

Article

Explainable Semi-Supervised Learning Framework for Alzheimer's Disease Prediction Using SHAP-Based Feature Selection and Cost-Sensitive CatBoost

Abdallah El Chakik ¹, Bilal Nakhal ¹ and Ghalia Nassreddine ^{2,*}

¹ Department of Mathematics and Computer Science, Faculty of Science, Beirut Arab University, Beirut 1107 2809, Lebanon; a.alshakik@bau.edu.lb (A.E.C.); b.nakhal@bau.edu.lb (B.N.)

² Computer and Information Systems Department, Rafik Hariri University, Meshref 2010, Lebanon

* Correspondence: nassreddinega@rhu.edu.lb

Abstract

Alzheimer's disease (AD) remains a critical global health challenge for which early diagnosis is essential for effective intervention. However, AD detection is a challenging and complex task due to the scarcity of labeled clinical data, class imbalance, and lack of prediction interpretability in healthcare systems. Although current machine learning frameworks are promising, they are frequently based on fully supervised environments and lack transparency, which limits their ability to be adopted in real-world clinical scenarios where labeled data are often scarce. This study proposes a new lightweight, semi-supervised, SHAP-enhanced CatBoost (LSSSC) framework that covers the challenges related to insufficient label data, interpretability of prediction models, and performance in a single learning paradigm. The proposed solution provides two major innovations: (i) a hybrid semi-supervised learning algorithm that incorporates the use of confidence-aware pseudo-labeling with cost-sensitive learning, allowing the effective use of the unlabeled data and alleviating the problem of class imbalance, and (ii) an intrinsic SHAP-based feature selection mechanism that assists not only in increasing interpretability but also in minimizing the dimensionality of the feature variables. The suggested approach is tested on a real AD dataset of 2149 patients and 34 diverse clinical, demographic, and lifestyle characteristics. Experimental findings show that LSSSC outperforms traditional supervised and ensemble frameworks, with an accuracy of 95.35%, macro-F1 score of 0.9490, and AUC of 0.9513. Moreover, the combination of SHAP and LIME provides global and local interpretability, permitting a meaningful understanding of the model's predictions. These contributions make LSSSC a strong, scalable, and interpretable framework for early AD prediction and promise much with respect to its application to real-world clinical decision-support systems.



Academic Editors: Domenico De Berardis and Alessandro Castorina

Received: 29 April 2026

Revised: 4 July 2026

Accepted: 14 July 2026

Published: 15 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Alzheimer's disease; semi-supervised learning; explainable AI; SHapley Additive exPlanations; Local Interpretable Model-agnostic Explanations; feature selection; cost-sensitive learning; CatBoost

1. Introduction

Alzheimer's disease (AD) is one of the most common neurodegenerative disorders globally, with an estimated 5.3 million patients in the USA, including 5.1 million who are older than 65 years and roughly 200,000 who are younger than 65 years. Nowadays, a person in the USA is diagnosed with AD every 67 s. By 2050, it is expected that such a

diagnosis will be made every 33 s, and by then, there will be around 1 million new cases each year [1]. Dementia is one of the leading causes of death among all diseases according to the World Health Organization. It is also one of the major causes of impairment in the elderly population worldwide [2]. The pathophysiological mechanisms underlying AD are complex and involve an accumulation of amyloid-beta plaques and tau protein tangles, which affect neuronal networks and lead to a progressive loss of cognitive function, memory deficits, and difficulties in performing daily activities. While the apolipoprotein E (APOE) ϵ_4 allele and aging are the primary factors that affect the onset of Alzheimer's disease, other genetic, environmental, and lifestyle factors play roles in its onset. Recent studies suggest that women have an increased risk of developing AD and a more rapid rate of cognitive decline than men, suggesting that there are important biological distinctions in the onset and progression of the disease, including hormonal, brain-structure, and metabolism differences [3]. Timely detection and intervention are critical to improving patient outcomes, as approaches to treatment are most likely to be most effective in early stages of the disease, before neuronal damage is irreversible [4]. However, early AD detection is a laborious and costly process that requires the gathering of a lot of data, the use of advanced predictive technologies, and collaboration with an experienced doctor. Automated systems offer improved accuracy relative to human evaluations and are immune to human error; thus, they can be applied in medical decision-support systems. Many recent studies have utilized MRI scans, biomarkers, and quantitative data derived from MRI scans to investigate AD [5,6].

Conventional diagnostic methods for AD mostly depend on clinical evaluations, cognitive assessments, and costly neuroimaging modalities like positron emission tomography (PET) and magnetic resonance imaging (MRI) [7]. Nevertheless, these techniques frequently detect the disease only after considerable neuronal damage has transpired, thereby constraining therapeutic efficacy. AD International reports that most individuals with dementia globally do not have an official diagnosis, which excludes them from treatment and care. Existing diagnostic procedures require advanced techniques and high expenses, preventing healthcare providers from performing accessible screenings for patients who need testing in underserved areas with no access to specialized medical staff [8]. The development of predictive models became possible through recent technological progress, which enabled the analysis of complex multimodal datasets that contain demographic data, as well as cognitive assessment results, lifestyle information, and biomarker measurements. Research studies combining speech patterns with noninvasive biomarkers have achieved successful results with respect to the identification of cognitive decline through the application of new approaches [9].

Artificial intelligence (AI) and machine learning (ML) technologies have surfaced as effective instruments for tackling diagnostic problems by facilitating early prediction of many diseases, including AD, through the use of accessible clinical data [10]. The authors of [11] proposed a subject-level 3D Convolutional Neural Network (CNN) framework enhanced by preprocessing techniques to improve detection and classification performance across different AD levels. The authors used a structural MRI (SMRI) dataset collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database to test the performance of their proposed approach. The authors of [12] proposed a new approach for AD detection using deep learning techniques based on CNNs and MRI datasets. The authors used a pre-trained CNN model to extract features from brain MRI images and classified them into AD or non-AD categories. The authors of [13] proposed a new approach for AD detection that begins with the mapping of the brain's anatomical parts (hippocampus, white matter, and grey matter). The respective volumes were calculated from 3D MRI images. Image segmentation and calculation of volume were performed using two software

programs: Analyze Direct and Insight Segmentation and Registration Toolkit (ITK) Snap. Calculated volumes of the anatomical parts, incorporated with other demographic factors and Mini-Mental State Examination (MMSE) scores, were fed to different ML algorithms for AD detection. The extracted features were also combined randomly for other more complex analyses using ML classifiers like random forest, gradient boost, decision tree, and K nearest neighbor (KNN). The promising results of AI models for AD prediction face major barriers for medical use because their unexplained decision systems make them difficult to understand. Research studies today focus on improving ML accuracy, but they ignore the need to understand model outcomes. Healthcare providers continue to exhibit reluctance to depend on “black-box” algorithms, as their decision-making processes are neither comprehensible nor verifiable against clinical expertise. The lack of comprehension of model results leads to patient safety risks because medical staff must break rules to follow regulations or abandon their ethical standards as professionals. Explainable AI (XAI) frameworks become vital due to their ability to offer good explications for model predictions while preserving their high predictive accuracy [14].

Recently, the diagnosis of AD has also improved, with an increasing emphasis on “biomarker-driven” diagnosis, which allows for earlier and more accurate identification of pathological changes. Blood-based biomarkers like plasma-phosphorylated tau species (p-tau217 and p-tau231) have been shown to have high diagnostic accuracy and a good correlation with cerebral tau pathology and are promising minimally invasive alternatives to CSF analysis [15]. Likewise, activation of astroglia and neuroinflammation in the early stages of AD have been identified as being associated with the expression of glial fibrillary acidic protein (GFAP) [16]. In addition to fluid biomarkers, neuroimaging methods including amyloid positron emission tomography (PET) and tau PET are considered gold-standard methods for in vivo visualization of amyloid plaque and neurofibrillary tangle deposition [17]. The above developments are aligned under the AT(N) framework, which categorizes biomarkers into amyloid (A), tau (T), and neurodegeneration (N) categories and gives a biological definition of AD that is standardized. Taken together, these advances point to a paradigm shift toward diagnosis based on biology and suggest that computational and machine learning approaches must be able to complement and augment biomarker-based clinical decisions [18].

In addition, understanding the main factors influencing the initiation and advancement of AD is essential for accurate detection and prevention. These factors encompass lifestyle choices, healthcare information, genetic factors, and environmental effects. A more profound understanding of these elements can clarify the disease’s causes and assist in identifying at-risk people. Such an understanding is crucial for the formulation of tailored preventative strategies and interventions that may postpone symptom onset or impede disease development. Furthermore, analysis of these indicators can augment early detection initiatives, facilitating prompt treatment and enhancing the quality of life of individuals impacted by AD. A thorough comprehension of the factors of Alzheimer’s disease can guide public health policy and enhance responses to this escalating global health concern [19,20]. The authors of [21] proposed a new ML-driven predictive approach for AD detection that uses XAI for more interpretability. To enhance model robustness, class-balancing was handled using different techniques, like the Synthetic Minority Over-sampling Technique (SMOTE), SMOTE combined with Edited Nearest Neighbors (SMOTEENN), and Adaptive Synthetic Sampling (ADASYN). A feature selection technique was performed using chi-square and Recursive Feature Elimination (RFE) to select the most relevant factors. The proposed ensemble model, combining LightGBM (LGBM) and Random Forest (RF) with chi-square feature selection and utilizing soft voting, achieved the highest test accuracy. Additionally, the authors used SHAP (SHapley Additive exPlanations) and LIME

(Local Interpretable Model-agnostic Explanations) to interpret the model's decision-making process.

Gaps in Existing Methods and Motivation of This Study

Despite the advances made in the identification of AD using ML and DL methods, there are still several key issues that have not been addressed in previous studies.

First, current methods, such as the use of 3D CNN-based methods and transfer learning models [11,12], heavily depend on fully supervised learning and require large-scale labeled data. Like previous methods, classical machine learning techniques with handcrafted volumetric and clinical features and classifiers like random forest or k-nearest neighbors [13] require complete annotations, which are, in medical domains, expensive and difficult to acquire.

Second, these methods demonstrate good predictive performance but require high-dimensional neuroimaging data and may not make the most of simple, easily accessible clinical, demographic, and lifestyle factors. This restricts their use in cost-sensitive, resource-limited healthcare systems.

Third, existing studies focus mostly on the handling of class imbalance with synthetic oversampling techniques like SMOTE, SMOTEENN, and ADASYN [21]. However, these techniques can create artificial distributions of data points, add additional noise, and cause overfitting, especially for small datasets or datasets that are non-homogeneous, such as medical datasets. This issue has raised concerns of model robustness and generalization in real-world clinical settings.

Fourth, interpretability is still the major challenge. Several approaches have been proposed for Explainable AI (XAI), like SHAP and LIME; however, these methods are normally used as post hoc explanation tools instead of being part of the learning process. Therefore, most models continue to remain black boxes, restricting clinical trust and understanding.

Fifth, new research involving multimodal risk factors has emphasized a complex interplay between lifestyle, genetic, environmental, and healthcare risk factors in determining the course of AD [19,20]. However, most currently available computational models systematically incorporate these different kinds of risk factors in a single predictive model and are, thus, unable to account for the full spectrum of disease progression.

Lastly, there is no comprehensive framework to combine the above four concepts, as well as optimized gradient-boosting models. Existing hybrid ML-XAI methods like those described in [21] are fully supervised and use external imbalance correction methods instead of embedded learning methods.

Merging the five complementary components into a single pipeline, the proposed Lightweight Semi-Supervised SHAP-enhanced CatBoost (LSSSC) framework differs from state-of-the-art AD prediction models that use ML in a fully supervised fashion as follows:

- It uses Optuna hyperparameter optimization for CatBoost;
- It embeds feature selection using SHAP importance prior to using SS learning;
- It adopts pseudo-labeling based on confidence when utilizing unlabeled clinical data;
- It adopts cost-sensitive learning based on probability calibration and F1-optimal threshold selection rather than synthetic oversampling;
- It offers explainability complementary to that of global and local SHAP and LIME.

Such coordinated integration enables efficient use of unlabeled data and boosts interpretability, computational efficiency, and robustness, setting the proposed framework apart from the currently available supervised XAI pipelines for Alzheimer's disease detection.

This paper is structured into four primary sections. The proposed methodology is delineated in Section 2. The outcomes of this method are depicted and analyzed in Section 3. This study concludes in Section 4.

2. Proposed Approach

This section presents the proposed approach for AD prediction, called Lightweight Semi-Supervised SHAP-enhanced CatBoost (LSSSC). The main parts of the proposed approach are illustrated in Figure 1 and Algorithm 1. The LSSSC approach includes five main phases.

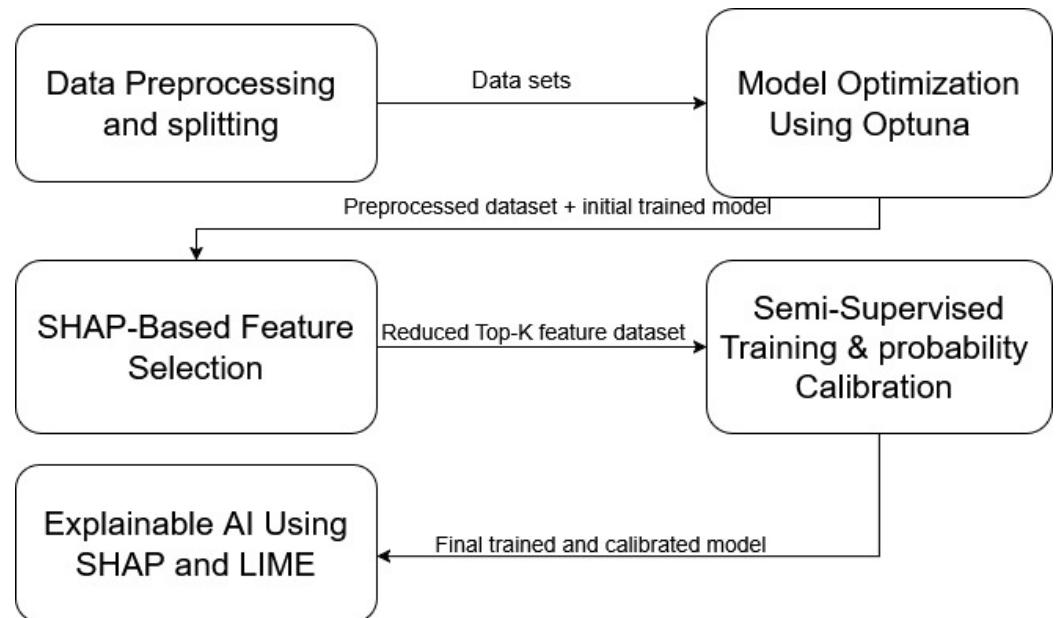


Figure 1. Flowchart of the proposed LSSSC framework for AD prediction.

Algorithm 1 Lightweight Semi-Supervised ShAP-enhanced CatBoost (LSSSC) for AD detection with Cost-Sensitive Thresholding

Part I: Data Preprocessing and Splitting

- Load and clean dataset
- Separate features X and target y (Diagnosis)
- Split data into training labeled, validation labeled, training unlabeled and testing sets
- Handle class imbalanced issue from training labeled set.

Part II: Model Optimization

- Optimize CatBoost Model using Optuna and ROC-AUC values and retrieve best hyperparameters
- Train and evaluate CatBoost model using optimized parameters

Part III: SHAP-Based Feature Selection

- Compute SHAP values on validation data
- Select top- K most influential features based on mean absolute SHAP values.
- Reduce training, validation, test, and unlabeled datasets accordingly

Part IV: Semi-Supervised Retraining and Probability Calibration

- Train CatBoost using top- K features
- Predict class probabilities for unlabeled samples
- Apply confidence-based pseudo-labeling
- Assign sample weights and augment training data
- Retrain model on augmented dataset
- Calibrate probabilities
- Select F1-optimal decision threshold
- Evaluate final model on test set

Part V: Explainable AI (XAI) using SHAP and LIME

- Compute SHAP values for the final model
- Analyze global feature importance
- Generate local explanations using SHAP
- Apply LIME for complementary local interpretability

2.1. Data Preprocessing and Splitting

The dataset used in this study was collected from an AD dataset publicly available on Kaggle [22]. It includes records from 2149 patients with 34 features, encompassing a wide range of health, social, lifestyle, and demographic attributes, with a particular focus on factors associated with cognitive health and AD diagnosis:

- Identification variables: a unique patient identifier and the physician in charge.
- Demographic attributes: age, gender, ethnicity, and education level. These factors are known for their influence on cognitive reserve and disease risk.
- Lifestyle-related: body mass index (BMI), smoking and alcohol consumption, physical activity, diet, and sleep quality. These factors reflect modifiable risk factors.
- Medical history: family history of AD, chronic conditions, blood pressure, and a cholesterol profile. These variables are important to both genetic predisposition and vascular contributions to cognitive decline.
- Cognitive and functional status: MMSE, functional assessments, and activities of daily living (ADL). These measures represent symptom-related features, including memory complaints, behavioral changes, and indicators of cognitive decline.
- Diagnosis variable: target outcome. This variable distinguishes between normal cognition and AD status.

Let D be the patients' collected data:

$$D = (\mathbf{x}_i, y_i)_{i=1}^N, \quad (1)$$

where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the feature vector, including lifestyle, demographic, and other factors, of the i th patient, and $y_i \in \{0, 1\}$ represents the corresponding diagnostic label (AD status).

In the preprocessing stage, non-informative and identifier-based attributes such as PatientID were excluded to avoid data leakage and to keep the model centered on clinically relevant attributes. The processed data were subsequently decomposed into a feature matrix ($\mathbf{X}$) and a target vector ($\mathbf{y}$). Categorical features of the $\mathbf{X}$ were selected and indexed:

$$\mathcal{C} = \{j \mid X_{:,j} \text{ is categorical}\} \quad (2)$$

where $X_{:,j}$ denotes the j -th column of $\mathbf{X}$. This indexing indicates which features represent discrete categorical variables, allowing CatBoost to handle categorical features using target-based statistics, thereby reducing the need for separate encoding techniques such as one-hot or label encoding [23,24].

After that, a stratified split on the dataset was applied to ensure unbiased evaluation, leading to two subsets: a training set (80%) and a testing set (20%). This split will preserve the class distribution. After that, the training set was further divided into labeled (50%) and unlabeled (50%) subsets to allow for semi-supervised learning. In addition, the labeled subset was split into training (75%) and validation (25%) sets. These two subsets were used later for model training and hyperparameter optimization. To handle the class-imbalance issue, class weights were computed from the labeled training set based on a balanced weighting strategy [25]:

$$w_c = \frac{|\mathcal{D}_{\text{train}}^{(l)}|}{|\mathbf{y}| \cdot |\{i \mid y_i = c\}|}, \quad \forall c \in \mathbf{y} \quad (3)$$

where $\mathcal{D}_{\text{train}}^{(l)}$ denotes the labeled training dataset, $|\mathbf{y}|$ is the total number of classes, and $|\{i \mid y_i = c\}|$ corresponds to the number of training samples belonging to class c . Based on this equation, higher weights are given to the minority classes and lower weights to majority classes, thereby reducing the effects of class imbalance during model training.

2.2. Model Optimization

To optimize the predictive performance of the CatBoost classifier for AD prediction, a systematic hyperparameter-tuning strategy was employed using Optuna. It is a Bayesian optimization framework based on sequential model-based optimization [26]. The process aims to find the optimal model configuration, which produces the highest discrimination power while keeping model complexity at suitable levels and avoiding overfitting. The optimization process concentrated on the top CatBoost hyperparameters, which control model capacity and learning dynamics and regularization strength. The optimized hyperparameters, along with their corresponding search ranges, appear in Table 1.

Table 1. CatBoost hyperparameters optimized using Optuna.

Parameter	Description
iterations	Boosting trees
depth	Tree depth
learning_rate	Learning rate
l2_leaf_reg	L2 regularization
loss_function	Loss function
eval_metric	Evaluation metric
class_weights	Imbalance handling
early_stopping_rounds	Early stopping

The goal of the optimization process is to select the hyperparameter configuration (θ^*) that provides the maximum Receiver Operating Characteristic Area Under the Curve (AUC) on the validation set [27]:

$$\theta^* = \arg \max_{\theta \in \Theta} \text{AUC}(y_{\text{val}}, \hat{p}_{\theta}(x_{\text{val}})) \quad (4)$$

where x_{val} and y_{val} represent the feature matrix and the corresponding true labels of the validation set, respectively. θ illustrates the CatBoost hyperparameter vector, and AUC is given by Equation (17). $\hat{p}_{\theta}(x)$ represents the predicted posterior probability produced by the CatBoost model parameterized by θ . It is given by

$$\hat{p}_{\theta}(x) = P(y = 1 \mid x; \theta). \quad (5)$$

The optimized parameters obtained from the process were used to train the first supervised CatBoost model, which served as the starting point for SHAP-based feature selection and semi-supervised learning and explainability analysis.

2.3. SHAP-Based Feature Selection

This paper employs SHapley Additive exPlanations (SHAP) values to develop a robust feature selection method aimed at identifying the primary influencing factors to mitigate model complexity. The SHAP technique integrates various methodologies from cooperative game theory to establish a unified framework that enables users to comprehend model predictions via both global and feature-specific analyses [28]. Let $X_{\text{val}} \in \mathbb{R}^{N \times d}$ represent the validation feature matrix with d features. The global importance score was calculated for each feature as the mean absolute SHAP value across all validation samples after applying the optimized CatBoost model [29]:

$$I_j = \frac{1}{N} \sum_{i=1}^N |\phi_{i,j}|, \quad (6)$$

where $\phi_{i,j}$ denotes the SHAP value of feature j for sample i . After that, features were ordered in descending order according to their importance scores (I_j). Thus, the top- K most influential features were selected and retained for subsequent modeling stages. The training, validation, test, and unlabeled datasets were reduced accordingly to include only the selected features. The system updated its categorical feature indices to follow the native categorical variable handling of CatBoost. The SHAP-driven feature selection method enhances model interpretability through dimension reduction, which leads to improved computational performance without affecting predictive accuracy. The selected features constitute the basis for the subsequent semi-supervised learning and explainability analysis.

2.4. Semi-Supervised Retraining and Probability Calibration

Let $\mathcal{D}^{(u)} = \{x_i\}_{i=1}^M$ denote the unlabeled dataset. The retrained model was employed to compute an estimation of the posterior class probabilities ($\hat{p}_\theta(y | x_i)$) for each unlabeled sample. Pseudo-labels were assigned using class-specific confidence thresholds to minimize noise propagation [30]:

$$\tilde{y}_i = \arg \max_{c \in \{0,1\}} \hat{p}_\theta(y = c | x_i), \quad (7)$$

$$x_i \in \mathcal{D}^{(u)*} \iff \hat{p}_\theta(y = \tilde{y}_i | x_i) \geq \tau_{\tilde{y}_i}, \quad (8)$$

where $\tau_1 = 0.85$ and $\tau_0 = 0.75$ are confidence thresholds for the positive ($c = 1$) and negative ($c = 0$) classes, respectively. These asymmetric thresholds alleviate class imbalance and reduce the risk of erroneous pseudo-labeling. Based on prediction confidence, each pseudo-labeled sample was assigned a weight using

$$w_i = \hat{p}_\theta(y = \tilde{y}_i | x_i). \quad (9)$$

The augmented training set was then built based on

$$\mathcal{D}_{\text{aug}} = \mathcal{D}_{\text{train}} \cup \{(x_i, \tilde{y}_i, w_i)\}_{x_i \in \mathcal{D}^{(u)*}}. \quad (10)$$

The CatBoost model was retrained on $\mathcal{D}_{\text{aug}}$, incorporating weighted samples and employing early stopping to optimize validation performance. The reliability of predicted probabilities improved through the application of isotonic regression as a post hoc calibration method. Given validation predictions ($\hat{p}_\theta(x)$), a calibrated probability function ($g(\cdot)$) was learned such that

$$\hat{p}_{\text{cal}}(x) = g(\hat{p}_\theta(x)), \quad (11)$$

where $g(\cdot)$ is a monotonic, non-parametric mapping optimized on the validation set. The model used a cost-sensitive decision threshold, which optimized the F1 score instead of using a fixed classification threshold. Let $P(t)$ and $R(t)$ denote precision and recall at threshold t . The optimal threshold (t^*) was determined as follows [31]:

$$t^* = \arg \max_{t \in [0,1]} \frac{2 P(t) R(t)}{P(t) + R(t)}. \quad (12)$$

Finally, the calibrated model and optimal threshold were applied to the test set to compute the AUC and Macro-F1 scores given in Equations (16) and (17).

Model Evaluation

The performance of the proposed model is evaluated based on accuracy, precision, recall, F1 score, and AUC. Let TP , TN , FP , and FN represent true positives, true negatives, false positives, and false negatives, respectively; the performance metrics are given by

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (14)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (15)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

The ROC curve represents the trade-off between the true-positive rate (TPR) and the false-positive rate (FPR). It is defined as

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}) \quad (17)$$

where

$$\text{TPR} = \frac{TP}{TP + FN} \quad (18)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (19)$$

2.5. Explainable AI (XAI) Using SHAP and LIME

SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) techniques [32] were used in this study to guarantee the transparency and interpretability of the predictions generated by the proposed model. SHAP may overstate feature significance in strongly correlated datasets. However, LIME explanations might vary considerably based on perturbation sampling methods and the quality of local approximations. Thus, the findings must be interpreted with care and supported by clinical expertise. SHAP waterfall plots offer a local explanation of single predictions by showing how each feature contributes to the deviation of the model output from a baseline value. Starting with the expected model prediction, features are sequentially cumulative according to their SHAP values, illustrating positive and negative contributions that increase or decrease the predicted outcome. This visualization points out the most important features for a specific instance and enables transparent interpretation of the decision-making process of complex models. On the other side, LIME describes individual predictions by capturing the behavior of a complex model in the local neighborhood of a given instance using a simple, interpreted surrogate model. By slightly altering the input features and examining how the model's prediction responds, LIME reveals which features have the greatest impact on a given decision. This approach offers clear, instance-level explanations and remains independent of the underlying model architecture. Using these two techniques will facilitate a thorough evaluation of feature significance and impact, both globally and locally, and increase trust and therapeutic applicability [33].

3. Results and Discussion

This section presents and discusses the results of applying the LSSSC approach on the dataset described in Section 2.1.

3.1. Performance and Complexity Evaluation

Table 2 shows a comparison of the performance of the proposed approach and that of four different ML models using the same pipeline illustrated in Figure 1 for AD prediction. It also includes the number of most significant features selected by each ML model, training time, peak memory, and floating-point operations (FLOPs). The evaluation was performed based on AUC and macro-F1 to guarantee both discriminative power and balanced class-wise performance, alongside computational efficiency metrics. Among all ML models, LSSSC achieves the best overall performance, reaching an AUC of 0.9513 and a macro-F1 score of 0.9490 with only nine selected features. These results point out the ability of the LSSSC method to capture complex nonlinear relationships in clinical data while conserving robustness and interpretability using SHAP-based feature ranking and probability calibration. In addition, the RF model shows excellent predictive capability, with an AUC value of 0.9401 and a macro-F1 value of 0.9437. However, RF performance was achieved at the expense of significantly higher computational cost, as illustrated by the longest training time and the high number of FLOPs. On the other hand, the decision tree and KNN models necessitate fewer computational resources but lead to lower macro-F1 scores, demonstrating weaker and less balanced classification performance. Considering Green AI principles, LSSSC excelled by offering high predictive accuracy without requiring excessive computational resources, carrying the practicality and sustainability of the proposed clinical decision-support framework.

As illustrated in Figure 2, the LSSSC approach achieves significant label efficiency. A sharp performance rise occurs when the percentage of labeled training data increases from 5% to 20%, with the AUC achieving a near-plateau value of 0.938. In addition, the model reaches around 98% of its peak performance based on only 20% of the available labels.

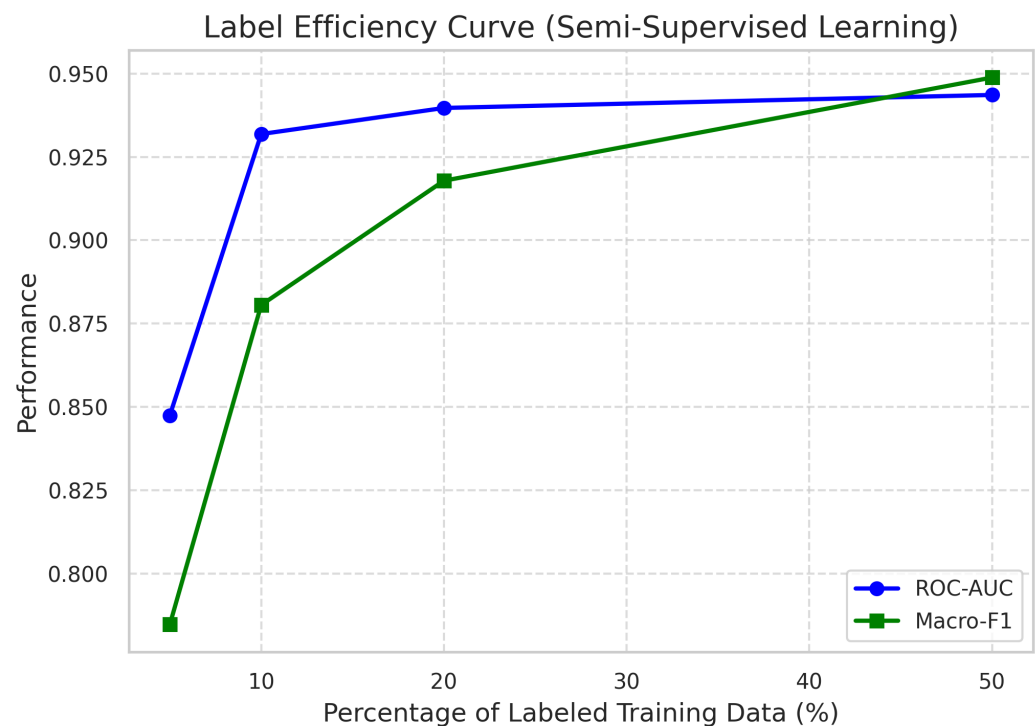


Figure 2. Label efficiency analysis of the semi-supervised model.

Table 3 illustrates the ablation results, showing that each component plays a different role in model performance. The baseline CatBoost model achieves satisfactory initial performance. Neither SHAP-based feature selection nor pseudo-labeling independently

improved the baseline under the current experimental setting. However, their integration with calibration and threshold optimization produced the strongest result, suggesting a complementary interaction that requires confirmation through repeated experiments. Thus, the full proposed model outperforms the rest across all metrics: precision = 0.9504, recall = 0.9476, accuracy = 0.9535, and F1 score = 0.9490, which shows that feature selection and semi-supervised learning are more robust and generalizing.

Additionally, the LSSSC model performs better based on the AUC value when few labeled data are available. However, with abundant labels, the macro-F1 score increases significantly and eventually surpasses ROC-AUC once half of the data is labeled. This figure points out the effectiveness of the semi-supervised approach, especially in conditions where getting high-quality labeled data is costly and time-consuming, such as in the detection of AD.

Table 2. Comparison between LSSSC and ML classification models for AD prediction.

Model	k	ROC-AUC	Macro-F1	Train Time (s)	Peak Memory (MB)	Energy/FLOPs
Random Forest	6	0.9401	0.9437	3.854	0.335	104,355,840 FLOPs
Decision Tree	11	0.9443	0.9132	0.063	0.096	13,760 FLOPs
KNN	5	0.9230	0.8622	0.008	0.066	21,500 FLOPs
LSSSC	9	0.9513	0.9490	1.097	0.069	13,374,720 FLOPs

Table 3. Ablation study results for different model configurations.

Model	Precision	Recall	Accuracy	F1 Score
Baseline CatBoost	0.94	0.94	0.94	0.9396
SHAP Feature Selection Only	0.93	0.94	0.94	0.9344
Semi-Supervised Only	0.93	0.93	0.93	0.9296
Full Proposed Model	0.9504	0.9476	0.9535	0.9490

Table 4 shows the classification report of the LSSSC model, illustrating strong predictive performance for AD detection. The proposed model achieves excellent precision, recall, and F1 scores for both classes. For the negative class (no AD), LSSSC achieves a precision of 0.9607, a recall of 0.9676, and an F1 score of 0.9642. However, for the positive class (AD), it reaches a precision of 0.9400, a recall of 0.9276, and an F1 score of 0.9338. These results demonstrate that the proposed model is able to correctly detect and identify two classes with a small misclassification error. These results show its effectiveness in clinical decision-making. The overall accuracy of the model is equal to 95.35%, confirming that the model works reliably on the dataset. Furthermore, the macro-F1 score is 0.9490, and the weighted F1 score is 0.9534, showing the model's balanced performance.

Table 4. Classification report of the LSSSC model.

Class	Precision	Recall	F1 Score
No AD	0.9607	0.9676	0.9642
AD	0.9400	0.9276	0.9338
Accuracy		0.9535	
Macro avg	0.9504	0.9476	0.9490
Weighted avg	0.9534	0.9535	0.9534

The ROC curve (Figure 3) represents the trade-off between the true-positive rate (sensitivity) and the false-positive rate (specificity). The LSSSC model achieves an AUC of 0.9513, showing near-optimal ranking performance and robustness to class imbalance.

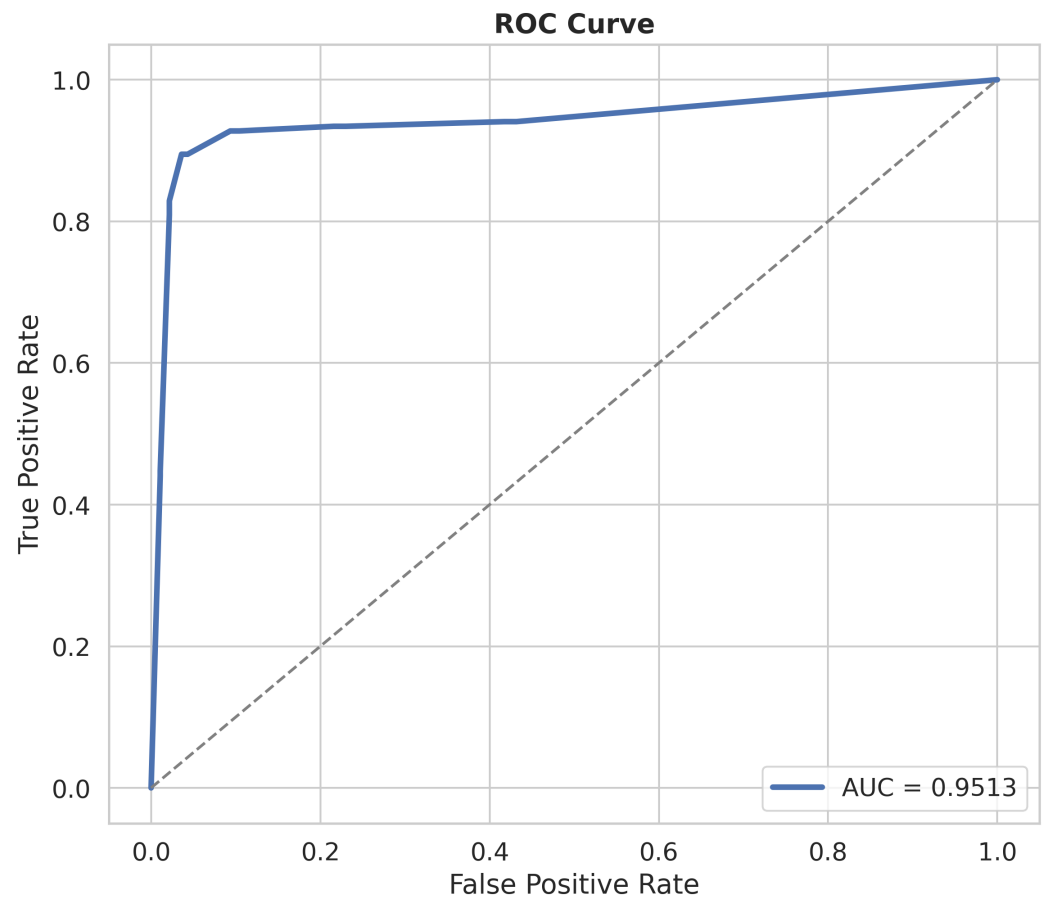


Figure 3. ROC curve for the LSSSC model.

3.2. SHAP Interpretability

The global interpretability analysis using SHAP is illustrated in Figure 4, revealing that functional and cognitive assessments are the most influential factors in the model's decision-making process. The functional assessment appeared as the most important factor, followed by Activities of Daily Living (ADL), memory complaints, behavior problems, and the MMSE (Mini-Mental State Examination) score. However, traditional physiological risk factors like age, educational level, and cholesterol (LDL) show lower impacts, suggesting that the proposed model focuses more on symptoms and cognitive decline than on demographic or biometric factors when detecting AD.

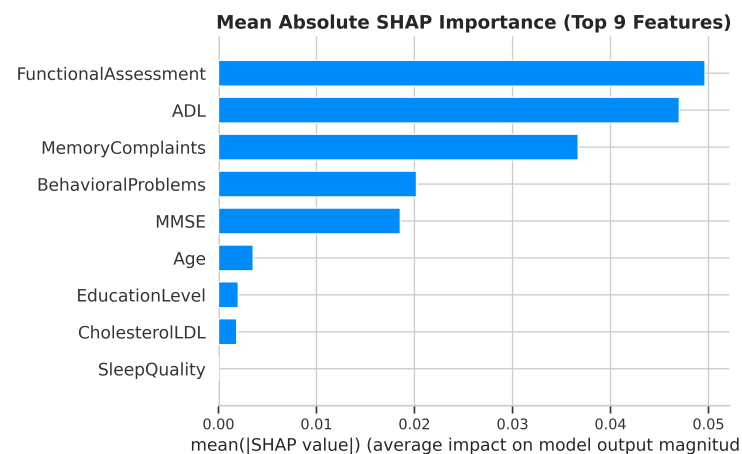


Figure 4. SHAP global feature importance.

The global interpretability of the proposed LSSSC model is represented in Figure 5, confirming the results provided by Figure 4. This figure confirms that Activities of Daily Living (ADL), functional assessment, and memory complaints are the most influential factors for AD detection. The lower scores in ADL and functional assessment, illustrated by blue points, are associated with higher SHAP values, revealing a high possibility of an AD diagnosis. Moreover, high values for memory complaints and behavioral problems, represented by red points, indicate that the model's output supports a positive classification. This figure also confirms that clinical features like BMI and LDL cholesterol are assigned values close to zero SHAP, showing that these factors have a limited role in the model's predictions compared to the high impact of cognitive and functional features.

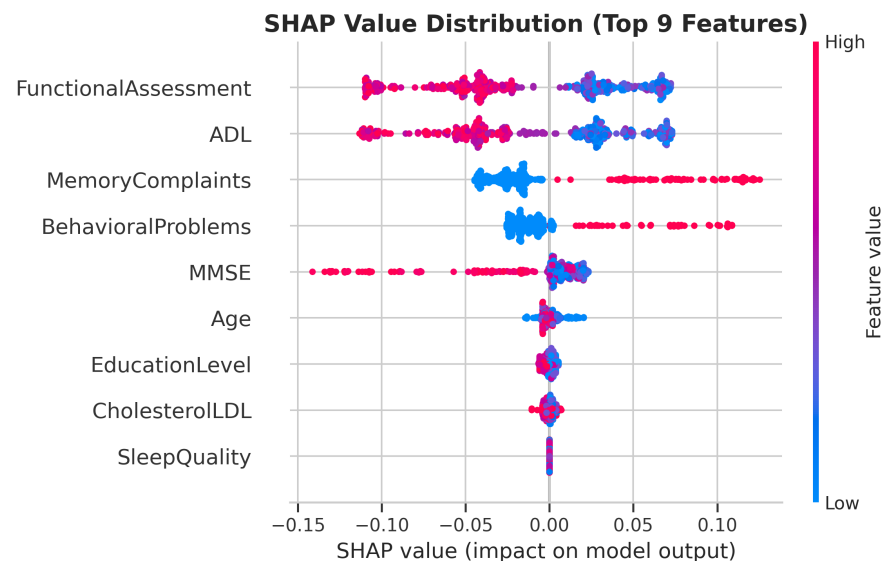


Figure 5. SHAP summary plot for the LSSSC model, illustrating the global feature importance and the directional impact of clinical variables on AD detection.

3.3. LIME Interpretability

The LIME results illustrated in Figure 6 provide a local explanation for a single prediction, where the model output is $f(x) = -1.969$. In this specific case, the patient's functional assessment score of 9.768 emerges as the primary factor reducing the likelihood of an AD diagnosis, contributing a strong negative weight of -1.12 . Additionally, the absence of memory complaints (value = 0) further decreases the probability of a positive AD classification, with a contribution of -0.52 . Conversely, features such as an ADL score of 4.359 and an MMSE score of 19.922 contribute positively to the prediction, with weights of $+0.32$ and $+0.17$, respectively, indicating clinical signals that push the prediction toward the base value. Overall, this figure demonstrates that the LSSSC model enhances interpretability by enabling clinicians to access critical risk assessment information through its component-based framework, clearly illustrating how individual clinical indicators influence patient-level risk evaluations.

The local interpretability of the LSSSC model is further described in the LIME bar chart illustrated in Figure 7. This figure elucidates the justification for an individual prediction in a form that supports clinical decision-making. This plot decomposes the model output into the impact of individual features evaluated at specific threshold values, with each bar indicating the associated importance of a feature and the color showing whether it supports (green) or contradicts (red) the predicted outcome. This figure also supports the results illustrated in Figures 5 and 6. Indeed, clinical factors like $ADL \leq 2.31$ and $FunctionalAssessment \leq 2.43$ show the high positive influence, offering clinicians a

strong indication that functional impairment is a primary driver of the model's decision in this case. However, the MemoryComplaints ≤ 0.00 factor emerges as a main negative contributor, pointing out that the absence of self-reported memory complaints lowers the model's estimated possibility of an AD diagnosis. By explicitly linking model outputs to clinically interpretable decision rules, LIME allows for transparent, patient-level reasoning and permits clinicians to critically evaluate the possibility of automated predictions. Thus, it strengthens trust and supports informed clinical decision-making.

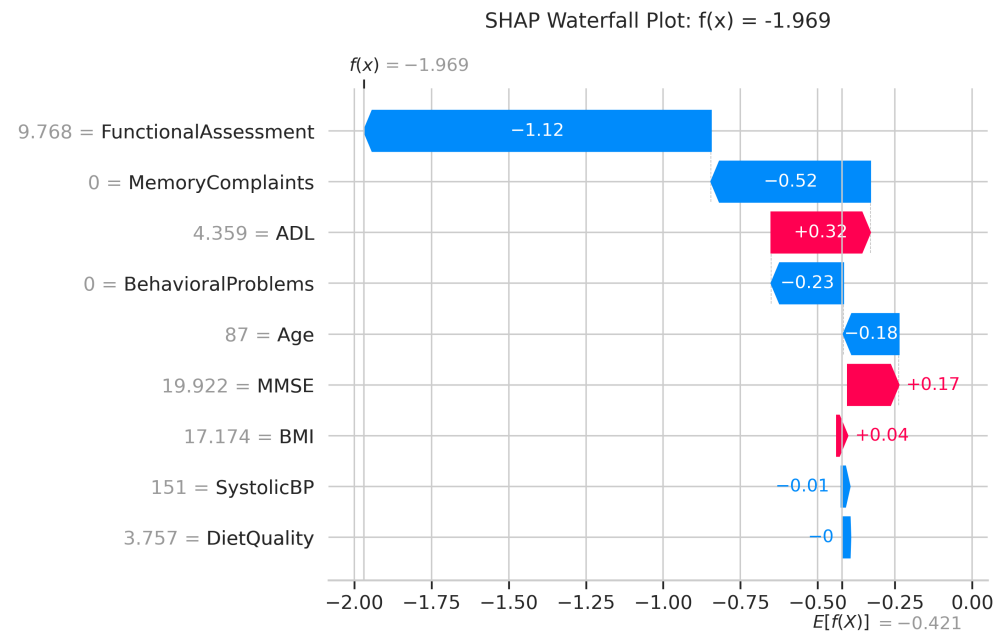


Figure 6. LIME waterfall plot offering a local explanation for a specific patient instance. The plot decomposes the model's prediction ($f(x) = -1.969$) from the base expected value ($E[f(X)] = -0.421$), highlighting individual feature contributions.

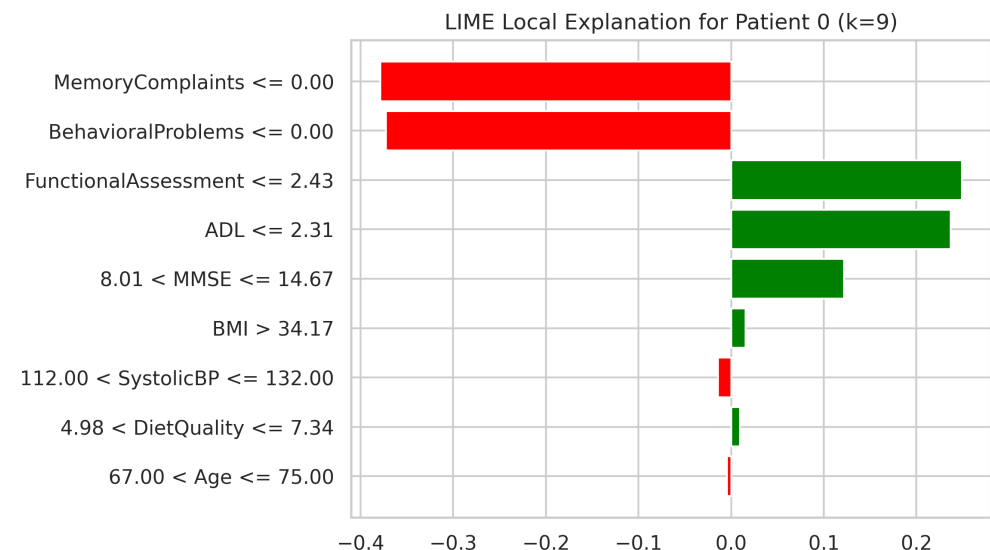


Figure 7. LIME local explanation bar chart for a selected test patient. Green bars represent feature ranges that support the model's prediction, while red bars represent ranges that contradict it.

3.4. Comparison with Existing Systems

Table 5 compares the proposed LSSSC model with several existing supervised approaches, including KNN, support vector machine (SVM), logistic regression (LR), XGBoost, a stacked ensemble (base model: KNN, SVM, RF, and LR; meta model: LR), and gradient

boosting with chi-square feature selection. The ensemble and gradient-boosting methods achieve high accuracy results, but the recall performance of gradient boost remains low, which might limit its ability to identify clinical cases. The LSSSC system achieves a slightly higher level of performance for accuracy (95.35%), precision (95.04%), recall (94.76%), and F1 score (94.90%) through its semi-supervised method. The LSSSC approach operates with both labeled and unlabeled data to minimize its requirement for fully labeled datasets, while its component-based output generates results that clinicians can understand for better decision-making.

Table 5. Classification performance comparison between existing methods and the proposed semi-supervised LSSSC model.

Method	Accuracy	Precision	Recall	F1 Score
KNN [34]	0.75	0.76	0.75	0.72
SVM [34]	0.85	0.85	0.85	0.85
LR [34]	0.84	0.84	0.84	0.84
XGBoost [34]	0.94	0.94	0.94	0.94
Stacked Ensemble [34]	0.94	0.94	0.94	0.94
Gradient Boosting + Chi-Square FS [21]	0.940	0.940	0.840	0.90
LSSSC (proposed)	0.9535	0.9504	0.9476	0.9490

3.5. Green AI Efficiency

The relationship between the predictive F1 score and sustainable computing operations is illustrated in Figure 8. This figure shows how the macro-F1 score and Green AI efficiency vary based on the number of SHAP-selected features (k). After $k = 5$, the macro-F1 scores remain high, while green AI efficiency shows significant performance changes. Thus, the results prove that the size of the feature dimension produces an unpredictable effect on system resource usage. The model reaches its peak Green AI efficiency of 76% for $k = 9$ with a macro-F1 value around 0.95. With $k > 9$, model green AI efficiency starts to decrease, reaching its minimal value of 50%. This chart demonstrates that SHAP-based feature selection adopted in the LSSSC, especially the model, offers valuable results, especially for $k = 9$, that provide a high-performing “green” model for AD detection.

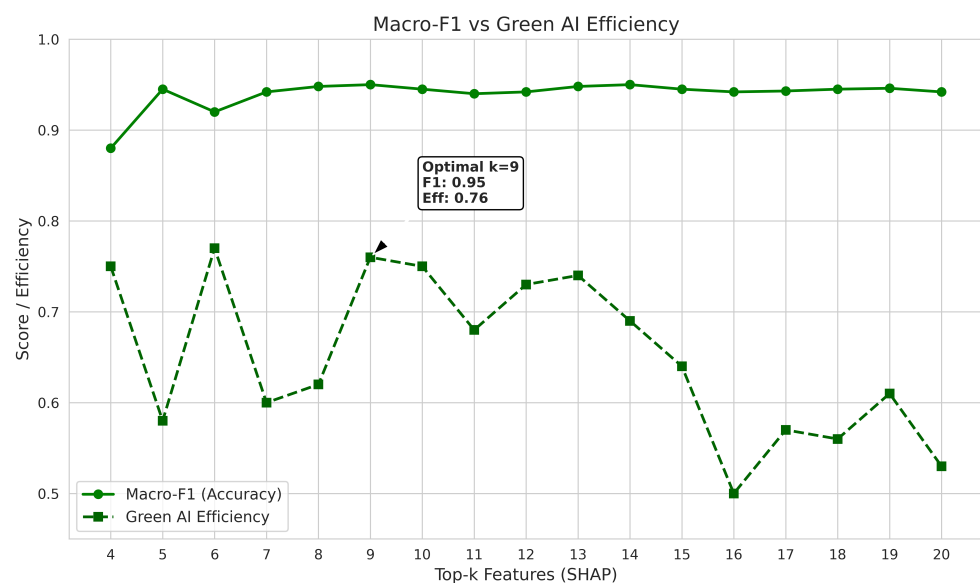


Figure 8. Trade-off analysis between macro-F1 score and green AI efficiency across varying top- k SHAP-selected features.

3.6. Clinical Translation and Future Applicability

This projected LSSSC framework has strong predictive capability that can potentially be utilized as an effective decision-support tool for the diagnosis of early-stage AD. Furthermore, it provides inherent interpretability through SHAP and LIME explanation methodologies; hence, this framework is likely to have great potential as a tool for clinicians to use when making decisions on whether or not to diagnose an individual with early-stage dementia caused by an underlying medical condition. However, there are several critical considerations regarding the integration of the LSSSC model into routine clinical practice. Even though the model has produced promising results on publicly available data, further external validation using multi-site and real-world health-system data will be critical to establish the model as robust, generalizable, and inclusive of a diverse population. This validation constitutes a main step towards acceptance by the clinical community. Furthermore, the model should align correctly with existing diagnostic standards, especially the AT(N) framework, ensuring that it uses the most recently validated versions of biomarkers for AD, including plasma p-tau and various neuroimaging modalities, among others. Additionally, a series of clinical trials is required to thoroughly evaluate the model's performance in actual diagnostic settings and to survey any potential influence the model may have on the clinician's decision-making process.

Lastly, the use of explainable AI methodologies, in conjunction with the proposed model, should enhance the level of transparency of the model, which is critical for clinician trust and acceptance by regulatory authorities. Future work will involve multimodal data integration efforts, external validation of the model, and optimization efforts towards models designed for deployment. These will be the foundations for successful, safe, and effective implementation of these tools within actual health systems.

3.7. Limitations of the Study

Although the proposed LSSSC scheme is found to have been successfully implemented, certain shortcomings must be pointed out. First, it was developed and tested on an open-source dataset from Kaggle, which may contain natural selection bias and may not be representative of real-world clinical populations. These datasets might not provide a comprehensive representation of demographic diversity, multi-center variability, or variations in clinical practice that could impact the model's generalizability when applied in actual healthcare settings.

Second, the dataset contains a large number of demographic, lifestyle, and cognitive features but does not include the biomarkers for Alzheimer's disease that have recently been established, including plasma p-tau₂₁₇, p-tau₂₃₁, GFAP, amyloid PET, tau PET, and the AT(N) biological framework. This lack of clinically validated biomarkers could restrict the applicability of the proposed model to state-of-the-art diagnostics.

Lastly, though semi-supervised learning and class weighting make the algorithm more robust in the presence of a small amount of labeled data, an external evaluation with multi-center clinical data would be needed to assess the robustness and clinical utility of the proposed algorithm. These limitations will be overcome in future work by including multimodal data and biomarker data to improve model generalization and clinical relevance.

4. Conclusions

This paper introduces a novel method for AD prediction that combines cost-sensitive learning, SHAP-based embedded feature selection, and confidence-based pseudo-labeling to leverage both labeled and unlabeled data. The proposed method also incorporates SHAP and LIME explanations to guarantee transparent and clinically interpretable predictions.

The method is based on a semi-supervised learning technique that reduces dependence on large amounts of labeled data. To deal with the class-imbalance challenge, a class-weighting method that conserves data integrity by giving higher penalties to minority classes is adopted. The proposed approach further provides robust and clinically relevant explanations by detecting the most influential features through SHAP-based importance scores embedded within the CatBoost model. Therefore, it guarantees stable and interpretable predictions. The LSSSC framework exhibits strong predictive performance by reaching an accuracy of 95.35%, a macro-F1 score of 0.9490, and an AUC of 0.9513. The proposed method retains a reasonable training time, memory consumption, and number of floating-point operations (FLOPs). By integrating global SHAP analyses with local LIME explanations, the proposed approach allows healthcare professionals to interpret model outputs at both population and individual patient levels. Overall, the proposed LSSSC framework provides a scalable and computationally efficient approach for early Alzheimer's disease detection by effectively leveraging semi-supervised learning while maintaining transparency and clinical interpretability.

However, in the current study, no neurologist, neuroscientist, or clinical AD specialist was formally involved. Clinical variables and model interpretations were therefore evaluated against the published literature rather than through direct expert review. Future work will include interdisciplinary collaboration for feature validation, interpretation, external evaluation, and prospective clinical assessment. Additionally, future research will focus on extending this framework to incorporate longitudinal and multimodal data, with the aim of enabling more personalized risk assessment and supporting robust, real-world clinical decision-making.

Author Contributions: Conceptualization, G.N. and A.E.C.; methodology, G.N.; software, G.N.; validation, A.E.C. and B.N.; formal analysis, A.E.C.; investigation, G.N.; resources, A.E.C.; data curation, B.N.; writing—original draft preparation, A.E.C. and G.N.; writing—review and editing, B.N.; visualization, G.N.; supervision, A.E.C.; project administration, B.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are openly available in the public domain. These data were derived from the following resources available online: <https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset>, accessed on 10 July 2026.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine learning
AD	Alzheimer's disease
LSSSC	Lightweight semi-supervised SHAP-enhanced CatBoost
XAI	Explainable AI
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
AI	Artificial intelligence
CNN	Convolutional neural network
SMOTE	Synthetic Minority Over-sampling Technique
ADASYN	Adaptive Synthetic Sampling
AUC	Area under the curve

TP	True positive
TN	True negative
FP	False positive
FN	False negative
TPR	True-positive rate
LR	Logistic regression
KNN	K nearest neighbor
SVM	Support vector machine

References

1. Alzheimer's Association. Alzheimer's Disease Facts and Figures. Alzheimer's Disease and Dementia. 2025. Available online: <https://www.alz.org/alzheimers-dementia/facts-figures> (accessed on 10 July 2026).
2. World Health Organization. Dementia. 3 July 2026. Available online: <https://www.who.int/news-room/fact-sheets/detail/dementia> (accessed on 10 July 2026).
3. Shawkatova, I.; Javor, J. Alzheimer's Disease: Recent Developments in Pathogenesis, Diagnosis, and Therapy. *Life* **2025**, *15*, 549. <https://doi.org/10.3390/life15040549>.
4. Liss, J.L.; Seleri Assunção, S.; Cummings, J.; Atri, A.; Geldmacher, D.S.; Candela, S.F.; Devanand, D.P.; Fillit, H.M.; Susman, J.; Mintzer, J.; et al. Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: A review and synthesis. *J. Intern. Med.* **2021**, *290*, 310–334. <https://doi.org/10.1111/joim.13244>.
5. Kavitha, C.; Mani, V.; Srividhya, S.R.; Khalaf, O.I.; Tavera Romero, C.A. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. *Front. Public Health* **2022**, *10*, 853294. <https://doi.org/10.3389/fpubh.2022.853294>.
6. Uddin, K.M.M.; Alam, M.J.; Uddin, M.A.; Aryal, S. A Novel Approach Utilizing Machine Learning for the Early Diagnosis of Alzheimer's Disease. *Biomed. Mater. Devices* **2023**, *1*, 882–898. <https://doi.org/10.1007/s44174-023-00078-9>.
7. Zhao, Y.; Guo, Q.; Zhang, Y.; Zheng, J.; Yang, Y.; Du, X.; Feng, H.; Zhang, S. Application of Deep Learning for Prediction of Alzheimer's Disease in PET/MR Imaging. *Bioengineering* **2023**, *10*, 1120. <https://doi.org/10.3390/bioengineering10101120>.
8. Gustavsson, A.; Norton, N.; Fast, T.; Frölich, L.; Georges, J.; Holzapfel, D.; Kirabali, T.; Krolak-Salmon, P.; Rossini, P.M.; Ferretti, M.T.; et al. Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's Dement.* **2022**, *19*, 658–670. <https://doi.org/10.1002/alz.12694>.
9. Sharma, S.; Guleria, K.; Tiwari, S.; Kumar, S. A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. *Meas. Sens.* **2022**, *24*, 100506. <https://doi.org/10.1016/j.measen.2022.100506>.
10. El Abed, A.; Nassreddine, G.; Nakhal, B.; El Chakik, A. AI-powered strategies for early detection and prevention of severe diseases. In *Transformative Technologies*; CRC Press: Boca Raton, FL, USA, 2026; pp. 16–34.
11. Abbasian, P.; Hammond, T.A. A Comprehensive Analysis of Early Alzheimer Disease Detection from 3D sMRI Images Using Deep Learning Frameworks. *Information* **2024**, *15*, 746. <https://doi.org/10.3390/info15120746>.
12. Kumar, S.; Singh, N.P.; Brahma, B. AI-Based Model for Detection and Classification of Alzheimer Disease. In Proceedings of the 2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMi), Gwalior, India, 10–11 December 2023; pp. 1–6. <https://doi.org/10.1109/cvmi59935.2023.10464469>.
13. Biswas, R.; Gini, J. R. Multi-class classification of Alzheimer's disease detection from 3D MRI image using ML techniques and its performance analysis. *Multimed. Tools Appl.* **2023**, *83*, 33527–33554. <https://doi.org/10.1007/s11042-023-16519-y>.
14. Vimbi, V.; Shaffi, N.; Mahmud, M. Interpreting artificial intelligence models: A systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Inform.* **2024**, *11*, 10. <https://doi.org/10.1186/s40708-024-00222-1>.
15. Teunissen, C.E.; Verberk, I.M.; Thijssen, E.H.; Vermunt, L.; Hansson, O.; Zetterberg, H.; van der Flier, W.M.; Mielke, M.M.; Del Campo, M. Blood-based biomarkers for Alzheimer's disease: Towards clinical implementation. *Lancet Neurol.* **2022**, *21*, 66–77. [https://doi.org/10.1016/S1474-4422\(21\)00361-6](https://doi.org/10.1016/S1474-4422(21)00361-6).
16. Pelkmans, W.; Shekari, M.; Brugulat-Serrat, A.; Sánchez-Benavides, G.; Minguillón, C.; Fauria, K.; Molinuevo, J.L.; Grau-Rivera, O.; González Escalante, A.; Kollmorgen, G.; et al. Astrocyte biomarkers GFAP and YKL-40 mediate early Alzheimer's disease progression. *Alzheimer's Dement.* **2023**, *20*, 483–493. <https://doi.org/10.1002/alz.13450>.
17. Bao, W.; Xie, F.; Zuo, C.; Guan, Y.; Huang, Y.H. PET Neuroimaging of Alzheimer's Disease: Radiotracers and Their Utility in Clinical Research. *Front. Aging Neurosci.* **2021**, *13*, 624330. <https://doi.org/10.3389/fnagi.2021.624330>.
18. Kobayashi, G.; Hirata, K.; Ono, M.; Kasuga, K.; Takado, Y. The 2024 NIA-AA biological definition of Alzheimer's disease: Linking biomarkers to clinical practice. *Front. Dement.* **2026**, *5*, 1736297. <https://doi.org/10.3389/frdem.2026.1736297>.
19. Ayodele, T.; Rogaeva, E.; Kurup, J.T.; Beecham, G.; Reitz, C. Early-Onset Alzheimer's Disease: What Is Missing in Research? *Curr. Neurol. Neurosci. Rep.* **2021**, *21*, 4. <https://doi.org/10.1007/s11910-020-01090-y>.

20. Rani, P.; Lamba, R.; Sachdeva, R.K.; Kumar, K.; Iwendi, C. A machine learning model for Alzheimer's disease prediction. *IET-Cyber-Phys. Syst. Theory Appl.* **2024**, *9*, 125–134. <https://doi.org/10.1049/cps2.12090>.
21. Mahamud, E.; Assaduzzaman, M.; Islam, J.; Fahad, N.; Hossen, M.J.; Ramanathan, T.T. Enhancing Alzheimer's disease detection: An explainable machine learning approach with ensemble techniques. *Intell.-Based Med.* **2025**, *11*, 100240. <https://doi.org/10.1016/j.ibmed.2025.100240>.
22. El Kharoua, R. Alzheimer's Disease Dataset. Kaggle.com. 2024. Available online: <https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset> (accessed on 10 July 2026).
23. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for big data: An interdisciplinary review. *J. Big Data* **2020**, *7*, 94. <https://doi.org/10.1186/s40537-020-00369-8>.
24. Hamid, M.; Hajjej, F.; Alluhaidan, A.S.; bin Mannie, N.W. Fine tuned CatBoost machine learning approach for early detection of cardiovascular disease through predictive modeling. *Sci. Rep.* **2025**, *15*, 31199. <https://doi.org/10.1038/s41598-025-13790-x>.
25. Das, S.; Mullick, S.S.; Zelinka, I. On Supervised Class-Imbalanced Learning: An Updated Perspective and Some Key Challenges. *IEEE Trans. Artif. Intell.* **2022**, *3*, 973–993. <https://doi.org/10.1109/TAI.2022.3160658>.
26. Lai, L.-H.; Lin, Y.-L.; Liu, Y.-H.; Lai, J.-P.; Yang, W.-C.; Hou, H.-P.; Pai, P.-F. The Use of Machine Learning Models with Optuna in Disease Prediction. *Electronics* **2024**, *13*, 4775. <https://doi.org/10.3390/electronics13234775>.
27. Wang, L.; Zhang, C.; Wang, W.; Deng, T.; Ma, T.; Shuai, P. Slope Stability Assessment Using an Optuna-TPE-Optimized CatBoost Model. *Eng* **2025**, *6*, 185. <https://doi.org/10.3390/eng6080185>.
28. Makumbura, R.K.; Mampitiya, L.; Rathnayake, N.; Meddage, D.P.P.; Henna, S.; Dang, T.L.; Hoshino, Y.; Rathnayake, U. Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results Eng.* **2024**, *23*, 102831. <https://doi.org/10.1016/j.rineng.2024.102831>.
29. Hamilton, R.I.; Papadopoulos, P.N. Using SHAP Values and Machine Learning to Understand Trends in the Transient Stability Limit. *IEEE Trans. Power Syst.* **2023**, *39*, 1384–1397. <https://doi.org/10.1109/tpwrs.2023.3248941>.
30. Qu, A.; Wu, Q.; Yu, L.; Li, J.; Liu, J. Class-Specific Thresholding for Imbalanced Semi-Supervised Learning. *IEEE Signal Process. Lett.* **2024**, *31*, 2375–2379. <https://doi.org/10.1109/lsp.2024.3449852>.
31. Ding, H.; Yand, X.; Qi, T. Hybrid Machine Learning Models Based on CATBoost Classifier for Assessing Students' Academic Performance. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 94. <https://doi.org/10.14569/ijacsa.2024.0150709>.
32. Salih, A.M.; Raisi-Estabragh, Z.; Galazzo, I.B.; Radeva, P.; Petersen, S.E.; Lekadir, K.; Menegaz, G. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Adv. Intell. Syst.* **2024**, *7*, 2400304. <https://doi.org/10.1002/aisy.202400304>.
33. Donmez, T.B.; Kutlu, M.; Mansour, M.; Yildiz, M.Z. Explainable AI in action: A comparative analysis of hypertension risk factors using SHAP and LIME. *Neural Comput. Appl.* **2024**, *37*, 4053–4074. <https://doi.org/10.1007/s00521-024-10724-y>.
34. Soladoye, A.A.; Aderinto, N.; Osho, D.; Olawade, D.B. Explainable machine learning models for early Alzheimer's disease detection using multimodal clinical data. *Int. J. Med. Inform.* **2025**, *204*, 106093. <https://doi.org/10.1016/j.ijmedinf.2025.106093>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.