Alzheimer's Disease Detection Using Optimized Vision Transformer

Nasrallah Asem Al-Sultani, Alaa Taima Albu-Salih, Osama Majeed Hilal
College of Computer Science and Information Technology, Department of Computer Science, University of Al-Qadisiyah, 58001 Al-Diwaniyah, Iraq
{cm.post23.17, alaa.taim, osama.m}@qu.edu.iq

Keywords: Alzheimer's Disease, Vision Transformer, Hippopotamus Optimization Algorithm, OViT, ADNI Dataset.

Abstract: Alzheimer's disease (AD) is a complex, progressive neurodegenerative condition that affects millions of people worldwide, making early diagnosis critical for effective treatment and clinical management to improve quality of life. In this study, we present an automated classification framework based on the Vision Transformer (ViT) model optimized with a modified hippopotamus optimization algorithm (M-HOA). Unlike traditional models that rely solely on ViTs or convolutional networks, the M-HOA algorithm is used to fine–tune key hyperparameters of the ViT model, improving feature extraction and classification accuracy. The model was evaluated on the ADNI dataset, which covers three diagnostic categories (AD, MCI, and NC). Experiments demonstrated that the proposed M-HOA-ViT model outperforms both the baseline and optimized ViT architectures, achieving a classification accuracy of 97.90%. The results indicate that integrating metaheuristic optimization with ViT significantly improves diagnostic accuracy, providing a robust and scalable approach for the early detection of Alzheimer's disease.

1 INTRODUCTION

Alzheimer's disease (AD) is a leading cause of dementia, predominantly affecting older adults. Symptoms of AD include cognitive impairment , difficulty performing daily tasks, loss of speech, and memory loss [1].In the initial phases of Alzheimer's disease, memory impairment is moderate; however, in the latter stages, individuals lose the capacity to engage with their environment [2]. Research from the Alzheimer's Association predicts that the number of individuals afflicted with the illness will exceed 130 million by 2050. While there is currently no treatment for Alzheimer's disease, early identification helps mitigate its prevalence [3]. Moderate cognitive impairment (MCI), an initial phase of AD, affects approximately 4-20% of individuals aged 50 and above [4]. Since pharmacological interventions can only delay the onset of its severe stages [5], Early detection and preventive strategies play a crucial role in preserving autonomy and alleviating social and emotional challenges [6].

There are several brain scanning techniques to determine whether a person has Alzheimer's disease, such as Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI) [7]. However, MRI-based methods have gained importance by diagnosing subtle structural changes in the brains of affected individuals [8].

Machine learning and deep learning, two subfields of AI, have shown promise in solving issues that traditional methods cannot, and have greatly assisted in creating highly accurate systems[9]. Therefore, a computerized system to detect Alzheimer's disease in the early stages is essential. A number of deep learning methods, including the CNN, have been used and show good performance [10].

Recent advances in computer vision have prompted researchers to explore innovative architectures. The vision transformer (ViT) is one such new technology that has generated significant interest from researchers. The idea of vision transformers are inspired by natural language processing (NLP) transformers, which are notable for their ability to handle text sequences[11]. In the vision transformer model, images are interpreted as a set of fixed-size patches rather than a grid of individual pixels. In doing so, the ViT model can leverage self-attention mechanisms to efficiently capture global dependencies and long – range connections within an image. This capability is

particularly useful because accurate diagnosis depends on the ability to understand context across multiple parts of an image[12]. They have become popular in the domain for their capacity to comprehend significant global interrelations among variables in the input space [13].A comprehensive systematic review that included an analysis of 36 studies indicated that ViTs often outperform or achieve comparable performance to CNNs on medical image classification tasks. Especially when using pre-trained models or sufficient data[14].

To improve the vision transformer model's ability to diagnose, considering the lack of research on using a metaheuristic algorithm for Alzheimer's disease, this paper proposes a vision transformer-based model using a modified Hippopotamus Optimization Algorithm (M-HOA). The goal is to identify the most appropriate hyperparameters for ViT.

Each section of this paper follows a specific order. Section 2 covers previous studies on ViT, Section 3 presents the proposed methodology, explores the ViT, outlines the stages of Hippopotamus Optimization Algorithm, and discusses its modification. Section 4 presents experimental results, Section 5 presents challenges, Section 6 suggests directions for future work, and Section 7 concludes the paper.

2 RELATED WORKS

Saman Saraf et al. [15] introduced OVITAD, an improved architecture that reduces the number of parameters, input size, number of attention vertices, and layers compared to traditional ViT, to enhance efficiency while reducing complexity, without sacrificing the model's ability to extract relevant patterns. Using the ADNI database, the model achieved 99.55% accuracy for sMRI images and 97% for rs-MRI images, surpassing the majority of deep learning models by decreasing trainable parameters by 30% relative to the baseline vision transformer model.

Uttam Khatri et al. [16] presented an improved methodology for applying Vision Transformer (ViT) to a small MRI dataset. Their study focused on how to improve the performance of ViT when data is limited. They used Shifted Patch Tokenization (SPT) to improve information capture and reduce spatial bias, and CoordConv Location Encoding (CPE) to enhance the model’s ability to perceive spatial relationships between different parts. Their experiments on ADNI MRI data showed an accuracy of 92.30%. Nevertheless, Lack of comparisons with advanced models and computational efficiency analysis. Odusami et al. [17] introduce a vision transformer model, which integrates a multi-fusion of (PET and MRI) data. The images were analyzed using a directional wavelet transform (DWT) approach, their model achieved an accuracy of 93.75%. In order to confirm the generalizability of the model, their study proposed improvements to the fusion parameters, larger and more diverse datasets. In [18], Ramesh Chandra Punia et al. Suggested a combination strategy to improve the categorisation of Alzheimer's disease. That combines explainable artificial intelligence (XAI), transfer learning, and vision transformers (ViT). InceptionV3, VGG19, ResNet50, and DenseNet121 are pre-trained deep learning models that they used with ViT to provide a more accurate classification model. This resulted in an accuracy of 96%. However, this resulted in a high computational cost due to ViT being combined with multiple models, which increases computational and memory requirements. In [19], Yue Yin et al. proposed the SMIL-DeiT model, an approach that combines Vision Transformer (DeiT-S), Self-Supervised learning (DINO), and Multiple Instance Learning (MIL) techniques. To improve classification, used the ADNI dataset containing 2032 T1-weighted MRI images. The model achieved 93.2% accuracy, outperforming traditional CNNs. Fei Huang et al. in [20] proposed a new approach to Alzheimer's disease diagnosis using Monte Carlo ensemble vision transformer (MC-ViT), where the model relies on combining ViT and Monte Carlo random sampling to improve the classification process. This approach differs from traditional ensemble learning methods that use multiple models, as it relies on only one model, but generates multiple classification decisions by sampling different input images. Structural magnetic resonance imaging (sMRI) images are divided into 3D patches, and then features are extracted from them using a 3D convolutional neural network (3D Patch Network). Next, Monte Carlo sampling is used to pick out the most important features. These are then sent to the vision transformer (ViT), which figures out how these features are connected in space and makes the classification more accurate. The model was tested on two medical databases, ADNI and OASIS-3 achieved an accuracy of 90%. In [21], Anuvab Sen et al. Proposed using Vision Transformer (ViT) with metaheuristic optimization algorithms such (DE, GA, PSO, and ACO) to improve Alzheimer's disease classification from MRI images. They used the ADNI dataset. The differential evolution (DE) based model achieved the best accuracy of 96.8%, outperforming

previous models. However, the data size was very small. Mohammed et al in [22]. Presented an innovative concept called the binary Vision Transformer (BiViT). This model combines parallel coding (PCE) and latent representation fusion (MLF) approaches to enhance case recognition accuracy using 2D images, enabling it to extract spatial and semantic features from multiple trajectories simultaneously. Their model demonstrated outstanding performance with an accuracy of 96.38%. However, the study compared only the CNN model, without evaluating performance against other transformer models. This is a limitation of the evaluation and calls for expansion in future studies.

3 METHODOLOGY

This paper presents a comprehensive framework based on deep learning techniques for diagnosing Alzheimer's disease. In this context, the proposed approach and the model architecture used in the classification process are detailed (Figure 1 illustrates the complete framework).

3.1 Dataset

The ADNI initiative is a pioneering research project launched in 2004 with the participation of prestigious academic and research institutions, providing a comprehensive registry of high-quality and accurate data[1]. The dataset used include images from samples from Alzheimer's, divided into three categories, the Table 1 shows the demographic and clinical distribution of three groups of participants: data include the number of individuals in each group, mean age, as well as mean scores on the CDR (Dementia Severity Rating) and MMSE (Mini-Mental Examination) tests.

Table 1: Details about the participants' demographics.

Group	Alzheimer's Disease (AD)	Normal Control (NC)	Mild Cognitive (MCI)
Number	610	1297	895
Male/Female	365/245	710/587	615/304
Age	57.±8.15	65.73 ±8.12	67.9 ±9.8
CDR	5.4±2.6	0.1±0.3	1.6±1.1
MMSE	20±4.30	29.17 ±1.22	27.3 ±3.71

[1] http://adni.loni.usc.edu

3.2 Image Preprocessing

We used preprocessing techniques to improve data quality and ensure consistency with the ViT model. A series of preprocessing steps was applied to brain MRIs. These steps aimed to improve image quality, standardize input dimensions for consistency with the ViT model, expand the dataset to improve generalization, and prepare it for efficient model training and evaluation. The steps are as follows:

1) Start
2) Background Removal Outside the Active Brain Region
3) Input Brain MRI Images
4) Scale Images to 224 x 244 Pixels (Linear Interpolation)
5) Normalize Pixels to [0,1] and Standardize with ImageNet21k Standards
6) Data Augmentation: Horizontal Flip (8%), Brightness/Contrast, Adjustment (±2%)
7) Data Splitting: 60% Training, 20% Validation, 20% Testing (Balanced Random Distribution).

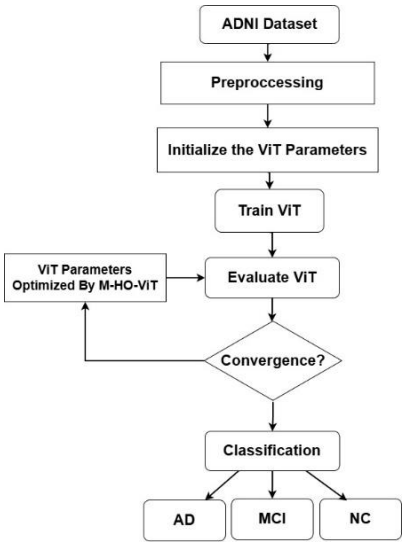

Figure 1: The structural diagram of the proposed model.

Figure 1 illustrates the steps involved in classifying Alzheimer's disease using ViT. The process starts with the ADNI dataset, followed by preprocessing steps to improve image quality and prepare it appropriately for the model, then training and optimizing ViT parameters using the M-HOA-ViT algorithm, and finally classifying the case.

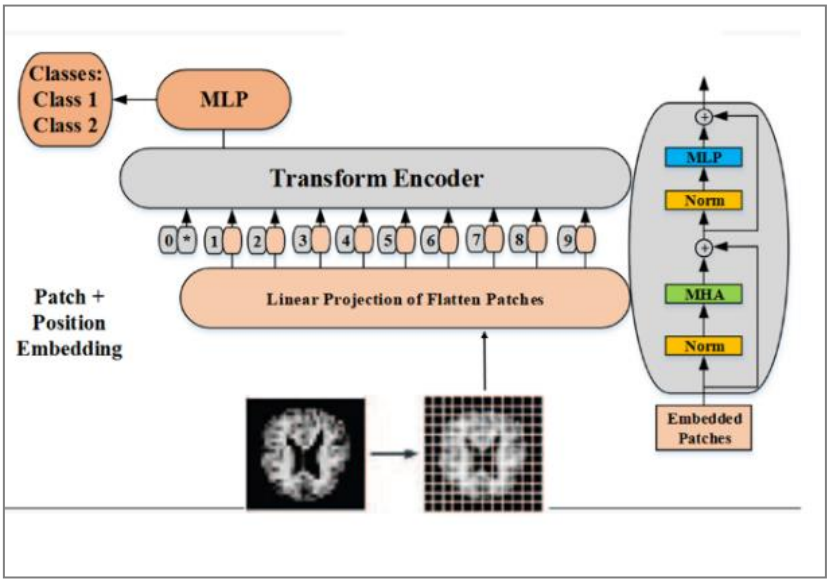

Figure 2: ViT model architecture.

3.3 Proposed Model

The proposed model is built on the Vision Transformer (ViT) architecture, which has proven remarkably successful in many computer vision tasks.

Metaheuristic algorithms are used to discover the optimal set of hyperparameters for the classification task, improve the model's accuracy. This integration aims to improve the model's efficiency, accuracy, and generalization ability. A comprehensive overview of the HOA algorithm, the ViT, and their optimization strategies is presented below.

3.3.1 ViT Model

The Vision Transformer is a modern image processing model based on the Transformers architecture originally developed for text processing [11]. An illustration of the overall structure is presented in Figure 2.

The model initially divides the images into small, uniform parts known as patches. These patches are then converted to digital representations (vectors) using a linear projection layer with positional information embedded in each patch to indicate its location within the image. The vectors are then transferred to the transformer encoder layer, which includes several sub-layers, a normalization layer (Norm), a multi-head attention (MHA) mechanism that helps understand important relationships between patches, and a multi-layer neural network (MPL) that extracts deep features from the image [12]. This method contributes to acquiring global contextual knowledge, which enhances its performance on complex tasks [23]. The diagram depicts the overall framework of the model.

3.3.2 Hippopotamus Optimization Algorithm (HOA)

Mohamed Hussein Amiri et al. In 2024 proposed a novel metaheuristic algorithm called Hippopotamus Optimization Algorithm (HOA) was proposed, inspired by the behaviors of hippopotamus in nature [24]. The algorithm consists of three main stages: exploitation, migration, and exploration.

This algorithm was chosen because it provides a balance between global and local search, achieves excellent experimental performance, and is well-suited for hyperparameter optimization problems in complex environments. Comparisons with other optimization algorithms have also demonstrated its superior performance:

A) Algorithm initialization: The algorithm starts by randomly initializing hippo locations, where these locations represent potential solutions to the optimization problem.

$$X_i: x_{i,j} = lb_j + r \times (ub_j - lb_j), i = 1,2,\dots,N; j = 1,2,\dots,m. \quad (1)$$

x_i refers to the position of the candidate solution $i\,th$ in the search space, where r is used as random number generated (between the range

0,1) to ensure random distribution within specified range of limits, while ub_j, lbj denote the lower and upper bounds for the $j\ th$ variable. n denote the population size (number of hippopotamuses in the herd), and m represent the total number of decision variables. Therefore, this information is then used to compile the population matrix according to (2).

$$x = \begin{matrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ \lfloor x_N \rfloor_{N\times m} \end{matrix} = \begin{matrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & . & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & . & \vdots & \ddots & \vdots \\ \lfloor x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \rfloor_{N\times m} \end{matrix} . \quad (2)$$

B) HOA as a mathematical model the simulation of the hippopotamus movement within the HOA is divided into three phases according to their natural patterns [24].

3.3.2.1 Phase 1: Exploration (Updating Hippopotamus in the Pond)

A hippopotamus herd is composed of several adult females, juveniles, multiple males, and a dominant male who leads the herd. The dominant male is selected based on the objective function's value. Hippopotamuses tend to cluster together, with the dominant male safeguarding the herd and its territory from potential dangers. The spatial position of male Hippopotamuses in the lake is represented mathematically by (3).

$$\chi_i^{\mathcal{M}\ \text{hippo}}: \underset{xi_j}{\text{M hippo}} = x_{ij} + y1 \cdot \left(\mathcal{D}\ \text{hippo} - I_1 x_{ij}\right) \quad (3)$$

In (3), xi^{mhippo} shows where the male hippo is and $Dhippo$ shows where the dominant hippo is. In this case, $Dhippo$ has the lowest cost in the current iteration. The random vector between 0 and 1 is denoted by $\overrightarrow{r_1}$,…,$\overrightarrow{r_4}$, $\overrightarrow{r_5}$, according to (4). The values I_1 and I_2 are integers between 1 and 2 (according to (3) and (6)). While in (4), e_1 and e_2 represent random integers that can be 0 or 1:

$$\curvearrowleft = \begin{matrix} I_2 \times \vec{r}_1 + (\sim \varrho_1) \\ 2 \times \vec{r}_2 - 1 \\ \vec{r}_3 \\ I_1 \times \vec{r}_4 + (\sim \varrho_2) \\ r_5 \end{matrix} , \quad (4)$$

$$T = \exp\left(-\frac{t}{Max_{\text{iterations}}}\right). \quad (5)$$

$$X_i^{FBhippo}: x_i^{FBhippo} = \quad (6)$$

$$= \begin{cases} x_{ij} + h_1 \cdot \left(D_{\text{hippo}} - I_2 MG_i\right) & , T > 0.6 \\ \Xi & , \text{else} \end{cases} .$$

$$\Xi = \begin{cases} x_{ij} + h_2 \cdot \left(MG_i - D_{\text{hippo}}\right), r_6 > 0.5 \\ lb_j + r_7\left(ub_j - lb_j\right), \text{else} \end{cases} . \quad (7)$$

In (6) and (7), express the location of female hippo or immature calves ($x_i^{FBhippo}$) in the herd. Hippo calves often stay close to their mother, but due to curiosity, they depart from the herd or their mums. if the T value is more than 0.6, it indicates that the calf has moved away from its mother according to (5). if r_6 An arbitrary value if the value ranging between 0 and 1, if (7) exceeds 0.5, it means that the hippo has moved away from its mother, but is still in or close to the group. This behavior is modeled for females and immature hippos using (6) and (7). The value (h_2 and h_1) Numbers are randomly chosen from five scenarios. While r_7 in (7) is a random number between 0 and 1, (8) and (9) represent an update of the positions of male and female hippos, or immature hippos, within the herd, while F_i represents the value of the objective function.

$$X_i = \begin{cases} X_i^{Mhippo} F_i^{Mhippo} & < F_i \\ & \text{else} \end{cases} \quad (8)$$

$$X_i = \begin{cases} X_i^{FBhippo} F_i^{\text{FBhippo}} \\ X_i \end{cases} \quad (9)$$

Using h vectors with scenarios $I1$ and $I2$, the algorithm improves the overall search process and increases the exploration efficiency, leading to better results.

3.3.2.2 Phase 2: Explore (the Hippopotamus' Defense Mechanism Against Predators)

A hippopotamus is also at risk. Quickly turning towards the animal and making loud noises scares it away is their main defence. (10) represents where the attacker is in the search space.

$$predator_j = lb_j + \vec{r_8}.\left(ub_j - lb_j\right), j = 1,2,\ldots, \quad (10)$$

where $\vec{r_8}$ denote to random vector from 0 to 1.

$$D^{\rightarrow} = |predator_j - x_{ij}|. \quad (11)$$

The (11) calculates the distance between the hippo i and the predator. In this phase, the hippo uses a defensive strategy based on $F_{predatorj}$ to protect itself. When $F_{predatorj}$, is lower than F_i, it indicates to the predator is extremely close, which prompts the hippo to quickly turn towards the threat and move in its direction to force it away. On the other hand, if $F_{predatorj}$ is higher, this suggests that the predator or any other potential threat is far from the hippo territory (as shown in (12)).

$$x_{ij}^{HippoR} = \begin{cases} RL^{\rightarrow} \oplus Predator_j + \left(\frac{l}{(c - d\ x \cos(2\pi g))}\right). \\ \left(\frac{1}{D^{\rightarrow}}\right) F_{predator_j} > F_i \end{cases} \quad (12)$$

x_i^{HippoR}represents the position of the hippo facing the predator. $RL^{\rightarrow}$ indicate to a random vector following a Lévy distribution, which simulates sudden shifts in the position of the predator during an attack on the hippo. This random motion is described by (13). The values are chosen sequentially within the range [0,1]. Here, v is a constant ($v = 1.5$), w is used as a representation of the gamma function, it is calculated based on (14).

$$Levey(\theta) = 0.05 \cdot \frac{w \cdot \sigma_w}{|v|^{\frac{1}{\theta}}} . \quad (13)$$

$$\sigma_w = \left[\frac{\Gamma(1+\vartheta)\sin\left(\frac{\pi\vartheta}{2}\right)}{\Gamma\left(\frac{(1+\vartheta)}{2}\right)\vartheta 2^{\frac{(\vartheta-1)}{2}}}\right]^{\frac{1}{\vartheta}} . \quad (14)$$

In (12), the variable f is a random number uniformly chosen between 2 and 4, c is uniformly selected between 1 and 1.5, and D is a uniform random number between 2 and 3. Additionally, g is a uniformly distributed random number between -1 and 1.

According to (15), if F_i^{HippoR} exceeds F, it indicates that the hippo has been killed, and a new one will replace it in the herd. If it des not exceed F, then the predator retreats, and the hippopotamus goes back to the herd. The second phase brought noticeable improvements to help avoid getting stuck in local minima.

$$x_i = \begin{cases} X_i^{Mhippo} F_i^{Mhippo} < F_i \\ \mathrm{x_i} F_i^{MhippoR} \geq F_i \end{cases} . \quad (15)$$

3.3.2.3 Phase 3: Exploitation (Hippopotamus Escapes from Predator).

When a hippopotamus encounters a predator or find its usual defensives are not enough. In these cases run toward the nearest lake or pond, as lions and spotted hyenas tend to avoid water. This strategy helps the hippopotamus secure a safer spot.

This process is modeled by equations (16) - (17). If the newly generated location yields a lower cost function value, it signifies that the hippo has found a safer and more optimal spot nearby and has moved there. Where, t (represents the current iteration), T represents the total number of allowed iterations (the maximum number).

$$lb_j^{local} = \frac{lb_j}{t}, ub_j^{local} = \frac{ub_j}{t}, t = 1,2,\dots,T. \quad (16)$$

$$x_i^{Hippo^{\varepsilon}} : x_{ij}^{Hippo^{\varepsilon}} = x_{ij} + r_{10} \cdot (lb_j^{local} + s \cdot \left(ub_j^{local} - lbj^{local}\right) . \quad (17)$$

In (17), x_i^{hippoS} represents the position of a hippopotamus that has been searching for the nearest safe spot. While (s_1) is a random vector chosen from three different scenarios, as defined by (s) in (18).

$$s = \begin{cases} 2 \cdot \vec{r_{11}} - 1 \\ r_{12} \\ r_{13} \end{cases} . \quad (18)$$

In (18), the term $\vec{r}_{11}$ is a random vector with values between (0,1). In contrast, r_{10} from (17) and r_{13} are stochastic variables selected from the interval (0,1). Additionally, another thing is that a normal distribution is used to pick a number r_{12} .

$$x_i = \begin{cases} x_i^{Hippo\varepsilon} F_i^{Hippo\varepsilon} < F_i \\ x_i F_i^{Hippo\varepsilon} \geq F_i \end{cases} . \quad (19)$$

3.3.3 HOA Modification

We improved the original Hippopotamus algorithm through two major improvements: the first is by using a Chaotic Logistic Map to improve the initialization distribution, and the second is by introducing an adaptive control factor α and an improved update equation to adjust the balance between exploration and exploitation and improve the convergence accuracy.

3.3.3.1 Logistic Map

The logistic map is a fundamental one-dimensional function extensively applied in modeling complex behaviors, studying biological population dynamics, and in cryptography [25]. The equation below mathematically represents the chaotic map used for generating pseudorandom sequences—specifically, the one-dimensional logistic map as (20).

$$x_n + 1 = rx_n(1 - x_n). \quad (20)$$

Here, n = 0, 1,2, … denotes the iteration and r is the control parameter, chosen from the interval (3.999,4) to ensure the sequences exhibit chaotic behavior.

3.3.3.2 Selective Divergence and Convergence Strategy

The three phases of the HOA are the predator escape stage. In the modified method, two main strategies are adopted: Selective Divergence and optimal solution attraction. An adaptive control factor α was introduced, which gradually reduces the impact of large jumps, preventing instability in later iterations. Simultaneously, the position update (22) and parameter B (fixed at 0.3) ensure convergence by guiding solutions toward the best-known solve.

$$\alpha = \min\left(\alpha_{\max}, \left(\frac{\| x_i - \text{Predator} \|}{\text{max\_distance}}\right) \times \left(1 - \frac{t}{T}\right)\right). \quad (21)$$

$$x_i^{HippoE} = x_i + a\,.(x_i - Predator) + B\,.(x_{best} - x_i). \quad (22)$$

4 EXPERIMENTAL RESULTS

In this section, we review the results from integrating the vision transformer with the optimization algorithm to diagnose the disease. Table 2 summarizes the details of the model used, while the optimized hyperparameters determined by the M-HOA algorithm for training the ViT model are summarized in Table 3.

Table 2: Details of parameters of the ViT model.

Property	Value
Model type	ViT-B/16 (Base Vision Transformer)
Trained with	ImageNet-21k
Input image size	224 x 224 pixels
Attention head per layer	12

While implementing the proposed system, we utilized a high-efficiency processing unit (Intel Core i7) paired with 32GB of RAM, ensuring seamless execution of training and classification processes. The system was built using Python 3.10 and the PyTorch framework.

Table 3: Hyperparameters used to train the Vision Transformer model.

Hyperparameters	Value
Batch size	28
Epoch	17
Dropout rate	0.19
Learning rate	$1x10^{-5}$
Number of heads	16

The results shown in Table 4 and Figure 3 illustrate the performance of the trained models. The analysis showed that the proposed model, known as (M-HOA-ViT), achieved the highest accuracy rate of 97.90%, and the (HOA-ViT) model recorded an accuracy rate of 96.55%, while (ViT) achieved the lowest accuracy rate of 95.74%. The final findings indicate that using HOA algorithm for hyperparameter selection results in a modest improvement in accuracy; in contrast, the modified HOA algorithm significantly enhances the model's performance.

Table 4: Details Results of the proposed system.

Model	Performance				
	Accuracy	Precision	Recall	F1-score	Test loss
ViT	95.74%	96.71%	80.88%	88.09%	0.0631
HOA-ViT	96.55%	94.89%	83%	88.55%	0.0602
M-HOA-ViT	97.90%	97.29%	87.29%	92.02%	0.0558

The final findings show that integrating the optimization algorithm with the ViT has significantly enhanced the model's performance, achieving an accuracy of 97.90% in diagnosing Alzheimer's disease. The model was trained for 17 epochs using MRI images classified into three categories representing different stages of the disease. Figure 4 shows the performance of the model during both training and testing epochs. While the confusion matrix is presented in Figure 5.

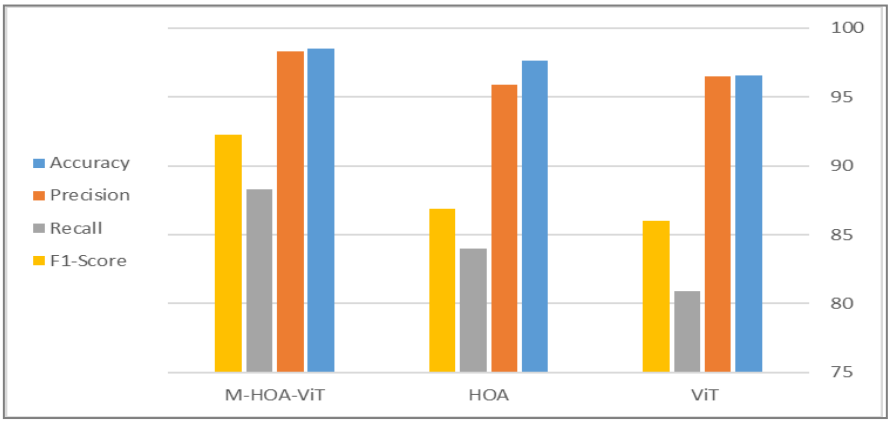

Figure 3: Result of the proposed model.

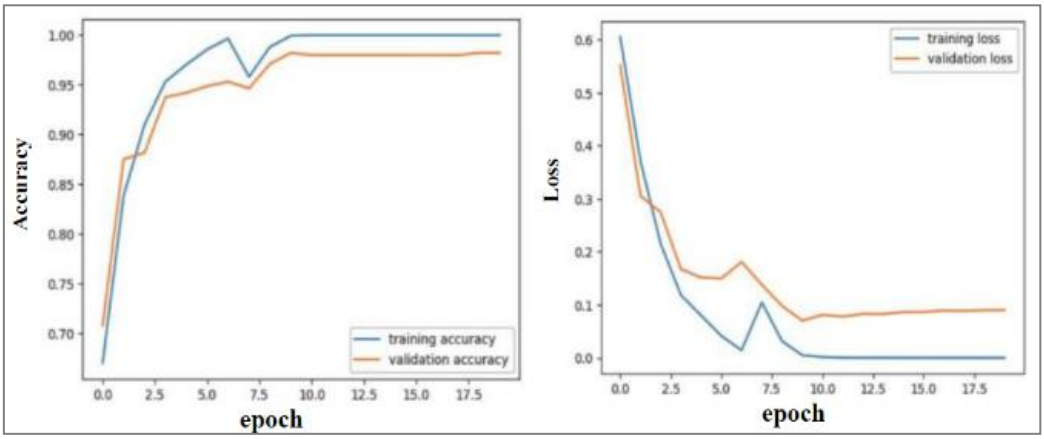

Figure 4: Accuracy and loss of the training and testing for M-HOA-ViT.

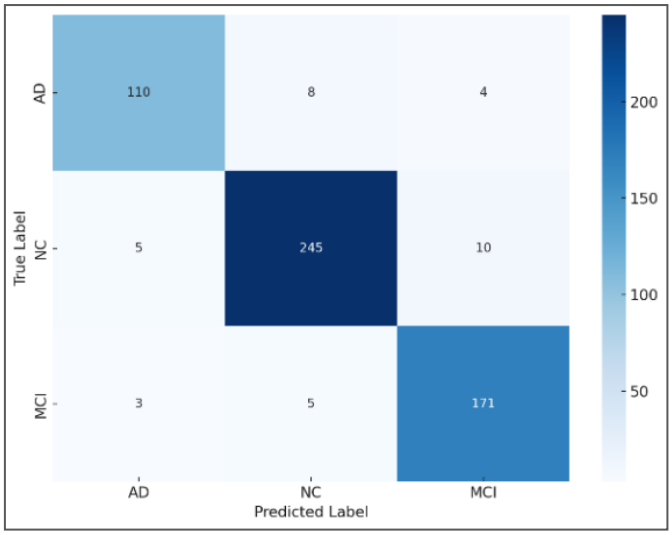

Figure 5: Confusion matrix.

The figure shows that the M-HOA-ViT model outperforms the other models with remarkable accuracy, reflecting its effectiveness in classifying various cases. Modifications to the HOA algorithm have improved the recall rate, helping to detect rare cases.

Table 5 provides a comparison of the diagnostic accuracy of various approaches for Alzheimer's disease detection. While most methods use binary classification, which may overlook early brain changes, our proposed three-class system offers improved sensitivity and early detection.

Table 5: Evaluation of the suggested model compared with previous Alzheimer's detection methods.

Authors	Algorithm/ Approach	Accuracy
Anuvab Sen et al 2023 [21]	Metaheuristic (DE, GA, PSO, ACO) algorithms with ViT	96.8% 91% 92% 94%
Xing mu et al 2023 [26]	CNN+ViT	97.43%
Yanjun Lyu et al 2022 [27]	TL+ViT	96.8%
Proposed model	M-HOA-ViT	97.90%

The table presents a comparative analysis of various models employing the vision transformer for classifying the disease. The DE-ViT model Sen et al [20] attained an accuracy of 96.8% with the application of a differential evolution (DE), which enhanced parameter optimization. The model presented by Mo et al [23] achieved high accuracy as a result of combining ViT with CNNs, which effectively extracted local and global characteristics together. While the TL+ViT in [24] achieved a competitive accuracy of 96.8%, using transfer learning techniques and a pre-trained model. The superiority of our model lies in its ability to precisely fine-tune learning parameters, which improves performance and reduces variability in results. Additionally, the use of M-HOA algorithm provides an effective balance between exploration and exploitation.

5 CHALLENGES

The key challenges encountered during the development and deployment of the model are summarized as follows:

- Data availability constraints: Due to strict privacy regulations, the brain imaging datasets accessible for this study are limited. This restriction could impede the model's ability to capture the full variability present in the broader population.
- Substantial computational overhead: The Vision Transformer requires considerable computational power and extended training durations, which can constrain iterative experimentation and may limit scalability in resource- limited environments
- Dependency on hyperparameter: The model's performance heavily depends on proper hyperparameter tuning, where small changes can lead to noticeable performance variations.

6 CONCLUSIONS

This study presents a new methodology that integrates the Vision Transformer (ViT) model with a modified Hippopotamus Optimization Algorithm (M-HOA-ViT). The modified algorithm incorporates logistic maps, a selective divergence strategy, and enhanced convergence mechanisms, which together significantly improve the model’s performance. Experimental evaluations show that the proposed method achieves an accuracy of 97.90%, enhancing its ability to detect subtle differences that indicate early brain changes. This, in turn, facilitates the timely detection of Alzheimer’s disease and contributes to improved therapeutic outcomes. However, we encountered some challenges during implementation, most notably limited data availability because privacy restrictions, the model's high computational capacity, and its high sensitivity to hyperparameter settings. This highlights the importance of developing less resource-intensive models and improving data accessibility to ensure the application of these techniques in real word medical environments.

7 FUTURE WORK

The increasing demand for accurate and efficient Vision Transformer (ViT) detection systems highlights the need for further advancements to fully harness their potential. By combining different imaging modalities (MRI) and PET, data gaps can be filled, diagnosis can be improved, and clinical and genetic data can also be used to enhance diagnostic accuracy. In addition, developing sophisticated attention mechanisms that can pick up on minute features in brain imaging should receive special

attention, particularly in brain areas like the frontal cortex and hippocampus, which are most impacted by disease.

REFERENCES

[1] H. Jahn, "Memory loss in Alzheimer's disease," Dialogues Clin. Neurosci., vol. 15, no. 4, pp. 445–454, Dec. 2013, doi: 10.31887/DCNS.2013.15.4/hjahn.

[2] Z. Zhang and F. Khalvati, "Introducing Vision Transformer for Alzheimer's Disease classification task with 3D input," Oct. 03, 2022, arXiv: arXiv:2210.01177. doi: 10.48550/arXiv.2210.01177.

[3] "2022 Alzheimer's disease facts and figures," Alzheimers Dement., vol. 18, no. 4, pp. 700–789, Apr. 2022, doi: 10.1002/alz.12638.

[4] Z. Breijyeh and R. Karaman, "Comprehensive Review on Alzheimer's Disease: Causes and Treatment," Molecules, vol. 25, no. 24, p. 5789, Dec. 2020, doi: 10.3390/molecules25245789.

[5] Q. Behfar, A. Ramirez Zuniga, and P. V. Martino-Adami, "Aging, Senescence, and Dementia," J. Prev. Alzheimers Dis., vol. 9, no. 3, pp. 523–531, Jul. 2022, doi: 10.14283/jpad.2022.42.

[6] M. Bruscoli and S. Lovestone, "Is MCI really just early dementia? A systematic review of conversion studies," Int. Psychogeriatr., vol. 16, no. 2, pp. 129–140, Jun. 2004, doi: 10.1017/S1041610204000092.

[7] A. Lakhan, T.-M. Grønli, G. Muhammad, and P. Tiwari, "EDCNNS: Federated learning enabled evolutionary deep convolutional neural network for Alzheimer disease detection," Appl. Soft Comput., vol. 147, p. 110804, Nov. 2023, doi: 10.1016/j.asoc.2023.110804.

[8] for the Alzheimer's Disease Neuroimaging Initiative, A. Chandra, G. Dervenoulas, and M. Politis, "Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment," J. Neurol., vol. 266, no. 6, pp. 1293–1302, Jun. 2019, doi: 10.1007/s00415-018-9016-3.

[9] A. A. Hasan, A. T. A. Salih, and A. Ghandour, "Lung Cancer Detection using Evolutionary Machine learning and Deep learning: A survey," in 2023 International Conference on Information Technology, Applied Mathematics and Statistics (ICITAMS), Al-Qadisyia, Iraq: IEEE, Mar. 2023, pp. 129–133. doi: 10.1109/ICITAMS57610.2023.10525500.

[10] S. E. Sorour, A. A. A. El-Mageed, K. M. Albarrak, A. K. Alnaim, A. A. Wafa, and E. El-Shafeiy, "Classification of Alzheimer's disease using MRI data based on Deep Learning Techniques," J. King Saud Univ. - Comput. Inf. Sci., vol. 36, no. 2, p. 101940, Feb. 2024, doi: 10.1016/j.jksuci.2024.101940.

[11] A. Vaswani et al., "Attention is all you need," Adv. Neural Inf. Process. Syst., vol. 30, 2017, Accessed: Feb. 24, 2025.

[12] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 2020, arXiv. doi: 10.48550/ARXIV.2010.11929.

[13] Y. Liu et al., "A Survey of Visual Transformers," IEEE Trans. Neural Netw. Learn. Syst., vol. 35, no. 6, pp. 7478–7498, Jun. 2024, doi: 10.1109/TNNLS.2022.3227717.

[14] S. Takahashi et al., "Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review," J. Med. Syst., vol. 48, no. 1, Sep. 2024, doi: 10.1007/s10916-024-02105-8.

[15] S. Sarraf, A. Sarraf, D. D. DeSouza, J. A. E. Anderson, M. Kabia, and The Alzheimer's Disease Neuroimaging Initiative, "OViTAD: Optimized Vision Transformer to Predict Various Stages of Alzheimer's Disease Using fMRI and Structural MRI Data," Brain Sci., vol. 13, no. 2, p. 260, Feb. 2023, doi: 10.3390/brainsci13020260.

[16] U. Khatri and G.-R. Kwon, "Training vision transformer with gradient centralization optimizer for Alzhemier's disease small dataset increase the diagnostic accuracy," Jan. 22, 2024. doi: 10.36227/techrxiv.170594593.30710633/v1.

[17] M. Odusami, R. Maskeliūnas, and R. Damaševičius, "Pixel-Level Fusion Approach with Vision Transformer for Early Detection of Alzheimer's Disease," Electronics, vol. 12, no. 5, p. 1218, Mar. 2023, doi: 10.3390/electronics12051218.

[18] R. C. Poonia and H. A. Al-Alshaikh, "Ensemble approach of transfer learning and vision transformer leveraging explainable AI for disease diagnosis: An advancement towards smart healthcare 5.0," Comput. Biol. Med., vol. 179, p. 108874, Sep. 2024, doi: 10.1016/j.compbiomed.2024.108874.

[19] Y. Yin, W. Jin, J. Bai, R. Liu, and H. Zhen, "SMIL-DeiT:Multiple Instance Learning and Self-supervised Vision Transformer network for Early Alzheimer's disease classification," in 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy: IEEE, Jul. 2022, pp. 1–6. doi: 10.1109/IJCNN55064.2022.9892524.

[20] F. Huang and A. Qiu, "Ensemble Vision Transformer for Dementia Diagnosis," IEEE J. Biomed. Health Inform., vol. 28, no. 9, pp. 5551–5561, Sep. 2024, doi: 10.1109/JBHI.2024.3412812.

[21] A. Sen, U. Sen, and S. Roy, "A Comparative Analysis on Metaheuristic Algorithms Based Vision Transformer Model for Early Detection of Alzheimer's Disease," in 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN), Dec. 2023, pp. 200–205. doi: 10.1109/CICN59264.2023.10402213.

[22] S. M. A. H. Shah, M. Q. Khan, A. Rizwan, S. U. Jan, N. A. Samee, and M. M. Jamjoom, "Computer-aided diagnosis of Alzheimer's disease and neurocognitive disorders with multimodal Bi-Vision Transformer (BiViT)," Pattern Anal. Appl., vol. 27, no. 3, p. 76, Sep. 2024, doi: 10.1007/s10044-024-01297-6.

[23] J. He et al., "TransFG: A Transformer Architecture for Fine-Grained Recognition," Proc. AAAI Conf. Artif. Intell., vol. 36, no. 1, Art. no. 1, Jun. 2022, doi: 10.1609/aaai.v36i1.19967.

[24] M. H. Amiri, N. Mehrabi Hashjin, M. Montazeri, S. Mirjalili, and N. Khodadadi, "Hippopotamus optimization algorithm: a novel nature-inspired optimization algorithm," Sci. Rep., vol. 14, no. 1, p. 5032, Feb. 2024, doi: 10.1038/s41598-024-54910-3.

[25] M. A. Murillo-Escobar, C. Cruz-Hernández, L. Cardoza-Avendaño, and R. Méndez-Ramírez, "A novel pseudorandom number generator based on pseudorandomly enhanced logistic map," Nonlinear Dyn., vol. 87, no. 1, pp. 407–425, Jan. 2017, doi: 10.1007/s11071-016-3051-3.

[26] X. Mu et al., "Alzheimer Classification Based on Convolutional Neural Network and Vision Transformer," in 2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Chengdu, China: IEEE, Nov. 2023, pp. 329–334. doi: 10.1109/ICICML60161.2023.10424819.

[27] Y. Lyu, X. Yu, D. Zhu, and L. Zhang, "Classification of Alzheimer's Disease via Vision Transforme" in Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments, Corfu Greece: ACM, Jun. 2022, pp. 463–468. doi: 10.1145/3529190.3534754.