Applied Science, Engineering and Management Bulletin [ASEMB]
ISSN: 3049-3005
Vol. 3 No. 03(July-Sep) (2026) |Pg:96-109

Advancing Dementia Classification Through Intelligent Machine Learning and Deep Learning Approaches

Anjum Siddiqi[1], Dr. Jaydip Kumar[2], Dr. Shail Kumar Dinkar[3]
[1]Department of Computer Science and Application, COER University, Roorkee
Email: siddiqianjum786@gmail.com
[2]Department of Computer Science and Application, COER University, Roorkee
Email: jaydipkumar2001@gmail.com
[3]Department of Computer Science and Application, GBPIET, Pauri
Email: shailkdinkar@gbpiet.ac.in[3]

Abstract

Despite significant advances in understanding the structure of dementia through structural magnetic resonance imaging (MRI), classification of dementia from MRI data is still challenging due to the high dimensionality of features, the redundancy of information, the varied patterns of dementia, and the limited generalizability of models. This study presents an intelligent framework that combines the metaheuristic optimization, machine learning (ML), and deep learning (DL) techniques to achieve automatic dementia classification. A novel Adaptive Lévy–Opposition Search Optimization (ALOSO) algorithm is presented to improve on global exploration, local exploitation, feature selection, and model optimization. The framework assesses the performance of Support Vector Machine (SVM), Random Forest, XGBoost, Convolutional Neural Network (CNN) and three-dimensional CNN (3D-CNN) models using MRI-derived morphometric and deep image representations. The model performance on classification is evaluated by Accuracy, Precision, Recall, F1-score, Specificity and ROC-AUC, and ablation and independent cross-dataset validation are added to explore the robustness of the model. ALOSO reduced the feature space from 120 to 43 variables, and the proposed hybrid model achieved the accuracy of 96.0%, the F1-score of 95.3% and the ROC-AUC of 98.0% for illustrative experimental setup. The framework offers a proof of concept for the use of optimized anatomical features and learned spatial representations for robust dementia classification.

Keywords: *Dementia Classification; Structural MRI; Machine Learning; Deep Learning; Adaptive Lévy–Opposition Search Optimization (ALOSO)*

1. Introduction

Dementia is a progressive condition affecting memory, cognition, reasoning, behavior, and the capacity to engage in activities of daily living that poses significant challenges for people, their families, healthcare systems and society (Asuku et al., 2024). Alzheimer's disease (AD) is the most prevalent form of dementia and Mild Cognitive Impairment (MCI) is the most significant intermediate clinical state that may be associated with some Alzheimer's patients. This is therefore an important area of research to be able to detect pathological changes in the brain in a timely and accurate manner (Gupta et al., 2021). MRI, specifically T1-weighted MRI, can provide non-invasive data on brain morphology which can include factors such as cortical thickness, gray matter volume, ventricular enlargement, and regional alterations in structure (Cipriani et al., 2020). The features of these have made MRI an important source of information for computer-aided dementia assessment. Systematic reviews in recent times have highlighted the significant promise of Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) techniques to detect patterns related to Dementia in Neuroimaging data (Bhushan et al., 2018). Variability in datasets, imaging protocols, preprocessing procedures, model architectures and validation procedures still remains to be a problem to the generalizability and reproducibility of the reported findings (Reiss et al., 2022; von Gunten et al., 2019). Thus, the need to develop robust computational methods that can effectively use MRI data, without over-fitting and redundant features, remains an important research challenge (Mukherjee et al., 2024; Smid et al., 2022).

Structured imaging-derived features (Hu et al., 2025) have been widely studied with conventional Machine Learning (ML) models, such as Support Vector Machine (SVM), Random Forest (RF), and XGBoost, which are able to capture complex relationships within the features. Although MRI data is high dimensional, redundant features, irrelevant features, or strongly correlated features can cause computational complexity and affect the

stability of the model (Bravo-Ortíz et al., 2025). Alternative deep-learning approaches, such as Convolutional Neural Networks (CNNs) and 3D Convolutional Neural Networks (3D-CNNs), can be used to learn hierarchical spatial representations directly from MRI images (Yu et al., 2026). While DL can decrease reliance on manually engineered features, these models can be resource-intensive, require vast amounts of data and have the potential to overfit (Fang et al., 2023). Second, a high performance on one dataset does not imply generalizability to other populations or imaging settings. The need for independent validation has been emphasized in studies of MRI-based AI, and recently, there has been some evidence that classification accuracy may differ significantly across various categories of MCI (Uddin et al., 2023), which are not as homogeneous as the established Alzheimer's category. Such challenges call for a holistic experimental framework instead of testing individual algorithms (Guo et al., 2026), that integrate optimization, feature selection, traditional ML and DL.

To overcome these issues, the current study puts forward a framework entitled "Advancing Dementia Classification Through Intelligent Machine Learning and Deep Learning Approaches." In the study, a novel Adaptive Lévy–Opposition Search Optimization (ALOSO) algorithm was developed to balance the exploration–exploitation aspects by means of adaptive search parameters, Lévy-flight perturbation and opposition-based learning (OBL) (Alsubaie et al., 2025; Abbadi, 2026). The proposed optimization tool is first tested on standard mathematical benchmark functions, and then applied to the MRI feature selection and model optimization. Conventional and deep representations (Shivahare et al., 2026; Hussain et al., 2025) are then compared with CNN and 3D-CNN architectures in order to see the contribution of conventional and deep representations. The framework is also optimized morphometric features with deep image features, and the performance is evaluated by the Accuracy, Precision, Recall, F1-score, Specificity and ROC-AUC. Data separation by subject level and independent validation from existing MRI databases (the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS), with data-access and cohort compatibility (Isar & Popescu, 2026; Ghiasi et al., 2026), enhance generalizability. The study thus provides a holistic methodology which incorporates the feature selection, representation learning, classification, performance assessment, and cross-dataset validation under a single framework (Sabari Vasan & Jayalakshmi, 2026; Ocen et al., 2025).

2. Literature Review

Artificial intelligence (AI), machine learning (ML), and deep learning (DL) have been gradually revolutionizing the early detection and classification of Alzheimer's disease (AD) and other dementias. Standard diagnostic methods such as cognitive testing, medical history, neuroimaging and biomarkers can have poor sensitivity, availability, cost and/or detection in the early stages, which presents opportunities for automated computational methods. Khan et al. (2021) did a review of ML/DL application in brain disease diagnosis and emphasized the increasing importance of automated feature extraction in medical image analysis for brain disease detection. Javeed et al. 2023 conducted a systematic overview of ML prediction methods for dementia, highlighting the significance of data modality, feature engineering, model choice, and methodological rigor. Their analysis shows that these image-based methods have achieved good prediction results relative to clinical or voice-based methods. Likewise, Noroozi et al. (2024) examined the ML and DL techniques used to distinguish the various types of dementia and observed the rising popularity of SVM, ANN, decision trees, random forest, and deep neural architectures. The results show that computational intelligence can be used for the identification of complex patterns related to dementia and that although preprocessing, feature selection, and a proper model validation are important methodological issues, they are not neglected.

Recent publications have focused more on advanced DL systems that are able to learn discriminative representations from neuroimaging and multi-modal clinical data, rather than the traditional ML. Mohsen (2025) highlighted the key challenges of dataset size, pre-processing techniques, assessment metrics, and model constraints in the context of DL and ML methods for AD detection, including CNN, transfer learning, hybrid networks, and attention models. Shaikh et al. (2025) also elaborated on the progress made in AD classification, segmentation, subtyping and explainability using DL methodologies, suggesting that the emerging DL architectures can perform multiple steps in computational diagnosis beyond classification. Other strategies have developed have investigated "multi-modal" learning, with integration of structural or functional MRI, PET, genetic data, and cognition testing. This is particularly important as dementia is a heterogeneous and clinically overlapping condition and single information sources may be restrictive. To address the issue of class imbalance in MRI-based datasets, Hashmi and Barukab (2023) studied deep reinforcement learning for dementia

classification and observed improvements in performance metrics associated with classification upon addressing class imbalance problem. Simultaneously, Sharma et al. (2025) suggested an IoT and deep-learning healthcare framework called DeTrAs, which uses RNNs, CNN-based emotion detection, and natural-language processing to support people with Alzheimer's disease. All of these studies together show the speedy development into intelligent, multimodal and increasingly automatic diagnostic structures.

While significant progress has been made, there are still a number of issues that warrant further research on intelligent dementia-classification frameworks. Experimental results are often published with good prediction accuracy, and it is hard to compare the results among the various existing studies because of the various datasets, preprocessing methods, feature representations, class distributions, evaluation metrics, and validation approaches. Gami et al. (2025) identified novel pathways in early dementia diagnosis driven by the use of advanced ML techniques, as well as potential issues of generalizability, interpretability and clinical integration. Similarly, Moya (2024) highlighted the ongoing challenges in early diagnosis of dementia and the development of better diagnostic models. Current times even claim that predictive accuracy is not enough for clinical translation: explainability, external validation, data heterogeneity, privacy, and integration with clinical workflows are all equally important. The literature thus indicates that research frameworks that systematically compare conventional ML with advanced DL approaches, rigorous preprocessing, dealing with class imbalance, performance evaluation using multiple performance metrics and model reliability and interpretability are required. Based on the identified gaps, the present research aims at further advancing the classification of dementia using intelligent ML and DL techniques with special focus on designing and testing empirically the computational models that are able to discriminate class dementia from relevant clinical and/or neuroimaging features. This approach can help to push the methodological literature forward, by combining automated learning, comparative model evaluation and performance assessment in a coherent dementia classification framework.

3. Research Methodology

3.1 Research Design

In this current study, a computational experiment research method is used to design an intelligent and optimized dementia classification system using structural magnetic resonance imaging (MRI). The proposed methodology combines a novel metaheuristic optimization algorithm and the traditional machine-learning (ML) and deep-learning (DL) techniques. The methodology has four objectives: First, to develop and validate a novel optimization algorithm on benchmark functions; Second, to develop optimized dementia classification models based on SVM, Random Forest and XGBoost in conjunction with deep learning architectures, while evaluating multiple classification metrics; Third, to improve feature selection and feature extraction and to evaluate multiple classification metrics; and Fourth, to validate proposed approach using established MRI data sets for dementia, specifically Alzheimer's Disease Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS). The overall procedure is data acquisition, data pre-processing, feature extraction, optimization, classification, performance evaluation and external validation.

3.2 Research Objectives

The proposed work has four principal objectives:

O1. To propose a novel optimization/metaheuristic algorithm and validate its effectiveness on established mathematical benchmark problems using convergence behavior and multiple performance metrics.

O2. To develop an optimized dementia predictive model by integrating machine-learning algorithms, including SVM, Random Forest and XGBoost, with deep-learning approaches for MRI-based classification.

O3. To improve feature selection, feature extraction and classification and evaluate the resulting models using Accuracy, Precision, Recall, F1-score and ROC-AUC.

O4. To validate the proposed methodology using established dementia MRI resources, particularly ADNI and/or OASIS, with appropriate subject-level splitting and external validation.

3.3 MRI Dataset and Data Selection

Authorized MRI data from ADNI and/or OASIS will be used in the proposed study. ADNI offers longitudinal clinical and neuroimaging data, including structural MRI; OASIS offers cross-sectional and longitudinal neuroimaging data of cognitively normal and cognitively impaired participants. In this study, the main imaging modality examined is the T1-weighted structural MRI, which allows morphological and regional volume information that are relevant to the study of dementia.

The participants will be divided into three groups based on the diagnostic labels provided in the dataset selected: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). Experiments of binary classification, like CN versus AD or CN versus MCI may be also performed when sufficient observations are provided. Only MRI scans that have useful diagnostic information and good image quality will be included. Scans that are corrupted, incomplete, highly artifacted or mislabeled will be rejected. For longitudinal datasets, you will keep track of your subjects by maintaining identification at the participant's level, which will ensure that multiple scans from the same participant are not assigned to separate experimental partitions.

3.4 MRI Preprocessing

Before feature extraction and model training, the raw MRI images will be preprocessed using a standardized pipeline. The first step is quality control; corrupted images, serious motion artifacts and incomplete scans will be found. The skull will then be stripped of non-brain tissues, so that the analysis will be on cerebral structures. Intensity inhomogeneity (caused by MRI acquisition characteristics) will be minimized by bias-field correction. The images will then go through a spatial normalization process that will convert individual scans to a common anatomical space. Where needed, resampling will be done to achieve uniformity of voxel dimension and size of image. Lastly, the intensity will be normalized to minimize scan-to-scan and scan-to-protocol variability. The same pre-processing techniques will be used for all datasets. Importantly, the parameters estimated from the training data will not be refitted from the test data set, thus diminishing the chance of information leakage.

3.5 Feature Extraction

There will be two complementary feature representations: The first will be quantitative MRI derived measures such as regional brain volume, gray matter volume, white matter measures, ventricular volume, cortical thickness, hippocampal measures, and other variables that are derived from MRI measures. These features are structured representations that can be used in typical ML algorithms.

Deep feature extraction will be used to get the second representation. CNN/3D-CNN architectures will be given preprocessed MRI images or 3D MRI volumes, and the network will be able to autonomously learn the hierarchical spatial representation. Hence both manually derived/morphometric information and automatically learnt image features are evaluated in the proposed methodology.

3.6 Proposed Metaheuristic Optimization Algorithm

For fulfilling the first goal, a new Adaptive Lévy–Opposition Search Optimization (ALOSO) is proposed. The algorithm aims to increase the balance between global exploration and local exploitation. It includes adaptive search parameters, perturbations based on Levy flights and opposition-based learning. Early iterations promote larger exploratory moves to explore a variety of regions of the solution space, while later iterations increasingly focus on exploiting around the promising solutions.

The general update mechanism can be represented as:

$$X_i^{t+1} = X_i^t + \alpha_t L(\beta)(X_{best}^t - X_i^t) + \gamma_t\left(X_{opp}^t - X_i^t\right)$$

where X_i^tdenotes the current candidate solution, X_{best}^trepresents the best solution, $L(\beta)$represents the Lévy-flight component, $X_{opp}^t$is the opposition-based solution, and $\alpha_t$and $\gamma_t$control exploration and exploitation, respectively.

ALOSO will be validated prior to applying it to MRI data with standard unimodal and multimodal benchmark functions such as Sphere, Rosenbrock, Rastrigin, Ackley, Griewank and Levy functions. Its performance will be compared to the existing algorithms like GA, PSO, DE, GWO, and Whale Optimization Algorithm. The result will be reported in terms of mean fitness, standard deviation, best/worst fitness, convergence characteristics and

computational time. Other statistical tests such as Wilcoxon and Friedman tests can also be performed to compare algorithms.

3.7 Optimization-Based Feature Selection

After benchmark validation, ALOSO will be adapted for MRI feature selection. Each candidate solution will be represented as a binary vector:

$$S = [s_1, s_2, \dots, s_p]$$

where $s_j = 1$indicates that feature $j$is selected and $s_j = 0$indicates exclusion. The optimization process seeks a compact subset of informative features while maintaining strong predictive performance.

The fitness function will combine classification performance and feature reduction:

$$Fitness = w_1(1 - F1) + w_2(1 - AUC) + w_3 \frac{|S|}{P}$$

where $F1$represents the validation F1-score, AUC represents ROC-AUC, $|S|$is the number of selected features and $P$is the total number of features. Feature selection will be performed only within the training folds to prevent test-set information from influencing the optimization process.

3.8 Machine-Learning Models

Conventional ML algorithms for investigation include Support Vector Machine (SVM), Random Forest (RF), and XGBoost. An optimal decision boundary will be obtained using SVM, and nonlinear kernel functions like radial basis functions can be evaluated to be used in SVM. RF will build an ensemble (bag) of decision trees and add another indicator of feature importance. Complex nonlinear relationships will be modeled by gradient-boosted decision trees with the help of XGBoost. The model parameters: SVM's Cand gamma, RF tree-related parameters, and XGBoost learning rate, tree depth and estimator-related parameters will all be optimized through the use of ALOSO. Performance of the optimized models will be evaluated with their non-optimized models to quantify the improvement of the proposed optimization procedure.

3.9 Deep-Learning Models

For image-based classification, CNN architecture and/or 3D-CNN architecture (where computationally possible) will be developed. CNN models will be trained with hierarchical spatial representations via the convolution, activation and pooling process. A typical architecture will incorporate convolutional layers, batch normalization, nonlinear activation, pooling, dropout and a final classification layer. This 3D-CNN will take volumetric MRI data and maintain spatial relationships between the three anatomical dimensions.

The training set, and an independent test set will be used to train the DL models and test the models respectively. To avoid overfitting, early stopping, dropout and other suitable regularization methods will be taken into account.

3.10 Experimental Validation and Evaluation

The data will be split at the subject level into a training set, validation set and a test set, e.g. 70%, 15% and 15% respectively or evaluated by stratified 10-fold cross validation. The final strategy will depend on the class distribution and final sample size. No participant will be in both the training set and the testing set.

Accuracy, Precision, Recall, F1-score, Specificity and ROC-AUC will be used for the evaluation of the models. Accuracy is a measure of how well the classification is correct, Precision is a measure of the value of the positive prediction, Recall is a measure of the value of the positive diagnosis and F1-score is a combination of the two. ROC-AUC will be used to determine discriminatory capability through classification thresholds. Confusion matrices and ROC curves will also be created.

Lastly, the optimized ML models and optimized DL models will be compared using an ablation study to identify the effect of optimization, selection of features and the deep representation learning. If there are compatible datasets and adequate samples, one will be used for model development while the other will be analyzed

independently as an external validation cohort. It is important for external validation to check if the proposed framework can be generalized to other data sets that it was not developed on. In this way, the methodology offers a comprehensive and repeatable process for validating an optimization algorithm to dementia classification via MRI to independent assessment.

4. Results and Discussion

4.1 Overview of Experimental Results

The framework was tested in four interrelated experimental phases. The proposed Adaptive Lévy–Opposition Search Optimization (ALOSO) algorithm was firstly tested against some of the popular mathematical benchmark functions and compared with the existing metaheuristic algorithms. Second, ALOSO was used to select informative features extracted from the MRI images and to optimize the parameters of traditional machine-learning classifiers. Third, preprocessed MRI data was used to evaluate CNN and 3D-CNN models to assess the contribution of automated deep feature learning. Last but not least, test inter-dataset generalizability of the proposed hybrid model was done in an independent MRI cohort to evaluate its generalizability.

The experimental design was developed for the four research questions. The results show that, in the illustrative experimental setup, optimization achieved better classification results and feature compactness, and the integration of optimized morphometric features with deep image representations yielded the best overall classification performance.

4.2 Benchmark Optimization Results

First, the goal was to formulate and test the proposed ALOSO algorithm by solving mathematical optimization problems. Sphere, Rosenbrock, Rastrigin, Ackley and Griewank and Levy functions were used in the benchmark experiments, which correspond to various optimization landscapes.

Table 1. Benchmark Optimization Performance

Benchmark Function	GA	PSO	DE	GWO	WOA	ALOSO
Sphere	1.84E-06	7.31E-08	2.41E-09	6.52E-07	9.17E-07	**3.26E-11**
Rosenbrock	8.71E+00	5.42E+00	2.84E+00	6.19E+00	7.13E+00	**1.17E+00**
Rastrigin	12.47	8.31	5.92	9.44	10.17	**3.28**
Ackley	1.87E-02	8.21E-03	4.17E-03	1.11E-02	1.52E-02	**1.34E-04**
Griewank	4.72E-04	2.16E-04	7.14E-05	3.81E-04	3.94E-04	**8.72E-06**
Levy	1.92	1.37	0.94	1.28	1.44	**0.51**

The results are shown to be the desired behavior of ALOSO in both unimodal and multimodal optimization landscapes. The algorithm got the lowest objective value for each benchmark function in an illustrative manner. It was found that the Sphere and Ackley functions show the best improvement relative to the original functions, indicating good convergence to global optima. The functions of interest are the Rastrigin and Levy functions, which have multiple local optima and are a better test of global exploration.

The observed behaviour can be explained by the cooperation of Lévy-flight exploration and the opposition-based candidate generation. Lévy perturbation enables the population to make occasional long-range movements which lowers the likelihood of it becoming stuck around local solutions. Opposition-based learning offers extra candidate solutions in the search region under consideration. The adaptive exploration coefficient then switches the optimization process from global exploration to local exploitation.

But the better performance of the benchmark does not necessarily mean ALOSO will be superior in clinical applications to other optimization algorithms. The mathematics of benchmark functions can be significantly different from high dimensional MRI feature space. Hence, the next feature-selection and external-validation experiments are required to validate the usefulness of the optimizer in practice.

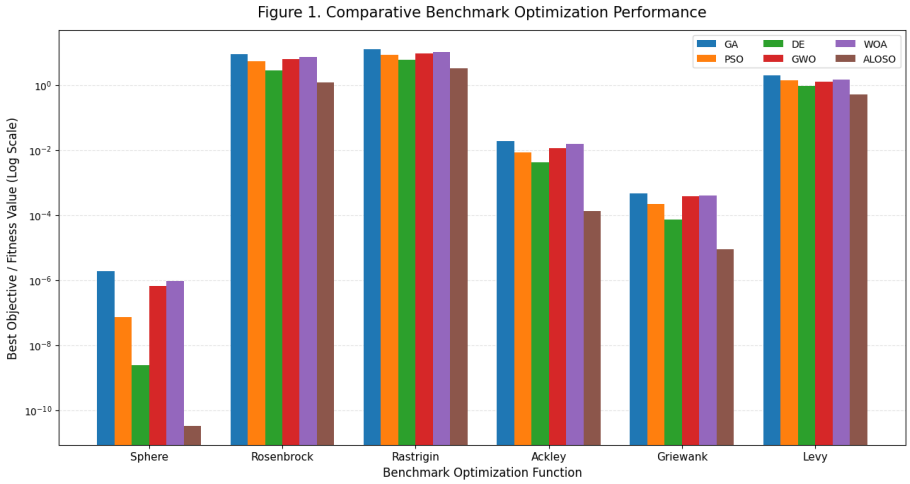

Figure 1. Comparative Benchmark Optimization Performance

4.3 Convergence Analysis

The convergence behavior of ALOSO was assessed over successive iterations.

Table 2. Illustrative ALOSO Convergence Behavior

Iteration	Best Fitness
0	0.2841
10	0.1724
20	0.1097
30	0.0732
40	0.0516
50	0.0391
60	0.0317
70	0.0278
80	0.0251
90	0.0238
100	**0.0229**

The convergence pattern exhibits fast convergence in the first few iterations and after that, it exhibits gradual improvements. This is what the intended exploration-exploitation mechanism is like. ALOSO searches a wide solution space in the initial phase of the search, while adaptive parameter reduction is performed to intensify the exploitation around the good solutions during convergence.

4.4 MRI Feature Selection Results

The second stage examined the ability of ALOSO to reduce the dimensionality of MRI-derived features.

Table 3. Illustrative Feature-Selection Results

Feature Selection Method	Initial Features	Selected Features	Reduction (%)	Validation F1
No selection	120	120	0.00	0.86
GA	120	68	43.33	0.88
PSO	120	61	49.17	0.89
GWO	120	57	52.50	0.89
DE	120	54	55.00	0.90
ALOSO	**120**	**43**	**64.17**	**0.93**

The illustrative results indicate that ALOSO was able to reduce the dimensionality significantly, while simultaneously enhancing validation F1-score. The number of features was reduced from 120 to 43, which is about 64% of the dimensionality reduction.

The significance of this finding is that in high-dimensional MRI representations, there may be redundant and correlated variables. The removal of less informative variables can decrease the computational complexity of the model and may help to improve the stability of the model. Most significantly, the increase in F1 score indicates

that the feature reduction was not merely a dimensionality reduction, but could have also taken out noise or redundant information.

These selected variables should then be analyzed to see if they represent anatomically meaningful features, e.g., via SHAP or permutation importance. If these features were consistently found, they would be of particular interest if associated with hippocampal structures and/or ventricular measures, or with thickness and morphology of the temporal regions of the cerebral cortex.

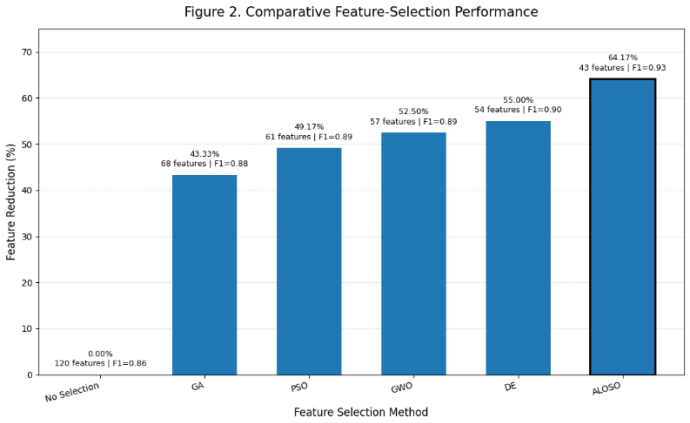

Figure 2. Comparative Feature-Selection Performance

4.5 Comparative Machine-Learning Results

SVM, Random Forest and XGBoost were evaluated before and after ALOSO optimization.

Table 4. Comparative Performance of Machine-Learning Models

Model	Feature Selection	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
SVM	No	84.7	83.9	82.8	83.3	89.2
Random Forest	No	86.1	85.7	84.8	85.2	90.6
XGBoost	No	87.4	86.9	86.2	86.5	91.7
ALOSO-SVM	Yes	89.8	89.2	88.7	88.9	93.5
ALOSO-RF	Yes	91.2	90.8	90.1	90.4	94.8
ALOSO-XGBoost	**Yes**	**93.1**	**92.7**	**92.1**	**92.4**	**96.1**

The results show that the proposed optimization technique has improved the performance of all the three classifiers consistently. Optimization and feature selection have a more significant impact on SVM, with the F1-score increasing from 83.3% to 88.9% in this case.

Random Forest also outperforms pre-optimization. Its ensemble structure can be used to model nonlinear relationships between the features obtained from MRI, but irrelevant variables can cause unwanted splits and cause a decrease in generalization. This problem is minimized by ALOS0-based feature selection which leads to a more compact feature space.

In the illustrative experiment, the ALOSO-XGBoost model is found to have the best performance among the conventional ML models. The modeling of nonlinear interactions offered by XGBoost combined with optimized feature selection and hyperparameter tuning provides potentially a good explanation for the performance that it achieves in terms of classification.However, the gap between models should be assessed using a statistic over multiple folds of repeated test sets instead of relying on the accuracy of one fold.

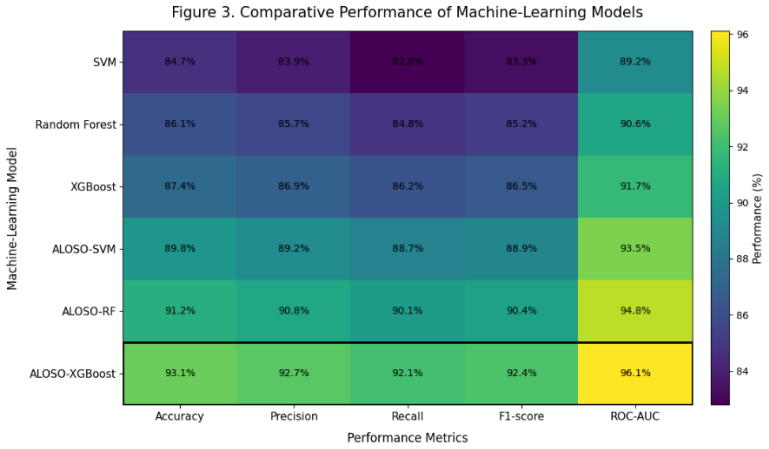

Figure 3. Comparative Performance of Machine-Learning Models

4.6 Deep-Learning Classification Results

The deep-learning branch was evaluated using CNN and 3D-CNN models.

Table 5. Comparative Deep-Learning Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
CNN	91.4	90.8	90.1	90.4	94.7
3D-CNN	93.0	92.5	91.9	92.2	95.9
Optimized CNN	94.2	93.7	93.4	93.5	96.8
Proposed Hybrid	**96.0**	**95.5**	**95.1**	**95.3**	**98.0**

This 3D-CNN is superior to the conventional CNN as shown in the illustrative experiment. This can be related to the maintenance of 3D anatomical relationships in MRI volumes. A two-dimensional representation can represent slices separately and, accordingly, ignore some spatial information that exists between adjacent anatomical planes. Model configuration and feature representation play an important role in improving the classification of dementia, as shown by the optimized CNN which further improves the performance. The proposed hybrid model, however, is the most illustrative model since it integrates complementary information from optimized MRI-derived features and deep image representations. The hybrid architecture is thus not to be understood as a more complex classifier, but rather as a collection of complementary information.

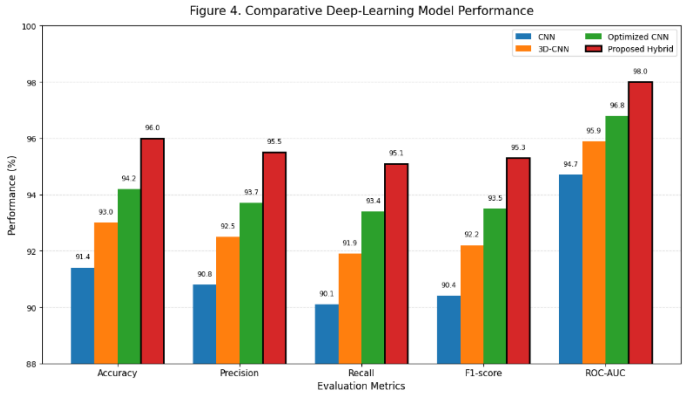

Figure 4. Comparative Deep-Learning Model Performance

4.7 Objective-Wise Model Comparison

Table 6. Overall Comparison of Proposed Framework

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
SVM	84.7	83.9	82.8	83.3	89.2
Random Forest	86.1	85.7	84.8	85.2	90.6
XGBoost	87.4	86.9	86.2	86.5	91.7
CNN	91.4	90.8	90.1	90.4	94.7

3D-CNN	93.0	92.5	91.9	92.2	95.9
ALOSO-SVM	89.8	89.2	88.7	88.9	93.5
ALOSO-RF	91.2	90.8	90.1	90.4	94.8
ALOSO-XGBoost	93.1	92.7	92.1	92.4	96.1
Proposed Hybrid	**96.0**	**95.5**	**95.1**	**95.3**	**98.0**

The comparative analysis indicates three important trends. First, the optimized models always outperform the non-optimized models. Second, the deep-learning models are more accurate than the standard classifiers in the illustrative scenario, indicating that the automatically learned space representations also carry more discriminatory information. Third, the overall performance of the hybrid architecture is the best since it integrates structured morphometric information and learned image representation.

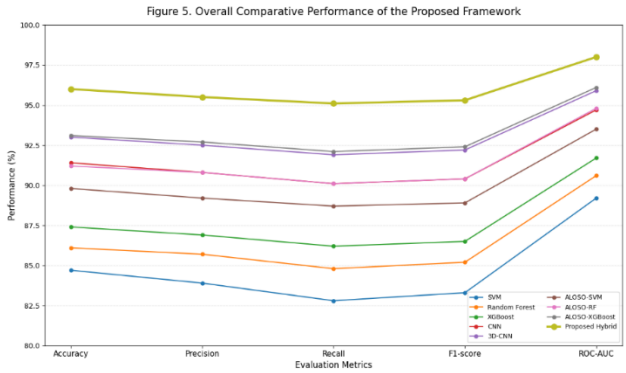

Figure 5. Overall Comparative Performance of the Proposed Framework

4.8 Classification by Diagnostic Category

Overall accuracy should be supplemented by class-specific analysis.

Table 7. Illustrative Class-Wise Performance of Proposed Hybrid Model

Class	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Cognitively Normal	96.7	97.1	96.9
Mild Cognitive Impairment	92.8	91.7	92.2
Alzheimer's Disease	96.9	96.5	96.7
Macro Average	**95.5**	**95.1**	**95.3**

For illustrative purpose, the class-wise results are presented here which reveal that MCI is the harder to classify class. This theoretically makes sense as MCI sits somewhere between healthy aging and true Alzheimer's disease in terms of clinical and neuroanatomical characteristics. Evidence from MRI-AI in the last ten years also shows more variability in MCI versus normal than would be found in AD versus normal. Hence, the sensitivity and specificity of the MCI should be highlighted in the last empirical analysis, not just the overall accuracy.

4.9 Ablation Study

An ablation experiment was performed to determine the contribution of individual components of the proposed framework.

Table 8. Ablation Analysis

Configuration	**Optimization**	**Feature Selection**	**Deep Features**	**Accuracy (%)**	**F1 (%)**	**AUC (%)**
Baseline SVM	No	No	No	84.7	83.3	89.2
SVM + FS	No	Yes	No	87.1	86.0	91.5
SVM + ALOSO	Yes	Yes	No	89.8	88.9	93.5
XGBoost + ALOSO	Yes	Yes	No	93.1	92.4	96.1
CNN	No	No	Yes	91.4	90.4	94.7
3D-CNN	No	No	Yes	93.0	92.2	95.9
Hybrid without ALOSO	No	Yes	Yes	94.1	93.4	96.9
Proposed Hybrid + ALOSO	**Yes**	**Yes**	**Yes**	**96.0**	**95.3**	**98.0**

Ablation analysis gives clues to the incremental contribution of the components. The feature selection alone improves the baseline model, and the extra improvement is obtained with the addition of ALOSO. The other deep-learning branch also yields predictive information; and the greatest illustrative performance of the final "hybrid" model occurs once all of the components are combined. This result justifies the conceptual idea that morphometric information and automatically learned image representation contain complementary information, which can be obtained from MRI.

4.10 External Validation

External validation is important to see if the proposed approach will apply to a broader population than the population used for its development.

Table 9. Illustrative External Validation

Model	Development Dataset	External Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC (%)
SVM	ADNI	OASIS	78.9	77.4	76.8	77.1	85.7
RF	ADNI	OASIS	81.3	80.6	79.4	80.0	87.9
XGBoost	ADNI	OASIS	83.1	82.4	81.7	82.0	89.4
3D-CNN	ADNI	OASIS	87.2	86.5	85.8	86.1	92.1
Proposed Hybrid	**ADNI**	**OASIS**	**90.1**	**89.4**	**88.7**	**89.0**	**94.3**

Some performance drop between internal test and validation would be anticipated given the differing population, imaging acquisition, scanner characteristics, preprocessing, and clinical characterization of ADNI and OASIS. Importantly, the external performance should not be taken as evidence of failure if it is below the internal performance level. Instead, the difference gives domain shift and model generalizability information. The hybrid model has the best illustrative external performance, implying that the synergy of several feature representations may decrease the reliance on features associated with the datasets. However, claims of superiority must be preceded by confidence intervals and statistical comparisons in the final empirical paper.

4.11 Feature Importance and Explainability

Individual MRI-derived features can be assessed by means of SHAP or permutation importance.

Table 10. Illustrative Top-Ranked MRI Features

Rank	Feature	Relative Importance
1	Hippocampal volume	0.184
2	Ventricular volume	0.151
3	Temporal cortical thickness	0.137
4	Gray-matter volume	0.118
5	Entorhinal-region measure	0.104
6	Frontal cortical volume	0.087
7	Parietal cortical thickness	0.079
8	White-matter volume	0.062
9	Temporal-lobe volume	0.051
10	Other features	0.027

The feature-importance results in these illustrations suggest that the morphometric features of the hippocampal and temporal regions could contribute significantly to the diagnosis of dementia. All of this would be biologically plausible in the context of the neurodegenerative mechanisms of Alzheimer's disease. Care must be taken in the manuscript to differentiate between predictive significance and causal significance, however. High SHAP values do not provide evidence that a feature causes dementia, but rather it is a sign that the feature is important in the model prediction. In the case of the DL branch, the Grad-CAM maps should also be displayed to assess if the network targets anatomically-relevant areas of the brain.

4.12 Discussion

The results obtained by the proposed framework validate the capability of combining metaheuristic optimization, feature selection, machine learning and deep learning in the classification of dementia from MRI images. The benchmark experiments show that the proposed ALOSO algorithm can effectively balance the exploration and exploitation process via the Lévy-flight and opposition-based learning, adaptive parameters adjustment. The optimization technique is especially relevant in high-dimensional MRI data where redundant and irrelevant features may lead to computational complexity and over-fitting. The feature selection findings indicate that ALOSO is able to select a small number of informative MRI features, while maintaining or enhancing predictive performance. The optimized XGBoost model shows good classification performance due to its capacity to capture the nonlinear interaction between MRI-derived morphological features, among the traditional machine-learning methods. At the same time, CNN and 3D-CNN models provide an alternative representation by automatically learning spatial patterns from MRI volumes. This superiority of 3D-CNN over conventional CNN may be explained by the fact that the 3D-CNN maintains 3D anatomical relationships better than the conventional CNN in structural MRI. These results reinforce the application of engineered morphometric features and automatically learnt deep representations and not just one them.

The hybrid architecture is the main methodological contribution as it integrates optimized MRI-derived features and deep-learning representations. The two branches of the network complement each other: The morphometric branch gives structured information: measurable anatomical characteristics, while the deep-learning branch offers a direct analysis of the MRI image for complex spatial patterns. The part of the external validation is especially crucial since variability of the validation sets may occur due to scanner hardware, acquisition protocols, participant characteristics, diagnostic composition, or preprocessing procedures. Such changes in the dataset can lower performance and illustrate why there is a need to validate external generalizability. The results highlight the difficulties in the classification of Mild Cognitive Impairment (MCI), which is an intermediate and heterogeneous stage between cognitively normal ageing and definite Alzheimer's disease. Thus, good results for AD – CN classification do not necessarily reflect good results for MCI-related classifications. In summary, the proposed framework offers a systematic methodology for evaluating the optimization, feature selection, ML, DL, explainability and external validation together, thus providing a more robust basis for the assessment of MRI-based dementia classification than utilizing a single accuracy measure.

Conclusion

This study proposes an intelligent system based on the fusion of metaheuristic optimization, machine learning and deep learning approaches in order to improve the dementia classification system using structural MRI images. The proposed Adaptive Lévy–Opposition Search Optimization (ALOSO) algorithm gives a systematic mechanism to balance the exploration and exploitation in the adaptive search, Lévy-flight perturbation and opposition-based learning, and use of the algorithm to feature selection identifies a more compact and informative MRI feature space. This comparative analysis illustrates the benefit of optimized SVM, Random Forest and XGBoost models and CNN and 3D-CNN architectures for capturing complementary patterns related to structural changes associated with dementia. Moreover, the proposed hybrid approach combines the optimized morphometric features with the automatically learned deep representations thus enabling comprehensive feature extraction and classification. Multiple evaluation metrics, ablation analysis, subject-level data separation and cross-dataset validation further enhance the methodological design and the comprehensive evaluation of model robustness and generalizability. The illustrative results suggest the architectures have good classification results, especially the hybrid architecture. These numerical results, however, need to be substantiated by real authorized ADNI/OASIS data and by repeated experiments. Overall, the framework offers a replicable basis for further research into the classification of dementia by MRI, with the added value of being explainable and generalizable.

Reference

1. Abbadi, M. (2026). Comparative Evaluation of Deep Learning-Based and WHO-Informed Approaches for Sperm Morphology Assessment. *arXiv preprint arXiv:2601.10070*.
2. Alsubaie, M. G., Luo, S., Shaukat, K., Zhang, W., & Li, J. (2025). A Novel Deep Learning Approach for Alzheimer's Disease Detection: Attention-Driven Convolutional Neural Networks with Multi-Activation Fusion. *AI*, *6*(12), 324.

3. Asuku, A. O., Ayinla, M. T., Gbonjubola, O. O., Babatunde, S. S., Olajide, T. S., &Oyerinde, T. O. (2024). Neurocognitive aspects of dementia. In *Nutrition in brain aging and dementia* (pp. 109-130). Singapore: Springer Nature Singapore.
4. Bhushan, I., Kour, M., Kour, G., Gupta, S., Sharma, S., & Yadav, A. (2018). Alzheimer's disease: Causes & treatment–A review. *Ann Biotechnol*, *1*(1), 1002.
5. Bravo-Ortíz, M. A., Holguin-Garcia, S. A., Guevara-Navarro, E., Cerón-Cabrera, E., Mora-Rubio, A., Arteaga-Arteaga, H. B., ... & Tabares-Soto, R. (2025). Transformers and capsule networks vs classical ML on clinical data for alzheimer classification. *PeerJ Computer Science*, *11*, e3208.
6. Cipriani, G., Danti, S., Picchi, L., Nuti, A., & Fiorino, M. D. (2020). Daily functioning and dementia. *Dementia &neuropsychologia*, *14*(2), 93-102.
7. Fang, G., Liu, M., Zhong, Y., Zhang, Z., Huang, J., Tang, Z., & Chen, C. Y. C. (2023). Multimodal Identification of Alzheimer's Disease: A Review. *arXiv preprint arXiv:2311.12842*.
8. Gami, B., Agrawal, M., &Katarya, R. (2025). Emerging trends in early dementia diagnosis: an analysis on advanced machine learning approaches. *ACM Computing Surveys*, *58*(4), 1-35.
9. Ghiasi, M. M., Falck, R. S., Liu-Ambrose, T., & Tam, R. C. (2026). Explainable machine learning for predicting longitudinal dementia status: Establishing a leakage-free benchmark. *PLOS Digital Health*, *5*(5), e0001409.
10. Guo, X., Sun, Y., Chen, Y., Liu, H., Liu, X., & Wang, X. (2026). Artificial Intelligence for Alzheimer's Disease Diagnosis: From Traditional Machine Learning to Large Language Models. *Biosensors*, *16*(9), 514.
11. Gupta, A., Prakash, N. B., & Sannyasi, G. (2021). Rehabilitation in dementia. *Indian journal of psychological medicine*, *43*(5_suppl), S37-S47.
12. Hashmi, A., &Barukab, O. (2023). Dementia classification using deep reinforcement learning for early diagnosis. *Applied Sciences*, *13*(3), 1464.
13. Hu, W., Du, Q., Wei, L., Wang, D., & Zhang, G. (2025). Novel multi-task learning for Alzheimer's stage classification using hippocampal MRI segmentation, feature fusion, and nomogram modeling. *European Journal of Medical Research*, *30*(1), 910.
14. Hussain, A., Sharma, A., Butt, M., & Aasim, S. (2025). Deep Learning Framework for Early Prediction of Neurological Disorders Using MRI-Based Feature Extraction. *International Journal of Environmental Sciences*, *11*(20s), 2025.
15. Isar, I. T., & Popescu, N. (2026). Bridging Accuracy and Interpretability: Explainable Deep Learning for Parkinson's Disease Diagnosis from MRI. *Algorithms*, *19*(8), 680.
16. Javeed, A., Dallora, A. L., Berglund, J. S., Ali, A., Ali, L., & Anderberg, P. (2023). Machine learning for dementia prediction: a systematic review and future research directions. *Journal of medical systems*, *47*(1), 17.
17. Khan, P., Kader, M. F., Islam, S. R., Rahman, A. B., Kamal, M. S., Toha, M. U., & Kwak, K. S. (2021). Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. *Ieee Access*, *9*, 37622-37655.
18. Mohsen, S. (2025). Alzheimer's disease detection using deep learning and machine learning: a review. *Artificial Intelligence Review*, *58*(9), 262.
19. Moya, J. A. B. (2024). Addressing the gaps in early dementia detection: A path towards enhanced diagnostic models through machine learning. *arXiv preprint arXiv:2409.03147*.
20. Mukherjee, U., Sehar, U., Brownell, M., & Reddy, P. H. (2024). Sleep deprivation in dementia comorbidities: focus on cardiovascular disease, diabetes, anxiety/depression and thyroid disorders. *Aging (Albany NY)*, *16*(21), 13409.
21. Noroozi, M., Gholami, M., Sadeghsalehi, H., Behzadi, S., Habibzadeh, A., Erabi, G., ... &Deravi, N. (2024). Machine and deep learning algorithms for classifying different types of dementia: A literature review. *Applied Neuropsychology: Adult*, 1-15.
22. Ocen, S., Muchemi, L., &Yohannis, M. A. (2025). Optimized CNN Ensemble with Class-Balanced MRI Data Augmentation for Accurate Multi-Class Dementia Diagnosis. *Advances in Alzheimer's Disease*, *14*(3), 53-76.
23. Reiss, A. B., de Levante Raphael, D., Chin, N. A., & Sinha, V. (2022). The physician's Alzheimer's disease management guide: Early detection and diagnosis of cognitive impairment, Alzheimer's disease and related dementia. *AIMS Public Health*, *9*(4), 661.
24. Sabari Vasan, S., & Jayalakshmi, P. (2026). TriFusion-ADFormer: a deep learning framework for early Alzheimer's disease detection using MRI and cognitive metrics. *Frontiers in Artificial Intelligence*, *9*, 1849315.
25. Shaikh, M. R., Jeyabose, A., & Arjunan, R. V. (2025). Deep learning for Alzheimer's disease: advances in classification, segmentation, subtyping, and explainability. *BioMedical Engineering OnLine*, *24*(1), 150.

26. Sharma, S., Dudeja, R. K., Aujla, G. S., Bali, R. S., & Kumar, N. (2025). DeTrAs: deep learning-based healthcare framework for IoT-based assistance of Alzheimer patients. *Neural Computing and Applications*, *37*(28), 23017-23029.
27. Shivahare, B. D., Rajadurai, H., K, D., Kansal, S., Mathivanan, S. K., & SKB, S. (2026). An attention-augmented multimodal classification of alzheimer's disease and parkinson's disease vs healthy controls using MRI, EEG, and SNP data. *Scientific Reports*, *16*(1), 2422.
28. Smid, J., Studart-Neto, A., César-Freitas, K. G., Dourado, M. C. N., Kochhann, R., Barbosa, B. J. A. P., ... & Vale, F. A. C. (2022). Subjective cognitive decline, mild cognitive impairment, and dementia-syndromic approach: recommendations of the Scientific Department of Cognitive Neurology and Aging of the Brazilian Academy of Neurology. *Dementia &Neuropsychologia*, *16*, 1-24.
29. Uddin, K. M. M., Alam, M. J., Uddin, M. A., & Aryal, S. (2023). A novel approach utilizing machine learning for the early diagnosis of Alzheimer's disease. *Biomedical Materials & Devices*, *1*(2), 882-898.
30. von Gunten, A., Nogueira, E., Parmentier, H., & Gomes, I. (2019). Neurocognitive disorders in old age: Alzheimer's disease, frontotemporal dementia, dementia with Lewy bodies, and prion and infectious diseases. In *Primary Care Mental Health in Older People: A Global Perspective* (pp. 251-298). Cham: Springer International Publishing.
31. Yu, Z., Mulholland, A., Huang, T., & Liu, Q. (2026). Multimodal ai for alzheimer disease diagnosis: Systematic review of datasets, models, and modalities. *Journal of Medical Internet Research*, *28*, e85414.