

Parkinson's Disease State Recognition from LFP: An Ensemble Model with Feature Optimization

Yanghuazi Liu^{1*}, Dechun Zhao¹, Yuchen Shen¹, Yi Chen¹, Yuan Zhang¹, You Lv¹, Siyu Chen¹

¹ School of Life Health Information Science and Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

Email Address

s230501009@stu.cqupt.edu.cn (Yanghuazi Liu)

*Correspondence: s230501009@stu.cqupt.edu.cn

Received: 14 January 2026; **Accepted:** 18 February 2026; **Published:** 2 March 2026

Abstract:

Parkinson's disease (PD) is a debilitating neurodegenerative disorder. While deep brain stimulation (DBS) alleviates symptoms, advancing closed-loop DBS requires real-time, accurate decoding of patient states from local field potential signals. Traditional machine learning models face a trade-off between accuracy and computational efficiency. This study proposes a feature-optimized ensemble framework. It first employs multi-objective optimization for feature selection to identify an optimal subset, reducing dimensionality while preserving discriminative information. Based on this, a feature-partitioned ensemble model under a hard voting framework is constructed, allocating distinct feature types to specialized base classifiers. Evaluation on a public PD dataset confirms the model's superiority, achieving a leading accuracy of 97.18%. The results show that the ensemble leverages diverse algorithms via hard voting, while its feature-partitioning design avoids the computational burden of processing full high-dimensional feature sets. SHAP interpretability analysis identified key decision-driving features, reaffirming the critical role of beta-band activity in PD state discrimination.

Keywords:

Local Field Potential, Parkinson's Disease, Deep Brain Stimulation, Feature Selection, Ensemble Model

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder whose prevalence is increasing with global aging. The Global Burden of Disease Study 2021 reported approximately 11.77 million cases worldwide, with significant rises in incidence and disability since 1990 [1]. Pathologically, PD involves dopaminergic neuron loss in the substantia nigra, leading to basal ganglia dysfunction and clinical symptoms including bradykinesia, tremor, rigidity, cognitive decline, and sleep disturbances [2]. Current treatments are symptomatic, with deep brain stimulation

(DBS) being a standard surgical therapy for advanced PD, typically applying high-frequency stimulation to targets like the subthalamic nucleus (STN) to alleviate symptoms [3]. However, conventional open-loop DBS requires extensive manual parameter adjustment, which is time-consuming and inefficient [4]. In response, closed-loop DBS has emerged as a promising alternative that uses real-time electrophysiological or kinematic biomarkers to automatically adjust stimulation, aiming to improve precision, reduce side effects, and conserve battery life [5]. Accurate and rapid state detection is thus critical for effective adaptive stimulation. The β -band of local field potential (LFP) is widely recognized as an electrophysiological biomarker for PD. Elevated beta oscillations in the STN of PD patients reflect pathological resting-state activity [6], which decreases with levodopa or DBS, correlating with motor improvement [7]. This has inspired closed-loop strategies like β -triggered DBS, which uses beta amplitude thresholds to control stimulation [8], highlighting LFP's value for adaptive neuromodulation.

Methodologically, early studies employed techniques like Hidden Markov Models [9] and SVM [10] to classify motor or behavioral tasks from LFP, achieving accuracies above 80%. Later work introduced methods like lp-norm Multiple Kernel Learning [11] to improve performance under low sampling rates. Subsequent research applied Random Forests to distinguish PD patients from controls with high accuracy (94.6%) [12], Logistic Regression for tremor detection (AUC 0.78) [13], and Linear Discriminant Analysis for dyskinesia prediction (70.2% accuracy) [14].

Recent advances include using AdaBoost for gait-based classification (>90% accuracy) [15], patient-specific SVM for closed-loop DBS in tremor [16], and XGBoost models correlating strongly with DBS outcomes [17]. Deep learning is also being explored for LFP analysis [18]. However, studies directly focusing on PD state recognition using LFP remain limited, and most rely on single-center datasets. A lack of large-scale, standardized public data constrains model generalizability and comparative research [19,20].

The main contributions of this study are summarized as follows:

Algorithm Benchmarking: We systematically evaluate common machine learning algorithms (k-NN, SVM, RF, NB, LDA, LR, GBDT) on a public PD LFP dataset, establishing a performance baseline for reproducible research.

Multi-Objective Feature Selection: We propose a novel feature selection framework using multi-objective optimization to simultaneously maximize feature stability, discriminative power, and subset compactness.

Feature-Partitioned Ensemble Learning: We design an ensemble model that partitions features, preserving the strengths of diverse classifiers while reducing computational cost from high-dimensional data.

Hyperparameter Optimization: We implement Bayesian optimization to tune key parameters (e.g., learning rate, tree depth) for maximizing model performance in LFP-based PD state classification.

2. Material and Methods

2.1. Dataset

The study utilized a public STN-LFP dataset from Parkinson's disease patients, released by Peter Brown's team at Oxford [21]. It included unilateral or bilateral recordings from 26 patients (60 s each), capturing ON/OFF medication states and 130

Hz DBS periods. Signals before and during DBS were stored separately. Each file contained variables ‘signal_base’, ‘signal_dbs’, and sampling rate ‘fs’ (2048 or 4096 Hz). Patient subset information is summarized in Table 1.

Table 1. Surgical and recording details of a subset of patients.

IDX	Gender	age	Illness duration	Main symptoms	Electrode site	Sampling rate(Hz)	Post-op days
1	male	54	6	Tremor,	Bilateral	2048	5
2	male	58	11	Tremor,	Bilateral	2048	5
3	male	60	6	Motor fluctuation,	Bilateral	2048	5
4	female	64	10	Bradykinesia,	Bilateral	2048	5
5	male	63	20	Tremor, Motor	Bilateral	2048	4
6	female	64	7	Bradykinesia,	Bilateral	2048	4
7	female	67	6	Tremor,	Right STN	2048	4
8	male	71	8	Freezing, Rigidity	Left STN	2048	3

2.2. Data preprocessing

To reduce inter-subject variability and improve generalizability, each patient’s signals were pooled and labeled as either diseased (off-DBS) or normal (on-DBS) [22]. As the DBS recordings contained significant stimulation artifacts (Figure 1) that degraded data quality, artifact removal was performed first. The PARRM method [23] was applied for this purpose, with results shown in Figure 2. Quantitative evaluation against a pre-treatment baseline recording from the same patient confirmed effective artifact suppression: the signal-to-noise ratio improved from -34.14 dB to -5.24 dB, and mean squared error decreased from 151,948.75 to 195.80. Minor residual noise persisted, but the processed signals closely resembled baseline neural activity. Following artifact removal, all data underwent uniform preprocessing: a 50 Hz notch filter removed power-line interference; a 4–100 Hz band-pass filter eliminated high-frequency noise while preserving physiological bands; and signals were downsampled to 1000 Hz to ensure consistency and comparability across recordings.

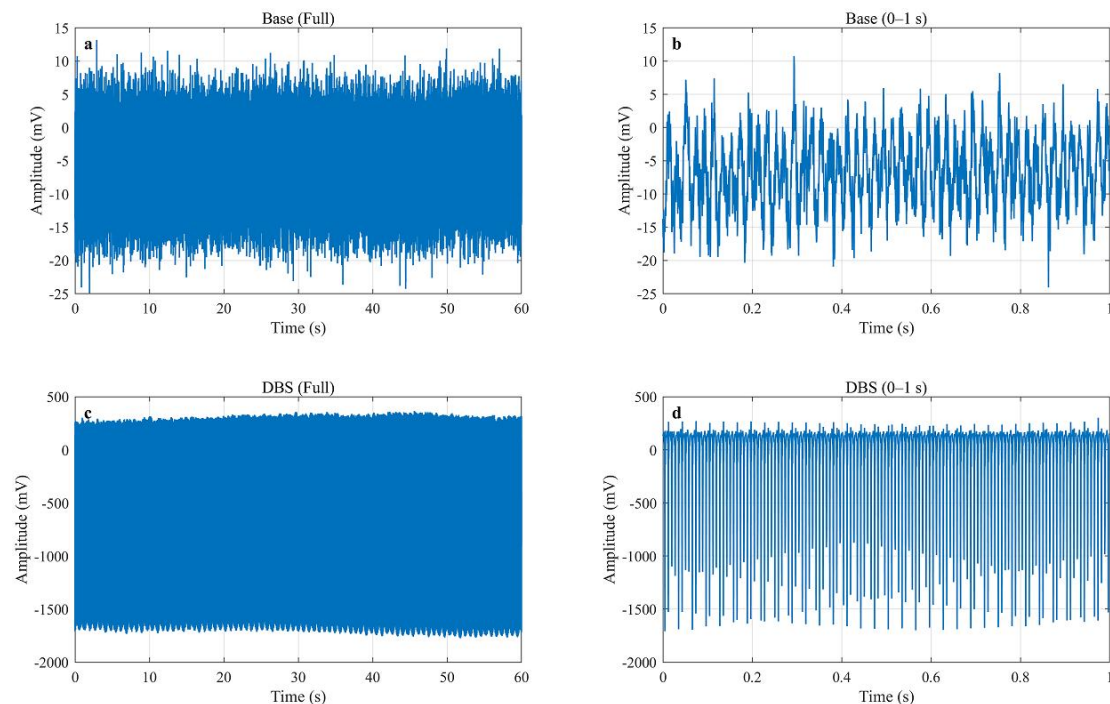


Figure 1 (a) & (b). Signal recordings before DBS therapy; (c) & (d): Signal recordings during DBS therapy.

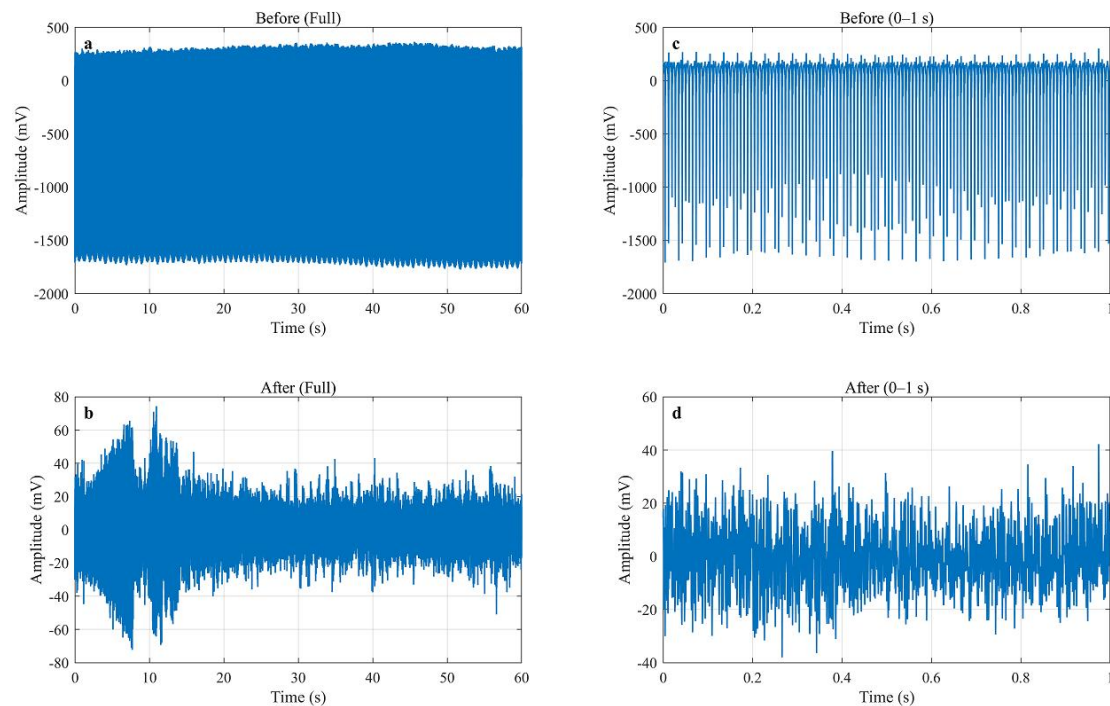


Figure 2(a) & (b): Raw signals; (c) & (d): Signals after artifact removal.

2.3. Classification models

This study systematically evaluates supervised binary models commonly used in existing literature for PD state classification, with the assessed algorithms including:

K-Nearest Neighbors (KNN): An instance-based learning algorithm that classifies each sample based on the majority label among its K closest neighbors in feature space. Although intuitive and easy to implement, KNN can become computationally intensive with large-scale datasets.

Support Vector Machine (SVM): SVM identifies the optimal hyperplane that maximizes the margin between different classes. Through the use of kernel functions, it can handle complex, nonlinearly separable data, rendering it particularly suitable for high-dimensional neural signal data such as LFPs.

Random Forest (RF): As an ensemble method, RF constructs multiple decision trees trained on bootstrapped data samples with random feature subsets. This strategy effectively reduces overfitting and enhances generalization performance, especially in high-dimensional regimes.

Naïve Bayes (NB): A probabilistic classifier grounded in Bayes' theorem, NB operates under the assumption of feature independence. Despite this simplification, it remains computationally efficient and performs competitively in high-dimensional classification tasks.

Linear Discriminant Analysis (LDA): LDA seeks a feature projection that maximizes inter-class separation while minimizing intra-class variance. It is widely used for dimensionality reduction and improving class discrimination in supervised settings.

Logistic Regression (LR): A linear model that estimates classification probabilities using a sigmoid activation function. While inherently designed for binary outcomes, LR can be extended to multi-class problems via strategies such as one-vs-rest.

Gradient Boosting Decision Trees (GBDT): This technique builds decision trees sequentially, with each new model trained to correct errors from previous ones. GBDT often achieves high predictive accuracy on complex nonlinear relationships and supports regularization to improve robustness.

2.4. Performance evaluation

In this study, classification performance was evaluated based on four basic outcomes defined by the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Based on these basic quantities, the following performance metrics were computed.

Accuracy (ACC): Quantifies the overall correctness of the classifier. A higher value indicates better overall predictive performance.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Recall (Rec): Measures the model's ability to identify all relevant positive instances. A lower value suggests a higher rate of missed detections.

$$Rec = \frac{TP}{TP + FN} \quad (2)$$

Precision (Pre): Measures the proportion of correctly identified positive instances among all cases predicted as positive.

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

F1 Score: Represents the harmonic mean of precision and recall. A higher value indicates a better balance between the two metrics.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Intersection over Union (IoU): Measures the spatial overlap between predicted and actual positive regions. Higher values indicate better alignment.

$$IoU = \frac{TP}{TP + FP + FN} \quad (5)$$

Balanced Accuracy (BAC): Computes the average of recall and specificity, providing a robust performance measure in scenarios with class imbalance.

$$BAC = \frac{1}{2} \times \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (6)$$

Matthews Correlation Coefficient (MCC): Evaluates classification quality by considering all four confusion matrix categories. Its value ranges from -1 (total disagreement) to 1 (perfect prediction).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (7)$$

Youden's Index(YI): Summarizes the diagnostic ability of a model by equally weighting sensitivity and specificity. A higher value indicates superior discriminative power.

$$\text{Youden's Index} = \text{Sensitivity} + \text{Specificity} - 1 \quad (8)$$

Yule's Q(YQ): Measures the association between predictions and actual observations. It ranges from -1 (complete disagreement) to 1 (perfect agreement).

$$\text{Yule's Q} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{(\text{TP} \times \text{TN}) + (\text{FP} \times \text{FN})} \quad (9)$$

2.5. Feature extraction and selection

2.5.1. Feature extraction

In machine learning classification tasks, manual feature extraction is critical to model performance, as the quality and quantity of features jointly determine the classification outcome. Feature extraction in this study was categorized into three types: spectral features, time-domain features, and nonlinear features (Table 2, Table 3, Table 4). Spectral features were derived from five frequency bands: theta, low-beta, high-beta, low-gamma, and the full signal, with multiple metrics computed per band. All processes yielded a total of 48 spectral, 26 time-domain, and 39 nonlinear features, providing multi-dimensional input for subsequent classification. The spectral feature set included the power spectral density (PSD), which is currently recognized as highly relevant to PD states in LFP signals.

Table 2. Spectral features.

Feature Name	Meaning
Power Spectral Density	Reflects the energy distribution of the signal across different frequency
Power	Represents the total energy within each frequency band, used to
Peak Power (Top)	Captures peak band energy, indicating frequency location and activity
Normalized Peak Power	Ratio of peak to total band power, showing dominant frequency

Table 3. Time-domain features.

Type	Feature Name	Meaning
Statistical	Maximum, minimum, mean,	Describes signal amplitude, fluctuation and
Frequency-	Zero-Crossing Rate, Willison	Describes signal frequency and rapid
Signal	Waveform Factor, Crest Factor,	Describes the nonlinear characteristics,
Hjorth	Activity, Mobility, Complexity	Represents signal energy, frequency, and
Differential Features	First-Order difference mean, Normalized First-Order difference	Characterizes the trend and acceleration features of the signal

Table 4. Nonlinear features.

Type	Feature Name	Meaning
Fractal Self-	Higuchi Fractal Dimension,	Reflects signal fractals, correlations and
Entropy	Sample Entropy, Fuzzy	Quantifies signal complexity and uncertainty
Chaos and	Maximum Lyapunov	Characterizes chaos, divergence rate, and
Recurrence	Recurrence Rate,	Reflects the recurrence, periodicity and
Energy	Relative Wavelet Energy,	Describes band energy distribution and

2.5.2. Feature selection

A critical stage in our methodology was feature selection. Following feature extraction, we obtained the raw dataset $X \in R^{n \times p}$ and the label vector $y \in R^n$ (comprising n samples and p features). To determine the optimal feature subset, this

study proposed a multi-objective optimization framework that simultaneously considers three objectives: feature stability, classifier performance, and feature set parsimony.

Step 1: Consensus Ranking-based Assessment of Feature Importance

The multi-objective optimization framework employed seven feature importance assessment methods to rank the original features. For each method $m \in Method$, it computed a feature importance score vector $Imp^m = [Imp_1^m, Imp_2^m, \dots, Imp_p^m]$, where Imp_j^m denotes the importance score of the j -th feature according to method m .

$$Method = \{ANOVA, Pearson, Spearman, ChiSquare, RF \text{ Feature Importance, XGBoost Feature Importance, LASSO}\} \quad (10)$$

Taking the ANOVA F-value as an example, its calculation formula is given by:

$$F_j = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{\sum_{c=1}^2 n_c (\bar{x}_{jc} - \bar{x}_j)^2 / (2 - 1)}{\sum_{c=1}^2 \sum_{i=1}^{n_c} (x_{jic} - \bar{x}_{jc})^2 / (n - 2)} \quad (11)$$

Where n_c denotes the number of samples in class c , $\bar{x}_{jc}$ represents the mean of the j -th feature within class c , and $\bar{x}_j$ is the overall mean of the j -th feature.

For each method $m \in Method$, a feature ranking $R^m = \text{argsort}(Imp^m)$, [descending sort] was obtained. Subsequently, the average rank $\bar{R}_j = \frac{1}{|M|} \sum_{m \in M} \text{position}(j \text{ in } R^m)$, was computed for each feature. Finally, the consensus ranking $R^{\text{consensus}} = \text{argsort}(\bar{R})$ was derived by applying the argsort function to the average ranks. All features were then reordered according to this consensus ranking, forming a new sequence sorted by descending importance.

Step 2: Resampling-based Stability Evaluation of Features

To evaluate the stability of the feature ranking, we employed $B=100$ resampling iterations using an alternating strategy of bootstrap and 80% subsampling. For every iteration b , the consensus ranking R_b was recomputed from the corresponding resampled dataset (X_b, y_b) .

For each candidate feature number $k \in [5, p_{\max}]$ (where $p_{\max} = \min(p, 50)$, means a maximum of 50 features and a minimum of 5), a stability score is computed:

$$S(k) = \frac{2}{B(B-1)} \sum_{i=1}^{B-1} \sum_{j=i+1}^B \left[1 - \frac{\frac{1}{|R_i \cap R_j|} \sum_{f \in R_i \cap R_j} |\text{rank}_i(f) - \text{rank}_j(f)|}{k-1} \right] \quad (12)$$

R_i, R_j : the top- k feature sets from the i -th and j -th resamples; $\text{rank}_i(f)$ as the rank (1 to k) of feature f in resample i , it assesses the ranking consistency of the most important features across varying data subsets. Values nearer to 1 signify greater stability.

Step 3: Multi-Classifer Based Performance Evaluation

The classification performance of various feature subsets was assessed using seven classifiers, $C = KNN, SVM, RF, NB, LDA, LR, MLP, GBDT$. For every candidate feature count k , the corresponding feature subset $X^{(k)} \in R^{n \times k}$ was constructed by selecting the top- k features from the importance ranking.

Subsequently, a ten-fold ($K=10$) stratified cross-validation was employed. For each classifier $c \in C$, the performance score was calculated by evaluating it on each feature subset.

$$P_c(k) = \frac{1}{K-2} \sum_{fold=2}^{K-1} Acc_{fold}(k) \quad (13)$$

To enhance robustness, the highest and lowest accuracy rates from each fold were discarded for every classifier. The final composite performance score was then computed as the average across all classifiers:

$$P(k) = \frac{1}{|C|} \sum_{c \in C} P_c(k) \quad (14)$$

Step 4: Feature Set Parsimony Assessment

To encourage feature set conciseness, this study defined the parsimony score for the third optimization objective as follows:

$$C(k) = \frac{\ln(1 + (p_{max} - k))}{\ln(1 + p_{max})} \quad (15)$$

This function is monotonically decreasing with the increase of the feature number k , thereby penalizing excessive features and promoting the selection of more compact feature subsets.

Step 5: Multi-Objective Optimization and Optimal Feature Number Selection

Integrating the three objectives $P(k)$, $S(k)$, $C(k)$, a weighted scoring function was constructed::

$$F(k) = w_p \cdot P(k) + w_s \cdot S(k) + w_c \cdot C(k) \quad (16)$$

The weight coefficients w_p (performance weight), w_s (stability weight), and w_c (parsimony weight) satisfied the constraint $w_p + w_s + w_c = 1$.

The optimal number of features was then determined by maximizing the composite score:

$$k^* = \arg \max_{k \in [5, p_{max}]} F(k) \quad (17)$$

The final selected feature subset comprised the top- k most important features, $F = f_1, f_2, \dots, f_{k^*}$, sorted according to the consensus ranking. This approach achieved a balance among stability, performance, and parsimony, ensuring that the selected feature subset possessed strong discriminative power while maintaining good robustness.

The feature selection results are shown in Figure 3. For time-domain, spectral, and nonlinear features, optimal classification performance required 11, 34, and 20 features respectively. Feature stability remained relatively steady for time-domain and nonlinear features. In performance evaluation, the GBDT model consistently performed robustly. However, when jointly considering stability, feature count, and performance, the highest comprehensive scores for time-domain, spectral, and nonlinear features were achieved at $k=6$ (score 0.88), $k=17$ (0.783), and $k=9$ (0.895), respectively. The final selected feature subsets are presented in Table 5, Table 6, Table 7.

Table 5. Selected time-domain features.

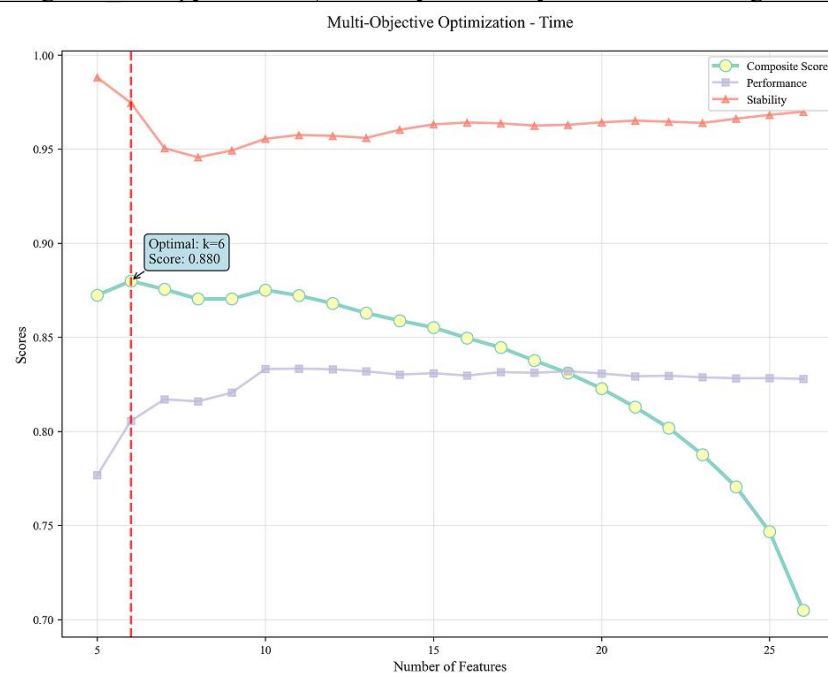
Time-domain Feature	Description
slope_sign_changes	Frequency of waveform fluctuations
hjorth_mobility	Distribution range of dominant frequencies
diff2_norm	Intensity of signal acceleration changes
zcr	Zero-crossing rate, signal oscillation frequency
diff1_norm	Intensity of signal variation rate
abs_mean	Average signal amplitude level

Table 6. Selected spectral features.

Spectral Feature	Description
psd_high_beta	PSD in high-beta band
power_high_beta	Power in high-beta band
peak_power_global_7	7th global-signal peak power
peak_power_low_beta_1	1st peak power in low-beta band
peak_power_high_beta_1,2	1st/2nd peak power in high-beta band
power_low_beta	Power in low-beta band
norm_peak_power_gamma_1,2	Normalized 1st/2nd peak power in gamma band
peak_power_global_6	6th global peak power
psd_low_beta	PSD in low-beta band
peak_power_gamma_1,2,3	1st/2nd/3rd peak power in gamma band
psd_gamma	PSD in gamma band
power_gamma	Power in gamma band
peak_power_global_5	5th global peak power

Table 7. Selected nonlinear features.

Nonlinear Feature	Description
higuchi_fd	Signal irregularity and complexity
svd_entropy	Signal complexity via singular value distribution
wavelet_entropy	Multi-scale complexity in wavelet domain
fuzzy_entropy	Signal sequence regularity
correlation_dim	Attractor geometric dimension
poincare_entropy	Phase space point distribution entropy
max_lyapunov	Sensitivity to initial conditions (chaos)
c0_complexity	Proportion of non-regular components
kolmogorov_entropy	Dynamical system information generation rate



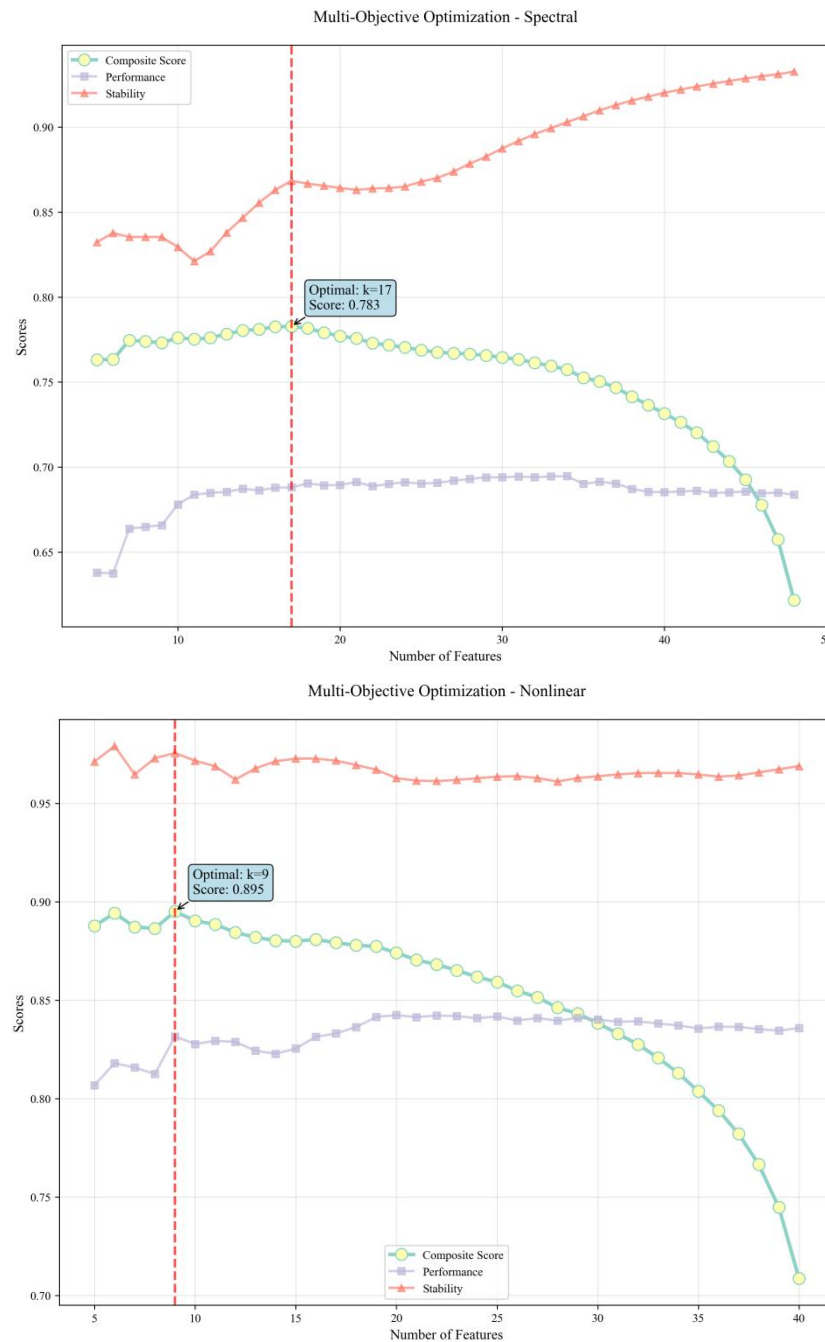


Figure 3. Performance of three feature types in multi-objective optimization (Top-left: time-domain; Top-right: spectral; Bottom: nonlinear).

2.5.3. Hyperparameter optimization

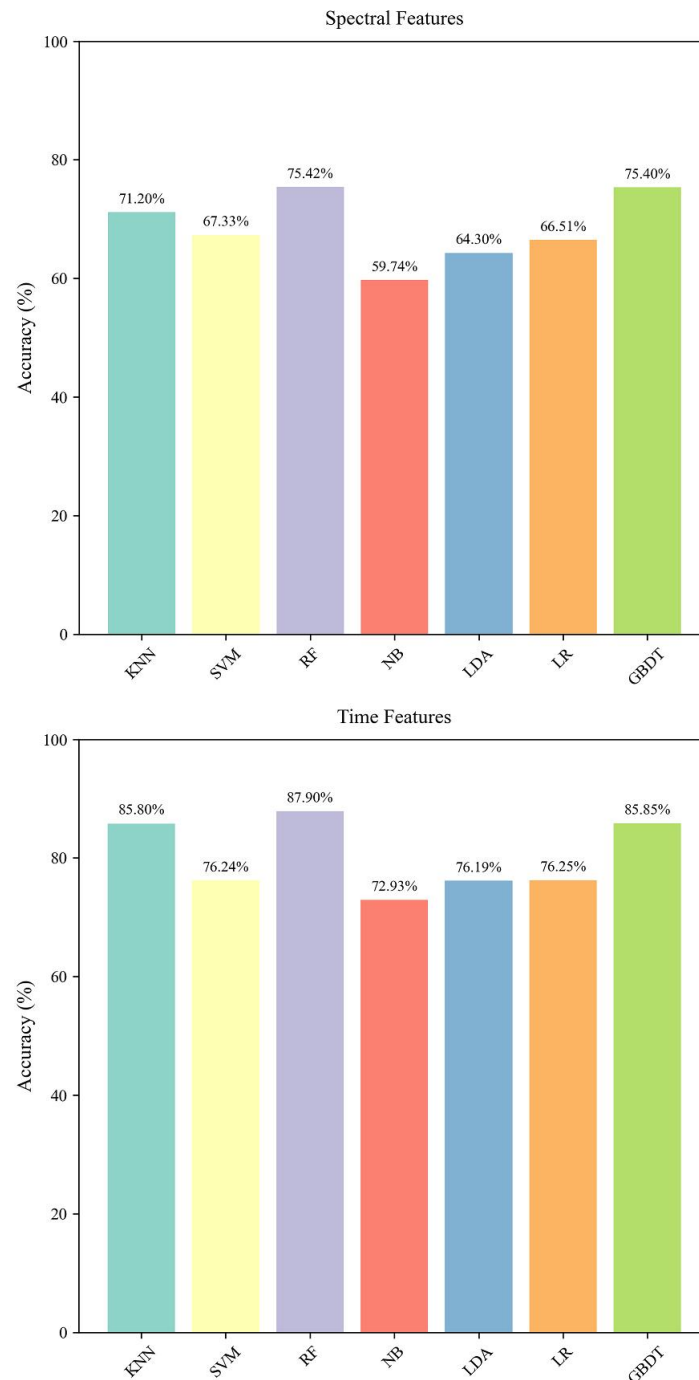
Hyperparameter optimization was performed using Bayesian optimization to maximize model performance. This approach constructs a Gaussian Process surrogate model and employs the Expected Improvement acquisition function to efficiently balance exploration and exploitation during the search. Classifier-specific hyperparameter spaces were defined (e.g., regularization parameter and kernel for SVM; tree count and depth for Random Forest), with each configuration evaluated via 10-fold cross-validation. Compared to grid or random search, Bayesian optimization achieves higher sample efficiency, converging to near-optimal settings with fewer evaluations.

3. Experiment and result

3.1. Single-classifier experiment

The experimental setup used Python on Windows 11 (Anaconda), with an NVIDIA GTX 2080 GPU and Intel i5-12400 CPU. A fixed random seed of 42 was set. Preprocessed signals were segmented using a 1-second sliding window with a 0.5-second stride, producing 6,236 labeled samples (diseased=1, normal=0). All models were evaluated via 10-fold cross-validation, using the optimal feature subset from Section 2.5.2 and hyperparameters tuned via Bayesian optimization (Section 2.5.3).

Figure 4 compares classifier performance across three feature types, with detailed results in Table 8, Table 9, Table 10. Figure 5 shows corresponding ROC curves, illustrating model discrimination ability for PD state classification.



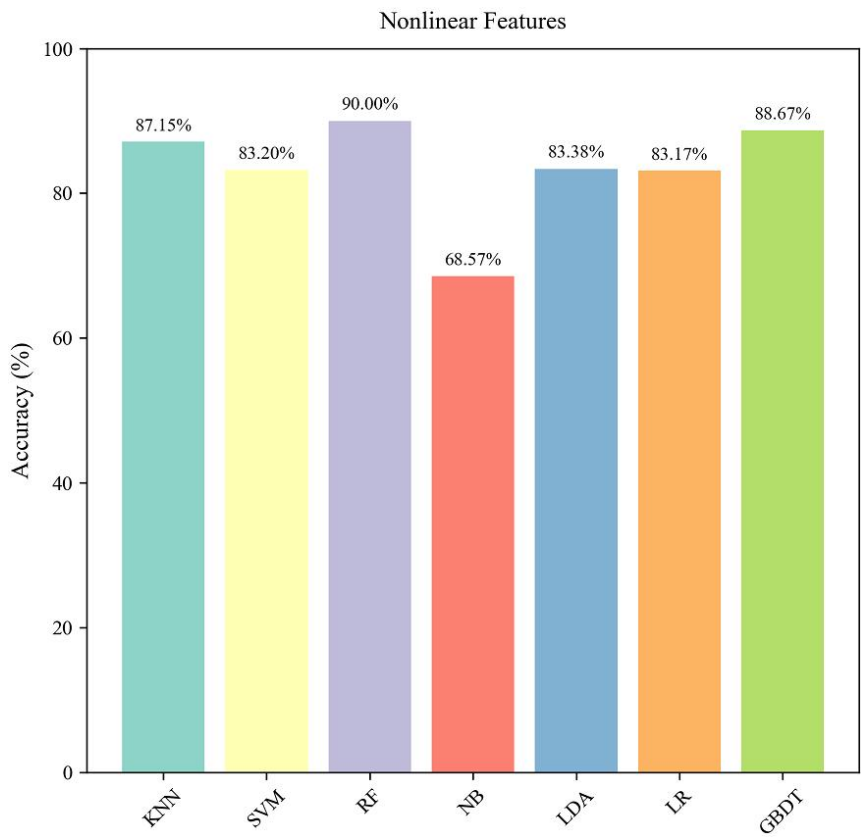
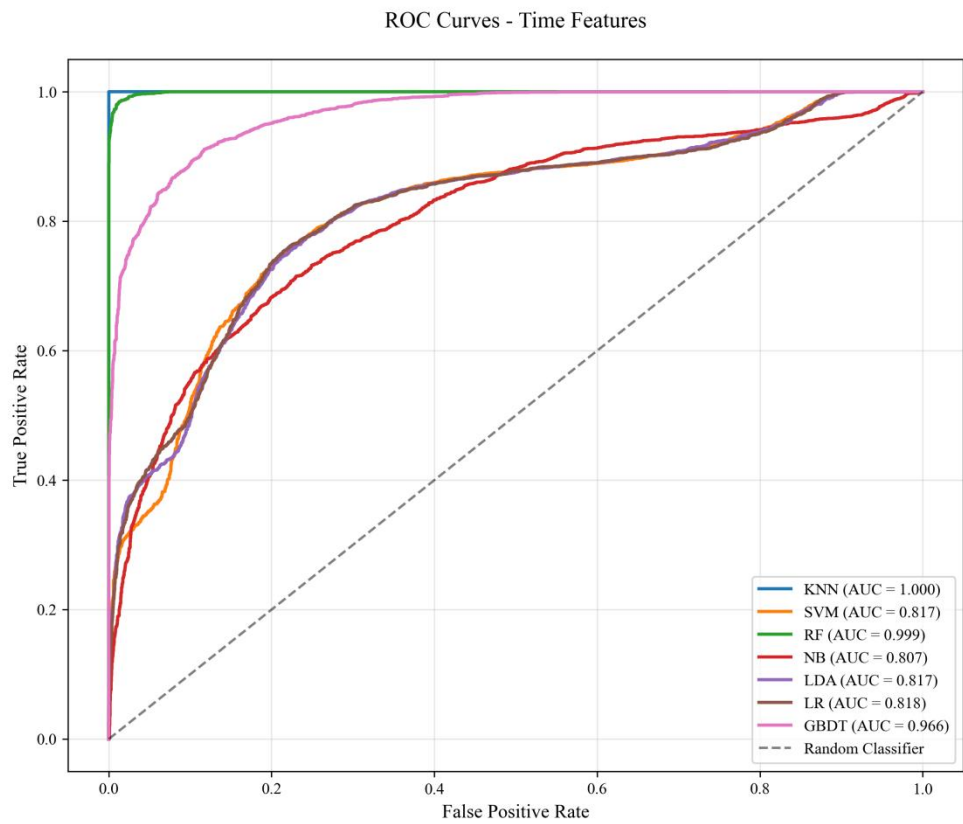


Figure 4. Performance of classifiers across three feature types.



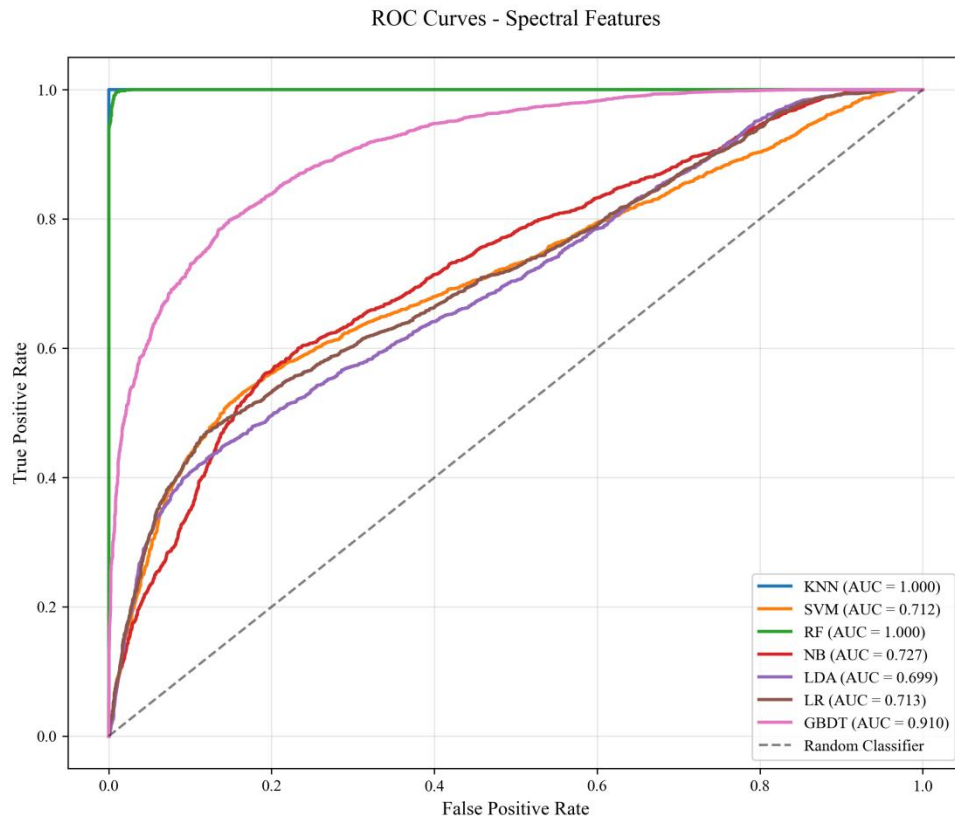


Figure 5. ROC Curves of classifiers for three feature categories (Top-left: time-domain; Top-right: spectral; Bottom: nonlinear).

Table 8. Classifier performance on time-domain features.

Approach	Acc	Rec	Pre	F1	IoU	BAC
KNN	85.80%±0.86%	86.68%±2.02%	85.22%±1.49%	85.92%±0.90%	75.32%±1.39%	85.80%±0.86%
SVM	76.24%±1.52%	80.94%±1.85%	74.06%±2.23%	77.31%±1.20%	63.03%±1.59%	76.24%±1.52%
RF	87.90%±1.04%	87.93%±2.36%	87.89%±1.00%	87.89%±1.16%	78.42%±1.85%	87.90%±1.04%
NB	72.93%±2.03%	74.39%±2.47%	72.37%±2.71%	73.32%±1.85%	57.92%±2.32%	72.93%±2.03%
LDA	76.19%±1.70%	78.72%±1.91%	75.02%±2.47%	76.79%±1.41%	62.34%±1.86%	76.19%±1.70%
LR	76.25%±1.39%	78.66%±1.83%	75.13%±2.23%	76.82%±1.12%	62.37%±1.47%	76.25%±1.39%
GBDT	85.85%±1.39%	85.02%±2.64%	86.47%±1.28%	85.71%±1.54%	75.03%±2.37%	85.85%±1.39%

Table 9. Classifier performance on spectral features.

Approach	Acc	Rec	Pre	F1	IoU	BAC
KNN	71.20%±1.54%	68.55%±2.55%	72.42%±1.88%	70.40%±1.70%	54.35%±2.01%	71.20%±1.54%
SVM	67.33%±1.31%	46.66%±2.61%	79.59%±2.21%	58.78%±2.20%	41.66%±2.21%	67.33%±1.30%
RF	75.42%±1.46%	72.47%±2.22%	77.04%±1.72%	74.66%±1.59%	59.59%±2.03%	75.42%±1.46%

NB	59.74%±1.1 3%	25.87%±2.4 6%	80.35%±3.1 6%	39.05%±2.8 7%	24.30%±2.1 9%	59.74%±1.1 0%
LDA	64.30%±1.4 6%	54.21%±2.1 5%	67.95%±2.0 0%	60.28%±1.7 4%	43.16%±1.8 0%	64.30%±1.4 6%
LR	66.51%±0.8 5%	54.82%±2.0 7%	71.60%±1.4 2%	62.06%±1.2 7%	45.00%±1.3 6%	66.51%±0.8 5%
GBDT	75.40%±1.1 8%	70.64%±2.2 4%	78.09%±1.2 6%	74.16%±1.4 4%	58.95%±1.8 4%	75.40%±1.1 8%

Table 10. Classifier performance on nonlinear features.

Approach	Acc	Rec	Pre	F1	IoU	BAC
KNN	87.15%±1.1 8%	89.03%±1.7 7%	85.89%±2.3 0%	87.39%±1.0 3%	77.62%±1.6 2%	87.15%±1.1 8%
SVM	83.20%±1.4 7%	83.47%±1.4 4%	83.05%±1.9 6%	83.25%±1.3 9%	71.33%±2.0 2%	83.20%±1.4 7%
RF	90.00%±1.4 5%	88.80%±1.7 7%	91.05%±2.3 2%	89.89%±1.4 1%	81.66%±2.3 3%	90.00%±1.4 5%
NB	68.57%±2.1 4%	64.38%±3.3 2%	70.34%±2.7 9%	67.17%±2.3 5%	50.62%±2.6 5%	68.57%±2.1 4%
LDA	83.38%±1.4 7%	81.96%±1.7 8%	84.38%±1.9 7%	83.14%±1.4 7%	71.17%±2.1 3%	83.38%±1.4 7%
LR	83.17%±1.4 5%	83.09%±1.2 8%	83.26%±2.0 3%	83.16%±1.3 5%	71.20%±1.9 6%	83.17%±1.4 4%
GBDT	88.67%±1.5 9%	86.75%±1.6 8%	90.26%±2.2 2%	88.45%±1.5 8%	79.33%±2.5 5%	88.67%±1.5 9%

To address potential feature insufficiency from separate categories and to enable broader comparison, all three feature types were merged into a combined set. Models retrained on this set achieved improved results (Figure 6, Table 11). The best-performing GBDT model attained an MCC of 0.8601, Youden's Index of 0.8598, and Yule's Q of 0.9888. Optimal hyperparameters from Bayesian optimization are listed in Table 12.

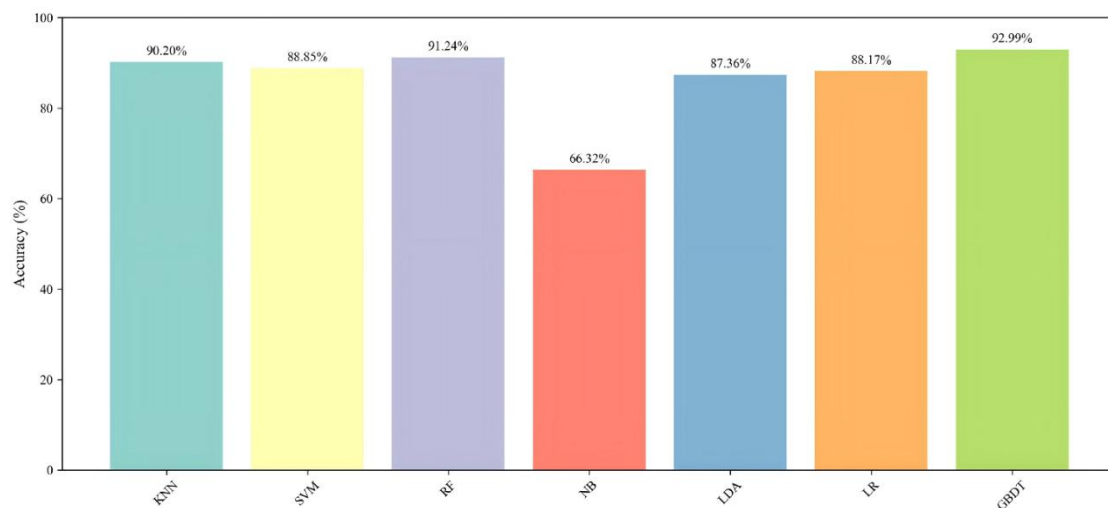


Figure 6. Accuracy comparison of classifiers on the combined feature set.

Table 11. Performance of classifiers on the combined feature set

Approach	Acc	Rec	Pre	F1	IoU	BAC
KNN	90.20%±1.3 2%	91.79%±2.4 4%	89.01%±1.8 3%	90.34%±1.3 2%	82.41%±2.2 1%	90.20%±1.3 2%

SVM	88.85%±1.1 1%	88.54%±1.4 5%	89.11%±1.5 5%	88.81%±1.1 0%	79.90%±1.7 9%	88.85%±1.1 1%
RF	91.24%±1.2 5%	91.27%±1.9 0%	91.25%±1.8 3%	91.24%±1.2 4%	83.92%±2.1 0%	91.24%±1.2 5%
NB	66.32%±1.2 7%	41.37%±2.5 9%	82.61%±2.0 9%	55.08%±2.4 0%	38.04%±2.2 8%	66.32%±1.2 6%
LDA	87.36%±1.0 3%	84.82%±1.8 1%	89.38%±1.4 4%	87.02%±1.1 0%	77.04%±1.7 2%	87.36%±1.0 3%
LR	88.17%±0.9 1%	87.61%±1.2 7%	88.64%±1.4 8%	88.11%±0.8 8%	78.76%±1.4 1%	88.17%±0.9 1%
GBDT	92.99%±0.9 9%	93.33%±1.8 4%	92.72%±1.2 4%	93.01%±1.0 1%	86.94%±1.7 8%	92.99%±0.9 9%

Table 12. Model parameter.

Approach	Hyperparameter combination
KNN	n_neighbors=15, p=1, weights=uniform
SVM	C=8.15, kernel=linear, gamma=auto
RF	n_estimators=181, max_depth=14, min_samples_split=5, min_samples_leaf=4
NB	var_smoothing=1.12e-8
LDA	solver=svd
LR	C=10.0, penalty=l2, solver=liblinear
GBDT	n_estimators=129, learning_rate=0.110, max_depth=5

3.2. Ensemble algorithm based on model fusion

Analysis of single-classifier results indicated that while methods like KNN, SVM, and RF showed discriminative ability, their performance varied considerably. To balance real-time applicability with improved classification accuracy, a model fusion strategy integrating multiple base classifiers was adopted. Among ensemble methods, voting is widely used for classification due to its simplicity and interpretability [25].

3.2.1. Design of the hard-voting ensemble model

To address inconsistent single-classifier performance and enhance overall discriminative capability, a hard voting ensemble strategy was adopted. Ensembles with 3, 5, and 7 base classifiers were systematically constructed on the combined feature set using the parameters from Table 12. Results (Figure 7) showed no significant accuracy gain with more classifiers, while training time increased—indicating a diminishing-return effect beyond three classifiers. The three-classifier configuration was therefore selected as optimal, achieving an MCC of 0.8694, Youden's Index of 0.8691, and Yule's Q of 0.9903. Further analysis (Figure 8) confirmed that most three-classifier combinations effectively integrated complementary strengths of base models, improving generalization through collaborative decision-making and yielding more robust predictive performance.

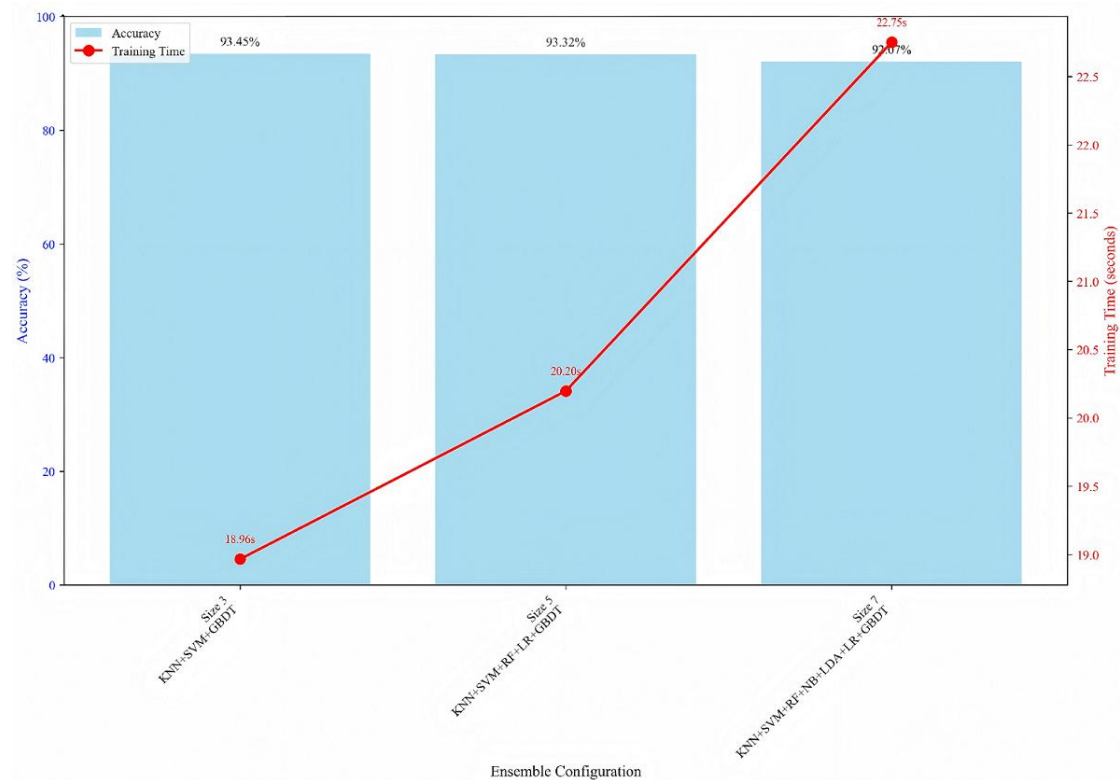


Figure 7. Comparison of optimal combinations with 3/5/7 base classifiers.

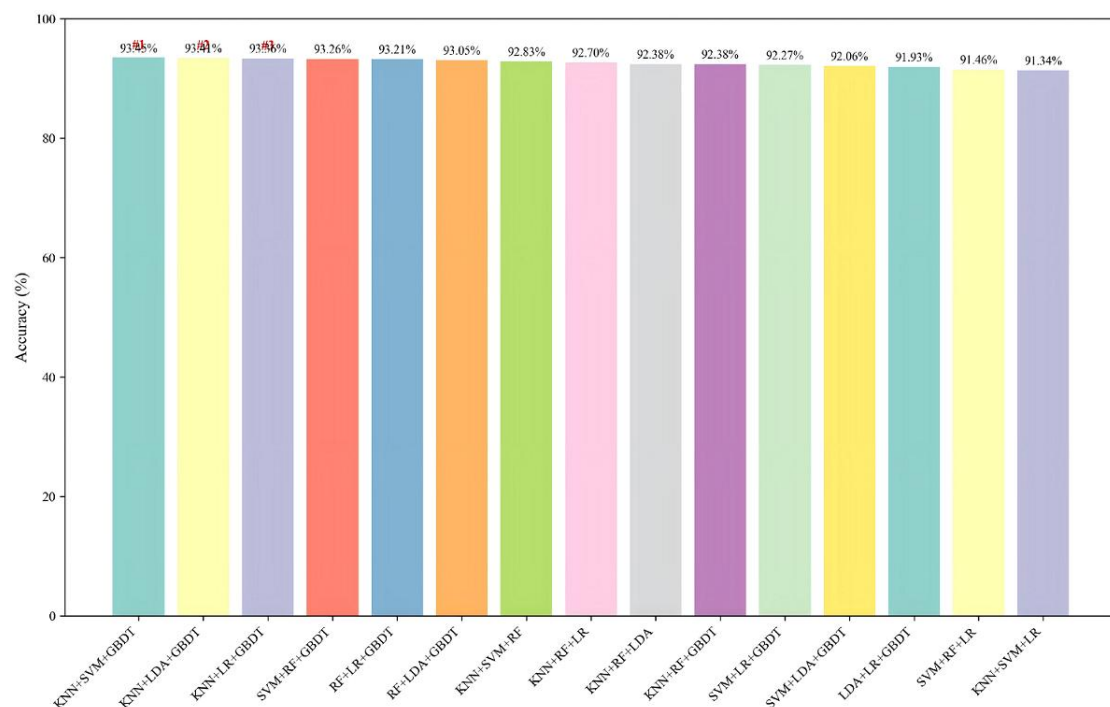


Figure 8. Top Performing combinations under 3 base classifiers.

3.2.2. Feature-partitioned ensemble model

Building on the finding that hard voting integrates classifier strengths, this study structurally optimized the 3-classifier ensemble by proposing a feature-partitioned model. Its core idea allocates the combined feature set to different base classifiers by physical meaning (spectral, time-domain, nonlinear), integrated via hard voting

(Figure 9). Unlike traditional ensembles where all classifiers process the full set, this scheme specializes the feature space. Dimensionality analysis illustrates: traditional approaches require three classifiers each processing 32 features (total dimension 96), while the partitioned scheme assigns 6, 17, and 9 features respectively, keeping the total at 32 and reducing computational complexity.

Experimental results (Figure 10) show all partitioned combinations achieved excellent performance with substantially lower training time, confirming the scheme avoids computational cost from dimension expansion while preserving accuracy. Detailed optimization (Table 13) identified KNN (time-domain) + RF (spectral) + GBDT (nonlinear) as the best combination. For a fair comparison, all literature methods were reevaluated under our experimental setup, with only their best configurations compared against ours (Table 14). The feature-partitioned ensemble consistently outperformed all baselines. Moreover, the optimal combination was not unique—most pairings surpassed both single classifiers and the original ensemble, indicating that the performance gain is systematic rather than accidental.

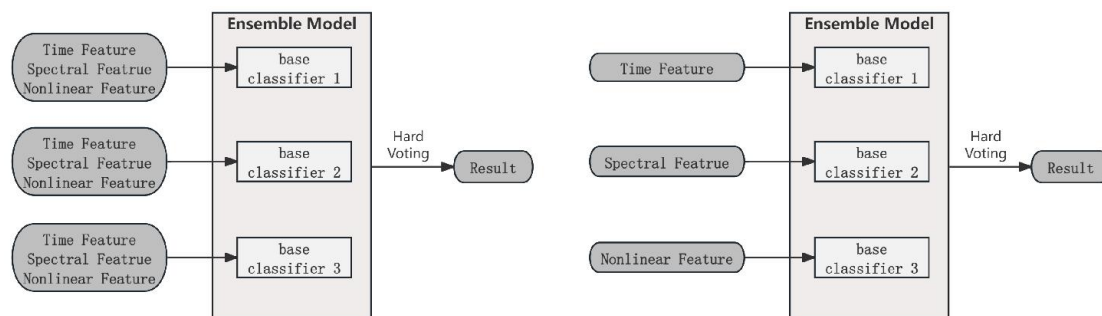


Figure 9. Original ensemble model (left); Feature-partitioned ensemble model (right).

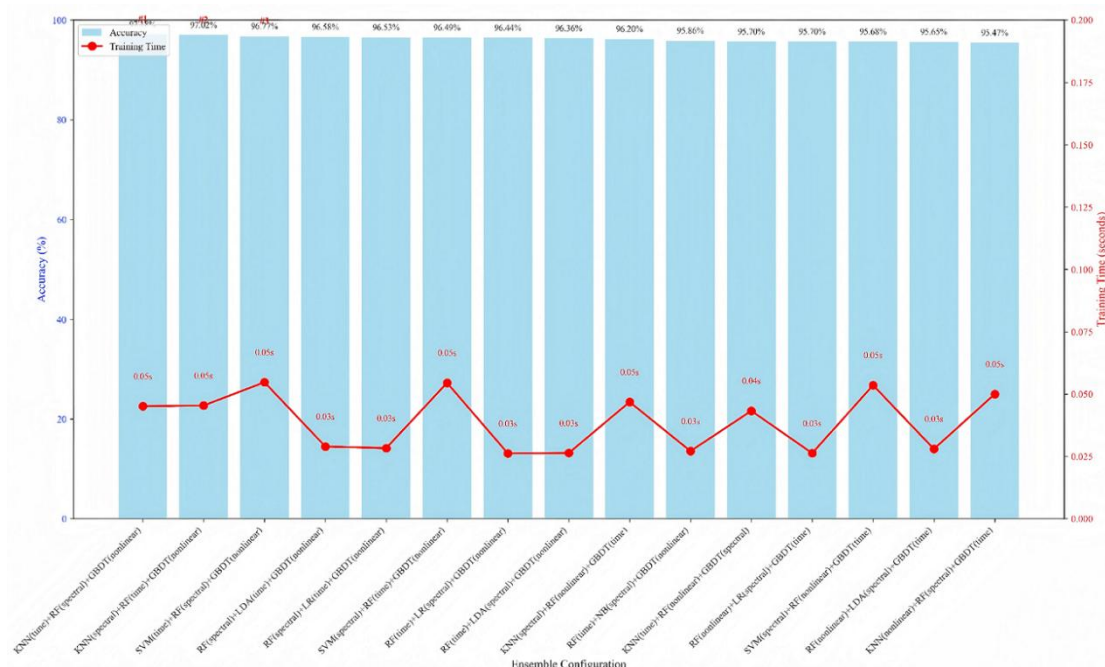


Figure 10. Optimal feature-partitioned combination.

Table 13. Performance of top-five combinations.

Approach	Acc	Rec	Pre	F1	IoU	BAC
KNN(time)+RF(spectral)+GBDT(nonlinear)	97.18%±0.58%	96.47%±1.09%	97.86%±0.73%	97.15%±0.59%	94.47%±1.12%	97.18%±0.58%
KNN(spectral)+RF(time)+GBDT(nonlinear)	97.02%±0.87%	96.50%±1.19%	97.50%±0.83%	97.00%±0.89%	94.19%±1.67%	97.02%±0.87%
SVM(time)+RF(spectral)+GBDT(nonlinear)	96.77%±0.74%	96.31%±1.13%	97.22%±0.72%	96.76%±0.75%	93.73%±1.40%	96.77%±0.74%
RF(spectral)+LDA(time)+GBDT(nonlinear)	96.58%±0.67%	95.92%±1.18%	97.21%±0.66%	96.56%±0.69%	93.35%±1.28%	96.58%±0.67%
RF(spectral)+LR(time)+GBDT(nonlinear)	96.53%±0.64%	95.80%±1.21%	97.23%±0.61%	96.51%±0.66%	93.25%±1.23%	96.53%±0.64%

Table 14. Comparison with other methods.

Approach	Acc	Rec	Pre	F1	IoU	BAC	MC C	YI	YQ
SVM	88.85%±1.11%	88.54%±1.45%	89.11%±1.55%	88.81%±1.10%	79.90%±1.79%	88.85%±1.11%	0.7772	0.7770	0.9689
RF	91.24%±1.25%	91.27%±1.90%	91.25%±1.83%	91.24%±1.24%	83.92%±2.10%	91.24%±1.25%	0.8252	0.8248	0.9818
LDA	87.36%±1.03%	84.82%±1.81%	89.38%±1.44%	87.02%±1.10%	77.04%±1.72%	87.36%±1.03%	0.7483	0.7471	0.9606
LR	88.17%±0.91%	87.61%±1.27%	88.64%±1.48%	88.11%±0.88%	78.76%±1.41%	88.17%±0.91%	0.7637	0.7635	0.9648
XGBoost	92.84%±0.92%	93.23%±1.66%	92.54%±1.31%	92.87%±0.93%	86.70%±1.62%	92.84%±0.92%	0.8572	0.8569	0.9883
AdaBoost	87.15%±1.29%	86.14%±1.96%	87.99%±2.27%	87.02%±1.24%	0.7704 ± 0.0193	87.15%±1.29%	0.7437	0.7430	0.9580
Ours	97.18%±0.58%	96.47%±1.09%	97.86%±0.73%	97.15%±0.59%	94.47%±1.12%	97.18%±0.58%	0.9437	0.9435	0.9985

3.2.3. Feature Importance Analysis

SHAP (SHapley Additive exPlanations) [26] is a feature importance method based on cooperative game theory, which quantifies each feature's contribution to model predictions by fairly distributing the “payoff” of the collaborative output. To interpret feature contributions in the classification model, this study applied SHAP analysis to the optimal ensemble (KNN-time + RF-spectral + GBDT-nonlinear).

Comparing SHAP value distributions across the three classifiers (Figure 11) revealed a multi-level decision mechanism: GBDT utilized nonlinear features to analyze signal complexity and chaotic properties; KNN focused on statistical waveform characteristics via time-domain features; RF emphasized spectral energy distribution patterns. Key discriminative features included the number of slope sign changes (time-domain), peak gamma-band power followed by beta-band power (spectral), and the Higuchi fractal dimension (nonlinear). This not only reaffirmed the established relevance of beta activity but also highlighted the value of other bands and signal complexity, demonstrating the model's ability to capture multi-scale characteristics of PD LFP signals.

Clinically, these data-driven insights provide a validated feature set to enhance intraoperative signal assessment and postoperative efficacy evaluation in DBS therapy, supporting the development of personalized treatment strategies.

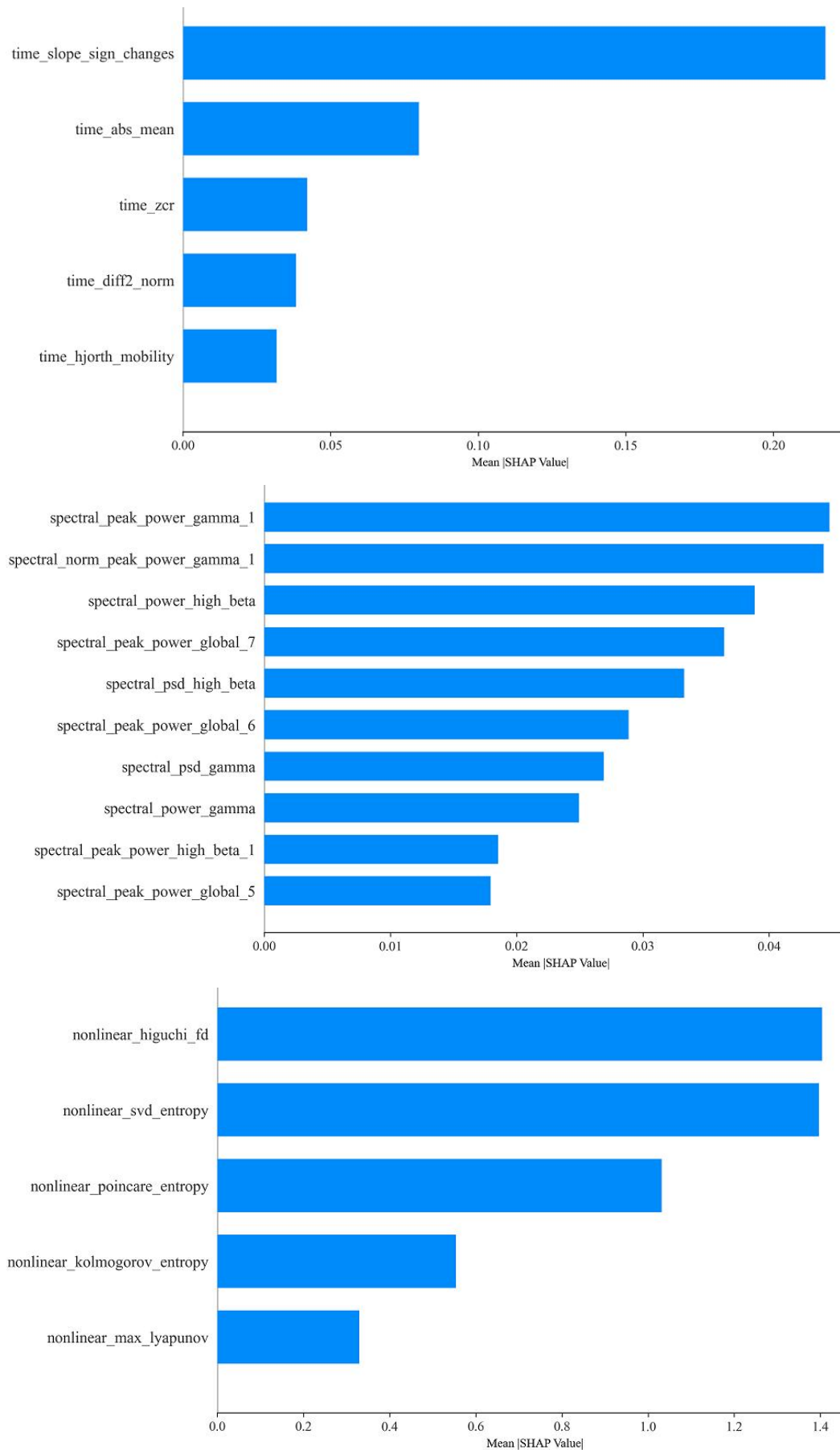


Figure 11. Feature importance visualization.

4. Discussion

In DBS research, identifying reliable biomarkers for PD state recognition remains challenging. While existing methods often use posture monitoring or surface EMG, their long-term application is constrained by patient discomfort and wearability issues from external sensors [27]. Implanted DBS systems allow LFP recording during stimulation, enabling continuous neural monitoring without extra hardware [28]. However, current LFP studies mainly focus on DBS targeting or specific behavioral tasks, with limited attention to automated, general-purpose PD state recognition [29].

This study utilized one of the few public datasets containing systematically recorded continuous LFP from PD patients before and after DBS. This addresses the limitations of previous single-center, small-sample studies that used heterogeneous devices and protocols (e.g., Shah et al. [13]; He et al. [16]; Bakstein et al. [30]), which have historically hindered model generalizability and cross-study comparability.

Ensemble analysis revealed diminishing accuracy gains beyond three base classifiers, alongside increased computational cost. The three-classifier structure thus offered an optimal performance-efficiency balance. Building on this, we proposed a feature-partitioned ensemble model that maintained robust, superior performance across multiple feature-classifier configurations. Under a unified evaluation framework with Bayesian hyperparameter tuning, all methods were systematically optimized. As summarized in Table 14, reference implementations achieved the following accuracies: SVM [10][16] $88.85\% \pm 1.11\%$; LDA [14] $87.36\% \pm 1.03\%$; LR [13] $88.17\% \pm 0.91\%$. Among ensemble benchmarks, XGBoost [17] reached 92.84% and AdaBoost [15] 87.15%. Although XGBoost outperformed other references, it remained below our proposed model, which achieved 97.18% accuracy.

Despite promising results, LFP-based PD classification still faces data limitations. The scarcity of large-scale, standardized multi-center datasets restricts model generalizability across sites and devices. Future work should prioritize curating and releasing multi-center, multi-platform datasets to support robust validation and enhance clinical adaptability. Longitudinal datasets tracking patients over time would also be valuable to capture disease progression and treatment dynamics. Integrating cross-sectional breadth with longitudinal depth is essential for developing adaptive closed-loop DBS systems capable of personalized neuromodulation in real-world practice.

5. Conclusions

This study proposes a feature-partitioned ensemble framework for accurate PD state classification using LFP signals from DBS systems. Extensive validation on public datasets shows the model outperforms conventional single classifiers and established ensemble methods. The advantage stems from two key innovations: a multi-objective optimization-based feature selection method that identifies discriminative, compact feature subsets, and a specialized ensemble architecture that combines classifier strengths while avoiding the computational cost of traditional approaches. SHAP interpretability analysis validates the model's decision patterns and reaffirms the clinical relevance of beta-band activity in PD discrimination. This work provides a robust algorithmic foundation for advancing intelligent neuromodulation and enabling personalized PD therapy.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Luo, Y.; Qiao, L.; Li, M.; Wen, X.; Zhang, W.; Li, X. Global, regional, national epidemiology and trends of Parkinson's disease from 1990 to 2021: findings from the Global Burden of Disease Study 2021. *Frontiers in aging neuroscience*, 2025, 16, 1498756. DOI: <https://doi.org/10.3389/fnagi.2024.1498756>.
- [2] Caproni, S.; Colosimo, C. Diagnosis and differential diagnosis of Parkinson disease. *Clinics in geriatric medicine*, 2020, 36(1), 13-24. DOI: <https://doi.org/10.1016/j.cger.2019.09.014>.
- [3] Hartmann, C.J.; Fliegen, S.; Groiss, S.J.; Wojtecki, L.; Schnitzler, A. An update on best practice of deep brain stimulation in Parkinson's disease. *Therapeutic advances in neurological disorders*, 2019, 12, 1756286419838096. DOI: <https://doi.org/10.1177/1756286419838096>.
- [4] Bronstein, J.M.; Tagliati, M.; Alterman, R.L.; Lozano, A.M.; Volkmann, J.; Stefani, A.; DeLong, M.R. Deep brain stimulation for Parkinson disease: an expert consensus and review of key issues. *Archives of neurology*, 2011, 68(2), 165-165. DOI: <https://doi.org/10.1001/archneurol.2010.260>.
- [5] Oliveira, A.M.; Coelho, L.; Carvalho, E.; Ferreira-Pinto, M.J.; Vaz, R.; Aguiar, P. Machine learning for adaptive deep brain stimulation in Parkinson's disease: closing the loop. *Journal of Neurology*, 2023, 270(11), 5313-5326. DOI: <https://doi.org/10.1007/s00415-023-11873-1>.
- [6] Whitmer, D.; De Solages, C.; Hill, B.; Yu, H.; Henderson, J. M.; Bronte-Stewart, H. High frequency deep brain stimulation attenuates subthalamic and cortical rhythms in Parkinson's disease. *Frontiers in human neuroscience*, 2012, 6, 155. DOI: <https://doi.org/10.3389/fnhum.2012.00155>.
- [7] Neumann, W.J.; Degen, K.; Schneider, G.H.; Brücke, C.; Huebl, J.; Brown, P.; Kühn, A.A. Subthalamic synchronized oscillatory activity correlates with motor impairment in patients with Parkinson's disease. *Movement Disorders*, 2016, 31(11), 1748-1751. DOI: <https://doi.org/10.1002/mds.26759>.
- [8] Little, S.; Pogosyan, A.; Neal, S.; Zavala, B.; Zrinzo, L.; Hariz, M.; Brown, P. Adaptive deep brain stimulation in advanced Parkinson disease. *Annals of neurology*, 2013, 74(3), 449-457. DOI: <https://doi.org/10.1002/ana.23951>.
- [9] Jiang, H.; Zhang, J.J.; Hebb, A.; Mahoor, M.H. (2013, November). Time-frequency analysis of brain electrical Signals for behavior recognition in patients with Parkinson's disease. In 2013 Asilomar conference on signals, systems and computers (pp. 1843-1847). IEEE. DOI: <https://doi.org/10.1109/ACSSC.2013.6810621>.
- [10] Niketeghad, S.; Hebb, A.O.; Nedrud, J.; Hanrahan, S.J.; Mahoor, M.H. (2014, August). Single trial behavioral task classification using subthalamic nucleus

- local field potential signals. In 2014 36th annual international conference of the IEEE engineering in medicine and biology society (pp. 3793-3796). IEEE. DOI: <https://doi.org/10.1109/EMBC.2014.6944449>.
- [11] Golshan, H.M.; Hebb, A.O.; Hanrahan, S.J.; Nedrud, J.; Mahoor, M.H. (2016, August). A multiple kernel learning approach for human behavioral task classification using STN-LFP signal. In 2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 1030-1033). IEEE. DOI: <https://doi.org/10.1109/EMBC.2016.7590878>.
- [12] Kuhner, A.; Schubert, T.; Cenciarini, M.; Wiesmeier, I.K.; Coenen, V.A.; Burgard, W.; Maurer, C. Correlations between motor symptoms across different motor tasks, quantified via random forest feature classification in Parkinson's disease. *Frontiers in neurology*, 2017, 8, 607. DOI: <https://doi.org/10.3389/fneur.2017.00607>.
- [13] Shah, S.A.; Tinkhauser, G.; Chen, C.C.; Little, S.; Brown, P. (2018, July). Parkinsonian tremor detection from subthalamic nucleus local field potentials for closed-loop deep brain stimulation. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2320-2324). IEEE. DOI: <https://doi.org/10.1109/EMBC.2018.8512741>.
- [14] Wang, T.; Shoaran, M.; Emami, A. (2018, April). Towards adaptive deep brain stimulation in Parkinson's disease: Lfp-based feature analysis and classification. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 2536-2540). IEEE. DOI: <https://doi.org/10.1109/ICASSP.2018.8462472>.
- [15] Kuhner, A.; Wiesmeier, I.K.; Cenciarini, M.; Maier, T.L.; Kammermeier, S.; Coenen, V.A.; Maurer, C. Motion biomarkers showing maximum contrast between healthy subjects and parkinson's disease patients treated with deep brain stimulation of the subthalamic nucleus. a pilot study. *Frontiers in neuroscience*, 2020, 13, 1450. DOI: <https://doi.org/10.3389/fnins.2019.01450>.
- [16] He, S.; Baig, F.; Mostofi, A.; Pogosyan, A.; Debarros, J.; Green, A.L.; Tan, H. Closed-loop deep brain stimulation for essential tremor based on thalamic local field potentials. *Movement Disorders*, 2021, 36(4), 863-873. DOI: <https://doi.org/10.1002/mds.28513>.
- [17] Hirschmann, J.; Steina, A.; Vesper, J.; Florin, E.; Schnitzler, A. Neuronal oscillations predict deep brain stimulation outcome in Parkinson's disease. *Brain Stimulation*, 2022, 15(3), 792-802. DOI: <https://doi.org/10.1016/j.brs.2022.05.008>.
- [18] Hosny, M.; Zhu, M.; Gao, W.; Fu, Y. A novel deep learning model for STN localization from LFPs in Parkinson's disease. *Biomedical Signal Processing and Control*, 2022, 77, 103830. DOI: <https://doi.org/10.1016/j.bspc.2022.103830>.
- [19] Houston, B.; Thompson, M.; Ko, A.; Chizeck, H. A machine-learning approach to volitional control of a closed-loop deep brain stimulation system. *Journal of neural engineering*, 2019, 16(1), 016004. DOI: <https://doi.org/10.1088/1741-2552/aae67f>.
- [20] Reck, C.; Florin, E.; Wojtecki, L.; Krause, H.; Groiss, S.; Voges, J.; Timmermann, L. Characterisation of tremor-associated local field potentials in the subthalamic nucleus in Parkinson's disease. *European journal of neuroscience*, 2009, 29(3), 599-612. DOI: <https://doi.org/10.1111/j.1460-9568.2008.06597.x>.

- [21][dataset]Wiest, C.; Torrecillos, F.; Pogosyan, A.; Baig, F.; Pereira, E.; Morgante, F.; Tan, H. (2022). STN local field potential recordings from awake patients with Parkinson's, ON and OFF meds, and during 130 Hz DBS. DOI: <https://doi.org/10.5287/bodleian:mzJ7YwXvo>.
- [22]Zhao, D.; Luo, Z.; Yao, M.; Wei, L.; Qin, L.; Wang, Z. State identification of Parkinson's disease based on transfer learning. *Technology and Health Care*, 2024, 32(6), 4097-4107. DOI: <https://doi.org/10.3233/THC-231929>.
- [23]Dastin-van Rijn, E.M.; Provenza, N.R.; Calvert, J.S.; Gilron, R.E.; Allawala, A.B.; Darie, R.; Borton, D.A. Uncovering biomarkers during therapeutic neuromodulation with PARRM: Period-based Artifact Reconstruction and Removal Method. *Cell reports methods*, 2021, 1(2). DOI: <https://doi.org/10.1016/j.crmeth.2021.100010>.
- [24]Lee, S.B.; Kim, Y.J.; Hwang, S.; Son, H.; Lee, S.K.; Park, K.I.; Kim, Y.G. Predicting Parkinson's disease using gradient boosting decision tree models with electroencephalography signals. *Parkinsonism & Related Disorders*, 2022, 95, 77-85. DOI: <https://doi.org/10.1016/j.parkreldis.2022.01.011>.
- [25]Alqahtani, A.F.; Ilyas, M. An ensemble-based multi-classification machine learning classifiers approach to detect multiple classes of cyberbullying. *Machine Learning and Knowledge Extraction*, 2024, 6(1), 156-170. DOI: <https://doi.org/10.3390/make6010009>.
- [26]Liu, Y., Liu, Z., Luo, X.; Zhao, H. Diagnosis of Parkinson's disease based on SHAP value feature selection. *Biocybernetics and Biomedical Engineering*, 2022, 42(3), 856-869. DOI: <https://doi.org/10.1016/j.bbe.2022.06.007>.
- [27]Lu, R.; Xu, Y.; Li, X.; Fan, Y.; Zeng, W.; Tan, Y.; Cao, X. Evaluation of wearable sensor devices in Parkinson's disease: A review of current status and future prospects. *Parkinson's Disease*, 2020, 2020(1), 4693019. DOI: <https://doi.org/10.1155/2020/4693019>.
- [28]Leung, C.H.S.; Simpson, H.D.; Thyagarajan, D. The Place of Local Field Potentials in Deep Brain Stimulation Programming for Parkinson's Disease: A Review. *Brain Sciences*, 2025, 15(2), 116. DOI: <https://doi.org/10.3390/brainsci15020116>.
- [29]Shokrpour, S.; MoghadamFarid, A.; Bazzaz Abkenar, S.; Haghi Kashani, M.; Akbari, M.; Sarvizadeh, M. Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges. *NPJ Parkinson's Disease*, 2025, 11(1), 187. DOI: <https://doi.org/10.1038/s41531-025-01025-9>.
- [30]Bakstein, E.; Burgess, J.; Warwick, K.; Ruiz, V.; Aziz, T.; Stein, J. Parkinsonian tremor identification with multiple local field potential feature classification. *Journal of neuroscience methods*, 2012, 209(2), 320-330. DOI: <https://doi.org/10.1016/j.jneumeth.2012.06.027>.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)