes, M., Barnard, S., Detyniecki, K., Hegde, M., Hixson, J., Haut, S., Kälviäinen, R., & French, J. (2020). Focal nonmotor versus motor seizures: The impact on diagnostic delay in focal epilepsy. *Epilepsia*, 61(12), 2643–2652.

- [7] Najmusseher & Banu, P. K. N. (2024). *BEED: Bangalore EEG Epilepsy Dataset* [Dataset]. UCI Machine Learning Repository. <https://archive.ics.uci.edu/dataset/1134/bee+d:+bangalore+eeg+epilepsy+dataset> (accessed 24 April 2026).
- [8] Najmusseher, Banu, P. K. N., Azar, A. T., & Kamal, N. A. (2024). Impact of multi-domain features for EEG based epileptic seizures classification. In *Proceedings of the 10th International Conference on Advanced Intelligent Systems and Informatics 2024* (Lecture Notes on Data Engineering and Communications Technologies, Vol. 220, pp. 317–329). Cham: Springer.
- [9] Ahmad, I., Yao, C., Li, L., Chen, Y., Liu, Z., Ullah, I., Shabaz, M., Wang, X., Huang, K., Li, G., Zhao, G., Samuel, O. W., & Chen, S. (2024). An efficient feature selection and explainable classification method for EEG-based epileptic seizure detection. *Journal of Information Security and Applications, 80*, 103654.
- [10] Najmusseher & Banu, P. K. N. (2025). Feature engineering for epileptic seizure classification using SeqBoostNet. *International Journal of Computing and Digital Systems, 17*(1), 1–15.
- [11] Najmusseher, Banu, P. K. N., & Janardhan, D. C. (2024). Automated epileptic seizure classification using adaptive fast Fourier transform with non-uniform sampling and improved deep belief network. *International Journal of Intelligent Engineering Informatics, 12*(4), 460–512.
- [12] Obaido, G. & Esenogho, E. (2026). Evaluating EEG-based seizure classification using foundation and classical ensemble models. *Applied Sciences, 16*(7), 3120.
- [13] Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.
- [14] Chen, T. & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- [15] Vanabelle, P., De Handschutter, P., El Tahry, R., Benjelloun, M., & Boukhebouze, M. (2020). Epileptic seizure detection using EEG signals and extreme gradient boosting. *The Journal of Biomedical Research, 34*(3), 228–239.
- [16] Wu, J., Zhou, T., & Li, T. (2020). Detecting epileptic seizures in EEG signals with complementary ensemble empirical mode decomposition and extreme gradient boosting. *Entropy, 22*(2), 140.
- [17] Meng, X.-B., Gao, X. Z., Lu, L., Liu, Y., & Zhang, H. (2016). A new bio-inspired optimisation algorithm: Bird swarm algorithm. *Journal of Experimental and Theoretical Artificial Intelligence, 28*(4), 673–687.
- [18] Yang, X.-S. & Deb, S. (2009). Cuckoo search via Lévy flights. In *Proceedings of the World Congress on Nature & Biologically Inspired Computing (NaBIC 2009)* (pp. 210–214). IEEE.
- [19] Lundberg, S. M. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.
- [20] Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games, 2*(28), 307–317.
- [21] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.
- [22] Doshi-Velez, F. & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
- [23] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.
- [24] Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.).
- [25] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). New York, NY, USA: ACM.