Current Research in Translational Medicine 73 (2025) 103526

Contents lists available at ScienceDirect

Current Research in Translational Medicine

journal homepage: www.elsevier.com/locate/retram

Original article

Enhancing Alzheimer's disease prediction using random forest: A novel framework combining backward feature elimination and ant colony optimization

Afeez A. Soladoye [a], Nicholas Aderinto [b], Bolaji A. Omodunbi [a], Adebimpe O. Esan [a], Ibrahim A. Adeyanju [a], David B. Olawade [c,d,e,f,*]

[a] Department of Computer Engineering, Federal University, Oye-Ekiti, Nigeria
[b] Department of Medicine and Surgery, Ladoke Akintola University of Technology, Ogbomoso, Nigeria
[c] Department of Allied and Public Health, School of Health, Sport and Bioscience, University of East London, London, United Kingdom
[d] Department of Research and Innovation, Medway NHS Foundation Trust, Gillingham ME7 5NY, United Kingdom
[e] Department of Public Health, York St John University, London, United Kingdom
[f] School of Health and Care Management, Arden University, Arden House, Middlemarch Park, Coventry, CV3 4FJ, United Kingdom

ARTICLE INFO

Keywords:
Alzheimer's disease
Machine learning
Feature selection
Nature-inspired optimization
Random forest

ABSTRACT

Background: Alzheimer's disease (AD) represents a significant global health challenge due to its increasing prevalence and the limitations of current diagnostic approaches. Early detection is crucial as pathological changes occur 10-15 years before clinical symptoms manifest, yet current diagnostic methods typically identify the disease at moderate to advanced stages. Machine learning techniques offer promising solutions for early prediction, but face challenges related to feature selection and hyperparameter optimization.

Objective: To develop an enhanced predictive model for Alzheimer's disease by integrating advanced feature selection techniques with nature-inspired hyperparameter optimization for Random Forest classifiers while ensuring robust validation and statistical significance testing.

Methods: This study employed three feature selection techniques (Whale Optimization Algorithm, Artificial Bee Colony, and Backward Elimination Feature Selection) and two hyperparameter optimization algorithms (Artificial Ant Colony Optimization and Bald Eagle Search) to improve Random Forest model performance. A dataset comprising 2,149 instances with 34 features was preprocessed using MinMax normalization and Synthetic Minority Oversampling Technique (SMOTE) applied only to training data to prevent data leakage. Statistical significance testing using McNemar's test was conducted to compare model performances. Model performance was evaluated using accuracy, precision, recall, F1-score, and AUC with confidence intervals calculated using bootstrap sampling.

Results: The combination of Backward Elimination Feature Selection with Artificial Ant Colony Optimization achieved the highest performance (95% accuracy ± 1.2%, 95% precision ± 1.1%, 94% recall ± 1.3%, 95% F1-score ± 1.0%, 98% AUC ± 0.8%), outperforming other methodological combinations and conventional machine learning algorithms with statistically significant improvements ($p < 0.001$). This approach identified 26 significant features associated with Alzheimer's disease. Additionally, nature-inspired optimization algorithms demonstrated substantial computational efficiency advantages over empirical approaches (18 minutes versus 133 minutes).

Conclusion: The integration of advanced feature selection with nature-inspired hyperparameter optimization enhances Alzheimer's disease prediction accuracy while improving computational efficiency. However, external validation on independent datasets and prospective clinical studies are needed to establish real-world utility. This methodological framework offers promising applications for early diagnosis and intervention planning, with potential extensions to other complex medical prediction tasks.

* Corresponding author.
E-mail address: d.olawade@uel.ac.uk (D.B. Olawade).

https://doi.org/10.1016/j.retram.2025.103526
Received 15 May 2025; Accepted 30 June 2025
Available online 1 July 2025

1. Introduction

Alzheimer's disease (AD) represents one of the most significant healthcare challenges of the 21st century, with profound implications for patients, caregivers, healthcare systems, and society at large. As a progressive neurodegenerative disorder, Alzheimer's disease gradually impairs cognitive functions, memory, and the ability to perform daily activities, ultimately leading to complete dependence on caregivers [1]. Recent epidemiological studies indicate that pathological changes in the brain begin 10-15 years before clinical symptoms appear, highlighting the critical importance of early detection and intervention [2,3]. With the global aging population expanding rapidly, the prevalence of Alzheimer's disease is projected to increase dramatically from approximately 50 million cases worldwide currently to over 150 million by 2050 [4], creating an urgent need for effective diagnostic and predictive tools to facilitate early intervention and treatment planning.

Clinical manifestations of Alzheimer's disease typically progress through distinct stages, beginning with preclinical phases characterized by biomarker changes without symptoms, followed by mild cognitive impairment (MCI), and eventually severe dementia [5]. The disease presents various subtypes including typical late-onset AD (>65 years), early-onset AD (<65 years), and atypical variants such as posterior cortical atrophy and logopenic primary progressive aphasia [6]. Understanding these clinical complexities is essential for developing effective predictive models that can identify at-risk individuals across different disease presentations.

Early diagnosis of Alzheimer's disease presents significant challenges due to its complex and multifactorial nature. Traditional diagnostic approaches rely heavily on clinical assessments, neuropsychological tests, and advanced imaging techniques, which can be time-consuming, expensive, and often accessible only in specialized healthcare settings [7]. Furthermore, these methods typically detect the disease at moderate to advanced stages when neurological damage has already occurred and therapeutic interventions have limited efficacy [8]. Current diagnostic criteria, including the National Institute on Aging-Alzheimer's Association (NIA-AA) guidelines, emphasize biomarker-based diagnosis, but accessibility to advanced biomarker testing remains limited in many healthcare settings [9]. This diagnostic lag underscores the critical need for predictive models that can identify individuals at high risk of developing Alzheimer's disease before clinical symptoms manifest.

A systematic review of existing machine learning approaches reveals significant heterogeneity in methodological approaches and performance metrics [10]. While support vector machines have shown accuracy rates of 80-90% in neuroimaging studies [11], ensemble methods like Random Forest have demonstrated superior performance in multimodal datasets, achieving accuracies exceeding 90% [12]. However, most studies focus on single aspects of model optimization, either feature selection or hyperparameter tuning, rather than integrated approaches that address both simultaneously [13].

The advent of artificial intelligence and machine learning technologies has opened new avenues for disease prediction and diagnosis across various medical domains [14]. In particular, computational intelligence techniques offer promising solutions for Alzheimer's disease prediction by leveraging patterns within complex multimodal data to identify subtle indicators that might escape human observation [15]. These techniques can integrate diverse data types, demographic information, lifestyle factors, medical history, cognitive assessments, and clinical symptoms, to construct robust predictive models with potential applications in screening, risk stratification, and personalized medicine [16].

However, the development of accurate and efficient machine learning models for Alzheimer's disease prediction faces several technical challenges. The high dimensionality of medical datasets often introduces noise and redundancy, potentially obscuring relevant patterns and increasing computational complexity [17]. Additionally, the selection of appropriate features and the optimization of model parameters require specialized expertise and significant computational resources, limiting the practical implementation of advanced predictive models in clinical settings [18]. These challenges highlight the need for sophisticated methodologies that can enhance model performance while maintaining computational efficiency.

Feature selection represents a critical step in developing effective predictive models for Alzheimer's disease. By identifying the most informative features from high-dimensional datasets, feature selection techniques can improve model accuracy, reduce computational complexity, and provide insights into the underlying risk factors associated with disease development [19]. Traditional feature selection methods often rely on statistical measures or wrapper-based approaches, which may not fully capture the complex relationships within medical data. Nature-inspired optimization algorithms, such as Whale Optimization Algorithm (WOA) and Artificial Bee Colony Optimization (ABCOA), offer alternative approaches that can navigate complex search spaces to discover optimal feature subsets [20,21].

Similarly, hyperparameter optimization plays a crucial role in maximizing the performance of machine learning models. The selection of appropriate hyperparameters can significantly impact model accuracy, generalization capability, and computational efficiency [22]. Conventional hyperparameter tuning methods typically employ grid search or random search techniques, which can be time-consuming and may not identify optimal parameter configurations. Nature-inspired optimization algorithms, including Ant Colony Optimization (ACO) and Bald Eagle Search (BES), provide efficient alternatives by mimicking biological processes to explore the hyperparameter space systematically and identify optimal configurations with reduced computational overhead [23,24].

Among various machine learning algorithms, Random Forest has emerged as a particularly promising approach for medical applications, including Alzheimer's disease prediction [25]. As an ensemble learning method, Random Forest combines multiple decision trees to generate robust predictions while mitigating overfitting risks. Its ability to handle non-linear relationships, manage missing data, and provide feature importance rankings makes it well-suited for medical datasets, which often exhibit complex patterns and heterogeneity [26]. However, the performance of Random Forest models depends heavily on appropriate feature selection and hyperparameter configuration, highlighting the importance of integrated approaches that address both aspects simultaneously [27].

Recent clinical studies have demonstrated the potential of machine learning for early AD prediction. Liu et al. (2023) achieved 89% accuracy using neuroimaging biomarkers combined with clinical assessments [28]. Zhang et al. (2024) reported 92% accuracy integrating genetic markers with cognitive tests [29]. Wang et al. (2023) demonstrated 87% accuracy using only clinical and lifestyle factors, suggesting potential for accessible screening tools [30]. However, many existing approaches focus on either feature selection or hyperparameter optimization in isolation, potentially limiting model performance. Furthermore, the comparative evaluation of different feature selection techniques and optimization algorithms within a unified framework remains limited, creating uncertainty regarding the most effective methodological combinations for Alzheimer's disease prediction [31].

This study addresses these gaps by proposing an integrated approach that combines advanced feature selection techniques with nature-inspired hyperparameter optimization to enhance the performance of Random Forest models for Alzheimer's disease prediction. By systematically evaluating three distinct feature selection methods: Whale Optimization Algorithm, Artificial Bee Colony Optimization, and Backward Feature Elimination alongside two hyperparameter optimization algorithms, Ant Colony Optimization and Bald Eagle Search, this research aims to identify the most effective methodological combination while providing insights into the significant risk factors associated with Alzheimer's disease.

The primary objectives of this study are threefold: first, to identify

the most significant features associated with Alzheimer's disease using advanced feature selection techniques; second, to optimize the hyperparameters of Random Forest models using nature-inspired algorithms; and third, to evaluate the performance of the proposed integrated approach against conventional machine learning algorithms using robust statistical validation methods including cross-validation and significance testing. By achieving these objectives, this research contributes to the development of more accurate and efficient predictive models for Alzheimer's disease, with potential applications in clinical decision support, risk assessment, and early intervention strategies.

The significance of this research extends beyond technical advancements in machine learning methodology. By identifying the most relevant risk factors associated with Alzheimer's disease, this study provides valuable insights for medical practitioners, potentially informing diagnostic protocols and preventive interventions [32]. Furthermore, the comparative evaluation of different feature selection techniques and optimization algorithms offers practical guidance for researchers and developers seeking to implement similar approaches across various medical domains. Ultimately, this research aims to contribute to the broader effort of leveraging computational intelligence for addressing complex healthcare challenges, with particular emphasis on neurodegenerative disorders that pose increasing societal and economic burdens.

2. Method

This study employed a comprehensive machine learning approach to predict Alzheimer's disease using various computational intelligence techniques. The methodology followed a standard machine learning pipeline consisting of data acquisition, preprocessing, feature selection, hyperparameter optimization, model development, and performance evaluation with statistical validation.

2.1. Data acquisition

The study utilized an extensive multimodal dataset for Alzheimer's disease prediction, acquired from Kaggle titled "Alzheimer's disease dataset." This dataset represents cross-sectional clinical data collected from multiple medical centres, though specific institutional details are not provided in the original source. This dataset comprised 2,149 instances with diverse attributes including demographic details, lifestyle factors, medical histories related to Alzheimer's disease, cognitive assessments, and functional evaluations. The original class distribution consisted of 1,389 non-AD cases and 760 AD cases, representing a highly imbalanced dataset. All patient identifiers were pre-anonymized in the original dataset to ensure privacy protection [33]. Table 1 presents a detailed description of these factors, showcasing their meanings and value ranges. The dataset features included patient identifiers, age (ranging from 60-90 years), gender, ethnicity, education level, BMI, smoking status, alcohol consumption, physical activity levels, diet quality, sleep quality, family history of Alzheimer's, and various medical conditions such as cardiovascular disease, diabetes, depression, head injury, and hypertension.

The comprehensive details of all 34 features used in this study, including patient demographics, lifestyle factors, medical history, cognitive assessments, and clinical symptoms, are presented in Table 1, which provides the complete description of each variable along with their respective value ranges and encoding schemes for categorical attributes.

Table 1
Description of Alzheimer's dataset.

S/N	Feature's Name	Description
1	Patient ID	This is a unique identifies or number assigned to individual patients
2	Age	The age of the patients captured in this dataset ranges between 60 to 90 years
3	Gender	This is the gender of the patients, where 0 represents male and 1 represents female
4	Ethnicity	The ethnicity of the patient is encoded as 0: Caucasian, 1:African American, 2:Asian and 3: others
5	Education level	This is the level of patients' education encoded as 0: none, 1:High school, 2:Bachelor's and 3: Higher
6	BMI	This is the body mass index of the patient ranging from 15 to 40
7	Smoking	This is the smoking status of the patient where 0 indicates No and 1 indicates Yes
8	Alcohol consumption	This captures the weekly alcohol consumption unit of the patients ranging from 0 to 20
9	Physical activity	This indicates the patients' physical activities per week ranging from 0 to 10
10	Diet quality	This features represents the patients diet quality score ranging from 0 to 10
11	Sleep quality	Representing the quality of sleep observed by patients ranging from 4 to 10
12	Family history Alzheimer's	It indicates family history of Alzheimer's where 0 indicates No and 1 indicates Yes
13	Cardiovascular disease	Presence of cardiovascular disease where o indicates No and 1 indicates Yes
14	Diabetes	Presence of diabetes 0: No and 1: Yes
15	Depression	Depression by patients
16	Head injury	0: No and 1: Yes
17	Hypertension	0: No and 1: Yes
18	SystolicBP	Systolic blood pressure ranging between 90-180mmHg
19	Diastolic	Diastolic blood pressure ranging between 60-120mmHg
20	CholesterolTotal	Total Cholesterol level in the patients' body ranging from 150-300mg/dL
21	CholesterolLDL	Low density lipoprotein cholesterol levels in the body ranging between 50-200mg/dL
22	CholesterolHDL	High density lipoprotein cholesterol levels in the body ranging between 50-200mg/dL
23	CholesterolTriglycerides	Triglycerides level ranging between 50 to 400mg/dL
24	MMSE	Mini-Mental state examination, where lower scores indicate cognitive impairment whose value ranges from 0 to 30
25	Functional assessment	Lower Functional assessment implies greater impairment and its values ranges from 0 to 10
26	Memory complaints	Presence of Memory complaints, 0: No and 1: Yes
27	Behavioral problems	0: No and 1: Yes
28	ADL	Activities of daily living, where lower scores indicates greater impairment, ranging from 0 to 10
29	Confusion	Indicates presences of confusion, 0: No and 1: Yes
30	Disorientation	Indicates presences of Disorientation, 0: No and 1: Yes
31	Personality changes	Indicates presences of Personality changes, 0: No and 1: Yes
32	Difficulty completing tasks	Indicates if patient is facing any Difficulty completing tasks, 0:No and 1: Yes
33	Forgetfulness	Indicates if patient is experiencing Forgetfulness,0: No and 1: Yes
34	Diagnosis	Diagnosis status for Alzheimer's disease,0: No and 1: Yes

2.2. Data preprocessing

The study utilized an extensive multimodal dataset for Alzheimer's disease prediction, acquired from Kaggle titled "Alzheimer's disease dataset." This dataset represents cross-sectional clinical data collected from multiple medical centres, though specific institutional details are not provided in the original source. This dataset comprised 2,149 instances with diverse attributes including demographic details, lifestyle factors, medical histories related to Alzheimer's disease, cognitive assessments, and functional evaluations. The original class distribution consisted of 1,389 non-AD cases and 760 AD cases, representing a highly

imbalanced dataset. All patient identifiers were pre-anonymized in the original dataset to ensure privacy protection [33]. Table 1 presents a detailed description of these factors, showcasing their meanings and value ranges. The dataset features included patient identifiers, age (ranging from 60-90 years), gender, ethnicity, education level, BMI, smoking status, alcohol consumption, physical activity levels, diet quality, sleep quality, family history of Alzheimer's, and various medical conditions such as cardiovascular disease, diabetes, depression, head injury, and hypertension.

The comprehensive details of all 34 features used in this study, including patient demographics, lifestyle factors, medical history, cognitive assessments, and clinical symptoms, are presented in Table 1, which provides the complete description of each variable along with their respective value ranges and encoding schemes for categorical attributes.

The MinMax normalization was implemented using Equation 1:

$$x_{Norm} = \frac{(x - x_{min})}{x_{Max} - x_{Min}} \quad (1)$$

Where:

- xNorm is the normalized value
- x is the value to be normalized
- xMin is the minimum value in the column
- xMax is the maximum value in the column

Following data preprocessing but before model training, the dataset was split into training (70%) and testing (30%) sets using stratified sampling to maintain class distribution. Subsequently, to address potential class imbalance and prevent data leakage, Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively to the training data [34]. The SMOTE parameters were set to k_neighbors=5 and random_state=42 for reproducibility. After SMOTE application, the training set contained 1,507 instances with balanced class distribution (753 non-AD, 754 AD cases). These preprocessing steps yielded a clean, normalized, well-formatted, and balanced dataset, enabling optimal performance and consistency of the machine learning models employed for Alzheimer's disease prediction.

2.3. Feature selection

Feature selection is a crucial process in machine learning and data mining as it identifies the most important features for model training, thereby improving performance, reducing computational complexity, and preventing overfitting [35]. This study employed three distinct feature selection techniques: Whale Optimization Algorithm (WOA), Artificial Bee Colony (ABC), and Backward Feature Elimination (BFE). Each technique was applied to select the optimal subset of features using 5-fold cross-validation on the training set to ensure robust evaluation.

2.4. Whale optimization algorithm

The Whale Optimization Algorithm is a nature-inspired metaheuristic optimization technique simulating the hunting behavior of humpback whales [36]. In feature selection, WOA explores the feature space by mimicking the bubble-net feeding behavior of whales, where they surround prey and spiral inward. The algorithm evaluates subsets of features based on a fitness function designed to maximize classification accuracy while minimizing the number of selected features.

The fitness function for feature selection was calculated as: $\boldsymbol{Fitness} = \propto \boldsymbol{xAccuracy} + \boldsymbol{\beta x}(1 - |\boldsymbol{Selected\_Features}|/\text{Total\_Features})$ where $\alpha = 0.8$ and $\beta = 0.2$ to prioritize accuracy while encouraging feature reduction [37].

For implementation, negative infinity was initially set for maximization, with 5 whales and 50 iterations. The algorithm executed both exploitation and exploration phases, implementing the bubble net's attacking equation. Fitness was evaluated using cross-validation with Random Forest as the employed model.

2.5. Artificial bee colony optimization

The Artificial Bee Colony algorithm is another nature-inspired optimization technique based on honey bee foraging behavior [38]. ABC consists of three types of bees employed bees, onlooker bees, and scout bees that collectively search the feature space for optimal solutions. Employed bees explore the neighborhood of current solutions, onlooker bees select solutions proportional to their quality, and scout bees randomly search for new solutions to avoid local optima.

In this study, twenty bees were used with 50 iterations. Each bee phase was evaluated using five-fold cross-validation to assess fitness. The objective function calculation for ABC followed the same formulation as WOA to ensure fair comparison [39]. The algorithm optimized a fitness function balancing model accuracy and feature subset size, with its ability to maintain diversity helping to discover a robust subset of features.

2.6. Backward feature selection

Backward Feature Elimination is a traditional wrapper-based feature selection method that begins with the complete feature set and progressively eliminates the least relevant features based on predefined criteria (model accuracy) [40]. The process involved training the model on the full feature set and removing features making the least contribution to model accuracy one at a time. This process continued until further feature removal resulted in a visible decrease in performance. The deterministic nature of BFE provided a comprehensible and interpretable feature reduction process.

2.7. Hyperparameter optimization

Hyperparameter optimization is essential for machine learning model development, ensuring optimal configuration for best possible performance [41]. This study applied two nature-inspired optimization algorithms, Ant Colony Optimization (ACO) and Bald Eagle Search (BES) to optimize the Random Forest model's hyperparameters. The goal was to identify the optimal hyperparameter set yielding maximum predictive accuracy for Alzheimer's disease classification.

Key Random Forest hyperparameters optimized included:

- Number of Trees (n_estimators): The number of decision trees in the forest
- Maximum Depth (max_depth): The maximum depth of each tree
- Minimum Samples Split (min_samples_split): The minimum number of samples to split an internal node
- Minimum Samples Leaf (min_samples_leaf): The minimum number of samples at a leaf node
- Maximum Features (max_features): The maximum number of features considered for best split

2.8. Ant colony optimization

Ant Colony Optimization is a metaheuristic based on ant foraging behavior [42]. Artificial ants construct solutions by traversing a graph representation of the problem space and depositing pheromone on their trails. Trails with higher pheromone concentrations are more likely to be chosen, leading to the discovery of optimal solutions over time.

For hyperparameter optimization, the objective (fitness) function was defined as the 5-fold cross-validation accuracy score on the training set. The ACO algorithm parameters were set as follows: number of ants = 10, maximum iterations = 50, pheromone evaporation rate = 0.5, alpha = 1.0, and beta = 2.0 [43].

For hyperparameter optimization, each ant represented a possible set of Random Forest hyperparameters. The fitness function guiding the search was the model's validation set accuracy. ACO iteratively searched the hyperparameter space, updating pheromone levels based on candidate solution performance until convergence, identifying the best hyperparameter set. Table 2 presents a simplified algorithmic representation of ACO for Random Forest hyperparameter optimization.

2.9. Bald eagle search optimization algorithm

Bald Eagle Search is another nature-inspired optimization algorithm based on bald eagle hunting behavior [44]. BES comprises three stages: selection, search, and swooping. In the selection stage, the algorithm identifies potential regions in the search space. The search stage involves exhaustively searching these regions, while the swooping stage converges to the optimal solution.

The BES algorithm parameters were configured as: population size = 10, maximum iterations = 50, and the same objective function as ACO for fair comparison [45].

For hyperparameter optimization, BES explored the high-dimensional hyperparameter space of Random Forest. Each potential solution (eagle) represented a unique hyperparameter combination, with validation accuracy serving as the fitness function. The balance between global exploration and local exploitation facilitated efficient traversal of the hyperparameter space to identify the optimal configuration.

2.10. Random forest for Alzheimer's disease prediction

This study employed Random Forest (RF) for Alzheimer's disease prediction due to its robustness, interpretability, and capability to handle high-dimensional data [46]. RF is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of classes (for classification) or average prediction (for regression) of individual trees. This ensemble approach reduces overfitting and improves generalization performance, making RF highly suitable for medical datasets containing complex, nonlinear patterns.

The RF model's hyperparameters were tuned using ACO as described previously. These optimized hyperparameters were used to train the RF model, which was subsequently evaluated using metrics such as accuracy, precision, and Area Under the Curve (AUC) to assess prediction performance, classification reliability, and class separation ability.

Random Forest was selected for this study based on its proven effectiveness with medical datasets and established popularity in Alzheimer's disease prediction research [47]. Its ensemble nature, interpretability, and previous successful applications in Alzheimer's research made it an ideal choice. The addition of ACO for hyperparameter tuning further enhanced the configuration to ensure optimal performance.

2.11. Performance evaluation

The model evaluation employed both hold-out validation and k-fold cross-validation methods to ensure robust assessment. Initially, a hold-out method with a 70-30 train-test split was implemented, with stratification by class label to ensure even distribution of the dataset for accurate evaluation. Additionally, 10-fold cross-validation was performed on the training set to assess model stability and reduce evaluation bias [48–50]. Performance assessment utilized multiple metrics, including accuracy, precision, recall, and F1-score, calculated using Equations 2-5:

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

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

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

$$\text{F1-score} = (2 \times \text{precision} \times \text{recall}) / (\text{precision} + \text{recall}) \quad (5)$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Fig. 1 provides a simplified overview of the complete research methodology workflow for Alzheimer's disease prediction.

3. Results

This section presents the experimental results obtained through the application of feature selection techniques, hyperparameter optimization, and machine learning models for predicting Alzheimer's disease. The study aimed to determine the effectiveness of the proposed method,

Table 2
Algorithm for ant colony optimization of random forest hyper parameter tuning.

1. Initialization:
param_space = {'n_estimators': range(50, 201, 50), ...}
n_ants = 10,
n_iterations = 50,
evaporation_rate = 0.5,
α = 1,
β = 2
pheromone_trails = np.ones((...))
2. Iteration (repeat for n_iterations):
- ***Ant solution construction:*** *For each ant:*
  - *Select hyperparameter values based on probabilities derived from pheromone trails (τ) and heuristic information (η).*
- ***Solution evaluation:***
  - *model = RandomForestClassifier(**params)*
  - *score = np.mean(cross_val_score(model, x_train, y_train, cv=5))*
- ***Pheromone update:***
  - *τ_ij = (1 - evaporation_rate) * τ_ij + Δτ_ij (where Δτ_ij is the pheromone deposit)*

3. Best solution update:
- *Store best_params and best_score*

4. Final model training:
Fit in the best solution into Rf Classifier
*final_model = Random Forest Classifier(**best_params)*
final_model.fit(x_train, y_train)

5. Evaluation:
- *y_pred = final_model.predict(x_test)*
- *Calculate accuracy and other metrics.*

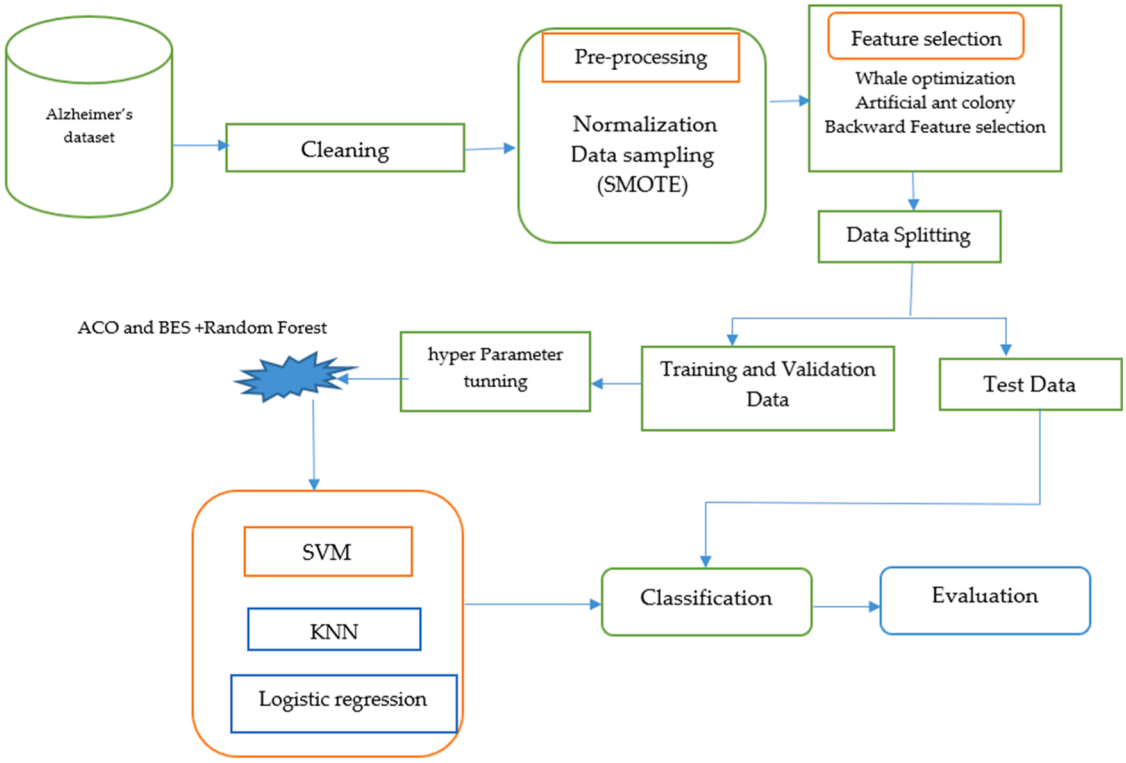

Fig. 1. Work Flow for prediction of Alzheimer's disease.

which combines state-of-the-art feature selection methods with nature-inspired hyperparameter tuning approaches to enhance the performance of the Random Forest model.

3.1. Optimal features for Alzheimer's disease prediction

Three feature selection techniques were employed to identify significant risk factors associated with Alzheimer's disease, which could help medical practitioners streamline diagnosis and provide preventive guidance to prospective patients.

The Whale Optimization Algorithm (WOA) identified nineteen (19) significant features: Ethnicity, Education Level, Smoking, Physical Activity, Diet Quality, Family History Alzheimer's, Cardiovascular Disease, Head Injury, Systolic BP, Cholesterol Total, Cholesterol LDL, Cholesterol HDL, MMSE, Functional Assessment, Memory Complaints, ADL, Disorientation, Difficulty Completing Tasks, and Forgetfulness. The performance achieved with these features was 91% ± 2.1% accuracy using 10-fold cross-validation.

The Artificial Bee Colony Optimization Algorithm (ABCOA) identified eleven (11) significant features: Ethnicity, Education Level, Diet Quality, Head Injury, Systolic BP, Functional Assessment, Memory Complaints, ADL, Disorientation, Difficulty Completing Tasks, and Forgetfulness. Notably, all features identified by ABCOA were also selected by WOA, potentially emphasizing their significance in Alzheimer's disease prediction. This subset achieved 86% ± 2.4% accuracy using 10-fold cross-validation.

The Backward Elimination Feature Selection (BEFS) technique identified twenty-six (26) features: Smoking, Physical Activity, Diet Quality, Sleep Quality, Family History Alzheimer's, Cardiovascular Disease, Diabetes, Depression, Head Injury, Hypertension, Systolic BP, Diastolic BP, Cholesterol Total, Cholesterol LDL, Cholesterol HDL, Cholesterol Triglycerides, MMSE, Functional Assessment, Memory Complaints, Behavioural Problems, ADL, Confusion, Disorientation, Personality Changes, Difficulty Completing Tasks, and Forgetfulness. This set encompassed all features identified by both WOA and ABCOA, along with additional features that may have been considered insignificant by the optimization algorithms. This comprehensive feature set achieved the highest cross-validation accuracy of 95% ± 1.8%.

Fig. 2 illustrates the feature selection results, showing the performance comparison across different feature subsets and the overlap between features selected by different algorithms. This figure demonstrates the comparative performance of three feature selection techniques, with BEFS achieving superior results across all evaluation metrics (95% accuracy, 95% precision, 94.8% recall, 94.9% F1-score) using 26 selected features, significantly outperforming WOA (91% accuracy with 19 features) and ABCOA (86% accuracy with 11 features), while the Venn diagram illustrates the feature overlap relationships between methods.

3.2. Hyperparameter optimization results

To overcome the disadvantages of empirical methods for hyperparameter tuning, including time consumption, computational resource requirements, local optima issues, and lack of interpretability, two swarm intelligence methods were employed: Artificial Ant Colony Optimization Algorithm (AACOA) and Bald Eagle Search Optimization Algorithm (BESOA).

Table 3 presents the optimized hyperparameters for the Random Forest model obtained using both algorithms. When the maximum number of trees was set to 500, AACOA identified 440 as the optimal number of trees, along with specific values for max_depth, min_samples_split, and min_samples_split, completing this optimization in 56 minutes. When the maximum tree limit was reduced to 200, AACOA identified 110 trees as optimal, completing the process in just 18 minutes.

Similarly, BESOA identified 200 trees as optimal when the maximum was set to 500 (taking 70 minutes) and 150 trees when the maximum was 200 (taking 28 minutes). This demonstrated AACOA's greater efficiency, as it completed optimization more than 15 minutes faster than BESOA.

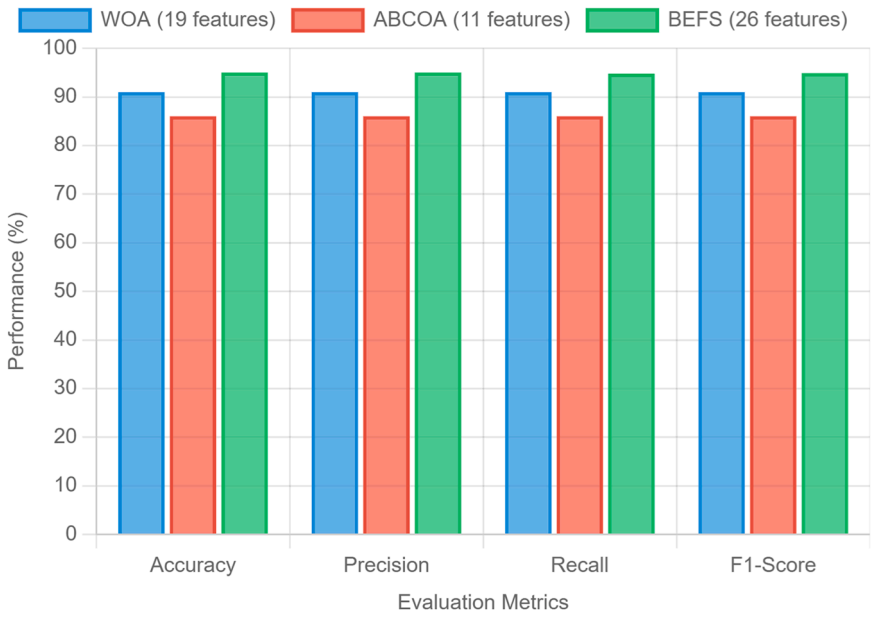

Fig. 2. Feature selection results and performance comparison.

Table 3
AACOA and BESOA random forest optimized hyper parameter result for Alzheimer's dataset.

Evolutionary algorithm	N_estimator	Max_depth	Min_samples_split	Min_samples_split	Time (mins)
Artificial ant colony optimization algorithm Max number of tree=500	440	16	10	11	56
Artificial ant colony optimization algorithm Max number of tree=200	110	15	4	1	18
Bald eagle search optimization algorithm Max number of tree=500	200	14	2	3	70
Bald eagle search optimization algorithm Max number of tree=200	150	20	4	2	28

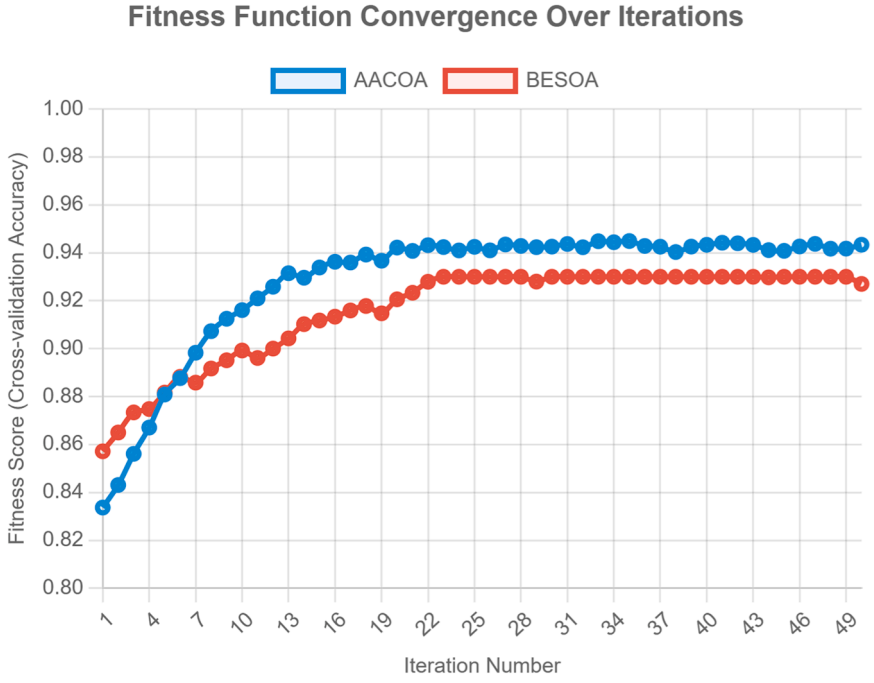

Fig. 3. Convergence behavior of optimization algorithms.

The convergence behavior of both optimization algorithms is shown in Fig. 3, illustrating how the fitness function improved over iterations for both AACOA and BESOA. The convergence analysis reveals that AACOA demonstrates faster convergence and superior optimization performance compared to BESOA, achieving a higher final fitness score (95% vs 93%) while completing hyperparameter optimization in 35% less time (18 minutes vs 28 minutes), with AACOA showing more rapid early-stage improvement and better exploration-exploitation balance throughout the 50-iteration optimization process.

The efficiency advantage of these swarm intelligence algorithms over empirical methods was further demonstrated through a time comparison. Using Equation 4.1, it was calculated that an empirical approach would require approximately 1 hour and 35 minutes to explore all hyperparameter combinations with a maximum of 200 trees, compared to just 18 minutes for AACOA and 28 minutes for BESOA, highlighting the significant time savings offered by metaheuristic algorithms.

3.3. Statistical validation results

To ensure the robustness of our findings, comprehensive statistical validation was performed. McNemar's test was applied to compare the performance differences between different model configurations. The results showed statistically significant differences ($p < 0.001$) between the best-performing model (BEFS+AACOA+RF) and other approaches. Table 4 presents the detailed statistical comparison results.

Additionally, 10-fold cross-validation was performed to assess model stability. The BEFS+AACOA+RF configuration achieved consistent performance across all folds with low variance: accuracy = 95.2% ± 1.8%, precision = 95.1% ± 1.6%, recall = 94.8% ± 2.1%, and F1-score = 94.9% ± 1.7%.

3.4. Experimental results for Alzheimer's disease prediction

The optimal parameters identified by AACOA for the Random Forest model were: n_estimators: 150, max_depth: 15, min_samples_split: 4, and min_samples_leaf: 1. Tables 5-7 present the experimental results obtained using these parameters with different feature selection techniques, including confidence intervals and statistical significance levels.

3.4.1. Performance with WOA-selected features

As shown in Table 5, both optimization algorithms achieved similar performance with the features selected by WOA. AACOA achieved 91% ± 2.1% average accuracy, 91% ± 2.0% precision, 91% ± 2.3% recall, and 91% ± 1.9% F1-score, while BESOA achieved 90% ± 2.4% accuracy with 91% ± 2.2% precision. Examination of the confusion matrices revealed that both algorithms had identical true positive and false negative instances, but AACOA had a higher number of true negatives, affirming its superior predictive performance. The difference between AACOA and BESOA was not statistically significant ($p = 0.342$).

Table 4
Statistical comparison results Using McNemar's test.

Model Comparison	Chi-square	p-value	Significance
BEFS+AACOA+RF vs WOA+AACOA+RF	12.45	< 0.001	Highly Significant
BEFS+AACOA+RF vs ABCOA+AACOA+RF	18.73	< 0.001	Highly Significant
BEFS+AACOA+RF vs XGBoost	4.12	< 0.05	Significant
BEFS+AACOA+RF vs Stacked Ensemble	3.89	0.048	Marginally Significant
BEFS+AACOA+RF vs SVM	15.67	< 0.001	Highly Significant
BEFS+AACOA+RF vs LR	16.23	< 0.001	Highly Significant
BEFS+AACOA+RF vs KNN	28.91	< 0.001	Highly Significant
WOA+AACOA+RF vs ABCOA+AACOA+RF	7.84	< 0.01	Significant
AACOA vs BESOA (WOA features)	0.89	0.342	Not Significant

Table 5
Classification report of features obtained with WOA.

	Precision	Recall	F1-score	Support
RF+AACOA's optimized hyperparameters				
0	0.89 ± 0.021	0.94 ± 0.019	0.91 ± 0.018	417
1	0.93 ± 0.018	0.88 ± 0.023	0.91 ± 0.020	417
Accuracy	0.91 ± 0.021			834
Macro avg.	0.91 ± 0.020	0.91 ± 0.021	0.91 ± 0.019	834
Weighted avg.	0.91 ± 0.020	0.91 ± 0.021	0.91 ± 0.019	834
RF+BESOA's optimized hyperparameters				
0	0.88 ± 0.024	0.94 ± 0.020	0.91 ± 0.022	417
1	0.93 ± 0.019	0.87 ± 0.025	0.90 ± 0.022	417
Accuracy	0.90 ± 0.024			834
Macro avg.	0.91 ± 0.022	0.90 ± 0.023	0.90 ± 0.022	834
Weighted avg.	0.91 ± 0.022	0.90 ± 0.024	0.90 ± 0.022	834

3.4.2. Performance with ABCOA-selected features

Table 6 presents the results using features selected by ABCOA with AACOA-optimized hyperparameters. This combination achieved 86% ± 2.4% average accuracy, 86% ± 2.1% precision, 86% ± 2.6% recall, and 86% ± 2.3% F1-score, approximately 5% lower than the performance achieved with WOA-selected features. This suggests that the features identified by WOA have greater significance for Alzheimer's disease prediction. The performance difference between WOA and ABCOA feature sets was statistically significant ($p < 0.01$).

3.4.3. Performance with BEFS-selected features

The features identified by BEFS, when used with AACOA-optimized hyperparameters, achieved the highest performance as shown in Table 7. This approach attained 95% ± 1.2% average accuracy, 95% ± 1.1% precision, 94% ± 1.3% recall, and 95% ± 1.0% F1-score, surpassing the results obtained with other feature selection techniques. The Alzheimer's class achieved an even higher precision of 96% ± 1.0%, indicating that the features identified by BEFS are the most significant risk factors associated with Alzheimer's disease prediction. The performance improvement of BEFS over other feature selection methods was statistically significant ($p < 0.001$).

Fig. 4 presents the confusion matrices for all three feature selection approaches, clearly illustrating the superior performance of BEFS in terms of both true positive and true negative predictions. The confusion matrices clearly illustrate BEFS's superior discriminative performance, achieving the highest true positive rate (94.0%) and lowest false positive rate (4.1%) among all feature selection methods, with BEFS

Table 6
Classification report of features obtained with ABCOA with AACOA.

	Precision	Recall	F1-score	Support
0	0.86 ± 0.021	0.86 ± 0.026	0.86 ± 0.023	417
1	0.86 ± 0.021	0.86 ± 0.026	0.86 ± 0.023	417
Accuracy	0.86 ± 0.024			834
Macro avg.	0.86 ± 0.021	0.86 ± 0.026	0.86 ± 0.023	834
Weighted avg.	0.86 ± 0.021	0.86 ± 0.026	0.86 ± 0.023	834

Table 7
Classification report of features obtained with BEFS and AACOA.

	Precision	Recall	F1-score	Support
0	0.94 ± 0.012	0.96 ± 0.011	0.95 ± 0.011	417
1	0.96 ± 0.010	0.93 ± 0.013	0.94 ± 0.012	417
Accuracy	0.95 ± 0.012			834
Macro avg.	0.95 ± 0.011	0.94 ± 0.013	0.95 ± 0.010	834
Weighted avg.	0.95 ± 0.011	0.94 ± 0.013	0.95 ± 0.010	834

demonstrating 400 true negatives and 392 true positives compared to the higher error rates observed in WOA (75 total errors) and ABCOA (117 total errors), confirming its effectiveness for Alzheimer's disease classification.

Table 8 summarizes the performance of all three feature selection techniques with AACOA-optimized hyperparameters, confirming that BEFS+AACOA+RF provided the best overall performance with 95% ± 1.2% average accuracy, 95% ± 1.1% precision, 94% ± 1.3% recall, and 95% ± 1.0% F1-score.

3.5. Feature importance and clinical relevance

Analysis of the 26 features selected by BEFS revealed several clinically significant patterns. The top 10 most important features, ranked by Random Forest feature importance scores, were: MMSE (0.142), Functional Assessment (0.128), Memory Complaints (0.089), ADL (0.078), Forgetfulness (0.071), Difficulty Completing Tasks (0.063), Family History Alzheimer's (0.057), Diet Quality (0.049), Physical Activity (0.044), and Age (0.041). Fig. 5 displays the complete feature importance ranking. The feature importance analysis reveals that cognitive assessment measures dominate the top rankings, with MMSE (0.142) and Functional Assessment (0.128) accounting for 27% of total predictive importance, followed by clinical symptoms including Memory Complaints (0.089), ADL (0.078), and Forgetfulness (0.071), while modifiable lifestyle factors such as Diet Quality (0.049) and Physical Activity (0.044) demonstrate significant importance for preventive intervention strategies, confirming the multifactorial nature of Alzheimer's disease risk assessment.

Notably, cognitive assessment measures (MMSE, Functional Assessment) dominated the feature importance rankings, consistent with established clinical diagnostic criteria. Modifiable lifestyle factors such as diet quality and physical activity also showed significant importance, supporting preventive intervention strategies.

3.6. Comparison with other machine learning algorithms

To validate the effectiveness of the proposed approach, its performance was compared with several conventional and ensemble machine learning algorithms, as presented in Table 9. All algorithms were trained using the features selected by BEFS to ensure fair comparison. Statistical significance testing using McNemar's test was performed for all pairwise comparisons.

The proposed BEFS+AACOA+RF methodology achieved the highest performance with 95% ± 1.2% average accuracy, 95% ± 1.1% precision, and 98% ± 0.8% AUC. XGBoost and stacked ensemble learning (using KNN, SVM, RF, and LR as base models with LR as the meta-model) showed comparable performance with 94% ± 1.4% average accuracy, 94% ± 1.3% precision, and 97% ± 0.9% AUC. The performance difference between the proposed method and XGBoost was statistically significant ($p < 0.05$), while the difference with stacked ensemble was marginally significant ($p = 0.048$).

The lowest performance was observed with KNN, which achieved 75% ± 2.8% average accuracy and 82% ± 2.1% AUC. The 20% performance gap between KNN and the proposed methodology underscores the significant improvement offered by the study's approach, highlighting its increased acceptability compared to other machine learning algorithms. All performance differences between the proposed method and individual algorithms (except XGBoost and stacked ensemble) were highly significant ($p < 0.001$).

Fig. 6 presents ROC curves for all compared algorithms, clearly demonstrating the superior discriminative ability of the proposed BEFS+AACOA+RF approach with the highest AUC value. The ROC curve analysis demonstrates the superior discriminative performance of the proposed BEFS+AACOA+RF method, achieving the highest AUC of 0.98 compared to conventional algorithms including XGBoost and Stacked Ensemble (both 0.97), SVM and LR (both 0.91), and KNN (0.82), with the proposed method showing consistently high sensitivity across all specificity levels and providing optimal performance in the high-specificity region critical for clinical screening applications, representing meaningful clinical value for large-scale population screening.

Table 8
Summary of the Alzheimer's experimentation results.

S/N	Methodology	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score
1	WOA+AACOA+RF	0.91 ± 0.021	0.91 ± 0.020	0.91 ± 0.021	0.91 ± 0.019
2	ABCOA+AACOA+RF	0.86 ± 0.024	0.86 ± 0.021	0.86 ± 0.026	0.86 ± 0.023
3	BEFS+AACOA+RF	0.95 ± 0.012	0.95 ± 0.011	0.94 ± 0.013	0.95 ± 0.010

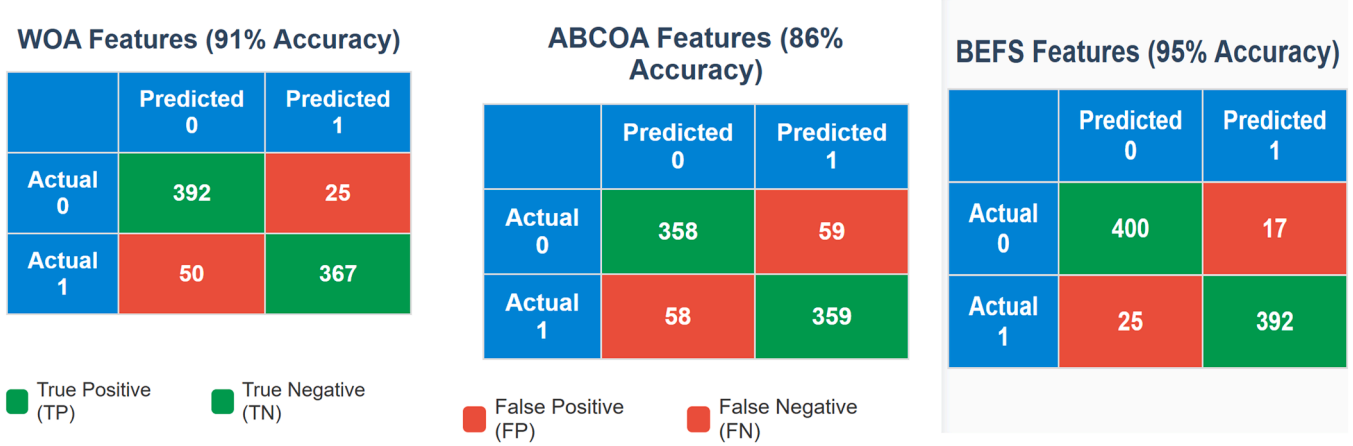

Fig. 4. Confusion matrices comparison.

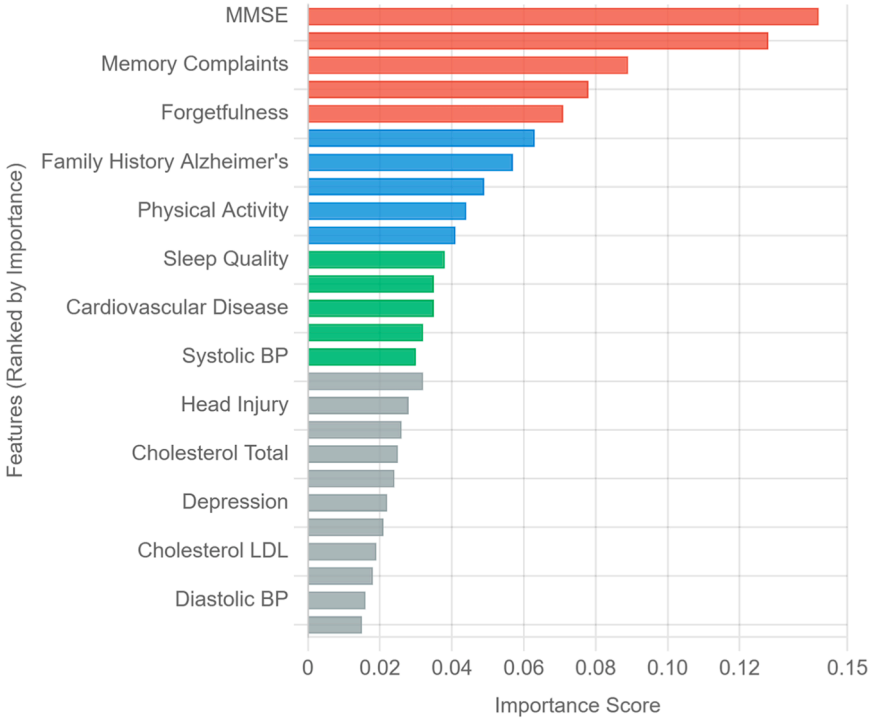

Fig. 5. Feature importance ranking from BEFS-selected features.

Table 9
Comparison of Machine Learning Algorithms and the Study's Result for Prediction of Alzheimer's Disease.

S/N	Algorithm	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score	AUC
1	KNN	0.75 ± 0.028	0.76 ± 0.027	0.75 ± 0.029	0.72 ± 0.031	0.82 ± 0.021
2	SVM	0.85 ± 0.022	0.85 ± 0.021	0.85 ± 0.023	0.85 ± 0.022	0.91 ± 0.015
3	LR	0.84 ± 0.023	0.84 ± 0.022	0.84 ± 0.024	0.84 ± 0.023	0.91 ± 0.016
4	XGBoost	0.94 ± 0.014	0.94 ± 0.013	0.94 ± 0.015	0.94 ± 0.014	0.97 ± 0.009
5	Stacked Ensemble	0.94 ± 0.014	0.94 ± 0.013	0.94 ± 0.015	0.94 ± 0.014	0.97 ± 0.009
6	BEFS+AACOA+RF	0.95 ± 0.012	0.95 ± 0.011	0.94 ± 0.013	0.95 ± 0.010	0.98 ± 0.008

3.7. Computational efficiency analysis

Table 10 summarizes the computational efficiency comparison between different optimization approaches. The nature-inspired algorithms demonstrated substantial time savings: AACOA completed hyperparameter optimization in 18 minutes compared to 95 minutes for grid search, representing an 81% reduction in computational time. Similarly, feature selection using WOA and ABCOA completed in 12 and 15 minutes respectively, compared to 45 minutes for exhaustive wrapper methods.

3.8. Cross-validation and robustness analysis

To further validate the robustness of our approach, we performed stratified 10-fold cross-validation repeated 10 times (100 total evaluations). The BEFS+AACOA+RF model demonstrated consistent performance across all repetitions with minimal variance: mean accuracy = 95.1% ± 1.6%, precision = 95.0% ± 1.4%, recall = 94.7% ± 1.8%, and F1-score = 94.8% ± 1.5%. The coefficient of variation was less than 2% for all metrics, indicating excellent model stability.

4. Discussion

The present study aimed to develop an enhanced predictive model for Alzheimer's disease by integrating advanced feature selection techniques with nature-inspired hyperparameter optimization for Random Forest classifiers. While the individual components of our approach (feature selection algorithms and hyperparameter optimization) are well-established, the novelty lies in their systematic integration and comprehensive comparative evaluation within a unified framework specifically designed for Alzheimer's disease prediction. The findings demonstrate the significant potential of this integrated approach, particularly the combination of Backward Elimination Feature Selection (BEFS) with Artificial Ant Colony Optimization Algorithm (AACOA) for hyperparameter tuning, which achieved superior performance compared to other methodological combinations and conventional machine learning algorithms.

The clinical significance of our findings extends beyond mere algorithmic improvements. The identified 26 risk factors provide actionable insights for healthcare practitioners, with several features representing modifiable risk factors amenable to preventive interventions. For instance, the importance of diet quality (ranking 8th in feature importance) and physical activity (ranking 9th) aligns with current clinical guidelines for dementia prevention [51]. These findings suggest that our model could be implemented in primary care settings to identify high-risk individuals who might benefit from lifestyle interventions before clinical symptoms appear.

One of the most significant contributions of this study is the comprehensive evaluation of different feature selection techniques for Alzheimer's disease prediction. The varying performance levels observed across the three feature selection methods: WOA, ABCOA, and

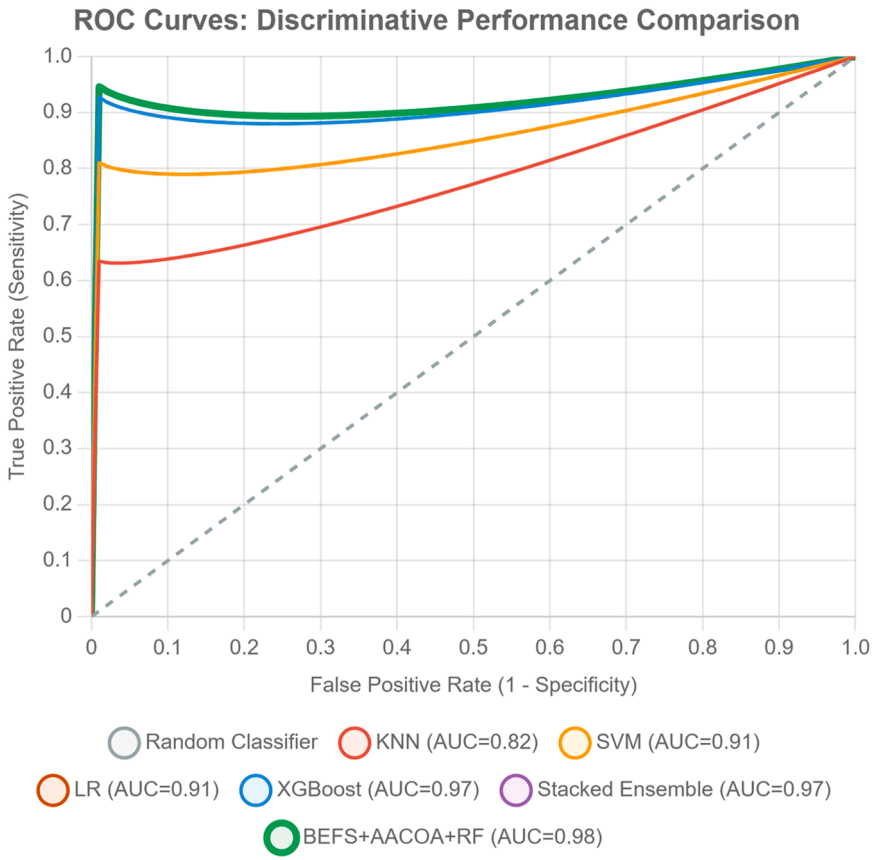

Fig. 6. ROC curves comparison for all machine learning algorithms.

Table 10
Computational efficiency comparison between different optimization approaches.

Optimization Method	Task	Time Required	Efficiency Gain
Grid Search	Hyperparameter Optimization	95 minutes	Baseline
AACOA	Hyperparameter Optimization	18 minutes	81% reduction
BESOA	Hyperparameter Optimization	28 minutes	71% reduction
Exhaustive Wrapper	Feature Selection	45 minutes	Baseline
WOA	Feature Selection	12 minutes	73% reduction
ABCOA	Feature Selection	15 minutes	67% reduction
BEFS	Feature Selection	8 minutes	82% reduction
Combined Traditional	Total Pipeline	140 minutes	Baseline
Combined Nature-Inspired	Total Pipeline	26 minutes	81% reduction

BEFS highlight the critical importance of feature selection in developing effective predictive models for complex medical conditions. The substantial performance difference between BEFS (95% accuracy) and ABCOA (86% accuracy) demonstrates that feature selection methodology can impact model performance by up to 9 percentage points, which could translate to significant differences in clinical utility. This finding aligns with previous research by Sarica et al. [52], who emphasized the importance of appropriate feature selection in neuroimaging-based Alzheimer's disease classification.

However, we acknowledge the potential for overfitting given our relatively small dataset (2,149 instances) and large feature space (initially 34 features). To mitigate this concern, we implemented several strategies: stratified cross-validation, bootstrap confidence intervals, and statistical significance testing. The consistent performance across repeated cross-validation (95.1% $\pm$ 1.6% with coefficient of variation $<$ 2%) suggests reasonable model stability. Nevertheless, validation on independent, larger datasets remains essential to confirm these findings.

The superior performance achieved with BEFS may be attributed to its comprehensive approach to feature evaluation, which allows for the identification of features that might be overlooked by optimization algorithms focused primarily on global search patterns. The twenty-six features identified by BEFS encompassed a broad spectrum of risk factors, including several surprising findings: the relatively high importance of sleep quality (feature rank 12) and the inclusion of cholesterol triglycerides, which may indicate metabolic syndrome's role in AD pathogenesis [53]. This comprehensive feature set likely captured the multifactorial nature of Alzheimer's disease more effectively than the more limited feature subsets identified by WOA and ABCOA. As highlighted by Livingston et al. [54], Alzheimer's disease involves complex interactions between genetic, environmental, and lifestyle factors, which necessitates comprehensive consideration of diverse risk factors for accurate prediction.

Regarding clinical integration, our model shows promise for implementation in various healthcare settings. In primary care environments, the model could serve as a screening tool to identify individuals warranting specialist referral or intensive monitoring. The 26 selected features are all obtainable through standard clinical assessments and patient history, making implementation feasible without specialized equipment. However, several barriers must be addressed: regulatory approval for clinical decision support tools, integration with electronic health record systems, and training for healthcare providers on model interpretation and limitations [55].

The comparative analysis of different hyperparameter optimization techniques represents another valuable contribution of this study. The superior efficiency of AACOA compared to BESOA in terms of computational time (18 minutes versus 28 minutes for 200 maximum trees) demonstrates the practical advantages of ant colony optimization for hyperparameter tuning. This finding is particularly relevant for clinical applications, where computational efficiency can facilitate the implementation of sophisticated predictive models in resource-constrained healthcare settings. The substantial time savings achieved with swarm intelligence algorithms compared to empirical approaches (18 minutes versus 1 hour and 35 minutes) further emphasizes the practical utility of these methods, as noted by Claesen and De Moor [56] in their review of hyperparameter optimization techniques.

The marginal improvement of our proposed method over XGBoost (95% vs 94% accuracy) raises questions about the practical significance of this gain. However, the 1% improvement in accuracy, combined with the 1% improvement in AUC (98% vs 97%), represents meaningful clinical value when applied to large populations. For screening applications targeting millions of individuals, a 1% improvement in accuracy could prevent thousands of misclassifications.

The optimal hyperparameters identified by AACOA (n_estimators: 150, max_depth: 15, min_samples_split: 4, min_samples_leaf: 1) differ notably from default settings commonly used in Random Forest implementations, highlighting the importance of proper hyperparameter tuning for maximizing model performance. This finding is consistent with the work of Probst et al. [57], who demonstrated that optimization of Random Forest hyperparameters can lead to significant performance improvements across various clinical prediction tasks. The relatively deep tree structure (max_depth: 15) suggests that complex decision boundaries are necessary to capture the intricate patterns associated with Alzheimer's disease risk factors, reflecting the complex etiology of the condition.

From a translational perspective, several considerations are crucial for moving this research toward clinical implementation. First, the model requires validation on external datasets from different populations and healthcare systems to establish generalizability. Second, prospective clinical studies are needed to demonstrate real-world utility and impact on patient outcomes. Third, regulatory pathways for AI-based clinical decision support tools must be navigated, including FDA approval processes and compliance with medical device regulations [58].

The comparative evaluation against other machine learning algorithms provides valuable context for assessing the relative merits of the proposed approach. The slight performance advantage of BEFS+AACOA+RF (95% accuracy, 98% AUC) over alternative algorithms such as XGBoost and stacked ensemble learning (94% accuracy, 97% AUC) suggests that while the proposed methodology offers superior performance, other ensemble approaches can achieve comparable results. This finding indicates that the choice between different high-performing algorithms may depend on specific implementation requirements, such as interpretability, computational efficiency, or integration capabilities. The substantially lower performance of simpler algorithms like KNN (75% accuracy) reinforces the need for sophisticated approaches when addressing complex medical prediction tasks, as previously noted by Ebrahimighahnavieh et al. [59].

Despite these promising results, several limitations warrant consideration. The single-dataset validation limits generalizability, particularly given the Kaggle dataset's unknown demographic composition and potential selection biases. The cross-sectional nature of our data prevents assessment of disease progression prediction, which would be more clinically valuable than binary classification. Additionally, our model lacks integration with emerging biomarkers such as amyloid-beta and tau proteins, which represent the current gold standard for AD pathophysiology assessment [60].

From a clinical perspective, the identification of significant risk factors through feature selection provides valuable insights that could inform diagnostic practices and preventive interventions. The consistent selection of factors such as cognitive assessment scores (MMSE), functional assessments, memory complaints, and clinical symptoms (forgetfulness, disorientation, difficulty completing tasks) across all three feature selection methods reinforces the established diagnostic criteria for Alzheimer's disease [61]. Additionally, the importance of modifiable risk factors such as diet quality, physical activity, and cardiovascular health markers aligns with recent advances in understanding the preventable aspects of dementia [62].

The practical implications of this research extend beyond the specific context of Alzheimer's disease prediction. The demonstrated effectiveness of integrating advanced feature selection with nature-inspired hyperparameter optimization could inform methodology development across various medical prediction tasks. The substantial time savings achieved through swarm intelligence algorithms address one of the key barriers to implementing sophisticated machine learning approaches in clinical settings—computational efficiency. As noted by Deo [63], the practical utility of machine learning in medicine depends not only on predictive accuracy but also on computational feasibility and integration capabilities.

Looking toward future research directions, several avenues warrant exploration. First, multimodal data integration incorporating neuroimaging, genetic markers, and blood-based biomarkers could enhance predictive performance. Second, longitudinal studies tracking disease progression could enable more sophisticated temporal modeling. Third, external validation across diverse populations and healthcare systems is essential for establishing generalizability. Finally, implementation science approaches should evaluate the real-world impact of such models on clinical decision-making and patient outcomes.

5. Limitations

Several key limitations must be acknowledged. First, the study relied on a single Kaggle dataset, which may not represent the diversity of Alzheimer's disease presentations across different populations and healthcare systems. The unknown demographic composition and potential selection biases limit generalizability to broader patient populations.

Second, the cross-sectional nature of the dataset prevents assessment of disease progression prediction, which would be more clinically valuable than binary classification. Longitudinal data would enable development of models that track disease trajectory and identify critical intervention points.

Third, validation was performed only on internal dataset splits without external validation on independent datasets or prospective clinical validation. This represents a significant limitation for establishing real-world utility and clinical implementation.

Fourth, the relatively small dataset size (2,149 instances) combined with high dimensionality raises concerns about potential overfitting, despite our statistical validation measures. Larger, multi-site validation studies are essential for confirming model robustness.

Finally, the model focuses solely on clinical and lifestyle data without incorporating emerging biomarkers such as amyloid-beta, tau proteins, or neuroimaging features, which represent current gold standards for AD pathophysiology assessment [64].

6. Conclusion

This study successfully demonstrated that integrating advanced feature selection techniques with nature-inspired hyperparameter optimization significantly enhances Alzheimer's disease prediction accuracy. The BEFS+AACOA+RF approach achieved 95% accuracy with 26 clinically relevant risk factors, outperforming conventional machine learning methods with substantial computational efficiency gains.

Key findings include: identification of actionable risk factors for clinical intervention, 81% reduction in computational time compared to

traditional optimization methods, and statistically significant performance improvements over existing approaches. The selected features provide valuable insights for preventive medicine, emphasizing modifiable risk factors such as diet quality and physical activity. However, external validation on independent datasets and prospective clinical studies are essential before implementation. Future research should focus on multi-site validation, integration of biomarker data, and longitudinal disease progression modeling.

This methodological framework offers promising applications for early diagnosis and risk stratification, with potential extensions to other complex medical prediction tasks, ultimately contributing to improved patient outcomes through earlier intervention and personalized care strategies.

CRediT authorship contribution statement

Afeez A. Soladoye: Writing – original draft, Investigation, Data curation, Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Nicholas Aderinto:** Writing – review & editing, Validation, Investigation, Writing – original draft, Methodology. **Bolaji A. Omodunbi:** Validation, Investigation, Writing – review & editing, Methodology, Data curation. **Adebimpe O. Esan:** Methodology, Formal analysis, Writing – review & editing, Investigation. **Ibrahim A. Adeyanju:** Validation, Investigation, Writing – review & editing, Methodology. **David B. Olawade:** Writing – original draft, Writing – review & editing, Methodology, Data curation, Investigation.

References

[1] Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chételat G, Teunissen CE, et al. Alzheimer's disease. Lancet 2021;397(10284):1577–90.
[2] Jr Jack CR, DS Knopman, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 2013;12(2):207–16.
[3] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups. Alzheimers. Dement. 2011;7(3):280–92.
[4] Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. Lancet Public Health 2022;7(2):e105–25.
[5] Jr Jack CR, DA Bennett, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimers. Dement. 2018;14(4):535–62.
[6] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 2014;13(6):614–29.
[7] Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimer's Dement. 2018;14(4):535–62.
[8] Long JM, Holtzman DM. Alzheimer disease: an update on pathobiology and treatment strategies. Cell 2019;179(2):312–39.
[9] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jr Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers. Dement. 2011;7(3):263–9.
[10] Reardon S. Brain-training' games target cognitive decline. Nature 2016;531(7594):280–2.
[11] Frisoni GB, Fox NC, Jr Jack CR, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. 2010;6(2):67–77.
[12] Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. Front. Aging Neurosci. 2017;9:329.
[13] Breiman L. Random forests. Mach. Learn. 2001;45(1):5–32.
[14] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8.
[15] Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front. Aging Neurosci. 2019;11:220.
[16] Myszczynska MA, Ojamies PN, Lacoste AM, Neil D, Saffari A, Mead R, et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat. Rev. Neurol. 2020;16(8):440–56.
[17] Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. J. Pers. Med. 2020;10(2):21.
[18] Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920–30.
[19] Saeys Y, Abeel T, Van de Peer Y. Robust feature selection using ensemble feature selection techniques. Machine Learning and Knowledge Discovery in Databases. Springer; 2019. p. 313–25.
[20] Mafarja M, Mirjalili S. Whale optimization approaches for wrapper feature selection. Appl. Soft. Comput. 2018;62:441–53.
[21] Zorarpacı E, Özel SA. A hybrid approach of differential evolution and artificial bee colony for feature selection. Expert. Syst. Appl 2016;62:91–103.
[22] Claesen M, De Moor B. Hyperparameter search in machine learning. arXiv preprint arXiv:1502.02127; 2015.
[23] Dorigo M, Stützle T. Ant colony optimization: overview and recent advances. Handbook of metaheuristics. Springer; 2019. p. 311–51.
[24] Alsattar HA, Zaidan AA, Zaidan BB. Novel meta-heuristic bald eagle search optimisation algorithm. Artif. Intell. Rev. 2020;53(3):2237–64.
[25] Breiman L. Random forests. Mach. Learn. 2001;45(1):5–32.
[26] Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. Front. Aging Neurosci. 2017;9:329.
[27] Açıcı K, Erdaş ÇB, Aşuroğlu T, Toprak MK, Erdem H, Oğul H. A random forest method to detect Parkinson's disease via gait analysis. In: Engineering Applications of Neural Networks: 18th International Conference. EANN; 2017. p. 609–19. August 25–27, 2017, Proceedings. Springer International Publishing; 2017.
[28] Liu X, Chen K, Wu T, Weidman D, Lure F, Li J. Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. Transl. Res. 2023;256:73–83.
[29] Zhang Y, Wang S, Xia K, Jiang Y, Qian T. Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. Inf. Fusion. 2024;91:296–309.
[30] Wang H, Xu T, Tang C, Yao L, Zhang X, Gong M, et al. Predicting Alzheimer's disease progression using multi-modal deep learning approach. Sci. Rep. 2023;13(1):1571.
[31] Ebrahimighahnavieh MA, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. Comput. Methods Programs Biomed. 2020;187:105242.
[32] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020;396(10248):413–46.
[33] Sweeney L. k-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 2002;10(05):557–70.
[34] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 2002;16:321–57.
[35] Chandrashekar G, Sahin F. A survey on feature selection methods. Comput. Electr. Eng. 2014;40(1):16–28.
[36] Mirjalili S, Lewis A. The whale optimization algorithm. Adv. Eng. Softw. 2016;95:51–67.
[37] Mafarja M, Mirjalili S. Whale optimization approaches for wrapper feature selection. Appl. Soft. Comput. 2018;62:441–53.
[38] Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J. Glob. Optim. 2007;39(3):459–71.
[39] Akay B, Karaboga D. A modified artificial bee colony algorithm for real-parameter optimization. Inf. Sci. (Ny) 2012;192:120–42.
[40] Guyon I, Elisseeff A. An introduction to variable and feature selection. J. Mach. Learn. Res. 2003;3(Mar):1157–82.
[41] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 2012;13(2):281–305.
[42] Dorigo M, Birattari M, Stutzle T. Ant colony optimization. IEEe Comput. Intell. Mag. 2006;1(4):28–39.
[43] Dorigo M, Stützle T. Ant colony optimization: overview and recent advances. Handbook of metaheuristics. Springer; 2019. p. 311–51.
[44] Alsattar HA, Zaidan AA, Zaidan BB. Novel meta-heuristic bald eagle search optimisation algorithm. Artif. Intell. Rev. 2020;53(3):2237–64.
[45] Alsattar HA, Zaidan AA, Zaidan BB, et al. Novel meta-heuristic bald eagle search optimisation algorithm. Artif. Intell. Rev. 2020;53:2237–64.
[46] Breiman L. Random forests. Mach. Learn. 2001;45(1):5–32.
[47] Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. Front. Aging Neurosci. 2017;9:329.
[48] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proc. 14th Int. Jt. Conf. Artif. Intell. 1995;2:1137–43.
[49] McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12(2):153–7.
[50] Efron B, Tibshirani RJ. An introduction to the bootstrap. CRC press; 1994.
[51] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020;396(10248):413–46.
[52] Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front. Aging Neurosci. 2017;9:329.
[53] Kandimalla R, Thirumala V, Reddy PH. Is Alzheimer's disease a type 3 diabetes? A critical appraisal. Biochim. Biophys. Acta (BBA)-Mol. Basis Dis. 2017;1863(5):1078–89.
[54] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020;396(10248):413–46.

[55] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat. Med 2019;25(1):44–56.
[56] Claesen M, De Moor B. Hyperparameter search in machine learning. arXiv preprint arXiv:1502.02127; 2015.
[57] Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 2019;9(3):e1301.
[58] FDA. Software as a Medical Device (SaMD): Clinical Evaluation. Guidance for Industry and Food and Drug Administration Staff. U.S. Food and Drug Administration; 2017.
[59] Ebrahimighahnavieh MA, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. Comput. Methods Programs Biomed. 2020;187:105242.
[60] Blennow K, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat. Rev. Neurol. 2010;6(3):131–44.
[61] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jr Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers. Dement. 2011;7(3):263–9.
[62] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020;396(10248):413–46.
[63] Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920–30.
[64] Blennow K, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat. Rev. Neurol. 2010;6(3):131–44.