International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://www.theaspd.com/ijes.php

Optimized Deep Feature Fusion For Alzheimer's Disease Detection Using AGWO And Pre-Trained Cnns

Sonali Bhosale[1] , Vaibhav V. Dixit[2]

[1]Department of Electronics and Teleommunication Engineering Sinhgad College of Engineering Smt. Kashibai Navale College of Engineering Affiliated to Savitribai Phule Pune University Pune, India deeptibhosale8611@gmail.com
Department of Electronics and Teleommunication Engineering Sinhgad College of Engineering RMD Sinhgad School of Engineering Affiliated to Savitribai Phule Pune University Pune, India vvdixit73@gmail.com

Abstract - *Timely intervention and therapy of Alzheimer's disease (AD) depend on an early and precise diagnosis. The sensitivity and specificity of traditional diagnostic methods are frequently compromised, particularly when used to various phases of cognitive decline. AD is a gradually advancing neurodegenerative condition that results in memory decline, cognitive difficulties, and shifts in behavior. Detecting AD at an early stage is crucial for slowing its progression and improving patients' quality of life. To address these challenges, deep learning and optimization-based methods offer significant potential in improving classification performance. To extract features, this study suggests a hybrid diagnostic framework that incorporates two pre-trained CNN models: VGG16 and InceptionV3, with Adaptive Grey Wolf Optimizer (AGWO) for optimal feature selection. The pipeline begins with data preprocessing techniques including augmentation, normalization, filtering, and scaling. Features extracted from both pre-trained models are fused and refined using AGWO to enhance discriminative power. The optimized features are then classified using an improved Multilayer Perceptron (MLP) to categorize subjects into Alzheimer's Disease stages. The proposed VGG16+InceptionV3+AGWO hybrid framework attained a high level of categorization accuracy of 98.62% on the test dataset, demonstrating its efficiency in identifying different stages of Alzheimer's Disease with remarkable precision. The integration of multiple pre-trained models with AGWO-based feature selection significantly enhances the performance of Alzheimer's Disease classification. The results validate the potential of the proposed framework as a robust tool for clinical decision support, encouraging further exploration and application in real-world diagnostic settings.*

Keywords: *Alzheimer's Disease, Feature Selection, Deep Learning, Hybrid Whale Optimization, MRI Classification.*

Keywords: *Alzheimer's Disease, Feature Selection and Extraction, Deep Learning, AGWO, IMLP.*

I. INTRODUCTION

A progressive neurological disease, Alzheimer's disease (AD) is typified by behavioral abnormalities, cognitive impairment, and memory loss. It is the most frequent cause of dementia, making up 60–80% of cases globally, and as the population ages, its prevalence is predicted to triple by 2050 [1]. Early AD diagnosis can greatly slow the progression of the disease and enhance patients' quality of life, especially when the patient is in the Mild Cognitive Impairment (MCI) stage [2]. Still, traditional diagnostic approaches such as cognitive assessments and clinical interviews often lack sensitivity and consistency, making them insufficient for early detection [3].

Researchers have looked to artificial intelligence (AI), in particular deep learning, for automatic and precise classification of AD using neuroimaging data to overcome the shortcomings of traditional diagnostic methods. Convolutional Neural Networks (CNNs) have shown remarkable success in extracting deep features from medical images, enabling automated diagnosis [4]. Yet, a single model may not fully capture complex structural changes in the brain. Ensemble and hybrid approaches that combine

multiple CNN architectures can offer complementary feature representations and improve generalization [5].

Furthermore, medical image data are often high-dimensional, and not all extracted features contribute equally to classification performance. Feature selection becomes crucial to reduce redundancy and improve model efficiency. Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Ant Colony Optimization (ACO) are examples of metaheuristic optimization algorithms that have been used for feature selection; nevertheless, they frequently have trouble with local optima or need optimizing several parameters [6], [7]. The Grey Wolf Optimizer, on the other hand, provides a better balance between exploration and exploitation, fewer parameters, and better convergence in complex search spaces. It is modeled after the hunting behavior and leadership structure of grey wolves [8]. In this study, an enhanced Multilayer Perceptron (MLP) was employed as the final classification layer to differentiate between AD stages. Despite their strength, traditional MLP architectures may overfit when handling high-dimensional medical imaging information. To address this, the enhanced MLP in our framework was optimized with dropout regularization, adaptive learning rates, and ReLU activation functions to improve generalization and convergence speed. Additionally, the use of AGWO-refined feature sets significantly reduced the input dimensionality, allowing the MLP to focus on the most informative patterns.

For the diagnosis of AD, several imaging techniques have been used, such as magnetic resonance imaging (MRI), computed tomography (CT), and Positron Emission Tomography (PET). While PET is useful for detecting amyloid plaques and metabolic changes, and CT is accessible for structural imaging, because of its high-resolution anatomical information and non-invasive nature, MRI is frequently used [9]. T1-weighted MRI is useful for observing structural shrinkage in important brain areas such as the cortex and hippocampus, which are crucial for the advancement of AD [10].

II. RELATED WORKS

Sudharsan et al. [11] proposed an early diagnosis method for AD using structural MRI and machine learning based models such as RELM, SVM, and IVM, focusing on distinguishing MCI from AD and healthy controls. Their approach uses a kernel-based transformation and a greedy score-based method to select key features, showing improved accuracy with RELM on the ADNI dataset. However, challenges like overfitting, limited interpretability, and the requirement for large, high-quality datasets are still restricted of such machine learning-based methods.

Nguyen et al. [12] proposed an ensemble learning method that combines XGBoost and a 3D-ResNet model to diagnose Alzheimer's disease using brain MRI examinations from the ADNI dataset. The method efficiently extracts relevant voxel features and demonstrates improved speed and performance, achieving a high AUC within 10 minutes. The study emphasizes the importance of proper test set generation to avoid data leakage and ensure result validity. However, the approach faces challenges such as high computational demands, potential overfitting, and the need for large, diverse datasets to effectively combine traditional and deep learning models.

A deep learning-based technique called AD-DL was presented by Sorour et al. [13] for the early identification of Alzheimer's disease using brain MRI data. The approach includes preprocessing, training, and evaluation phases, containing five DL models divided into two groups: those with and without data augmentation. These include CNN without augmentation, and with augmentation models such as CNNsLSTM with augmentation and VGG16-SVM with augmentation. The goal of the study is to maximize computing efficiency, F1-score, detection accuracy, recall, and precision. While the results are promising, limitations include dataset biases, performance variability from data augmentation strategies, and the high computational demands required for model training and tuning.

A deep learning-based ensemble method for earlier Alzheimer's disease diagnosis utilizing MRI scans was proposed by Fathi et al. [14]. The procedure entails gathering data, preprocessing it, creating

separate CNN models, and then combining the top six classifiers assessed on ADNI and a local dataset to create an ensemble. This ensemble strategy showed superior performance compared to many existing methods. However, it faces challenges such as high computational requirements, risk of overfitting due to model complexity, and the need for large and diverse datasets to ensure robust and generalizable outcomes.

To diagnose Alzheimer's disease early, Balaji et al. [15] developed a hybrid deep learning model that combines Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks using multimodal data, including MRI, PET, and neuropsychological test results. The model employs Adam optimization to enhance learning efficiency and demonstrates high precision in distinguishing cognitively normal individuals from those with early MCI. This demonstrates how deep networks may be used to find imaging biomarkers. However, The method has drawbacks, including high processing requirements and the requirement for a variety of datasets, and complexities in effectively integrating multiple data types and algorithms.

The growing demand for accurate and timely Alzheimer's disease diagnosis has spurred extensive research into computer-aided diagnostic systems, particularly those based on deep learning and feature optimization. Traditional clinical methods such as neuropsychological assessments and interviews often fail to detect subtle brain abnormalities during the initial stages of cognitive decline, leading researchers to explore neuroimaging and machine learning techniques for improved sensitivity and specificity [16].

Because Convolutional Neural Networks (CNNs) can extract high-level, discriminative characteristics from complex data, they have demonstrated significant promise in the categorization of medical images. Interestingly, transfer learning with pre-trained models such as VGG16 and InceptionV3 has proven successful in obtaining strong features from Alzheimer's categorization tasks using MRI data [17][18]. These models offer deep hierarchical feature representations that can capture subtle structural variations across persons with AD, mild cognitive impairment (MCI), and cognitively normal (CN).

However, relying solely on a single deep learning model may limit performance due to overfitting or limited generalization. Therefore, hybrid and ensemble methods that combine multiple CNN architectures have gained traction in recent years. For instance, Islam et al. [19] projected a CNNs based ensemble for Alzheimer's disease classification using MRI, which showed improved accuracy compared to individual models. Such architectures leverage complementary strengths of different networks, resulting in better feature diversity and diagnostic precision.

The high dimensionality of medical image features, however, presents a challenge to downstream classifiers. Redundant or irrelevant features can degrade model performance, especially in small datasets. To tackle this, metaheuristic optimization algorithms have been employed for feature selection. In this regard, methods such as Ant Colony Optimization (ACO) [22], Genetic Algorithm (GA) [21], and Particle Swarm Optimization (PSO) [20] have been investigated. While effective to some extent, these algorithms can be computationally expensive and susceptible to local minima, especially when searching in large feature spaces.

More recently, the Adaptive Grey Wolf Optimizer (AGWO) has emerged as a robust feature selection method, influenced by gray wolf hunting skills and leadership structures. It improves the standard Grey Wolf Optimizer algorithm by incorporating adaptive mechanisms for balancing exploration and exploitation during the search process [23]. When applied to biomedical data, studies have demonstrated that AGWO performs better than other optimization algorithms in terms of classification accuracy, global search capability, and convergence speed [24].

In terms of classifiers, Multilayer Perceptrons (MLPs) have been widely used in medical diagnostics due to their universal approximation capability. However, traditional MLPs may overfit when processing high-

dimensional feature vectors from neuroimaging. Therefore, researchers have integrated regularization techniques, such as dropout and adaptive learning rates, along with ReLU activation functions to enhance generalization [25]. In combination with effective feature selection, an optimized MLP can serve as a lightweight and accurate classifier for multiclass AD diagnosis.

The integration of deep feature fusion with AGWO-based optimization and an improved MLP classifier represents a potential approach to raising the sensitivity and specificity of AD diagnostic models. Similar hybrid frameworks have been applied successfully in other medical imaging domains, but there remains a research gap in their systematic application to multi-stage Alzheimer's disease classification, particularly using T1 weighted MRI data.

III. PROPOSED METHODOLOGY

The Pre-trained Hybrid AGWO-MLP Framework for Alzheimer's disease classification follows a structured pipeline demonstrated in figure 1 to enhance diagnostic accuracy.

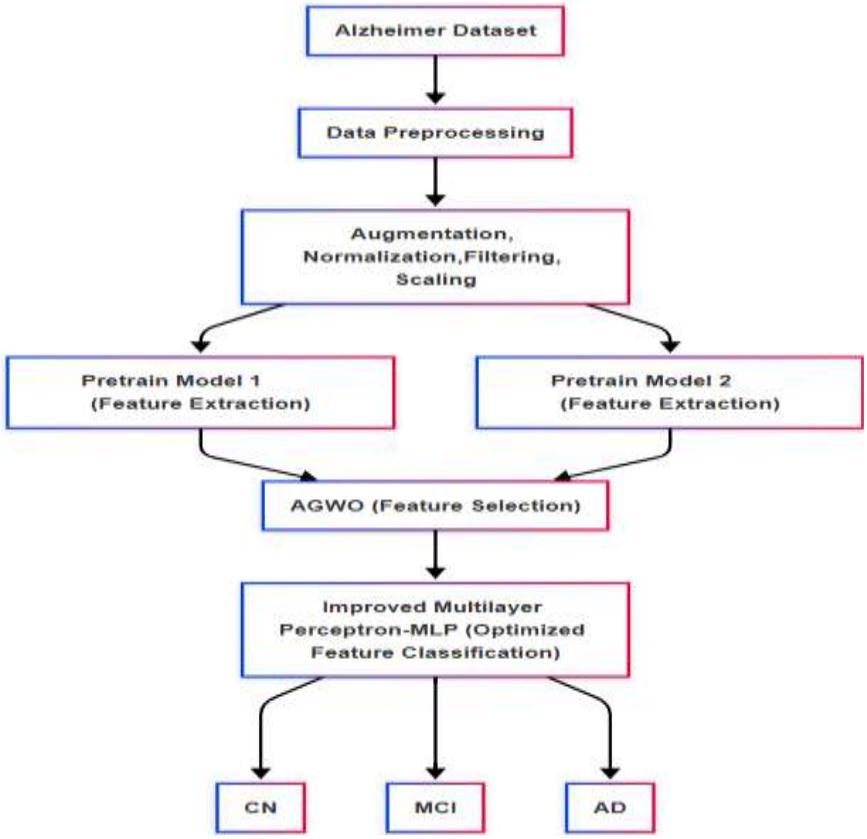


Figure.1 AGWO-MLP Framework for Alzheimer's Disease Detection

The proposed methodology for Alzheimer's disease classification is a hybrid deep learning and optimization-based framework designed to accurately classify subjects into five stages: Alzheimer's Disease (AD), Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Mild Cognitive Impairment (MCI). As illustrated in the flowchart, the process begins with an Alzheimer's MRI dataset that undergoes comprehensive data preprocessing, including skull stripping, resizing, filtering, and normalization. To further enhance model robustness, preprocessing includes data augmentation, intensity normalization, and feature scaling.

For feature extraction, two powerful pretrained convolutional neural networks—VGG16 and InceptionV3—are utilized. VGG16, known for its deep but simple structure with sequential convolutional layers, extracts detailed spatial texture features. InceptionV3, on the other hand, captures multi-scale and

hierarchical patterns using its advanced inception modules. These diverse features are then concatenated to form a rich feature representation.
To refine these features, the Adaptive Grey Wolf Optimizer (AGWO) is employed for feature selection. AGWO mimics the leadership hierarchy and hunting strategy of grey wolves to eliminate redundant and irrelevant features, thereby improving the computational efficiency and classification performance.
The optimized feature set is then fed into an Improved Multilayer Perceptron (MLP), which is designed with enhancements such as batch normalization and dropout layers for improved generalization. The final MLP classifier performs multi-class classification and accurately categorizes the input images into the five stages: AD, CN, EMCI, LMCI, and MCI.
This integrated pipeline—combining dual deep feature extractors, intelligent feature selection, and a robust classifier—offers a highly efficient and accurate solution for multi-stage Alzheimer's disease diagnosis. Such approaches are in line with recent advances in deep learning-based medical diagnostics and align with best practices observed in literature [4].

IV. IMPLEMENTATION

DATASET

The Alzheimer's Disease Neuroimaging Initiative (ADNI), a significant longitudinal, multicenter study focused on creating clinical, imaging, genetic, and biochemical indicators for the early identification and progression of Alzheimer's disease, served as the basis for this dataset, which was gathered from Kaggle. The collection consists of structural MRI pictures from both **male and female subjects**, with the majority of participants being aged **60 years and above**, reflecting the typical age range affected by neurodegenerative disorders. ADNI provides standardized and high-quality data that has been extensively validated and used in numerous medical AI research studies.

As shown in figure 2 Dataset Sample Brain MRI Images which contain 11,866 MRI scans in total, divided into five classes, Alzheimer's Disease (AD), Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Mild Cognitive Impairment (MCI)—make up the dataset used in this work. Many different stages and situations associated with the diagnosis of Alzheimer's disease are represented by these classes. The images are distributed across class-specific directories and are separated into training and testing subsets, approximately 80% for training and 20% for testing of the dataset allocated. Specifically, the AD class contains 2,420 images (1,936 for training and 484 for testing), CN has 2,123 images (1,698 training, 424 testing), EMCI includes 1,920 images (1,536 training, 384 testing), LMCI holds 2,637 images (2,109 training, 527 testing), and MCI comprises 2,766 images (2,212 training, 553 testing). Although the dataset is relatively balanced, some variation exists, with MCI having the highest image count and EMCI the lowest. To validate the integrity and consistency of the dataset, sample images from each class were visualized, and a distribution plot was generated, aiding in understanding class representation and identifying potential class imbalances.

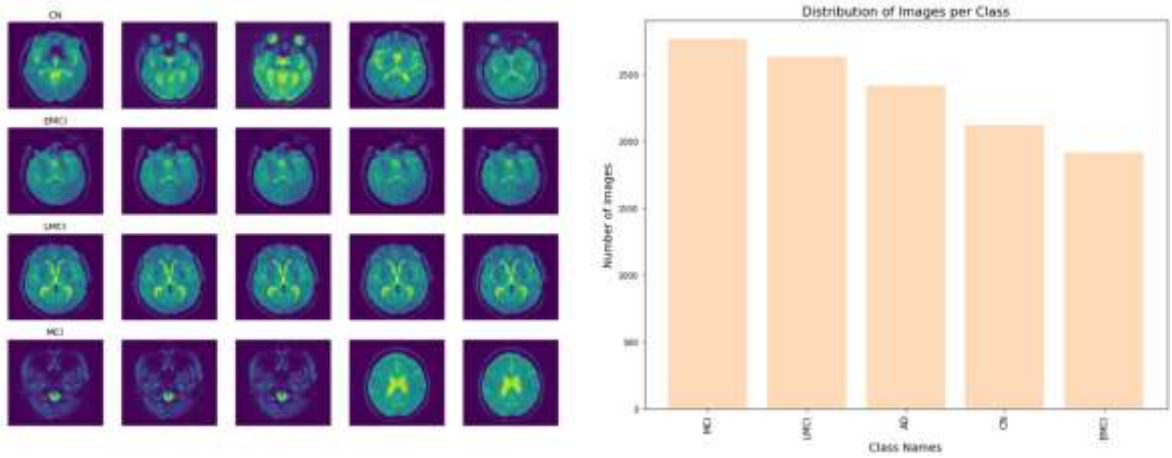


Figure. 2. Dataset Exploration: Sample Brain MRI Images and Class Distribution in Alzheimer's Disease Dataset

FEATURE EXTRACTION USING PRETRAINED MODELS

Two powerful pre-trained CNN architectures VGG16 and InceptionV3 are employed in parallel for feature extraction. The Alzheimer's dataset is used to refine these models, which were initially trained on ImageNet, in order to incorporate particular features.

The University of Oxford's Visual Geometry Group presented the VGG16 deep convolutional neural network architecture. It consists of 16 weight layers arranged uniformly, containing 3 fully connected layers and 13 convolutional layers. Just 3x3 convolutional kernels with stride 1 and padding 1 are used in the model. ReLU activations and 2x2 max pooling layers are then used for downsampling. VGG16 stacks multiple convolutional layers before each pooling operation, enabling the network to pick up ever-more intricate feature representations. The simplicity and depth of the architecture enable it to extract fine-grained and spatially rich hierarchical features, which are well-suited for medical imaging tasks. In recent research, VGG16 has attained a classification accuracy of 98.18% on a lung cancer CT image dataset, outperforming other models for instance ResNet50 and InceptionV3 On training and testing performance [26]. Within the suggested framework, to maintain the spatial representations crucial for categorizing Alzheimer's disease stages, features are taken from the final convolutional block (prior to all the connected layers) of the pre-trained VGG16 model, which is subsequently improved using the Alzheimer's MRI dataset.

InceptionV3, developed by Google, is a deeper and more efficient convolutional architecture that enhances feature extraction through its inception modules. Each module uses filters of different sizes (1×1, 3×3, and 5×5) to execute simultaneous convolutions. The resulting feature maps are then concatenated. The model can capture local as well as global features in the same layer because to this approach. The network also incorporates dimensionality reduction through 1×1 convolutions and employs auxiliary classifiers to improve gradient flow and reduce overfitting. A recent study demonstrated the effectiveness of InceptionV3 in an ensemble framework for lung cancer detection, where it contributed to improved classification performance due to its multi-scale feature extraction capabilities [27]. In the current methodology, the fine-tuned InceptionV3 extracts complementary features from the Alzheimer's MRI dataset, which are then concatenated with features from VGG16 to create a comprehensive hybrid feature vector for further classification.

FEATURE SELECTION

Principal Component Analysis (PCA) is often used statistical technique for dimensionality reduction. In order to capture the most variance in the data, it reduces the initial set of correlated characteristics

into a smaller set of uncorrelated components called principal components. PCA efficiently lowers the dimensionality while maintaining the most important information by projecting the data onto these major components. In situations where high-dimensional data may result in overfitting or increased computing complexity, this approach is especially helpful. Current studies have exhibited The effectiveness of PCA in preprocessing steps for various machine learning tasks, including medical image analysis and bioinformatics applications [28].

$$Z_{PCA} = Z \,.\, W$$

Where Z: $Standardized\ input\ data\ matrix$,
W: $Matrix\ of\ topk\ eigenvectors\ (Principal\ components)$ and
Z_{PCA}: $Transformed\ data\ in\ reduced\ dimension$.

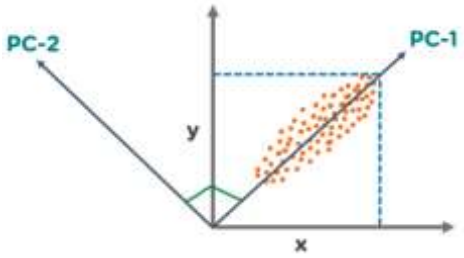


Figure. 3 Reduces dimensionality by projecting data onto principal components that capture the most variance.

In the given figure 3, there are multiple points marked on a two-dimensional plane. There are two main parts. PC1 is the main principal component that captures the highest amount of variation in the data. PC2 is a different principal component that is perpendicular to PC1.

Whale Optimization Algorithm (WOA) is inspired by nature metaheuristic algorithm that mimics the social conduct of humpback whales, particularly their bubble-net hunting tactic. WOA has been effectively applied to feature selection difficulties because of its balance between exploration and exploitation capabilities. In a recent study, an improved version of WOA was proposed for choosing feature in Internet of Things (IoT) applications. This enhanced WOA incorporated a disordered Hénon map mechanism and an adaptable coefficient vector to improve convergence speed and avoid local optima. The suggested technique achieved best performance regarding classification accuracy and dimensionality reduction in contrast to traditional feature selection algorithms [29].

A whale optimization algorithm, depicted in figure 4, was directly inspired by the distinctive bubble-net feeding strategy employed by humpback whales for hunting. It is a hunting technique specific to this species that involves diving beneath a school of tiny fish or krill and then making and releasing air bubbles as it approaches the surface in a spiral pattern around the shoal. Figure 1 shows a representation of the bubble-net feeding hunting technique.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://www.theaspd.com/ijes.php

Figure 4. Visualization of bubble-net feeding hunting method. The humpback swims in a spiral below the prey, creating bubbles that trap them.

Simulates the **bubble-net** strategy to balance global search and local refinement shown in Equation

$$\vec{X}(t+1) = \vec{X}^* - \vec{A} \cdot \left|\vec{C} \cdot \vec{X}^* - \vec{X}(t)\right|$$

Where $\vec{X}(t+1)$: $Updated\ whale\ position$, $\vec{X}^*$: Position of best solution (prey), $\vec{X}(t)$: Current whale position, $\vec{A} = 2a \cdot \vec{r} - a$: Adaptive coefficient for exploitation/exploration, $\vec{C} = 2 \cdot \vec{r}$: Random weight vector, $\vec{r}$: $Random\ vector\ in$ $[0,1]$ and a: $Linearly\ decreases\ from$ 2 to 0

Grey Wolf Optimizer (GWO) is a bio-inspired optimization algorithm that mimics the hunting mechanism and leadership hierarchy of grey wolves, as illustrated in Figure 5, which shows the schematic diagram of the gray wolf population hierarchy and predation processes. GWO has gained popularity in tasks involving feature selection because of its simplicity besides effectiveness in navigating complex search spaces. A new feature selection framework according to GWO was developed for mammogram image analysis, where the most essential elements were chosen by combining it with basic set theory. The hybrid approach demonstrated improved classification accuracy and reduced feature subset size, highlighting GWO's potential in medical image processing applications [30].

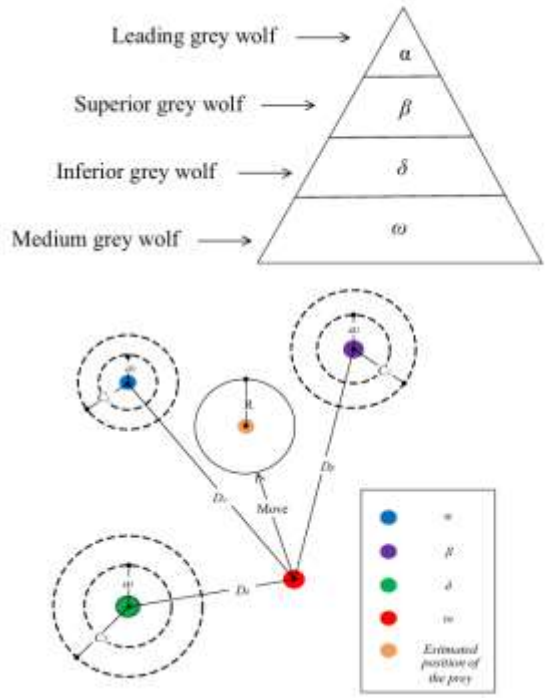


Figure 5. Schematic diagram of gray wolf population hierarchy and predation processes.

$$\vec{X}(t+1) = \frac{\vec{X}_\alpha + \vec{X}_\beta + \vec{X}_\delta}{3} - \frac{\vec{A}_1 \cdot \vec{D}_\alpha + \vec{A}_2 \cdot \vec{D}_\beta + \vec{A}_3 \cdot \vec{D}_\delta}{3}$$

Where $\vec{X}(t+1)$: $New\ position\ of\ a\ search\ agent$, $\vec{X}_\alpha, \vec{X}_\beta, \vec{X}_\delta$: Positions of top 3 wolves (solutions), $\vec{A}_i = 2a \cdot \vec{r} - a$: $Adaptive\ coefficient$, $\vec{D}_\alpha = \left|\vec{C}_1 \cdot X_\alpha - \vec{X}\right|$: distance to leader, $\vec{c}_i = 2 \cdot \vec{r}_i$: Random coefficient and a: $Decreasing\ parameter\ from$ 2 to 0. **Purpose**: Updates the position based on leaders, simulating **hunting habits** of grey wolves.

Adaptive Grey Wolf Optimizer (AGWO) is an advanced variant of the standard GWO that introduces adaptive mechanisms to enhance optimization performance. AGWO dynamically adjusts its exploration and exploitation parameters according to the candidate solutions' fitness history, allowing for more efficient convergence. This adaptability addresses the limitations of fixed parameter settings in traditional GWO, leading to better performance in feature selection tasks. A study introduced AGWO for feature selection, demonstrating its superiority over conventional GWO regarding the quality of the solution and the rate of convergence across various benchmark datasets [31].

FITNESS FUNCTION

The key idea behind this feature selection technique lies in its **fitness function**, which balances classification error and the amount of chosen features:

$$Cost = \alpha \times Error\ Rate + \beta \times \left(\frac{Number\ of\ Selected\ Features}{Total\ Features}\right)$$

Where $\alpha = 0.99, \beta = 0.01$, Error Rate is computed using K-Nearest Neighbors (KNN) classifier via cross-validation.

GWO POSITION UPDATE EQUATIONS

Each grey wolf updates its position based on the top three options found so far Alpha (X_α), Beta (X_β) and Delta (X_δ). For each dimension d in a candidate Solution:

Compute distance vectors:

$$D_\alpha = |C_1 \cdot X_\alpha - X_i|, \quad D_\beta = |C_2 \cdot X_\beta - X_i|, \quad \mathrm{D_\delta = |C_3 \cdot X_\delta - X_i|}$$

Updated Candidate Positions:

$$\mathrm{X_1 = X_\alpha - A_1 \cdot D_\alpha, \quad X_2 = X_\beta - A_2 \cdot D_\beta, \quad X_3 = X_\delta - A_3 \cdot D_\delta}$$

Final updated positions:

$$\mathrm{X_i = \frac{X_1 + X_2 + X_3}{3}}$$

$\mathrm{A = 2a \cdot r - a}$, where a linearly decreases from 2 to 0 during iterations.

$\mathrm{C = 2r}$, where r is a random vector in [0,1].

- **BINARY CONVERSION**

After the continuous update, the positions are binarized to select features using a threshold (typically 0.5):

$$\mathrm{X_{i,d}^{bin} = \begin{cases} 1, \text{if } X_{i,d} > \text{threshold} \\ 0, \text{otherwise} \end{cases}}$$

FITNESS EVALUATION AND FEATURE SUBSET EXTRACTION

The fitness of every binary mask is assessed by the above cost function. Best positions of top 3 wolves are updated according to the fitness. At the end of iterations, the final binary vector of the alpha wolf gives the optimal feature subset. The classification performance and feature reduction are summarized.

5. IMPLEMENTATION

5.1 Algorithm of AGWO

Standard GWO is suited for continuous optimization tasks, while Adapted GWO is tailored for feature selection in machine learning, ensuring feature subset validity and optimizing classification performance.

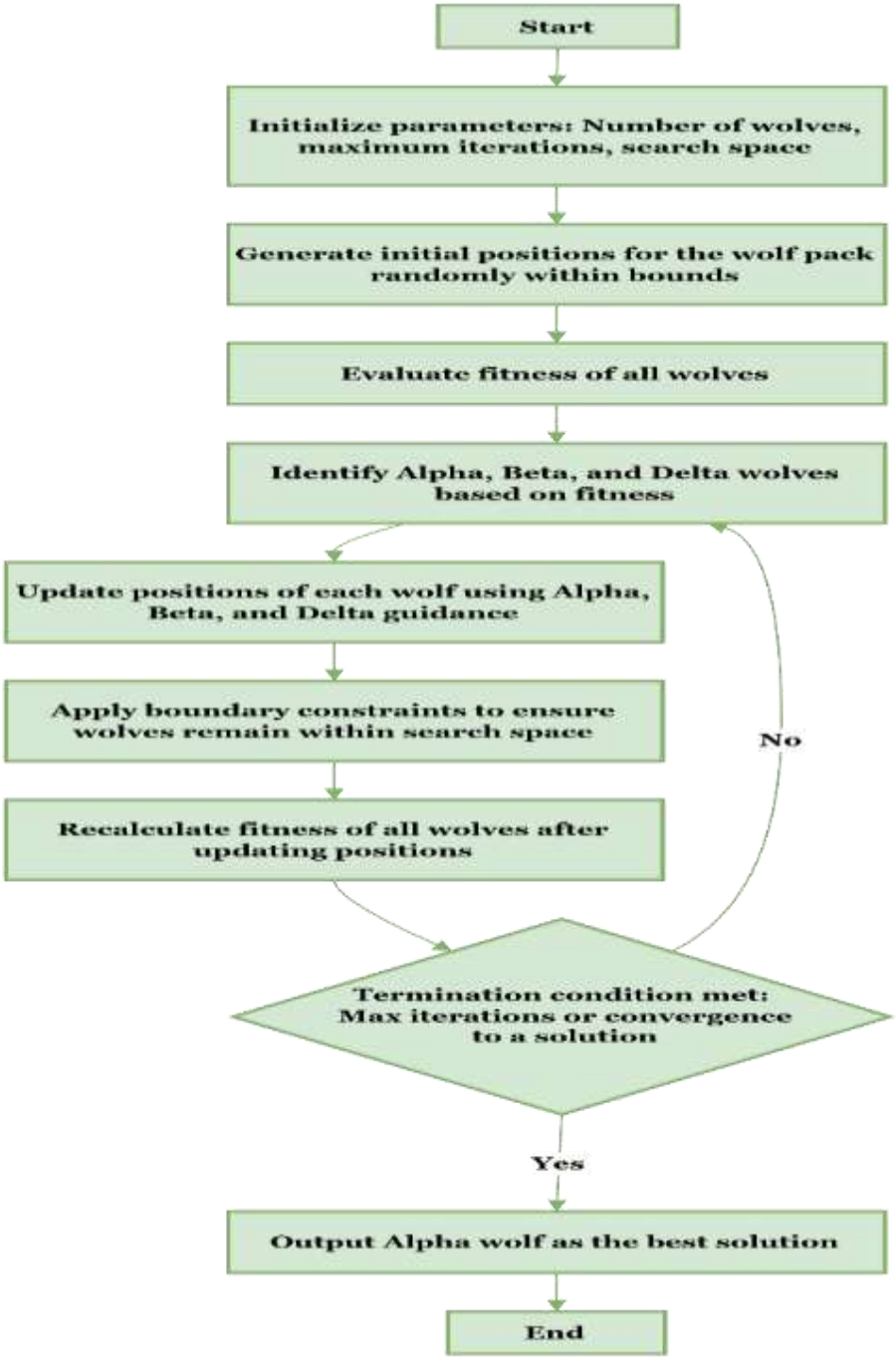


Figure 6. Flowchart of Adapted Grey Wolf Optimizer

AGWO optimizes the feature subset for improved classification performance by following a set of clearly defined stages, as shown in the segment of Figure 6. Initializing parameters, such as the maximum number of iterations, the number of wolves (population size), and randomly initialized feature subsets, is the first step in the process. Every wolf in the population is a possible remedy, or a distinct subset of characteristics.

The wolves' continuous position vectors are converted into binary form, where "1" indicates picked features and "0" indicates excluded ones, because feature selection is fundamentally a binary problem.

Each wolf's fitness is then assessed using a composite objective function that takes into account the number of features chosen to guarantee computational efficiency as well as classification accuracy, which is determined using a K-Nearest Neighbors (KNN) classifier. Alpha (α), Beta (β), and Delta (δ) are the three top-performing wolves, and they lead the rest of the pack.

Mathematical models that mimic the social hierarchy and cooperative hunting behavior of grey wolves are used to update the positions of the surviving wolves. Following this upgrade, feasible feature values are maintained by applying a boundary constraint. To make sure that gains are recorded throughout iterations, the wolves' new positions are once more translated to binary form and reassessed for fitness. Until the termination condition is satisfied either by reaching the maximum number of iterations or by not seeing a discernible improvement in fitness the algorithm continues iteratively. Lastly, the optimal feature set for use in subsequent classification tasks is chosen from the feature subset linked to the Alpha wolf, which represents the best answer discovered.

This approach effectively reduces feature dimensionality while maintaining or enhancing classification accuracy, making it well-suited for medical image analysis scenarios involving complex data distributions.

IMPROVED MULTI-LAYER PERCEPTRON (IMLP) MODEL

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	201,856
batch_normalization_94 (BatchNormalization)	(None, 256)	1,024
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 512)	131,584
batch_normalization_95 (BatchNormalization)	(None, 512)	2,048
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 256)	131,328
batch_normalization_96 (BatchNormalization)	(None, 256)	1,024
dropout_2 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 128)	32,896
batch_normalization_97 (BatchNormalization)	(None, 128)	512
dropout_3 (Dropout)	(None, 128)	0
dense_4 (Dense)	(None, 5)	645

Total params: 582,917 (2.22 MB)
Trainable params: 580,613 (2.21 MB)
Non-trainable params: 2,304 (9.00 KB)

Hyperparameter	Value
Optimizer	Adam
Learning Rate	0.0005
Loss Function	categorical_cross entropy
Metrics	Accuracy
Batch Size	32
Epochs	50
ReduceLROnPlateau (Monitor)	Validation Loss (val_loss)

Figure 7. Hyperparameters and Architecture of an MLP

The hyperparameters and architecture of an MLP (Multi-Layer Perceptron) model as shown in figure 7. The network consists of multiple dense (fully connected) layers, batch normalization, and dropout layers to enhance training stability and prevent overfitting. The model uses the Adam optimizer with a learning rate of 0.0005, a categorical cross-entropy loss function, and tracks accuracy as the evaluation metric. It is trained with a batch size of 32 for 50 epochs. The ReduceLROnPlateau callback is applied, monitoring

validation loss to adjust the learning rate dynamically. The total number of parameters in the model is 582,917, with 580,613 trainable parameters.

V. RESULTS AND DISCUSSION

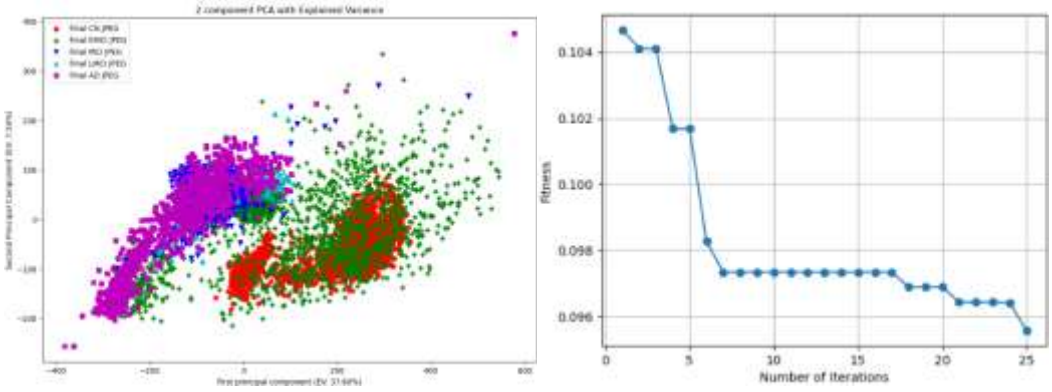
Data Processing

Class Names	Total Images	Training Images	Testing Images
AD	2420	1936	484
CN	2123	1698	424
EMCI	1920	1536	384
LMCI	2637	2109	527
MCI	2766	2212	553

Class Names	Total Images	Training Images	Testing Images
AD	2766	2212	553
CN	2766	2212	553
EMCI	2766	2212	553
LMCI	2766	2212	553
MCI	2766	2212	553

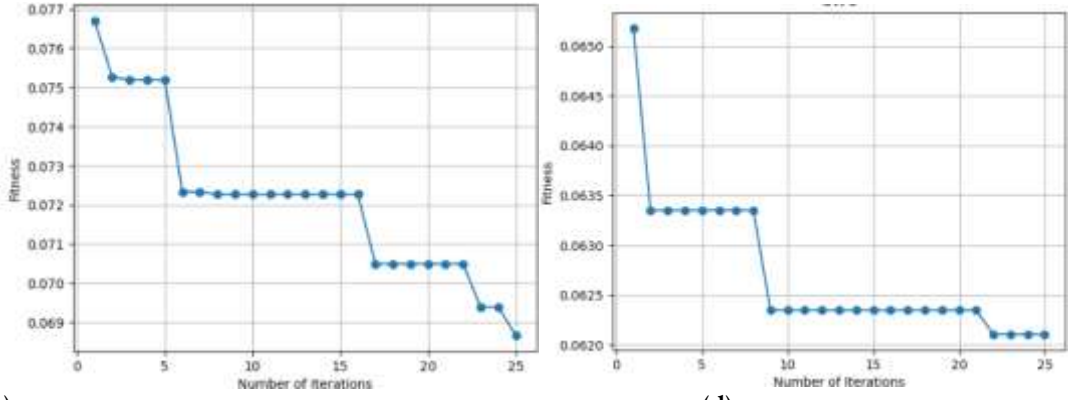
Figure 8. Distribution of dataset images before Augmentation and after Augmentation

ADNI Dataset illustrated in figure 8 consists of MRI brain scan images categorized into five classes: CN (Cognitively Normal), EMCI (Early Mild Cognitive Impairment), LMCI (Late Mild Cognitive Impairment), MCI (Mild Cognitive Impairment), and AD (Alzheimer's Disease). The dataset initially had an imbalanced distribution of images across classes, with varying numbers of training and testing images. To address this, data augmentation was applied, leading to a uniform distribution of 2,766 images per class, with 2,212 for training and 553 for testing. This augmentation enhances model performance by ensuring better generalization and balanced learning across all classes.



(a) (b)

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://www.theaspd.com/ijes.php

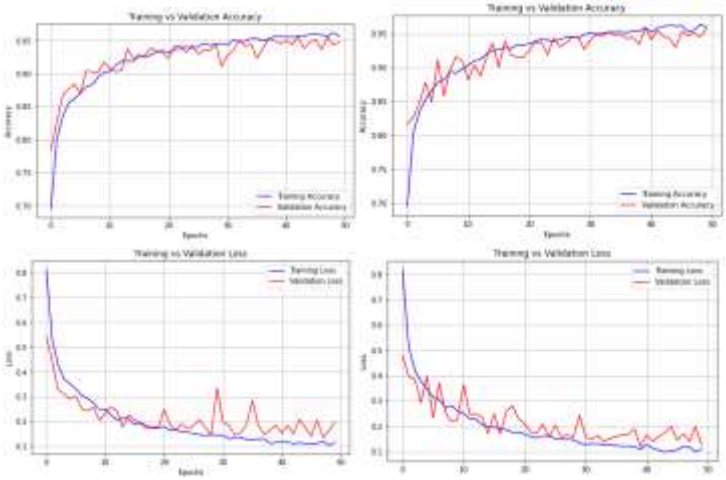


(c) (d)

Figure 9. Fitness Convergence of (b) WOA, (c) GWO, and (d) AGWO Compared to (a)PCA for Feature Selection

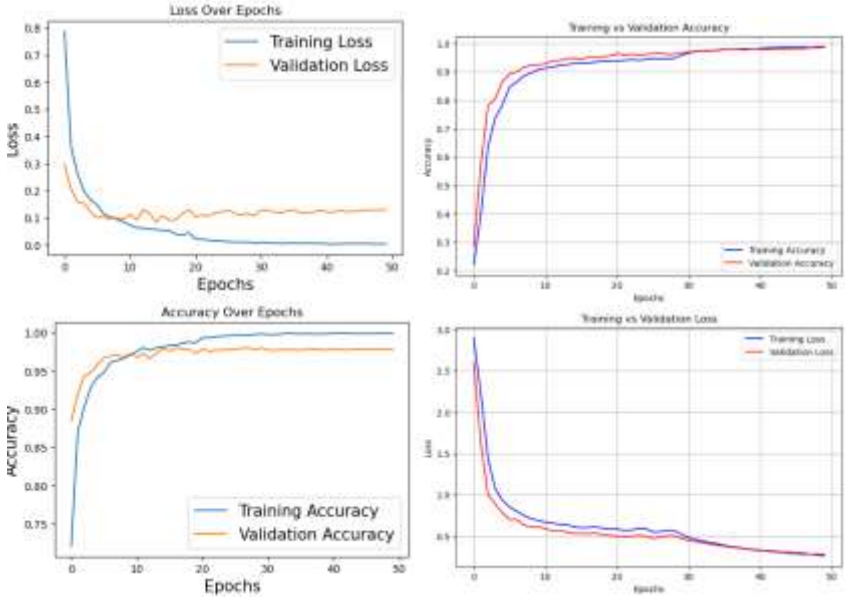
The demonstration in figure 9, the top-left PCA plot shows how the data is distributed after reducing it to two principal components. Each color and marker type represents a different Alzheimer's disease stage (CN, EMCI, MCI, LMCI, AD). The clusters are well separated, especially between AD and CN, which suggests that PCA successfully captures important patterns for classification, although some overlap remains between intermediate stages (like EMCI and MCI). The other three plots (fitness vs. number of iterations) show how different optimization algorithms WOA (Whale Optimization Algorithm), GWO (Grey Wolf Optimizer), and AGWO (Adaptive Grey Wolf Optimizer) improve the model's feature selection process over 25 iterations. A rapid drop-in fitness in the early iterations for all methods shows that they quickly find better feature subsets. However, AGWO reaches the lowest and most stable fitness value (~0.0620), compared to WOA (~0.0960) and GWO (~0.0690), which means AGWO selects the most optimal features, leading to a more accurate and generalized model.

In short, PCA provides good separation, and among the optimization methods, AGWO clearly achieves the best and most efficient feature selection, improving overall classification strength.



(a) Vgg16+InceptionV3+PCA **(b) Vgg16+InceptionV3+WOA**

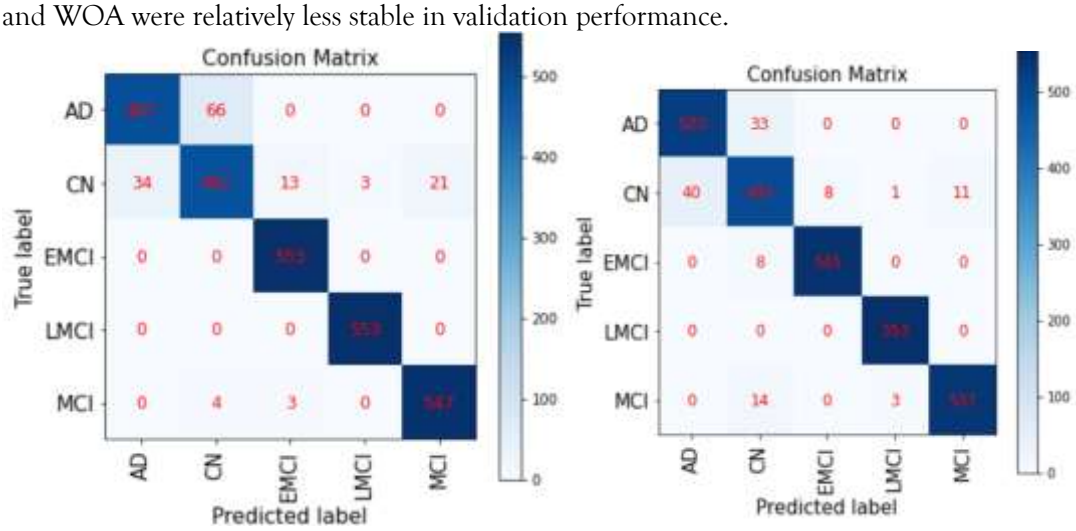
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://www.theaspd.com/ijes.php



(b) Vgg16+InceptionV3+GWO **(d) Vgg16+InceptionV3+AGWO**

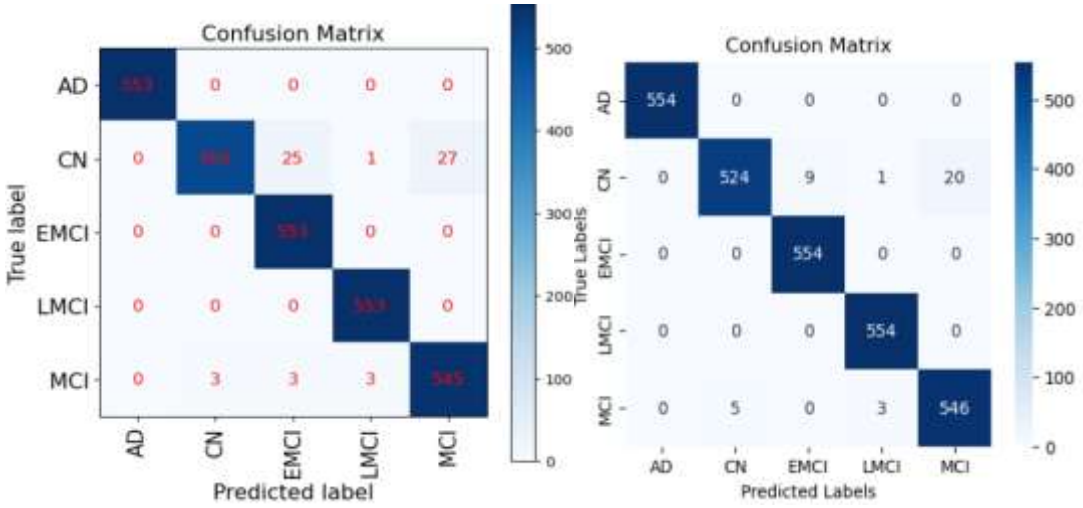
Figure 10. Illustrates the performance metrics, Panel (a) (b) (c) (d) shows different models training (accuracy and loss) and , while panel (b) displays the validation (accuracy and loss).

The graphs shown in figure 10 the training and validation performance of the VGG16+InceptionV3 model when combined with PCA, WOA, GWO, and AGWO for feature selection. For GWO and AGWO, the training and validation accuracy curves rise rapidly and remain close to each other around 98–100%, while the training and validation loss curves steadily decrease without major fluctuations. This smooth and consistent behavior indicates that GWO and AGWO helped the model generalize well without significant overfitting or underfitting. On the other hand, with PCA and WOA, although the training accuracy improves and reaches high values (around 95%), the validation accuracy fluctuates slightly more. Also, the validation loss curves for PCA and WOA show more spikes compared to GWO and AGWO, meaning the models struggled a bit to maintain stability on unseen data, which hints at possible slight overfitting. Overall, AGWO performed the best, followed closely by GWO, while PCA and WOA were relatively less stable in validation performance.



(a) **(b)**

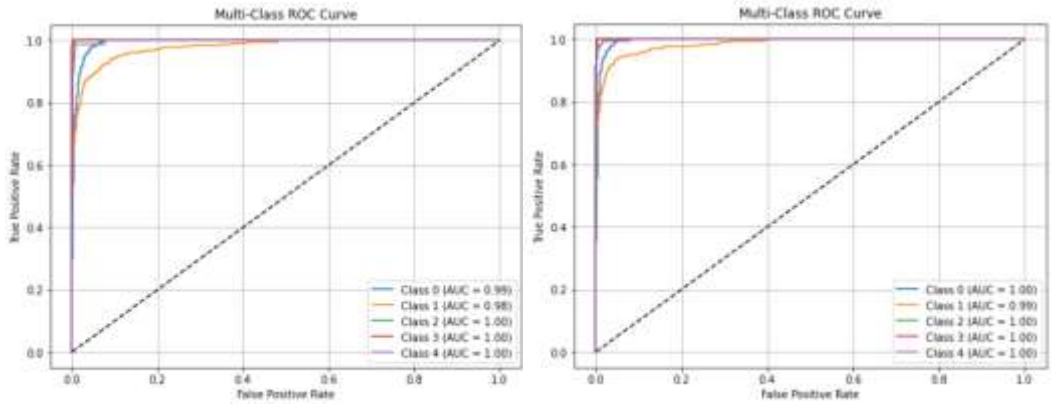
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://www.theaspd.com/ijes.php



(c) (d)

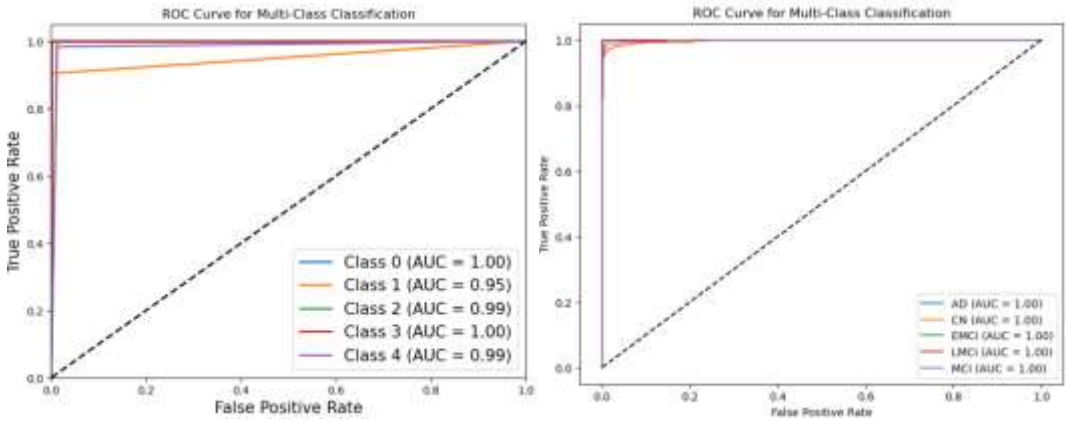
Figure 11. Confusion matrix of models on test data (a) Vgg16+InceptionV3+PCA; (b) Vgg16+InceptionV3+WOA; (c) Vgg16+InceptionV3+GWO; (d) Vgg16+InceptionV3+AGWO(proposed model).

The attached confusion matrices figure 11, compare the performance of VGG16+InceptionV3 models combined with different feature selection techniques: PCA, WOA, GWO, and AGWO. Without feature selection, the model shows noticeable confusion between AD and CN classes. Applying WOA improves AD classification but still has moderate CN confusion. PCA leads to almost perfect classification for AD, EMCI, and LMCI, though CN is often misclassified as MCI. GWO further enhances AD and EMCI predictions but leaves minor confusion between CN and MCI. Finally, AGWO achieves the best overall performance, with nearly perfect classification across all five classes and only minimal misclassification, particularly between CN and MCI, indicating that AGWO significantly boosts the model's generalization and accuracy.



(a) (b)

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://www.theaspd.com/ijes.php



(c) (d)

Figure 12. Alzheimer's Stage-Wise Classification Performance Measured by ROC Curves Using (a) PCA, (b) WOA, (c) GWO, and (d) proposed model (Vgg16+InceptionV3+AGWO)

Figure 12 shows he attached ROC curves evaluate the multi-class classification performance of VGG16+InceptionV3 models combined with PCA, WOA, GWO, and AGWO feature selection methods. In all models, the AUC (Area Under Curve) values are impressively high, mostly ranging from 0.95 to 1.00, indicating excellent discriminative ability. PCA and WOA show slight variations, with some classes having AUCs around 0.98–0.99, while GWO achieves near-perfect AUCs (mostly 0.99–1.00). AGWO outperforms all, achieving a perfect AUC of 1.00 across all five classes (AD, CN, EMCI, LMCI, MCI), demonstrating flawless classification capability and confirming that AGWO is the most effective feature selection strategy among the four approaches.

Table 1. Comparison of Training, Validation, and Testing Accuracy Along with Precision and Recall for Feature Selection Techniques on VGG16+InceptionV3 Features

Models	Train_Acc	Val_Acc	Test_Acc	Precision	Recall
Vgg16+InceptionV3+PCA	0.9628	0.9478	94.90%	0.95	0.95
Vgg16+InceptionV3+WOA	0.9728	0.9614	95.73%	0.96	0.95
Vgg16+InceptionV3+GWO	0.9994	0.9776	97.76%	0.97	0.97
Vgg16+InceptionV3+AGWO	0.9977	0.9918	98.62%	0.99	0.98

The provided table 1 clearly shows that among the models compared, Vgg16+InceptionV3+AGWO achieved the best overall performance with a training accuracy of 99.77%, validation accuracy of 99.18%, and the highest test accuracy of 98.62%, along with outstanding precision (0.99) and recall (0.98), indicating excellent generalization and very low error rates. Vgg16+InceptionV3+GWO also performed impressively, with slightly lower but still strong scores (train acc 99.94%, val acc 97.76%, test acc 97.76%, precision and recall both 0.97). In contrast, Vgg16+InceptionV3+WOA and PCA showed comparatively lower performances, with PCA achieving the lowest validation and test accuracy (94.78% and 94.90%, respectively), and WOA performing slightly better at 95.73% test accuracy. Precision and recall were also aligned with this trend. Overall, the models using optimization-based feature selection (GWO and AGWO) significantly outperformed the dimensionality reduction-based PCA and the WOA optimizer, proving the effectiveness of GWO and AGWO in enhancing model robustness and predictive power.

VI. CONCLUSION

This study proposed a novel hybrid diagnostic framework that integrates pre-trained deep learning models (VGG16 and InceptionV3) with the Adapted Grey Wolf Optimizer (AGWO) for enhanced feature selection, targeting early and accurate classification of Alzheimer's Disease (AD) stages. The framework begins with comprehensive data preprocessing and feature extraction, followed by optimal feature selection using AGWO, which is specifically adapted for binary optimization tasks crucial for feature subset selection. AGWO effectively balances classification performance and feature subset minimization by converting continuous wolf positions into binary representations and evaluating fitness based on a combination of classification error and feature size.

Additionally, an improved Multi-Layer Perceptron (MLP) classifier was employed, designed with multiple dense layers, batch normalization, and dropout, and optimized using the Adam optimizer with a dynamic learning rate adjustment via ReduceLROnPlateau. The model demonstrated excellent convergence properties, preventing overfitting while achieving high classification accuracy. The proposed VGG16+InceptionV3+AGWO framework attained a superior test accuracy of 98.62%, with precision and recall values of 0.99 and 0.98, respectively, outperforming other feature selection techniques such as PCA, WOA, and standard GWO. The comparative analysis of fitness convergence, confusion matrices, and ROC curves clearly established AGWO as the most efficient and stable feature selection strategy.

Overall, the integration of deep feature extraction with adapted metaheuristic optimization significantly boosts the classification performance, offering a powerful and reliable tool for clinical decision support in Alzheimer's Disease diagnosis. Future work will focus on expanding this framework to multimodal datasets, exploring real-time clinical applications, and integrating explainability modules to enhance trust and transparency in AI-assisted diagnostic systems.

REFERENCE

[1] Alzheimer's Association, "2023 Alzheimer's Disease Facts and Figures," Alzheimers Dement., vol. 19, no. 4, pp. 1–87, 2023.
[2] J. C. Morris, "Early-stage and preclinical Alzheimer disease," Alzheimer Disease & Associated Disorders, vol. 19, no. 3, pp. 163–165, 2005.
[3] M. S. Albert et al., "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and the Alzheimer's Association workgroup," Alzheimers Dement., vol. 7, no. 3, pp. 270–279, 2011.
[4] G. Litjens et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.
[5] S. Sarraf and G. Tofighi, "Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks," in 2016 IEEE 31st International Conference on Image Processing (ICIP), pp. 126–130, 2016.
[6] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization: An overview," Swarm Intelligence, vol. 1, no. 1, pp. 33–57, 2007.
[7] M. Dorigo and T. Stützle, Ant Colony Optimization. MIT press, 2004.
[8] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," Advances in Engineering Software, vol. 69, pp. 46–61, 2014.
[9] K. Thung et al., "Conversion of MCI patients to AD with structural MRI, functional MRI and machine learning," Frontiers in Aging Neuroscience, vol. 10, pp. 1–13, 2018.
[10] H. E. Payne et al., "MRI-based hippocampal volume as a biomarker of Alzheimer's disease progression: A systematic review," NeuroImage: Clinical, vol. 31, p. 102731, 2021.
[11] Sudharsan, M., & Thailambal, G. (2023). Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA). Materials Today: Proceedings, 81, 182-190.
[12] Nguyen, D., Nguyen, H., Ong, H., Le, H., Ha, H., Duc, N. T., & Ngo, H. T. (2022). Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease. IBRO Neuroscience Reports, 13, 255-263.
[13] Sorour, S. E., Abd El-Mageed, A. A., Albarrak, K. M., Alnaim, A. K., Wafa, A. A., & El-Shafeiy, E. (2024). Classification of Alzheimer's disease using MRI data based on Deep Learning Techniques. Journal of King Saud University-Computer and Information Sciences, 36(2), 101940.

[14] Fathi, S., Ahmadi, A., Dehnad, A., Almasi-Dooghaee, M., Sadegh, M., & Alzheimer's Disease Neuroimaging Initiative. (2024). A deep learning-based ensemble method for early diagnosis of Alzheimer's disease using MRI images. Neuroinformatics, 22(1), 89-105.
[15] Balaji, P., Chaurasia, M. A., Bilfaqih, S. M., Muniasamy, A., & Alsid, L. E. G. (2023). Hybridized deep learning approach for detecting Alzheimer's disease. Biomedicines, 11(1), 149.
[16] Jack, C. R., et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia, 14(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018
[17] Panwar, H., et al. (2021). A deep learning and Grad-CAM based color visualization approach for fast screening of COVID-19. Computers in Biology and Medicine, 124, 103986. https://doi.org/10.1016/j.compbiomed.2020.103986
[18] Suk, H. I., & Shen, D. (2013). Deep learning-based feature representation for AD/MCI classification. Medical Image Computing and Computer-Assisted Intervention, 17(Pt 2), 583–590. https://doi.org/10.1007/978-3-642-40811-3_73
[19] Islam, J., & Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Informatics, 5(2), 2. https://doi.org/10.1186/s40708-018-0080-3
[20] Zhang, Y., & Wang, S. (2015). A hybrid method for MRI brain image classification. Expert Systems with Applications, 42(13), 5613–5621. https://doi.org/10.1016/j.eswa.2015.02.035
[21] Harikumar, R., & Umamakeswari, A. (2020). Genetic algorithm based feature selection for Alzheimer's disease classification using MRI images. Materials Today: Proceedings, 33, 3953–3958. https://doi.org/10.1016/j.matpr.2020.06.578
[22] Abiyev, R. H., & Ma'aitah, M. K. S. (2018). Deep convolutional neural networks for chest diseases detection. Journal of Healthcare Engineering, 2018, 4168538. https://doi.org/10.1155/2018/4168538
[23] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. Advances in Engineering Software, 69, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007
[24] Jafari-Marandi, R., et al. (2020). Application of adaptive grey wolf optimizer with neural network for predictive modeling in medical datasets. Applied Soft Computing, 87, 105983. https://doi.org/10.1016/j.asoc.2019.105983
[25] Srivastava, N., et al. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.
[26] M. S. Alqahtani et al., "A hybrid deep learning approach for lung cancer classification from CT images using VGG16 and data augmentation," Information, vol. 12, no. 1, pp. 1–14, Jan. 2024. [Online]. Available: https://www.mdpi.com/2227-9709/12/1/18
[27] A. Bharati et al., "A deep ensemble learning-based CAD system for accurate lung cancer classification using CT imaging," BMC Medical Imaging, vol. 24, no. 1, pp. 1–14, Feb. 2024. [Online]. Available: https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-024-01345-x
[28] M. T. Alam et al., "Principal component analysis-based feature selection for cancer classification using deep learning," Sensors, vol. 24, no. 1, pp. 1–18, 2024.
[29] Y. Zhang et al., "Enhancing IoT feature selection: A two-stage approach via an improved whale optimization algorithm," Expert Systems with Applications, vol. 256, Dec. 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.122385
[30] S. Rani et al., "A novel feature selection framework based on grey wolf optimizer for mammogram image analysis," Neural Computing and Applications, vol. 33, pp. 14583–14602, 2021. [Online]. Available: https://doi.org/10.1007/s00521-021-06099-z
[31] M. A. Elaziz et al., "Adaptive grey wolf optimizer," Neural Computing and Applications, vol. 33, pp. 1–24, 2021. [Online]. Available: https://doi.org/10.1007/s00521-021-06885-9