IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 July 2022, accepted 27 July 2022, date of publication 5 August 2022, date of current version 11 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3196649

RESEARCH ARTICLE

Classification of Brain Volumetric Data to Determine Alzheimer's Disease Using Artificial Bee Colony Algorithm as Feature Selector

MÜMINE KAYA KELEŞ, (Member, IEEE), AND ÜMIT KILIÇ
Department of Computer Engineering, Adana Alparslan Türkeş Science and Technology University, 01250 Adana, Turkey

Corresponding author: Mümine Kaya Keleş (mkaya@atu.edu.tr)

This work was supported in part by the Scientific Research Projects Commission Unit of Adana Alparslan Turkes Science and Technology University under Grant 18332001, in part by the Alzheimer's Disease Neuroimaging Initiative—ADNI (National Institutes of Health) under Grant U01 AG024904, and in part of the Department of Defense (DOD) ADNI under Award W81XWH-12-2-0012.

ABSTRACT Alzheimer's disease is a degenerative disease that affects the age progression and causes the brain to be unable to fulfill its expected functions. Depending on the stage, the effects of Alzheimer's disease (AD) vary from forgetting the names of the surrounding people to not being able to continue daily life without assistance. To the best of our knowledge, there are currently no generally accepted diagnostic or treatment methods. In this study, a binary version of the artificial bee colony algorithm (BABC) is proposed as a feature selector for classifying AD from volumetric and statistical data of brain magnetic resonance images (MRIs). MRIs were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Volumetric and statistical data from the collected MRIs were obtained from an online system called volBrain. Then, for comparison, binary particle swarm optimization (BPSO), binary grey wolf optimization (BGWO), and binary differential evolution (BDE) were employed. For a comprehensive comparison, three algorithms, K-nearest Neighborhood (KNN), Random Forest (RF), and Support Vector Machine (SVM), are used as classifiers in feature selection progress. The results of this comparison demonstrate that BGWO outperforms BABC, which is a competitive method for this purpose. The outputs of the experiments show that all methods achieve their personal best by using RF as the classifier. Additionally, traditional data mining methods such as the Info Gain (IG), Gain Ratio (GR), Chi-square (CHI), and ReliefF methods were utilized for comparison. The results also demonstrate the superiority of the BABC over traditional methods. Another research point that this study focused on was to explore which parts of the brain are more relevant for AD diagnosis. The novelty of this study lies in the output of this point. Alongside the hippocampus and amygdala, the globus pallidus can also help in AD diagnosis.

INDEX TERMS Alzheimer's disease, artificial bee colony, data mining, feature selection, machine learning, magnetic resonance imaging, swarm intelligence.

I. INTRODUCTION

Alzheimer's disease (AD) is a neurological illness that is the most frequent form of dementia and results in the loss of brain functions such as reading, writing, thinking, and memorizing [1]. According to a report from 2018 [2], a new case was observed every three seconds. Dementia affects 50 million people worldwide, with that number estimated to increase to 152 million by 2050. In 2018, the global cost of dementia was predicted to be $1 trillion [2]. According to a report [1], the number of people with dementia in low- and middle-income nations is increasing at a significantly faster rate than in high-income countries. To date, no method is generally accepted for the diagnosis or treatment of AD. Nonetheless, early detection is crucial for reducing the effects of the disease and preserving the quality of patients' daily

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

lives. Traditional methods for the assessment of AD include the mini-mental state examination (MMSE), neurobiological and physical examinations, and patients' detailed disorder history. Some medical tests such as blood, urine, and genetic tests, are performed by physicians for assessment. Blood or urine tests are performed to assess other symptoms, such as vitamin and nutrient levels, infections, as well as liver, kidney, and thyroid function. Genetic tests were performed to obtain the family history of the dementia. As technology has developed, medical methods have improved. Brain scans are one of the methods used to diagnose AD. Brain scans can be used to detect dementia by monitoring the shrinkage and enlargement of brain parts [3]. These technologies produce and process the data. The processing of health data and extracting patterns from them has become an extensive area of medical research, such as the diagnosis of AD. Parts of the brain such as the hippocampus, cerebral cortex, and amygdala, which are related to planning, speaking, remembering, judgment, and thinking, are monitored to obtain clues. The shrinkage or enlargement of these parts depends on AD and its progression. The diagnosis of AD and the discovery of perceptible discrepancies in the data require not only experience and knowledge, but also conjoining and analyzing them with additional test results by experts [4].

One of the reasons for conducting this study is that a combination of data mining and artificial intelligence methods can prevent the time consumed by the traditional methods used for AD determination. The second reason is to help physicians working in this area focus on the proper regions by finding the brain region most relevant to Alzheimer Disease.

The main purpose of this study was to distinguish between magnetic resonance image (MRI) data grouped as Alzheimer's Disease-Healthy Control (AD-HC) with high accuracy. The second aim was to determine the brain parts that play an important role in discovering possible distinctions.

In this study and many others in the literature, MRI was used because it is non-invasive, cheaper, and easier to apply than other techniques. In the next section, the relevant literature is presented. Section III presents the materials and methods employed in this study. Section IV presents the results and discussion. The conclusions of this study have been added to Section V.

II. LITERATURE REVIEW

Data mining, which is the science of obtaining utilizable information from large datasets, has intersections with artificial intelligence, data management, databases, machine learning, and statistics [5]. In the medical field, many researchers have used data mining to diagnose and assess diseases. The following studies can be considered as general instances in this area.

Some taking attention general studies have focused on combining health data, data mining, and artificial intelligence. To be more specific about the topic of this article, works focused on dementia (especially AD) using feature selection are surveyed and reviewed in the following paragraphs.

Frontotemporal dementia, Lewy bodies, AD, and vascular dementia are some of the types of dementia. The most common type of dementia is AD [6]. There are studies employing data mining methods to diagnose AD, as shown in Table 1.

Plant *et al.* [7] conducted a study to identify AD-related brain regions using an SVM, Bayesian classifier, and voting feature intervals. Feature selection, classification, and clustering operations were applied to brain MRIs obtained from a private dataset to detect the most discriminative parts of the brain. They achieved an accuracy of 92% using the Bayesian method. According to their study, the prefrontal cortex, adjacent subcortical basal ganglia, hippocampal region, and posterior brain regions are regions exposed to this change.

Poulin *et al.* [8] also applied a study compared brain regions and detected disease-related areas. Their data were obtained from the ADNI and Open Access Series of Imaging Studies (OASIS). They used FreeSurfer software to obtain volumetric and statistical analysis results. Consequently, amygdala atrophy and hippocampal atrophy are disease-related areas at different stages of dementia severity.

MRI, FDG-PET, and CSF were combined by Zhang *et al.* [9] to determine differences between AD and normal groups. They used the ADNI dataset and SVM for classification. The accuracies were 86.2%, 90.6%, and 93.2% for MRI, MRI combined with FDG-PET, and all three modalities, respectively. They also ranked regions according to their SVM weights. Hippocampal formation as right, hippocampal formation as left, and right amygdala were the best three of the top 11 regions for classification.

For a similar purpose, Hinrichs *et al.* [10] used ADNI to analyze the progression of mild cognitive impairment. They used MRI and FDG-PET imaging modalities, and CSF, apolipoprotein E (APOE), and cognitive scores non-imaging modalities. Multi-Kernel Learning (MKL) and SVM were used for the classification task. MKL outperformed SVM by approximately 3%-4% in terms of accuracy. An accuracy of 87.6% was obtained with MRI and PET modalities, whereas an accuracy of 92.4% obtained with all APOE, CSF, PET, MRI, and cognitive measurements.

Westman *et al.* [11] used the ADNI dataset in their study to combine the MRI and CSF data. They used orthogonal partial least squares (OPLS) for classification. An accuracy of 91.8% was obtained for AD-HC classification.

Behesti & Demirel [12] used t-test to classify AD data. ADNI dataset was used. Voxel-based morphometry was employed and volumes of interest were produced. The created features were ranked using t-test scores. Then, the FC between the AD and HC groups was calculated to select the optimal number of discriminative features. SVM was used, and 96.3% accuracy was obtained using MRI-only data.

Zhu *et al.* [13] suggested a graph feature selection method and combined it with an SVM using MRI from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset for dementia diagnosis. The best result was 91.2% accuracy.

TABLE 1. Overview of the studies that use similar techniques as this paper (chronological).

Year	Authors	Used Methods	Modality	Dataset	Purpose
2010	Plant et. al. [7]	SVM, Bayesian Classifier, Voting Feature Intervals	MRI	Private dataset	Classification and finding most related brain region
2011	Poulin et al. [8]	linear regression analyses, t statistic	MRI, MMSE, Clinical Dementia Rating Sum of Boxes (CDR-S)	ADNI and OASIS	finding most related brain region
2011	Zhang et al. [9]	SVM	MRI, FDG-PET, and CSF	ADNI	Classification and finding most related brain region
2011	Hinrichs et al. [10]	MKL and SVM	MRI and FDG-PET imaging modalities, CSF, Apolipoprotein E (APOE) and Cognitive Scores	ADNI	Classification
2012	Westman et al. [11]	OPLS	MRI and CSF	ADNI	Classification
2016	Behesti & Demirel [12]	Propose a feature selection technique that uses t-test and Fisher Criterion. SVM is for classification	MRI	ADNI	Classification
2016	Zhu et al. [13]	Suggested a grap feature selection and combined it with SVM	MRI	ADNI	Classification
2017	Long et al. [14]	FreeSurfer	MRI	ADNI	Classification
2017	Sorensen et al. [15]	Linear Discriminant Analysis	MRI	ADNI, AIBL and CADDementia	Classification
2017	Behesti and Demirel [16]	T-test score, binary GA, and SVM	MRI	ADNI	Classification
2020	Hao et al. [17]	Multi-kernel SVM	MRI, FDG-PET	ADNI	Classification
2021	Zhang et al. [18]	A novel multiclass classifier based on SVM is proposed.	MRI, PET, and CSF	ADNI	Classification
2021	Buyrukoglu [19]	A predictive model was built applying various data mining algorithms	MRI	ADNI	Classification

Long *et al.* [14] also used ADNI in their study. They employed FreeSurfer software for pre-processing and feature extraction. SVM used for classification and 96.5% was calculated as accuracy for AD-HC classification.

Sorensen *et al.* [15] employed ADNI, AIBL (Australian Imaging Biomarkers and Lifestyle), and the CADDementia Challenge to collect data. Some measurements such as cortical thickness, volumetry, hippocampal shape, and texture of MRI biomarkers were combined in their study for the purpose of the diagnosis of mild cognitive impairment (MCI), HC, and AD. FreeSurfer software was used to extract features from the data. After the feature selection (FS) process, for classification, linear discriminant analysis was used. The outputs showed that hippocampal volume, ventricular volume, and hippocampal texture were the three most selected features in the study. As a result of the 3-class classification, 62.7% for ADNI and AIBL, and 63% for CADDementia Challenge were obtained with accuracy.

In another study [16], the method was slightly changed by Behesti and Demirel. After ranking features with the t-test score, a binary GA, whose part of the objective function is the FC, was used to identify the ideal feature subset. SVM was used, and 93% accuracy acquired for AD-HC classification.

Another study that employed MRI and FDG-PET was conducted by Hao *et al.* [17]. They proposed a novel multimodal FS method with consistent metric constraints to combine the information gathered from multimodal neuroimaging data. For the classification process, the ADNI dataset and multi-kernel SVM were used.

Zhang *et al.* [18] proposed a novel SVM-based multiclass classifier. They also used multimodalities, such as CSF, PET, and MRI from the ADNI dataset for classification and extensive analysis.

A recent study was conducted by Buyrukoglu [19]. In the feature selection phase, ensemble approaches were applied by the authors. A predictive model was built for the early diagnosis of AD, and RF, ANN, logistic regression, SVM, and naive Bayes were applied to the model. Having achieved 91% accuracy, RF surpassed other methods.

An overview of the literature that focuses on the same topic and similar techniques as this study can be seen in Table 1.

Some nature-inspired swarm intelligence algorithms such as the firefly algorithm (FFA) [20], [21], ant lion optimizer [22], grey wolf optimizer [23], whale optimizer [24], salp swarm algorithm [25], grasshopper optimization algorithm [26], butterfly optimization algorithm [27] have also been implemented as feature selectors. A comprehensive and informative comparison study for swarm algorithms including artificial fish swarm algorithms (AFSA), ant colony optimization (ACO), FFA, Artificial Bee Colony (ABC), PSO,

and bat algorithms (BA) for feature selection is carried out by Basir and Ahmad [28]. A more comprehensive and detailed analysis of nature-inspired methods for feature selection was conducted in 2020 by Sharma and Kaur [29].

III. MATERIALS AND METHODS

A. MATERIALS

1) VOLBRAIN ONLINE AUTOMATED BRAIN MRI VOLUMETRY SYSTEM

The volBrain, developed by Manjon and Coupe [30], aims to automatically obtain the brain's volumetric information using their MRI data. The system preprocesses raw MRI data, produces automatic brain segmentation from the data, and generates a report summarizing the volumetric results. Researchers can use the system worldwide. According to the website, the number of users has been increasing daily, as has the number of daily jobs processed in the system [31].

The volBrain system requires raw data that has not been exposed to any preprocessing. The system takes data in nifti (Neuroimaging Informatics Technology Initiative) format with an extension of .nii and handles it to generate result reports in the portable document format (PDF) and comma-separated value (CSV) formats. Previously collected MRI data were sent to the system, which produced a report comprising volumetric and statistical information for each MRI. The dataset used in this study was created by combining the reports produced.

The output of the volBrain system includes values such as the volume of each part of the brain in cm^3, percentage of the parts by total brain area, and asymmetry of the parts with respect to the left and right sides.

B. METHODS

1) FEATURE SELECTION

Feature Selection (FS) is a procedure for reducing redundant, inconsistent, and unnecessary features when constructing a model in a dataset. A subset of features was selected by evaluating them according to certain criteria. The set of features was reduced by eliminating inessential, unnecessary, irrelevant, and noisy features. As a result of this process, speeding up the classifier model and improving performance measurements are expected effects. Generally, FS methods can be divided into three categories, which are filters, wrappers, and embedded models. Filter techniques rely on the general characteristics of data to select the feature subset process. Additionally, the selected feature subsets were evaluated independently of the learning algorithm. A predefined learning algorithm is required as an evaluation criterion in wrapper techniques to evaluate selected feature subsets. Embedded models analytically acquire feature relevance from the objective of the learning model after incorporating variable selection as part of the training stage [32], [33].

Feature selection is an active research topic and has been applied to many fields such as genomic analysis [34], image retrieval [35], [36], text mining [37], [38], intrusion detection [39], image processing/computer vision [40], bioinformatics [41], [42], fault diagnosis [43].

2) ARTIFICIAL BEE COLONY (ABC) ALGORITHM AS A FEATURE SELECTOR

An Artificial Bee Colony is a swarm-based nature-inspired algorithm proposed by Karaboga [44]. A simulation of the honey bees' foraging behavior was performed using this algorithm. The simulation was conducted by dividing the bees into three categories: employed, onlooker, and scout bees. In the algorithm, bee foods are considered optimal solutions and the bees look for the best one. In ABC, half of the population consists of employed bees and the other half are onlooker bees. Employed bees whose food sources were exhausted were turned into scout bees. So it can be understood that each employed bee is assigned one food source. An onlooker bee decides to choose a food source by waiting in a dance area. Employed bees go to food sources that have been previously visited by itself. A random search is performed by a scout bee [45] on food that is considered exhausted. The main steps of the algorithm are shown in Fig. 1.

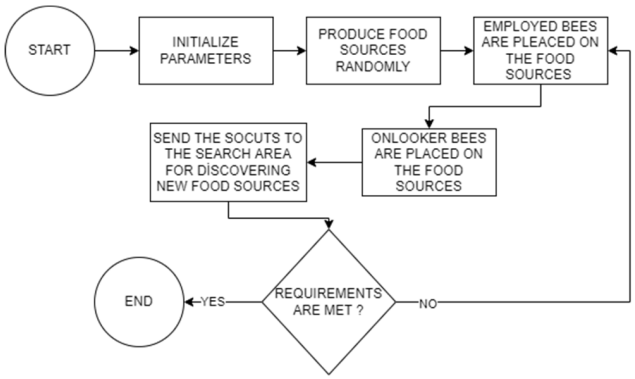


FIGURE 1. The flowchart of the ABC's main steps.

In Akay and Karaboga's study [46], ABC was explained as follows:

Initially, the algorithm randomly produced food sources. These food sources represent potential solutions to this problem. This process is performed using (1).

$$X_{ij} = X_j^{min} + rand(0, 1)(X_j^{max} - X_j^{min}) \tag{1}$$

where $i = 1,2,3 \ldots SN$ and $j = 1,2,3 \ldots D$, where D stands for the number of optimization parameters and SN stands for the number of food sources. Probable solutions, called food sources, are then employed for the search processes.

As previously mentioned, the number of employed bees is equal to the number of food sources. Employed bees are sent to food sources by assigning them one-to-one. Employed bees search for food sources and find their own neighborhoods. Subsequently, quality of the neighborhoods was assessed. Neighborhood exploration was performed using (2).

$$V_{ij} = X_{ij} + \theta_{ij}(X_{ij} - X_{kj}) \tag{2}$$

In this equation, j is a random integer in the range of [1, D]. k is an element of $\{1,2,3,\ldots \mathrm{SN}\}$ and is randomly chosen. θ_{ij} is a real random number that is uniformly distributed and is in the range $[-1,1]$. Following exploration, the fitness value for the minimization problem can be assigned to the V_{ij} using (3).

$$fitness_i = \begin{cases} \dfrac{1}{1+f_i}, & f_i \geq 0 \\ 1 + abs(f_i), & f_i < 0 \end{cases} \tag{3}$$

In the equation, f_i is the cost value. A greedy search is applied between X_i and V_i and the one with the best fitness value is selected. The new one is memorized and the old one is forgotten.

After all, information, including the fitness value of food sources, and the position of the food sources are shared with the onlooker bees. Onlooker bees choose a food source by assessing the fitness-related probabilities, as shown in (4). A positive feedback feature of the algorithm can be observed at this stage.

$$P_i = \frac{fitness_i}{\sum_{n=1}^{F} fitness_i} \tag{4}$$

If the probability value produced by (4) is greater than the random number generated by the algorithm for each source, modification is performed using (2) by an onlooker bee. The best is then chosen between the old and modified versions. If the solution is not improved, the related counter is incremented by 1; otherwise, the counter is set to 0.

Exhausted sources are checked whether there are any or not, after all employed bees and onlooker bees finished their tasks. If the counter is greater than the limit parameter of ABC, food source is considered exhausted. Finally, food source freshly created by the scout bees was replaced with the exhausted one. A pre-defined maximum cycle number, error tolerance, or fitness value can be used as termination criteria.

3) BINARY ABC FOR FEATURE SELECTION

Some changes must be made to convert ABC into a binary version and to use it as a feature selector. In the original algorithm, food sources are considered a probable optimal solution. In the binary version, these food sources were considered as a probable optimal feature subset for the solution, and the food sources are created as bit vectors. The size of the bit vector is N, where N denotes the total number of features. In this bit-vector representation, the value of the bit determines whether the related feature is a component of the feature subset. The feature assigned to the related position is a component feature subset if the value is one. A related feature is not included in the feature subset if the value is zero. An example of the bit vector representation of the features is presented in Table 2. The fitness value for probable solutions can be considered as their F-measure, accuracy, error rate, or the ratio between these and the number of selected features.

Considering the example in Table 2, it can be said that F1, F5, F7, and F8 are selected for the feature subset, where

TABLE 2. An example of bit vector representation for features.

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
1	0	0	0	1	0	1	1	0	0

F1 stands for the feature one so on. As stated above, the main goal of feature selection is to obtain maximum performance with the minimum number of features. The fitness function indicated in (5) was used to evaluate each probable feature subset.

The fitness function used in this study was based on the calculation of the probable solution classification error and the ratio of the number of selected features to that of all features. To obtain a classification error, the function requires a classifier for the mechanism. The KNN [47], RF and SVM classifiers were utilised for this purpose. The fitness function helps maintain a balance between the number of selected features and accuracy. This function and classifiers are used for all the BABC, binary grey wolf optimization (BGWO) [23], binary particle swarm optimization (BPSO) [48], and binary differential evolution (BDE) [49] algorithms. The reason for choosing them is its widespread use in literature.

$$fitness_i = \alpha ER + \beta \frac{|R|}{|N|} \tag{5}$$

where ER is the classification error rate produced by classifiers which is equal to accuracy-1. $|R|$ is the number of selected features and $|N|$ is the total number of features. α and β correspond to the error rate and importance of the feature subset length, respectively. α is in the range [0,1] and β =(1-α). The fitness function and the values of α and β were adopted from [22] and [24].

The main steps of the binary ABC algorithm for feature selection are as follows:

S1. Initialize the population randomly using (6) where $i=\{1,2,3,\ldots \mathrm{npop}\}$ and $j=\{1,2,3,\ldots N\}$. The variables npop and N represent the population (number of agents in the algorithms) and the total number of features, respectively. The variable $\mathrm{rand_j}$ is the generated random variables in the range $[0, 1]$. The value 0.5 is selected as the threshold to give the same chance for 1 and 0, at the beginning.

S2. Evaluate the fitness of the population using fitness function in (5).

S3. Generate random numbers between 0 and 1 for each feature. Then, the reverse bits according to (7) where $rand_j$ is the generated random variables in the range $[0, 1]$, $j=\{1,2,3,\ldots N\}$ and the MR is the pre-defined rate of modification. If the random variable is less than or equal to MR, flip the related bit by changing the value to 0 if it is 1 and to 1 if it is 0. Otherwise, leave the bit as it is.

S4. If the newly generated solution has better fitness, change it with the original and set the limit counter

as zero. If it does not, increase the limit counter which is used to detect exhausted sources.

S5. Calculate the probability for each solution using (8) where P_i is the probability of i^{th} solution, and fitness_i is the fitness value of i^{th} solution.

S6. Select the value i using Roulette Wheel Selection [50] method by employing P_i value.

S7. Apply a modification to i^{th} solution as mentioned in step (S3). If the produced solution is better, change them and set the limit counter to zero. Otherwise, increase the limit counter.

S8. If the limit counter of a probable solution is greater than the limit, create a new individual and change them. Then, set the limit counter as zero.

S9. Memorize the best solution.

S10. Repeat the steps between (S3) and (S9) until termination criteria are met.

S11. Steps number (S3) and (S4) represent the employed bee phase, steps between (S5) and (S7) stand for the onlooker bee phase, and step (S8) is for the scout bee phase.

$$solution_{ij} = \begin{cases} 1, rand_j > 0.5 \\ 0, otherwise \end{cases} \quad (6)$$

$$solution_{ij} = \begin{cases} Reverse\ the\ bit,\ rand_j < MR \\ Do\ not\ change, otherwise \end{cases} \quad (7)$$

$$P_i = \frac{fitness_i}{\sum_{n=1}^{npop} fitness_i} \quad (8)$$

In Table 3, parameter settings used for experiments can be seen.

TABLE 3. Parameter settings for experiments.

PARAMETERS	VALUE(S)
K-fold for cross validation	10
k for KNN	5
Number of iteration	100
Length of bit vector (*nvar*)	Number of features
Number of population (*npop*)	10
α for fitness function	0.99
Limit counter for BABC	0.6 x *nvar* x *npop*
Modification rate in ABC	0.2
Inertia weight in BPSO	Linearly decrease from 0.9 to 0.4
c1 in BPSO	2
c2 in BPSO	2
Maximum velocity in BPSO	6
Coefficient parameters in BGWO	Decreased linearly from 2 to 0

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EVALUATION METRICS

Binary classification was applied in this study, and the methods used were measured using the performance metrics described under this title.

The dataset contained two classes: AD and HC. The F-measure as shown in Equation (9), accuracy as shown in Equation (10), and error rate can be calculated considering the true positive (TP), which means that the actual class is HC, and the predicted label is HC; true negative (TN), which indicates that the actual label is AD and the predicted label is AD; false positive (FP), which means that the actual label is AD, but the predicted label is HC; and false negative (FN), which indicates that the actual label is HC, but the predicted label is AD as shown in Table 4.

TABLE 4. Confusion matrix.

		Predicted Class	
		HC	AD
Actual Class	HC	TP	FN
	AD	FP	TN

The fundamental evaluation metric for this study, the fitness value, was calculated as shown in (5). Using this fitness function, the classifier error rate and the number of selected features are balanced. This formula ensures that a solution with fewer features is selected when two solutions have similar error rates. The importance of the two metrics in the fitness formula, such as the number of features and error rate, can be determined using the alpha and beta values. MATLAB (Matrix Laboratory) was used for coding. The termination criterion is the maximum cycle number, which is equal to 100 for all algorithms.

$$F - measure = 2\,x\frac{RecallxPrecision}{Recall + Precision} \quad (9)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

In this study, the BABC was used as the feature selector. Binary versions of two swarm-based algorithms, BGWO [23] and BPSO [48], were employed to measure the differences in performance between the algorithms. Another approach used in the comparison was a successful hybrid implementation of differential evolution (BDE) and ABC for feature selection [49]. The powers of differential evolution and ABC were merged and used for feature selection by Zorarpaci and Ozel [49].

B. THE DATASET

MRI is a non-invasive imaging technology. It is used for diagnosis, disease detection, and treatment monitoring. This method creates a three-dimensional image of the body parts by capturing the patient's body from the coronal, sagittal, and axial planes. Fig. 2 depicts each plane of the brain MR image obtained using ADNI.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary

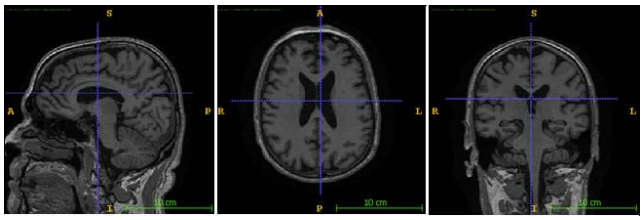


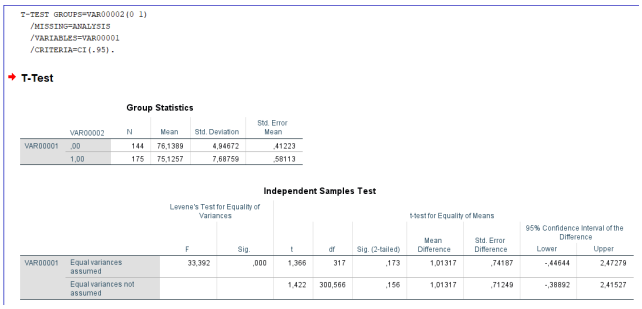
FIGURE 2. View of Sagittal (a), Axial (b) and Coronal (c) plane of the brain MRI.

goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). A total of 144 healthy persons grouped as healthy controls (HC) and 175 patients with AD were included. The data consisted of 152 men and 167 women aged between 55 and 91 years. The age statistics of the dataset, which are grouped by health status and gender, are shown in Table 5. The Independent Samples T-Test was applied to check whether there was a significant difference between classes in terms of age. As a result of the test as shown in Fig. 3, it was found that the significance value was 0.156 and thus did not meet the $p<0.05$ condition. Accordingly, when the results of the test and the averages of the groups were evaluated together, it was concluded that there was no significant difference between age information and Alzheimer's Disease. It is thought that the reason why there is no significant difference between age groups is due to the fact that the individuals in the sample are mostly between the ages of 62-91, that is, the averages of the classes are close to each other and between 75-76, there is no individual in the sample who can be described as young, and therefore they cannot be represented by any group.

TABLE 5. The age statistics of data.

Statistical Information	AD		HC	
	Female	Male	Female	Male
Minimum	55	56	62	60
Maximum	91	88	90	87
Mean	74.47	74.91	74.74	75.45
Standard Deviation	7.71	10.78	10.13	9.89
Number of Samples	83	92	69	75

It is stated that, in this study, ADNI data was not used directly. ADNI data were sent to the volBrain system as input to produce volumetric statistics of the MRIs scans. The dataset consisted of the output of the volBrain online system. This created dataset has 109 features, including the class feature which specifies whether or not the patient has AD. Details about volBrain are given in the section entitled ''volBrain online automated brain MRI volumetry system''.



```
T-TEST GROUPS=VAR00002(0 1)
  /MISSING=ANALYSIS
  /VARIABLES=VAR00001
  /CRITERIA=CI(.95).
```

T-Test

Group Statistics

	VAR00002	N	Mean	Std. Deviation	Std. Error Mean
VAR00001	,00	144	76,1389	4,94672	,41223
	1,00	175	75,1257	7,68759	,58113

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
VAR00001	Equal variances assumed	33,392	,000	1,366	317	,173	1,01317	,74187	-,44644	2,47279
	Equal variances not assumed			1,422	300,566	,156	1,01317	,71249	-,38892	2,41527

FIGURE 3. The independent samples T-Test results regarding age.

The dataset includes volumetric and statistical information of brain parts such as tissue, cerebrum, cerebellum, brainstem, lateral ventricles, caudate, thalamus, globus pallidus, hippocampus, amygdala, and accumbens.

C. RESULT TABLES

In this section, comparisons between the BABC and traditional methods, and the BABC and metaheuristics are presented, and tables are inferred. For all the tables, the average number of features was rounded to the nearest integer. Table 6 shows accuracy and F-measure of the BABC, info gain (IG), gain ratio (GR), chi-square (CHI), and ReliefF. Info Gain selects features by employing a decision tree structure [33]. Similarly, GR as well uses a tree structure for feature selection. It is an extension of IG and copes with the bias of IG using normalization [33]. ReliefF is a filter method for feature selection. It figures out a feature score for each feature and uses this score to select the most relevant features [33]. The measurements were performed using the same number of features. In terms of accuracy and F-measure, the BABC surpasses mentioned traditional methods.

For a comprehensive comparison, alongside KNN, different classifiers were used, such as RF and SVM, in the experiments. Accuracy and F-measure values for comparison of BABC and IG, GR, CHI, and ReliefF are given in Table 6. For the same number of selected features, BABC surpasses others. Traditional methods show similar performance.

TABLE 6. The accuracy and f-measure of BABC, IG, GR, CHI, and ReliefF.

Statistic	BABC	CHI	IG	GR	ReliefF
Avg. Num. of SF	52	52	52	52	52
Accuracy	0.839	0.799	0.799	0.799	0.809
F-measure	0.831	0.800	0.800	0.800	0.809

The best, worst, average, and standard deviation of the fitness and accuracy for each metaheuristic algorithm and each classifier are listed in Tables 7 and 8, respectively. These measurements were obtained after 10 runs. In addition, the average and standard deviation (SD) of the running time,

F-measure, and number of selected features (No. of SF) are depicted in Table 9. The convergence curves of fitness and accuracy for the utilised algorithms are shown in Figures 4, 5, 6, 7, 8 and 9.

TABLE 7. Best, worst, average, and standard deviation of fitness values for all algorithms and classifiers.

Statistic	Classifier	BABC	BGWO	BPSO	BDE
Best (No. of SF)	KNN	0.157 (57)	**0.115 (29)**	0.156 (42)	0.161 (70)
	RF	0.139 (58)	**0.085 (23)**	0.110 (53)	0.106 (67)
	SVM	0.163 (57)	**0.096 (29)**	0.134 (43)	0.142 (57)
Worst (No. of SF)	KNN	0.171 (42)	**0.139 (29)**	0.174 (52)	0.178 (80)
	RF	0.151 (58)	**0.104 (28)**	0.113 (49)	0.118 (72)
	SVM	0.163 (57)	**0.127 (30)**	0.140 (44)	0.158 (65)
Avg. (Avg. No of SF)	KNN	0.164 (52.2)	**0.130 (27)**	0.165 (51.8)	0.172 (69.7)
	RF	0.141 (58)	**0.096 (25.1)**	0.112 (51)	0.112 (69.5)
	SVM	0.163 (57)	**0.110 (27.1)**	0.137 (43.5)	0.154 (65.3)
Standard Deviation	KNN	**0.0038**	0.0078	0.0057	0.0052
	RF	0.0049	0.0072	**0.0013**	0.006
	SVM	**0**	0.015	0.003	0.006

The best results are indicated in bold style.

Table 7 lists the successes of the BGWO algorithm. It reaches the best average fitness value among all methods and classifiers by obtaining 0.130, 0.096, and 0.110 with selected 27, 25.1, and 27.1 features for KNN, RF, and SVM, respectively. For the average fitness value obtained by using KNN, the BABC is in second place with a 0.164 fitness value and an average number of features of 52.2. BPSO and BDE share second place when it comes to using RF. When the classifier was SVM, BPSO ranked second. Whereas BGWO has the best score for average fitness, BABC ranks first in the comparison of the standard deviations of fitness. Even though this demonstrates that the BABC algorithm is more stable than the others, the range of results of BABC's fitness is out of the range of BGWO. One likely cause for this stability may be the well-working exploration of bees whose foraging behaviour is implemented in the method. A possible explanation for the BGWO having the best average fitness is its detailed exploration and exploitation phase. The well-implemented social hierarchy and hunting mechanism of grey wolves make this method comprehensive. Other results, such as Best and Worst, are also shown in Table 7. It is inferred when all the results are considered that RF produces the best results for all methods.

Figures 4, 5 and 6 show the convergence curves of the fitness for all the algorithms using KNN, RF and SVM, respectively. The figures indicate the reflection of the exploration and exploitation mechanisms mentioned in the results. At the beginning of the experiment, all methods were in a competitive race, whereas BGWO attained a better convergence curve at the end.

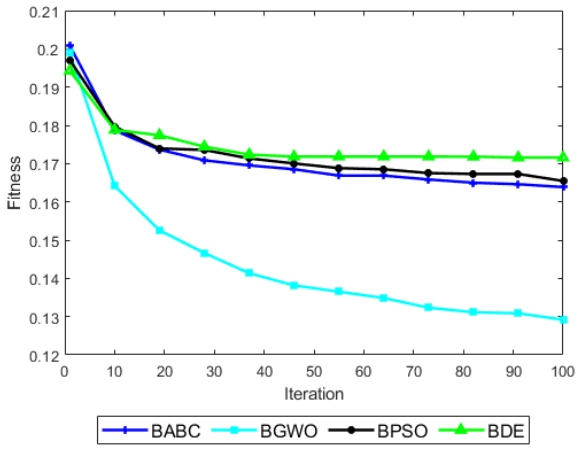


FIGURE 4. Convergence curves of fitness for all algorithms with KNN.

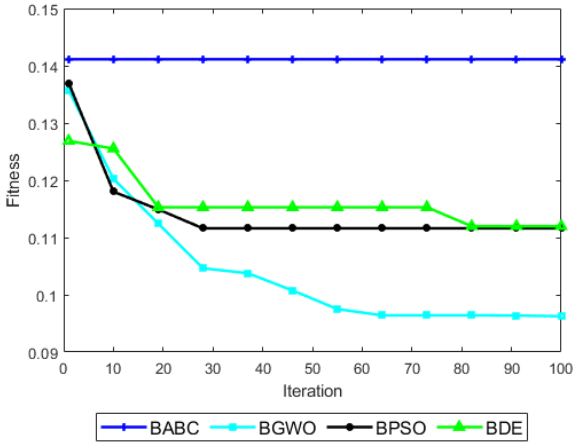


FIGURE 5. Convergence curves of fitness for all algorithms with RF.

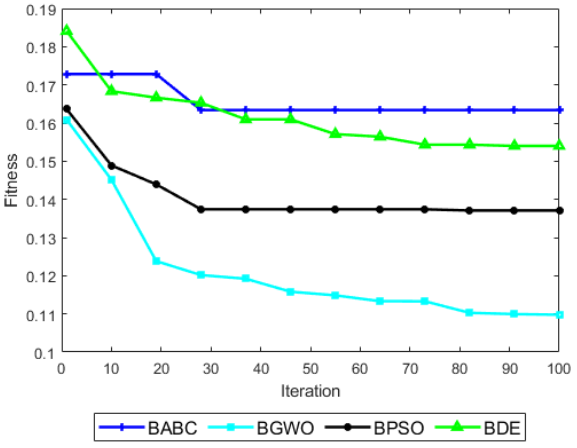


FIGURE 6. Convergence curves of fitness for all algorithms with SVM.

In Fig. 4, BABC is in second place while the winner is BGWO. In Fig. 5, BPSO and BDE have the same score and

TABLE 8. Best, worst, average, and standard deviation of accuracy values for all algorithms and classifiers.

Statistic		BABC	BGWO	BPSO	BDE
Best (No. of SF)	KNN	0.847 (57)	**0.887 (29)**	0.847 (42)	0.844 (70)
	RF	0.865 (58)	**0.916 (23)**	0.894 (53)	0.899 (67)
	SVM	0.840 (57)	**0.906 (29)**	0.868 (43)	0.862 (57)
Worst (No. of SF)	KNN	0.831 (42)	**0.862 (29)**	0.829 (52)	0.827 (80)
	RF	0.853 (58)	**0.897 (28)**	0.891 (49)	0.887 (72)
	SVM	0.840 (57)	**0.875 (30)**	0.863 (44)	0.847 (65)
Avg. (Avg. No of SF)	KNN	0.839 (52.2)	**0.872 (27)**	0.838 (51.8)	0.833 (69.7)
	RF	0.863 (58)	**0.905 (25.1)**	0.892 (51)	0.893 (69.5)
	SVM	0.840 (57)	**0.892 (27.1)**	0.866 (43.5)	0.851 (65.3)
Standard Deviation	KNN	**0.004**	0.008	0.0054	0.0047
	RF	0.005	0.007	**0.0015**	0.006
	SVM	**0**	0.015	0.003	0.006

The best results are indicated in bold style.

take the second row while in Fig. 6, BPSO stays the second best alone.

Table 8 includes the accuracy's best, worst, average, and standard deviation. The table points out that RF is the best classifier to use in terms of accuracy. As like in Table 7, BGWO surpasses its rivals. By using KNN, BABC takes second place. For SVM and RF, BPSO and BDE show better performance than BABC. However, the results are generally close to each other, except for BGWO.

Figures 7, 8 and 9 show the convergence curves for accuracy achieved using KNN, RF and SVM, respectively.

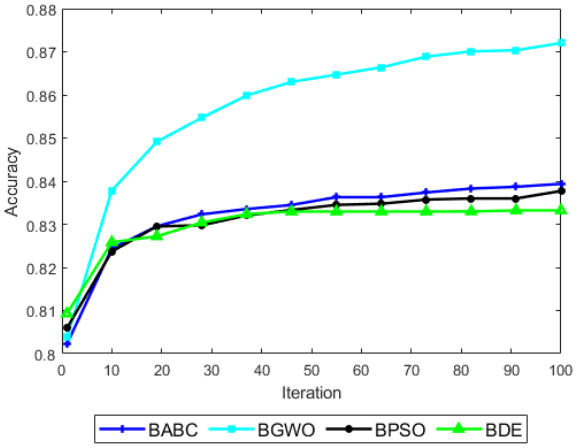


FIGURE 7. Convergence curves of accuracy for all algorithms with KNN.

BABC, BPSO, and BDE continued to increase slowly and almost horizontally after the 10^{th} iteration. In contrast, the BGWO proceeds to increase its performance steeply after the 10^{th} iteration.

In Fig. 7, the methods are close to each other at the beginning. As the iteration number increase, BGWO's performance shows higher performance. BABC is in the second row by having slightly better accuracy. It is obvious that BGWO's performance is the best in Fig. 8 and Fig. 9, too.

Table 9 shows the average and standard deviation (SD) of running time, F-measure, and the number of selected features (SF). The average number of SF was already included in Tables 7 and 8. The running time results are presented in seconds. The results are close to each other. The BDE ranks first for the average running time with KNN. BGWO is in the second row, and BABC takes third place. If RF and SVM are considered, BGWO and BABC have the best

TABLE 9. Average and standard deviation of running time (in seconds), f-measure, and No. of SF for all algorithms.

Statistic	Classifier	BABC	BGWO	BPSO	BDE
Avg. Running (sec.) Time	KNN	49.324	48.716	51.786	**48.192**
	RF	94.005	**5.58e+03**	6.1e+03	6.66e+03
	SVM	**67.264**	214.5	307.28	376.997
Std. Dev. of Running Time (sec.)	KNN	1.998	**0.884**	2.478	1.671
	RF	**12.366**	127.3	80.120	294.89
	SVM	6.324	24.164	**5.054**	17.283
Avg. F-Measure	KNN	0.831	**0.865**	0.830	0.826
	RF	0.850	**0.894**	0.881	0.882
	SVM	0.835	**0.883**	0.851	0.834
Std. Dev. of F-Measure	KNN	**0.005**	0.009	0.008	0.006
	RF	0.008	0.009	**0.002**	0.009
	SVM	**0**	0.0164	0.004	0.008
Avg. No. of SF	KNN	52.2	**27**	51.8	69.7
	RF	58	**25.1**	51	69.5
	SVM	57	**27.1**	43.5	65.3
Std. Dev. of SF	KNN	5.287	**4.216**	5.329	10.686
	RF	**0**	2.51	2.108	2.635
	SVM	**0**	2.92	0.527	5.208

The best results are indicated in bold style.

TABLE 10. Hippocampus, Amygdala, Globus Pallidus related regions selected by all methods in at least 8 out of 10 runs.

Related Regions	KNN				RF				SVM			
	BABC	BGWO	BPSO	BDE	BABC	BGWO	BPSO	BDE	BABC	BGWO	BPSO	BDE
Amygdala asymmetry				*			*	*	*		*	*
Amygdala's total volume		*			*			*				*
Globus pallidus asymmetry					*				*			
Hippocampus asymmetry	*			*	*				*			
Hippocampus' total volume				*	*			*	*			*
Percentage of amygdala's total volume							*	*	*			*
Percentage of globus pallidus' total volume					*							
Percentage of hippocampus' total volume	*		*	*			*		*	*		
Percentage of the left of hippocampus' total volume		*	*		*			*	*			*
Percentage of the right of hippocampus' total volume	*		*	*				*	*	*		*
Percentage of volume of the left of amygdala			*				*	*	*		*	*
Percentage of volume of the left of globus pallidus							*	*				*
Percentage of volume of the right of amygdala	*			*	*		*	*	*	*	*	*
Percentage of volume of the right of globus pallidus					*		*		*			
Total volume of globus pallidus				*					*			
Volume of the left of amygdala			*					*				*
Volume of the left of globus pallidus	*		*		*					*		
Volume of the left of hippocampus	*								*			
Volume of the right of amygdala				*								
Volume of the right of globus pallidus				*	*							
Volume of the right of hippocampus				*	*				*			

results, respectively. The first place is taken by BGWO when it comes to the standard deviation of running time using KNN, the average F-measure with all classifiers, the average number of SF with all classifiers, and standard deviation of SF using KNN. The BABC is declared the best in terms of the SD of the F-measure for KNN and SVM. BABC ranks third in terms of the average number of SF by obtaining 52.2, 58, and 57 using KNN, RF, and SVM, respectively. In the results of the average f-measure, BABC, BDE, and BPSO have the second-best results for KNN, RF, and SVM individually. Considering all tables and classifiers, it can be pointed out that all methods achieve their best results using RF as the classifier.

Another aim of this study was to identify the brain regions that are most associated with AD. The brain regions in the dataset were statistical and volumetric measurements of tissue (gray matter, white matter, cerebrospinal fluid, and brain), cerebrum, cerebellum, brainstem, lateral ventricles, caudate, putamen, thalamus, globus pallidus, hippocampus, amygdala, and accumbens. According to the outputs, the brain regions selected in at least eight out of 10 runs for each classifier were presented in Table 10. When the results were examined, it was inferred that the regions chosen by all four methods at least eight times over ten runs were hippocampus-related and amygdala-related.

As shown in Table 10, volume-based statistics of the hippocampus and its asymmetry relative to the whole brain can be used as indicator for the diagnosis of AD. Similarly, volume-based statistics of the amygdala and its degree of asymmetry can be a supportive criterion for detecting AD, according to the table. The hippocampus is responsible for both short-term and long-term memories. In addition, it plays a fundamental role in spatial memory. A primary role is played by the amygdala in the emotional response.

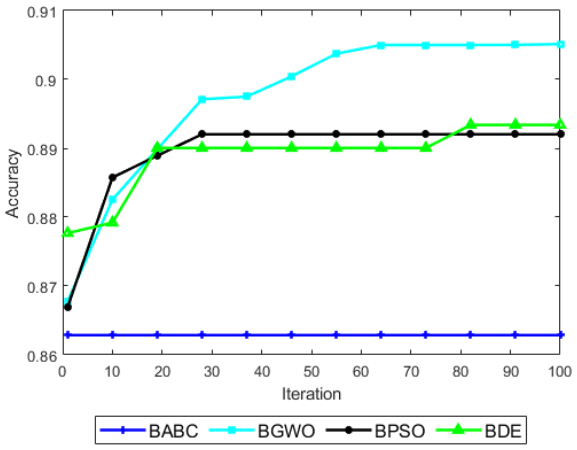


FIGURE 8. Convergence curves of accuracy for all algorithms with RF.

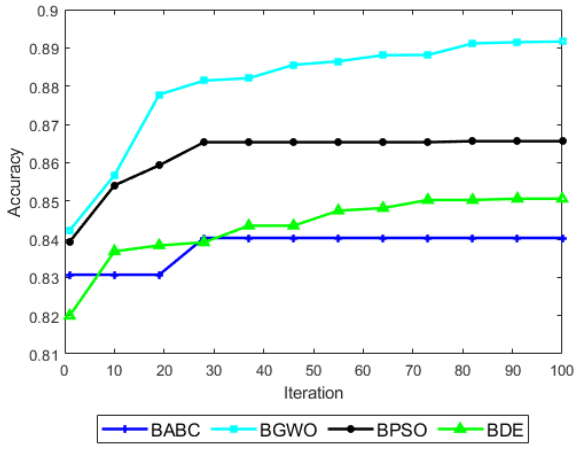


FIGURE 9. Convergence curves of accuracy for all algorithms with SVM.

The volume shrinkage of these regions causes difficulty in accomplishing the related tasks. These findings are consistent with previous studies [7], [8], [9], [15].

The globus pallidus is another attention-taking brain part in AD, according to the results of this study. Asymmetry of and the total volume's percentage of this part are selected only by BABC in the experiments. Besides BABC, other methods, as well, choose globus pallidus-related features such as right volume, left volume, and their percentages. The globus pallidus is liable to regulate voluntary movement. This damage gives rise to involuntary muscle tremors [51]. It was selected by methods at least eight times within ten runs. According to recent studies in the literature, it has been observed that changes in mitochondrial morphology in globus pallidus can also be seen in Alzheimer's disease [52], [53], [54]. The methods and the selected globus pallidus regions are presented in Table 10.

This section begins by describing the evaluation metrics and the dataset and comparing the outputs by illustrating the figures and the results table. The chapter ends with a clarification of the most commonly selected brain regions. The next section presents the conclusions of this study.

V. CONCLUSION

The main goal of the current study was to classify brain volumetric data by proposing a binary version of the ABC algorithm. To make a comprehensive comparison, traditional methods such as IG, GR, CHI, and ReliefF, and binary versions of three methods (BGWO, BPSO, and BDE) were used. In the experiments, all runs were performed using MATLAB software. The comparison between traditional data mining algorithms and BABC showed BABC's superiority in terms of accuracy and F-measure. The best, worst, average, and standard deviation of algorithms are presented in the tables. Each algorithm was run ten times to get an average performance. In addition, convergence curves of fitness and accuracy were plotted. The maximum number of cycles is used as the termination criterion. Three classifiers, namely KNN, RF, and SVM, were employed individually in the binary mechanism of the methods. In general, the binary grey wolf algorithm achieved the best score. In terms of the best, worst, and average fitness, BGWO was the first place for all classifiers. The average fitness values for the KNN, RF, and SVM were 0.13, 0.096, and 0.11, respectively. The fitness values obtained by the BABC were 0.64, 0.141, and 0.163 for KNN, RF, and SVM, respectively. In the evaluation of accuracy, the BGWO achieved superiority for the best and average values. In the fitness and accuracy results, the BABC has the minimum standard deviation, which symbolizes the lowest volatility, with the KNN and SVM. The convergence curves indicate that, except for BGWO, the other methods are in competition. Among all three classifiers, RF was the classifier on which all methods achieved their best personal results on. The average fitnesses acquired using RF were 0.096, 0.141, 0.112, and 0.0112 for BGWO, BABC, BPSO, and BDE, respectively. The average accuracy values with RF were 0.905, 0.863, 0.892, and 0.893 for BGWO, BABC, BPSO, and BDE, respectively.

The second aim of this study was to investigate brain regions that may be related to AD. The results of determining the most AD-related regions were in agreement with those in the literature. According to these results, the hippocampus and amygdala, which play important roles in memory and emotional responses, respectively, were the most relevant regions. In addition to these regions, the globus pallidus, which is responsible for unintended muscle shaking, calls for attention, and it can be seen as a finding different from the literature. When people with AD are observed, it can be noticed that they have difficulties managing the activities for which these regions are responsible.

One of the points that distinguishes this study from its peers is that it is one of the few studies to use the volBrain online system. The results of this study are important in terms of having an idea about the usefulness of using this online system.

The biggest limitation of this study was the limited sample size. Notwithstanding its relatively limited sample size, this study offers valuable insights into the determination of AD. It is unfortunate that obtaining appropriate and sufficient

health records for such research is one of the most difficult parts of the study. Employing more samples means many studies such as this. In future work, a combination of different methods and utilization of multimodal health data will be planned.

ACKNOWLEDGMENT

The authors thank to the Laboratory for Neuro Imaging at the University of Southern California, which was responsible for distributing the ADNI data.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Company Inc.; Meso Scale Diagnostics LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

[1] M. Prince, A. Wimo, M. Guercher, G.-C. Ali, Y.-T. Wu, and M. Prina, "The global impact of dementia: An analysis of prevalence, incidence, cost & trends," Alzheimer's Disease Int., London, U.K., World Alzheimer Rep. 2015, Sep. 2015.

[2] C. Patterson, "The state of the art of dementia research: New frontiers," Alzheimer's Disease Int., London, U.K., World Alzheimer Rep. 2018, Sep. 2018.

[3] (Aug. 2014). Dementia Australia. *Early Diagnosis of Dementia*. Accessed: May 2019. [Online]. Available: https://www.dementia.org.au/information/diagnosing-dementia/early-diagnosis-of-dementia

[4] S. Sarraf, D. D. DeSouza, J. A. E. Anderson, and G. Tofighi, "DeepAD: Alzheimer's disease classifcation via deep convolutional neural networks using MRI and fMRI," *BioRxiv*, to be published, doi: 10.1101/070441.

[5] D. Hand, M. Heikki, and P. Smyth, *Principles of Data Mining*, vol. 6. Cambridge, MA, USA: MIT Press, 2011, pp. 2–6.

[6] (2019.) World Health Organization. *Dementia*. Accessed: Jul. 2, 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia

[7] C. Plant, S. J. Teipel, A. Oswald, C. Böhm, T. Meindl, J. Mourao-Miranda, A. W. Bokde, H. Hampel, and M. Ewers, "Automated detection of brain atrophy patterns based on mri for the prediction of Alzheimer's disease," *NeuroImage*, vol. 50, no. 1, pp. 162–174, 2010.

[8] S. P. Poulin, R. Dautoff, J. C. Morris, L. F. Barrett, and B. C. Dickerson, "Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity," *Psychiatry Res., Neuroimag.*, vol. 194, no. 1, pp. 7–13, 2011.

[9] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.

[10] C. Hinrichs, V. Singh, G. Xu, and S. C. Johnson, "Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population," *NeuroImage*, vol. 55, no. 2, pp. 574–589, 2011.

[11] E. Westman, J.-S. Muehlboeck, and A. Simmons, "Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion," *NeuroImage*, vol. 62, no. 1, pp. 229–238, 2012.

[12] I. Beheshti and H. Demirel, "Feature-ranking-based Alzheimer's disease classification from structural MRI," *Magn. Reson. Imag.*, vol. 34, no. 3, pp. 252–263, Apr. 2016.

[13] Y. Zhu, Z. Zhong, W. Cao, and D. Cheng, "Graph feature selection for dementia diagnosis," *Neurocomputing*, vol. 195, pp. 19–22, Jun. 2016.

[14] X. Long, L. Chen, C. Jiang, and L. Zhang, "Prediction and classification of Alzheimer disease based on quantification of MRI deformation," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0173372.

[15] L. Sorensen, C. Igel, A. Pai, I. Balas, C. Anker, M. Lillholm, and M. Nielsen, "Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry," *Neuroimage Clin.*, vol. 13, pp. 470–482, Jan. 2017.

[16] I. Beheshti, H. Demirel, and H. Matsuda, "Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Comput. Biol. Med.*, vol. 83, pp. 109–119, Apr. 2017.

[17] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang, S. L. Risacher, A. J. Saykin, X. Yao, and L. Shen, "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101625.

[18] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Inf. Fusion*, vol. 66, pp. 170–183, Feb. 2021.

[19] S. Buyrukoğlu, "Early detection of alzheimer's disease using data mining: Comparision of ensemble feature selection approaches," *Konya Mühendislik Bilimleri Dergisi*, vol. 9, no. 1, pp. 50–61, 2021.

[20] H. Banati and M. Bajaj, "Fire fly based feature selection approach," *Int. J. Comput. Sci. Issues*, vol. 8, no. 4, p. 473, 2011.

[21] Y. Zhang, X.-F. Song, and D.-W. Gong, "A return-cost-based binary firefly algorithm for feature selection," *Inf. Sci.*, vol. 418, pp. 561–574, 2017.

[22] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016.

[23] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.

[24] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neuro Comput.*, vol. 260, pp. 302–312, Oct. 2017.

[25] I. Aljarah, M. Mafarja, A. A. Heidari, H. Faris, Y. Zhang, and S. Mirjalili, "Asynchronous accelerating multi-leader salp chains for feature selection," *Appl. Soft Comput.*, vol. 71, pp. 964–979, Oct. 2018.

[26] M. Mafarja, I. Aljarah, A. A. Heidari, A. I. Hammouri, H. Faris, A.-Z. Alam, and S. Mirjalili, "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems," *Knowl.-Based Syst.*, vol. 145, pp. 25–45, Apr. 2018.

[27] S. Arora and P. Anand, "Binary butterfly optimization approaches for feature selection," *Expert Syst. Appl.*, vol. 116, pp. 147–160, Feb. 2019.
[28] M. Basir and F. Ahmad, "Comparison on Swarm Algorithms for Feature Selections Reductions," *Int. J. Sci. Eng. Res*, vol. 5, pp. 479–486, 2014.
[29] M. Sharma and P. Kaur, "A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1–25, 2020.
[30] J. V. Manjón and P. Coupé, "Volbrain: An online MRI brain volumetry system," *Frontiers Neuroinf.*, vol. 10, p. 30, Jul. 2016.
[31] (2022). VolBrain. *VolBrain: Automated MRI Brain Volumetry System*. Accessed: May 19, 2019. [Online]. Available: http://volbrain.upv.es/users.php
[32] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, "Advancing feature selection research," ASU Feature Selection Repository, School Comput., Inform., Decis. Syst. Eng., Arizona State Univ., Tempe, AZ, USA, Tech. Rep. TR-10-007, Jan. 2010.
[33] K. B. Al Janabi and R. Kadhim, "Data reduction techniques: A comparative study for attribute selection methods," *Int. J. Adv. Comput. Sci. Technol.*, vol. 8, no. 1, pp. 2249–3123, 2018.
[34] I. Inza, P. Larrañaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in DNA microarray domains," *Artif. Intell. Med.*, vol. 31, no. 2, pp. 91–103, 2004.
[35] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*. London, U.K.: Pearson, 2004.
[36] D. L. Swets and J. J. Weng, "Efficient content-based image retrieval using automatic feature selection," in *Proc. Int. Symp. Comput. Vis. (ISCV)*, Nov. 1995, pp. 85–90.
[37] G. Forman, "An extensive empirical study of feature selection metrics for text classification," *J. Mach. Learn. Res.*, vol. 3, pp. 1289–1305, Mar. 2003.
[38] T. Liu, S. Liu, Z. Chen, and W.-Y. Ma, "An evaluation on feature selection for text clustering," in *Proc. 20th Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 488–495.
[39] W. Lee, S. J. Stolfo, and K. W. Mok, "Adaptive intrusion detection: A data mining approach," *Artif. Intell. Rev.*, vol. 14, no. 6, pp. 533–567, 2000.
[40] M. Muštra, M. Grgić, and K. Delač, "Breast density classification using multiple feature selection," *Automatika*, vol. 53, no. 4, pp. 362–372, Jan. 2012.
[41] N. Dessì, E. Pascariello, and B. Pes, "A comparative analysis of biomarker selection techniques," *BioMed Res. Int.*, vol. 2013, pp. 1–10, Sep. 2013.
[42] H. Abusamra, "A comparative study of feature selection and classification methods for gene expression data of glioma," *Proc. Comput. Sci.*, vol. 23, pp. 5–14, Jan. 2013.
[43] C. Liu, D. Jiang, and W. Yang, "Global geometric similarity scheme for feature selection in fault diagnosis," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3285–3595, 2014.
[44] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Dept. Comput. Eng., Erciyes Univ., Kayseri, Turkey, Tech. Rep. TR-06, Oct. 2005.
[45] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Apr. 2007.
[46] B. Akay and D. Karaboga, "A modified artificial bee colony algorithm for real-parameter optimization," *Inf. Sci.*, vol. 192, pp. 120–142, Jun. 2012.
[47] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statist.*, vol. 46, no. 3, pp. 175–185, Jan. 1992.
[48] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. Comput. Cybern. Simul.*, Oct. 1997, pp. 4104–4108.
[49] E. Zorarpacı and S. A. Özel, "A hybrid approach of differential evolution and artificial bee colony for feature selection," *Expert Syst. Appl.*, vol. 62, pp. 91–103, Nov. 2016.
[50] D. E. Goldber, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
[51] M. Gillies, J. Hyam, A. Weiss, C. Antoniades, R. Bogacz, J. Fitzgerald, T. Aziz, M. Whittington, and A. Green, "The cognitive role of the Globus pallidus interna; insights from disease states," *Exp. Brain Res.*, vol. 235, no. 5, p. 1455, 2017.
[52] S. Baloyannis, "Mitochondrial alterations in Alzheimer's disease," *J. Alzheimer's Disease*, vol. 9, no. 2, pp. 119–126, 2006.
[53] H. Cho, J.-H. Kim, C. Kim, B. S. Ye, H. J. Kim, C. W. Yoon, Y. Noh, G. H. Kim, Y. J. Kim, J.-H. Kim, C.-H. Kim, S. J. Kang, J. Chin, S. T. Kim, K.-H. Lee, D. L. Na, J.-K. Seong, and S. W. Seo, "Shape changes of the basal ganglia and thalamus in Alzheimer's disease: A three-year longitudinal study," *J. Alzheimer's Disease*, vol. 40, no. 2, pp. 285–295, Mar. 2014.
[54] W. Li, L. Yue, and S. Xiao, "Association between internet use, cognitive function, and globus pallidus volumes: A study among the elderly in Chinese communities," *Frontiers Public Health*, vol. 10, pp. 1–7, May 2022.

MÜMINE KAYA KELEŞ (Member, IEEE) was born in Adana, Turkey, in 1986. She received the B.Sc. and M.Sc. degrees in computer engineering from Çukurova University, Adana, in 2009 and 2011, respectively, and the Ph.D. degree from the Computer Sciences Division of Electrical and Electronics Engineering, Çukurova University, in 2015. She is currently an Assistant Professor at the Department of Computer Engineering, Adana Alparslan Türkeş Science and Technology University. Her research interests include data mining, database management systems, distance education, text and web mining, information retrieval, plagiarism detection, artificial intelligence, and its applications. She is a member of IEEE Computer Society and IEEE Computational Intelligence Society.

ÜMIT KILIÇ was born in Adana, Turkey, in 1992. He received the B.Sc. degree in computer engineering from Çukurova University, in 2016, and the M.Sc. degree in nanotechnology and engineering sciences from Adana Alparslan Türkeş Science and Technology University, in 2019, where he is currently pursuing the Ph.D. degree with the Department of Computer Engineering. He is a Research Assistant at the Department of Computer Engineering, Adana Alparslan Türkeş Science and Technology University. His research interests include data mining, artificial intelligence, and their applications.