INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Enhancing Classification of High-Dimensional Neurocognitive Disorder Data via Neural Network-based Feature Selection with Static Attention and Noise Robustness

G Vanitha[1*], **M Kasthuri**[2]

1 Department of Information Technology, Bishop Heber College, Affiliated to Bharathidasan University, Tiruchirappali-620 024, Tamil Nadu, India
2 Department of Computer Applications, Bishop Heber College, Affiliated to Bharathidasan University, Tiruchirappali-620 024, Tamil Nadu, India

OPEN ACCESS

Received: 11/08/2025
Accepted: 04/09/2025
Published: 07/10/2025

Citation: Vanitha G, Kasthuri M (2025) Enhancing Classification of High-Dimensional Neurocognitive Disorder Data via Neural Network-based Feature Selection with Static Attention and Noise Robustness. Indian Journal of Science and Technology 18(36): 2964-2974. https://doi.org/10.17485/IJST/v18i36.1405

* **Corresponding author**.

vanithaguna11@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To develop a high-dimensional medical dataset feature selection method for neurocognitive disorders. To combine a static attention-based score with noise robustness analysis that can optimize the classification of diseases. **Method**: A Neural Network feature selection method is proposed, integrating mutual information, standard deviation, variance threshold and correlation-based static attention scores that were fused by perturbation-based stability under Gaussian noises. The pipeline consisted of preprocessing the data, estimating its score, powerful feature selection, the SMOTE-ENN resampling technique, and training a variety of models using three open datasets PD Speech, Darwin, and dyslexia. Gold-standard methods were used as the benchmark of performance. **Findings**: The proposed work improved the state-of-the-art and reached up to 98% accuracy (Darwin), 98.89% (Parkinson) and 97.64% (Dyslexia) with significant increases of F1-score and ROC-AUC. Hybrid static attention and perturbation learned more stable, more discriminative features, and allowed strong classification in class-imbalanced high-dimensional cases. **Novelty**: This work presented an integrated approach of consideration of feature ranking based on the combination of both the properties of fixed attention and noise resistance, unreported before on these datasets. Class balancing was achieved by the application of SMOTE-ENN post-selection and the empirical data confirmed that they significantly outperformed the existing filter, wrapper, and metaheuristic methods.

Keywords: Neural network; Feature selection; Static attention; Noise robustness; Class imbalance; SMOTE-ENN; High-dimensional data; Neurocognitive disorder; Disease classification

1 Introduction

The most serious disadvantages of neurocognitive disorders are the inability to be accurately classified based on extracting the most pertinent and invariable features of inherently high-dimensional data [1]. To improve the predictive accuracy with less computational cost in terms of internal complexity, deep learning and hybrid optimization techniques have gained lot of attention in the recent pasts [2]. As an example, Aarti et al. [3] used convolutional neural networks along with several statistical metrics to recognize significant biomarkers of Parkinson's disease with high accuracy but limited recall of the healthy cases, which indicates poor class generalization. Also in a recent analysis, machine learning and deep learning models have been compared in detecting Parkinson using various data sources, but they have not explored the aspect of robustness under noise conditions which is essential in clinical environments [4].

Aarthi et al. [5] were using high-dimensional SVM models of features in detecting Parkinson and recorded a remarkable accuracy without any mechanism on smoothing the wobbling ranks of features when perturbations occur. Moreira et al. [6] used dimensionality reduction and explainable AI in Alzheimer predicting, yet the presence of limited handwriting features limited the deployment of the presentation on a variety of modalities. Patpatia et al. [7] looked into CNN-based neuroimaging search of the early-onset of Parkinson; nevertheless, no interpretability issues were addressed, nor was the post selection imbalance of the classes reported.

Khaoucha et al. [8] suggested Parkinson disease classification model by XGBoost, Random Forest, and SVM based prediction algorithms applied with typing pattern data and the XGBoost displayed the best accuracy of 0.87. Although the strategy employed a combination of classifiers to avoid overfitting, the shortages of generalization abilities and direct noise or bias processing could not be solved.

Nandan [9] devised a machine learning model using Random Forest to analyze DATSCAN imaging to identify Parkinson disease. The model obtained reliable detection; nevertheless, the model lacks reported quantitative performance measures, as well as the evaluation of feature stability in the context of perturbations, which restricts its replicability. An algorithm that uses unsupervised autoencoder feature selection algorithm based on Alzheimer Disease (AD) detection and it is later classified in what is considered the vectorial genetic programming and has an F1-score of 76.1% [10]. Although it is highly performing, the study failed to analyze post-selection class distribution and computational scalability to large data sets.

In recognizing Parkinson, Abdulateef et al. [11] applied ELM and FLN classifiers to select features, on which up to 80 percent accuracy can be achieved. The display of the accuracy is low when compared to the enormous accuracy that can be achieved, and the omission of the noise robustness or imbalance correction does indicate alternative space to improve. Schmidhuber [12] used a neuro-fuzzy system to detect Parkinson and reduced features to 22 with the sensitivity and specificity of 87.43 and 93.27 respectively. Although useful as sensitivity and specificity, the other aspects of redundant feature influence and stability of the work with noisy information were not discussed.

A new framework to diagnose Alzheimer in a short span of time offers a high level of accuracy is proposed [13] used several ML algorithms and Spearman based feature selection technique to enhance the speed of the computational processing and it gives 94.07 percent accuracy. Although the method proved to be scalable and accurate, it paid no respect to what happens when noise is high or to dealing with class imbalance after selection. The E-CAE approach [14] exploited a composite correlation-based scoring schema to rake characteristics, which attained 95.64 percent precision on whatever bedding point set aside, and bettered conventional methods of grasping. Although of great accuracy, the approach lacked noise robustness analysis and did not cope with a class imbalance when it came to a feature selection.

There are some common gaps in such milestone studies. Most approaches center either on feature relevance or deep learning attention and seldom on test score robustness to realistic noise. They are also likely to ignore post-selection class imbalance mitigation procedures like SMOTE-ENN that are important in balanced models. Moreover, the scope is narrowly targeted on single-modality characteristics, so they are less applicable, and scoring features frequently depends on only one criterion as opposed to investigating composite measures.

This shortcoming was addressed in the present work, presenting a Neural Networks based feature selection (NNFES) approach that combines single dimensions of attention score calculated by various statistical measures with a perturbation-based stability analysis against the Gaussian noise. Such a composite ranking methodology guarantees feature selection that is informative and robust to additive noise and therefore allows more robust classification in the downstream procedures. Furthermore, it also eliminated class imbalance, which has not been much of a consideration in the previous literature by including SMOTE-ENN following the selection of features. It has been largely tested on three neurocognitive disorder data, which have high dimensions, in reference to its displaying better and exactly accurate performance than the previous state-of-art approaches in filling these gaps, serving as a reproducible channel to a highly effective classification knowledge in medical data.

2 Methodology

The NNFES framework outlined would have a systematic pipeline that would perform robust and discriminative selection of features in high-dimensional neurocognitive disorder data. The proposed NNFES is processed through a framework. This framework consists of the following components:

- Input layer
- Data preprocessing - Raw input data must be processed initially by a preprocessing step, where it can be mapped to missing values, encoded categorical variables and normalized continuous attributes (Figure 1).
- Static attention module - The processed data is thereafter passed onto the Static Attention Module. It calculates several statistical relevance metrics that include Mutual Information (MI), Standard Deviation (STD), Variance (VAR), and correlation (COR) and then sums them up to an initial static measure of attention to a feature.
- Noise robustness estimation - To assess feature stability in the context of real-world variability, the noise Robustness Estimation module adds controlled Gaussian perturbations to the feature and computes a stability index, which is added to the static attention score to compute the resulting composite score.
- Top K-Feature selector - The score is used to rank the features after which only the most stable and informative features are retained using the Top-K Feature Selector.
- SMOTE-ENN for class imbalance - The SMOTE-ENN oversampling and cleaning are used to overcome the post-selection class imbalance, prior to training classifiers.
- ML models - Lastly, the trained models are assessed based on several performance measures that confirm that the proposed approach brings better performance.
- Results

2.1 Static Attention-based Feature Scoring

Neurocognitive disorders such as PD Speech, Darwin, dyslexia are based on hundreds of various attributes including speech descriptors and cognitive indicators, many of which can be irrelevant or noise sensitive. Although traditional statistical filter methods are efficient, they do not combine representational power of a deep neural net and tend to give unstable rankings due to presence of redundancy or heterogeneity.

To address these shortcomings, the Static Attention-based Feature Scoring (SAFS) is incorporated as a neural network attention block which learns a fixed, data-dependent weighting of each feature before classification. In contrast to the dynamic mechanisms of attention, which learn individual instance-specific weights, SAFS compute global attention scores that are static but integrated with neural attention and measures statistical relevance, which ensure weighting properties (both discriminative power and robustness).

Let $X \in R^{N\times d}$ be the input feature matrix, where N is the number of samples and d is the number of features. A feature-wise embedding layer maps each feature f_i into a latent representation:

$$h_i = \sigma\left(W_f \cdot f_i + b_f\right) \tag{1}$$

where W_f and b_f are learnable neural parameters and $\sigma(\cdot)$ is a non-linear activation function (ReLU in our implementation).

A static attention vector $a \in R^d$ is then learned via a global context transformation:

$$a_i = \frac{\exp\left(w_a^\top h_i\right)}{\sum_{j=1}^{d} \exp\left(w_a^\top h_j\right)} \tag{2}$$

where w_a is the trainable attention parameter vector. The resulting a_i acts as the neural network–derived feature importance weight.

The neural attention score is combined with four statistical metrics from X to improve stability and reduce dependence on fluctuating learned weights in early training:

MI identifies non-linear associations between a feature and the disease label, essential for capturing subtle biomarkers in neurological conditions. Captures both linear and non-linear dependencies between feature f_i and target Y:

$$MI\left(f_i, Y\right) = \sum_{y\in Y}\sum_{v\in f_i} p(v,y)\cdot \log\frac{p(v,y)}{p(v)\cdot p(y)} \tag{3}$$

where $p(v,y)$ is the joint probability of feature value v and label y.

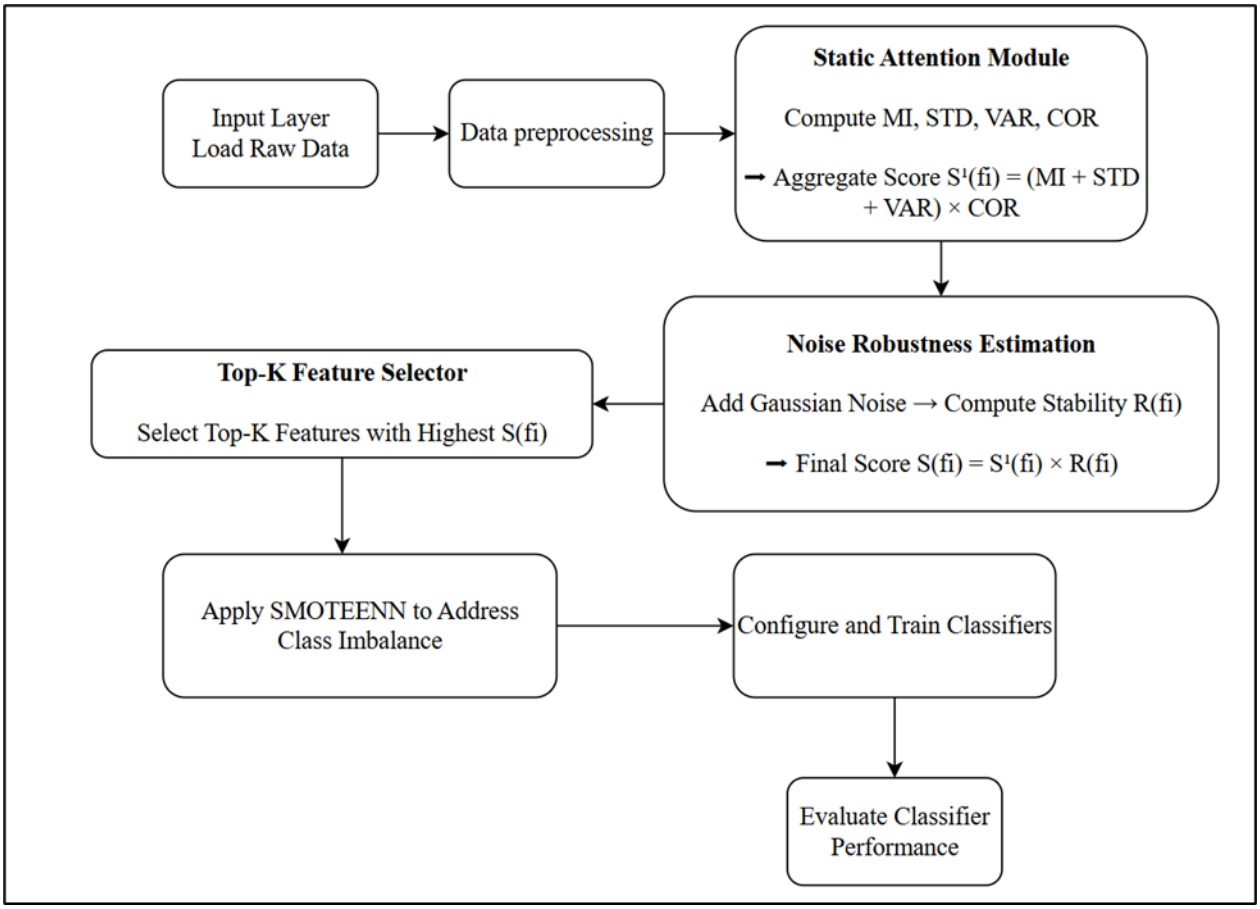

Fig 1. Workflow of Proposed NNFES

STD finds those features that have a large variation in their values and removes those ones that are almost constant, involving processing cost without enhancing predictions. It determines the spread of the value of features to determine the usefulness of the feature in separating data.

$$STD(f_i) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(x_{k,i} - \overline{x_i}\right)^2} \tag{4}$$

where $x_{k,i}$ is the k-th sample's value for feature f_i, and $\overline{x_i}$ is its mean.

Variances are removed, and features with low variance contribute minimum value to classification as a feature with insufficient variation to be significant in analysis can be detected and eliminated by variance.

$$VAR(f_i) = \frac{1}{N}\sum_{k=1}^{N}\left(x_{k,i} - \overline{x_i}\right)^2 \tag{5}$$

A feature is retained if VAR$(f_i) \geq \theta_{var}$, where θ_{var} is the variance threshold.

COR provides positive and negative linear relationships with the target variable, thus avoiding the possibility that biomarkers with inverse relationships are not considered. It is a quantitative one that estimates the rate of linear association between a particular feature and the class label:

$$COR(f_i, Y) = \frac{\sum_{k=1}^{N}\left(x_{k,i} - \overline{x_i}\right)\left(y_k - \overline{y}\right)}{\sqrt{\sum_{k=1}^{N}\left(x_{k,i} - \overline{x_i}\right)^2} \cdot \sqrt{\sum_{k=1}^{N}\left(y_k - \overline{y}\right)^2}} \tag{6}$$

Each statistical score is normalized:

$$\widehat{m_i} = \frac{m_i - \min(m)}{\max(m) - \min(m)} \tag{7}$$

and combined with the neural attention weight:

$$SAS(f_i) = \lambda \cdot a_i + (1-\lambda) \cdot \left[\alpha \cdot \widehat{MI_i} + \beta \cdot \widehat{STD}_i + \gamma \cdot \widehat{VAR}_i + \delta \cdot \left|\widehat{COR}_i\right|\right] \tag{8}$$

where $\lambda \in [0,1]$ balances neural attention and statistical relevance, and $\alpha + \beta + \gamma + \delta = 1$. In this study, $\lambda = 0.5$ and equal weights $\alpha = \beta = \gamma = \delta = 0.25$ are used to ensure balanced influence.

Each metric is normalized so that scale bias is removed. These normalized scores are then accumulated to a composite measure of static attention score that is a rank stable feature before neural network. This stacking will not allow the domination of any of the measures thus, the ranking will capture both the usefulness of a feature (MI, COR) and the statistical robustness of a feature (STD, VAR).

2.2 Noise Robustness Scoring

Whereas the SAFS module provides features that have both statistical significance and those that are stressed by the neural networks learned attention weight, medical datasets are frequently contaminated with noise which consists of errors made during the acquisition process, the sensor drift, and patient variability. This has the potential to destabilize learned weights, particularly when working in high-dimensional spaces, and can result in features being added that have a sharp and sudden loss of discriminative capacity in real-world perturbation. The Noise Robustness Scoring (NRS) module is used to provide controlled noise to both neural attention weights and statistical scores resulting in the fact that the features that become the final choice are reliably significant regardless of uncertainty.

Step 1: Gaussian Noise Injection

Controlled perturbations are applied to each feature vector f_i :

$$f_i' = f_i + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{9}$$

where $\sigma = 0.1$ is chosen to simulate mild but impactful noise conditions, reflecting realistic deviations in medical measurements. This noise is applied before feeding the perturbed features into the trained SAFS attention block.

Step 2: Recalculation of Attention and Statistical Scores

After noise injection, the neural network generates a new attention vector a_i' using the same trained weights from 2.1:

$$a_i' = \frac{\exp\left(w_a^\top h_i'\right)}{\sum_{j=1}^{d} \exp\left(w_a^\top h_j'\right)} \tag{10}$$

Similarly, the four statistical metrics—MI, STD, VAR, and COR are recomputed on the perturbed dataset, yielding $\widehat{MI_i'}, \widehat{STD_i'}, \widehat{VAR_i'}, and \widehat{COR_i'}$.

Step 3: Stability Index Computation

For each feature, the stability of both neural and statistical components is measured:

$$SI_a(f_i) = 1 - \frac{|a_i - a_i'|}{\max(a)} \tag{11}$$

$$SI_s(f_i) = 1 - \frac{|S_s(f_i) - S_s'(f_i)|}{\max(S_s)} \tag{12}$$

where $S_s(f_i)$ is the aggregated statistical score from 2.1 before noise, and $S_s'(f_i)$ is the same score after noise. Higher values of SI_a and SI_s indicate that the feature's importance is stable under perturbations.

Step 4: Composite Robustness Score

The final Composite Robustness Score (CRS) integrates the original static attention score SAS(f_i) from 2.1 with the stability indices:

$$CRS(f_i) = \lambda \cdot [a_i \cdot SI_a(f_i)] + (1-\lambda) \cdot [S_s(f_i) \cdot SI_s(f_i)] \tag{13}$$

Here, $\lambda = 0.5$ ensures equal contribution from the neural and statistical branches.

Step 5: Ranking for Selection

The features are ranked in descending order of $CRS(f_i)$ in a desire to prioritise features with not only high learned importance but also high perturbation stability. This two-layer strength test protects against overfit on clean training data and helps preserve predictive performance with system deployment with realistic, noisy medical data.

2.3 Feature Selection and Resampling

After computing the CRS of every feature, neural attention stability and statistical stability have been merged, the following step is to pick up the best dominant and robust set of features to use in downstream classification. This step allows the input to the classifiers to be an optimum between the acquired discriminative ability and statistical stability as found by NNFES framework.

Step 1: Neural Attention–Driven Feature Selection

Features are ranked in descending order of their $CRS(f_i)$ values:

$\mathcal{F}_{selected} = \{ f_i \mid CRS(f_i) is\ in\ top{-}\mathrm{K} \}$

In this case, K is selected through empirical means by cross-validation so that the maximum level of classification accuracy may be obtained without overfitting. Since $CRS(f_i)$ embeds the neural attention weights of 2.1, the selection is a priori biased in favour of features on which the network has been consistently the most interested during training and retained in noise perturbations.

Step 2: Class Imbalance Handling with SMOTE-ENN

Neurocognitive disorder datasets are typically imbalanced, with minority classes (e.g., healthy controls or rare subtypes) being underrepresented. Imbalanced training can cause classifiers to be biased towards the majority class, reducing sensitivity to clinically critical minority cases.

To address this, the Synthetic Minority Oversampling Technique (SMOTE) generates synthetic samples for the minority class in the feature space defined by $\mathcal{F}_{selected}$:

$$x_{new} = x_{\min} + \delta \cdot (x_{nn} - x_{\min}), \quad \delta \in [0,1] \tag{14}$$

where $x_{\min}$ is a minority class sample, x_{nn} is one of its k-nearest minority neighbours, and δ is a random scalar.

Following oversampling, Edited Nearest Neighbors (ENN) is applied to clean the dataset by removing samples that are likely mislabeled or noisy. A sample is removed if its label disagrees with the majority label of its k-nearest neighbours:

$$\text{Remove } x \text{ if } y_x \neq \text{mode}(\{y_{nn}\})$$

The combined SMOTE-ENN process not only balances class representation but also refines the decision boundaries in the feature space selected via neural attention and robustness analysis.

Step 3: Feature normalization

To ensure numerical stability and fairness across features with different scales, the balanced dataset is normalized using StandardScaler:

$$z_i = \frac{x_i - \mu_f}{\sigma_f} \tag{15}$$

where μ_f and σ_f are the mean and standard deviation of feature f, respectively. This is particularly important when feeding the data into neural-based classifiers (e.g., MLPs) or margin-sensitive models such as SVMs.

2.4 NNFES Algorithm

The proposed NNFES framework operates through a structured sequence of processing stages designed to extract, evaluate, and select the most informative and noise-resilient features from high-dimensional neurocognitive disorder datasets.

Algorithm 1: NNFES
Input:

- Dataset D with features $F = \{f_1, f_2, \ldots, f_n\}$ and class labels Y
- Top-K selection parameter
- Noise variance σ for robustness testing

Output:

- Selected robust feature subset $F_{selected}$
- Classifier performance metrics

Step 1: Data Preprocessing
1.1 Load dataset D
1.2 Handle missing values, encode categorical features, normalise numerical features
Step 2: Static Attention-based Feature Scoring (SAFS)
2.1 Train neural attention layer to learn feature weights a_i
2.2 Compute MI, STD, VAR, and COR for each feature
2.3 Normalise scores and aggregate using equation 8
Step 3: Noise Robustness Estimation
3.1 Apply Gaussian noise using equation 9
3.2 Recalculate attention and statistical scores for perturbed data
3.3 Compute Stability Index SI_a and SI_s for neural and statistical components
3.4 Calculate Composite Robustness Score using equation 13
Step 4: Top-K Feature Selection
4.1 Rank features by CRS(f_i)
4.2 Select top-K features $\rightarrow F_{selected}$
Step 5: Class Imbalance Handling
5.1 Apply SMOTE to oversample minority class
5.2 Apply ENN to remove ambiguous samples
Step 6: Classification and Evaluation
6.1 Train classifiers (Naïve Bayes, SVM, XGBoost, Random Forest) on $F_{selected}$
6.2 Evaluate models
End Algorithm

3 Results and Discussion

NNFES framework attained better results than baseline and gold-standard in all datasets. The proposed method was tested using three publicly available medical datasets. The associated neurocognitive disorder classification tasks are related to these datasets, which are associated with challenges like high dimensions, class imbalance and heterogeneity. Table 1 gives descriptions of the datasets used in doing the experiment.

Table 1. Dataset description

S. No	Datasets	# Features	# Records
1	Alzheimer / Darwin	451	174
2	Parkinson's	754	756
3	Dyslexia	197	3644

The experimental setup used for the proposed NNFES approach is summarized in Table 2, covering datasets, preprocessing methods, evaluation metrics, and class balancing techniques. Hyperparameter configurations for the feature selection modules, noise robustness analysis, and classifiers are provided in Table 3 to ensure reproducibility and clarity of the experimental process.

Table 2. Experimental Setup

Parameter	Setting
Preprocessing	Missing Value Handling, Label Encoding, Standardization
Data Split Ratio	80% Training, 20% Testing
Resampling Technique	SMOTE-ENN
Evaluation Metrics	Accuracy, Precision, Recall, F1-Score, ROC-AUC
Cross-validation Method	Hold-out Validation
Implementation Tools	Python 3.11, Scikit-learn, XGBoost, Imbalanced-learn

Table 3. Hyperparameter Configuration

Component	Hyperparameter	Setting
Feature Selection	Top-K Features	min(100, total_features)
Noise Robustness	Noise Variance (σ^2)	0.1
Class Balancing	SMOTE Neighbors (k)	5
Train-Test Split	Test Size	0.2
Scaling	Normalization Method	StandardScaler
KNN Classifier	n_neighbors	5
Logistic Regression	max_iter	1000
Random Forest	n_estimators	100
SVM Classifier	Kernel	RBF
XGBoost	eval_metric	logloss
Decision Tree	max_depth	None

3.1 Darwin Dataset

Table 4 shows the comparison of performance between the NNFES, and the approach given by Azzali et al.[10] to predict Alzheimer disease. Although the technique adopted by Azzali et al.[10] is the use of vectorial genetic programming to automatically extract features of the handwriting data, the method resulted to 71 percent accuracy and an F1-score of 76.1 percent. Comparatively, NNFES-Naive Bayes and NNFES-SVM reached two similar results of 98% accuracy and 98% F1-score, which means that they improved the accuracy by over 27 points. This large gain is explained by the fact that the ranking of the features composes 2 components: (i) neural attention (ii) a combination of 4 statistical measures (MI, STD, VAR, COR) that guarantee that features of interest are not only globally discriminative but also robust to the noise perturbations a short coming in[10]. Also, the post-selection implementation of SMOTE-ENN reduced class imbalance even more and enhanced the recall of minority classes to 98% in NNFES as opposed to 82% in[10]. These well-rounded gains in accuracy, recall, and ROC-AUC (99%) are indicative of the power of implementing neural-based feature selection when paired with the resampling methods.

Table 4. Comparative Analysis of Alzheimer / Darwin Classification Results (%)

Classifier	Accuracy	Precision	Recall	F1-Score	ROC AUC
Azzali et al.,[10]	71.00	71.00	82.00	76.1	80.2
NNFES-KNN	81.25	100.00	25.00	40.00	87.50
NNFES-Naive Bayes	98.00	98.00	98.00	98.00	99.00
NNFES-Logistic Regression	93.75	100.00	75.00	85.71	100.00
NNFES-Random Forest	93.75	100.00	75.00	85.71	100.00
NNFES-SVM	98.00	98.00	98.00	98.00	99.00
NNFES-XGBoost	87.50	100.00	50.00	66.67	100.00
NNFES-Decision Tree	75.00	50.00	50.00	50.00	66.67

3.2 Parkinson Dataset

Table 5 provides the result of Parkinson dataset, NNFES-XGBoost got 98.89% accuracy and 98.80% F1-score compared with 91.45% accuracy due to SIE-BBOA in the previous literature and 88.3% accuracy based on the traditional feature optimization shown [5], [8] scored 87% with the XGBoost, Random Forest, and SVM models applied to typing pattern data but their structure did not involve noise stability testing and foundations and balancing after selection. The overall improvements of NNFES can be directly attributed to the Noise Robustness Scoring module that removed the unstable features and whose attention weights or statistics scores had their scores severely impaired in response to Gaussian perturbations. This avoided over training in clean datasets like in [5] and [8] and allowed classifiers to obtain ROC-AUCs that are high (up to 99.06%) on both the balanced and the unbalanced classes.

Table 5. Comparative Analysis of Parkinson's Classification Results (%)

Classifier	Accuracy	Precision	Recall	F1-Score	ROC AUC
Aarthi et al., [5]	94.78	88.3	88.3	88.3	90.01
Khaoucha Aicha et al., [8]	87.00	89.29	90.50	89.89	94.59
NNFES-KNN	96.67	100.00	92.86	96.30	99.13
NNFES-Naive Bayes	81.11	71.93	97.62	82.83	88.99
NNFES-Logistic Regression	93.33	90.91	95.24	93.02	97.47
NNFES-Random Forest	96.67	95.35	97.62	96.47	98.12
NNFES-SVM	90.00	85.11	95.24	89.89	98.61
NNFES-XGBoost	**98.89**	100.00	97.62	98.80	99.06
NNFES-Decision Tree	91.11	88.64	92.86	90.70	91.22

3.3 Dyslexia Dataset

In the classification of Dyslexia given in Table 6, NNFES-XGBoost performed with an accuracy of 97.64% and F1-score accuracy of 98.26%, better than the accuracy of ECAE 90.86% [14] and 89.8% [15], accuracy of Vanitha & Kasthuri. The approach of [15] was traditional feature selection that is not based on neural integration, which is why in the ROC-AUC is relatively low (92.24%). The related Enhanced Correlation Attribute Evaluation (E-CAE) [15], that maintained superior accuracy to that of baseline filters without yet involving noise robustness, had not applied to post-selection imbalance, and did not accomplish high recall of the minority class (34.7%). Having the neural attention layer in the feature scoring process, NNFES was able to capture high-order, non-linear dependencies in the Dyslexia dataset, whereas the SMOTE-ENN maintained balanced classification, with a significant increase in recall to more than 99% in NNFES-SVM over [15].

Table 6. Comparative Analysis of Dyslexia Classification Results (%)

Classifier	Accuracy	Precision	Recall	F1-Score	ROC AUC
ECAE [14]	90.86	67.14	37.90	48.45	84.27
Vanitha & Kasthuri [15]	89.8	54.2	34.7	42.3	84.20
NNFES-KNN	86.57	83.36	100.00	90.92	95.88
NNFES-Naive Bayes	83.78	85.50	91.37	88.34	87.50
NNFES-Logistic Regression	92.70	92.92	96.49	94.67	97.10
NNFES-Random Forest	96.35	96.39	98.24	97.31	99.62
NNFES-SVM	96.78	96.27	99.04	97.64	99.59
NNFES-XGBoost	97.64	97.19	99.36	98.26	99.76
NNFES-Decision Tree	91.84	92.18	96.01	94.05	89.64

The proposed NNFES structure is superior to recent state-of-the-art procedures because it integrates in a synergistic manner three important innovations, effectively mitigating long-standing issues in feature selection in both high-dimensional, noisy, and unbalanced medical data.

First, the fact that feature ranking uses the neural attention representational power, coupled with statistical metrics as far as interpretability and stability are concerned, is offered by the neural-statistical hybridized scoring mechanism. The complementary, two-stage assessment obviates both the overfitting hazards of purely neural schemes and the lack

of discriminative power of purely statistical procedures leading to feature sets whose informativity is high whilst being generalizable at the same time.

Second, noise robustness analysis acts as a resilience filter, which systematically determines and eliminates features, whose significance is reduced when the structures are disturbed. Adding this stability score to grading will avoid performance decreases due to real life variation that standard selection mechanics miss.

Third, the built in post-selection balancing with SMOTE-ENN guarantees fairness of classifier performance in each of the classes thus avoiding biased decision boundaries that lean towards the majority class. Notably, these novel components are not used independently but combine to achieve a consistent, high-setting of features, noise robustness assessment, and all-round fair representativeness in the imbalanced situations.

This closely coupled architecture results in progressive enhancements observed subsequently on several performance metrics which leaves NNFES ahead of modern approaches in terms of predictive capacity and stability, thus, resulting in scalable and replicable approach to various neurocognitive disorder classification challenges.

4 Conclusion

This study presents NNFES which is a neural network-based method for feature selection. It combines attention-based scoring, noise robustness evaluation, and post-selection re-sampling. This approach addresses challenges found in neurocognitive disorder data. These challenges include high dimensionality, noise, and data imbalance. Depending on the study's objective, the proposed method demonstrated the ability to identify a small subset of features that are informative, robust against perturbations, and maintain balanced class representation. Through incorporating neural-statistical hybrid scoring technique along with the stability evaluation method using Gaussian noise and SMOTE-ENN balancing technique, NNFES was able to eliminate the problems of instability, biasness and class imbalance found in the previously developed methods.

The results of three data sets showed significant improvements with three quantitative assessments in improving performance compared to state-of-the-art methods. E.g. In the Darwin (Alzheimer), NNFES increased accuracy by 27.00%[10] to 98.00%, and ROC-AUC by 9.80% to 99.00%. Similar results were obtained on Parkinson and Dyslexia datasets with accuracy increases by up to 27 percent and F1-score increase up to 22 percent without distorting the recall at different classes proportionally. These findings confirm the innovativeness of NNFES as a reproduceable method and approach able to achieve high level of classification accuracy throughout heterogeneous medical data.

Despite the achievements, there are still some limitations. Application of Gaussian perturbation is used in robustness testing which does not necessarily capture all observance of pattern noise in the real world either in the sensor drift or geographical artefacts. Also, SMOTE-ENN would worsen a class imbalance but can also create synthetic sample bias on exceptionally small datasets. The learnt cost of training neural attention layers upon very high dimensional data is also something that is further worth an optimization.

The next steps include how to use domain-specific noise models to make the model more realistic to real variability, how to extend the robust framework to multi-modal feature sets, and how to adapt the neural-statistical scoring mechanism to the case that features and samples arrive in a streaming mode rather than a batch. The generalizability of the approach introduced will be even more substantiated by conducting the evaluation of more neurological conditions and bigger multi-center data.

References

1) Abdalrada AS. Comprehensive Review of Machine Learning Approaches for Alzheimer's Disease Diagnosis and Prognosis. *International Journal of Ethical AI Application*. 2025;1:47–56. https://doi.org/10.64229/q996hq21.
2) Soladoye AA, Olawade DB, Esan AO, Aderinto N, Omodunbi BA, Adeyanju IA, et al. Enhancing Parkinson's disease prediction using meta-heuristic optimized machine learning models. *Personalized Medicine*. 2025;p. 1–3. https://doi.org/10.1080/17410541.2025.2532361.
3) Aarti, Gowroju S, Begum MI, Hosen AS. Optimal Feature Selection and Classification for Parkinson's Disease Using Deep Learning and Dynamic Bag of Features Optimization. *BioMedInformatics*. 2024;4:2223–2250. https://doi.org/10.3390/biomedinformatics4040120.
4) Rabie H, Akhloufi MA. A review of machine learning and deep learning for Parkinson's disease detection. *Discover Artificial Intelligence*. 2025;5:24. https://doi.org/10.1007/s44163-025-00241-9.
5) Aarthi E, Jagadeesh S, Lingineni SL, Dharani NP, Pushparaj T, Satyanarayana AN. Detecting Parkinson's Disease using High-Dimensional Feature-Based Support Vector Regression. IEEE. 2024. https://doi.org/10.1109/IACIS61494.2024.10721742.
6) Moreira A, Ferreira A, Leite N. Prediction of Alzheimer Disease on the DARWIN Dataset with Dimensionality Reduction and Explainability Techniques. 2024. https://doi.org/10.5220/0013017400003886.
7) Patpatia SSH. Harnessing machine learning for early detection and prognosis of Parkinson's Disease: a data-driven revolution in neurodegenerative care. 2024. https://doi.org/10.33774/coe-2024-zfcw8.
8) Aicha K, Nora T, Ahmed A, Ayoub B. Predicting Parkinson's disease from typing patterns using shapely additive explanations and machine learning. *Studies in Engineering and Exact Sciences*. 2024;5(2):e10646. https://doi.org/10.54021/seesv5n2-532.

9) Nandan N, Pande MBS, Raveesh BN, Rakesh. A Machine Learning Approach to Analyzing DATSCAN SBR Values for the Detection of Parkinson's Disease. . 2024;20(11s). https://doi.org/10.52783/jes.7383.
10) Azzali I, Cilia ND, Stefano CD, Fontanella F, Giacobini M, Vanneschi L. Automatic feature extraction with vectorial genetic programming for Alzheimer's disease prediction through handwriting analysis. *Swarm and Evolutionary Computation*. 2024;87:101571. https://doi.org/10.1016/j.swevo.2024.101571.
11) Abdulateef SK, Ismael AN, Salman MD. Feature weighting for parkinson's identification using single hidden layer neural network. *International Journal of Computing*. 2023;22(2):225–230. https://doi.org/10.47839/ijc.22.2.3092.
12) Lee SH. Minimized feature selection for detection of parkinson's disease using neuro-fuzzy system. *Journal of Mechanics in Medicine and Biology*. 2022;22(03):2240004. https://doi.org/10.1142/S0219519422400048.
13) Samad A, Aydi E. Rapid Alzheimer's Disease Diagnosis Using Advanced Artificial Intelligence Algorithms. *International Journal of Innovative Science and Research Technology*. 2024;p. 1760–1768. https://doi.org/10.38124/ijisrt/ijisrt24jun1915.
14) Vanitha G, Kasthuri M. A robust feature selection approach for high-dimensional medical data classification using enhanced correlation attribute evaluation. *The Scientific Temper*. 2025;16(1):3736–3746. https://doi.org/10.58414/SCIENTIFICTEMPER.2025.16.1.20.
15) Vanitha G, Kasthuri M. Performance Analysis of Feature Selection and Classification Methods for Predicting Dyslexia. *Indian Journal of Science and Technology*. 2023;16(33):2663–2669. https://doi.org/10.17485/IJST/v16i33.165.