Received: 27 May 2021 | Revised: 6 September 2021 | Accepted: 9 September 2021
DOI: 10.1002/ima.22656

RESEARCH ARTICLE

WILEY

An efficient multiclass classifier for classification of Alzheimer's disease/mild cognitive impairment/Normal subjects

Lingraj Dora[1] | Sanjay Agrawal[2] | Rutuparna Panda[2] | Ajith Abraham[3]

[1]Department of Electrical and Electronics Engineering, VSSUT, Burla, India
[2]Department of Electronics and Telecommunication Engineering, VSSUT, Burla, India
[3]MIR Labs, Scientific Network for Innovation and Research Excellence, Washington, District of Columbia, USA

Correspondence
Rutuparna Panda, Department of Electronics and Telecommunication Engineering, VSSUT, Burla 768018, India.
Email: r_ppanda@yahoo.co.in

Funding information
Northern California Institute for Research and Education; Canadian Institutes of Health Research; National Institute of Biomedical Imaging and Bioengineering; National Institute on Aging; Department of Defense; National Institutes of Health, Grant/Award Number: U01 AG024904; Alzheimer's Disease Neuroimaging Initiative

Abstract
Typically, in sparse representation-based classifiers, the weight associated with each training sample is ignored, resulting in reduced accuracy. Moreover, individual binary classifiers solved a multiclass problem. It requires more time as multiple runs are needed to compute the accuracy. In this paper, we propose a novel optimal sparse representation-based classifier. It solves the ternary classification problem with improved accuracy in a single run. The ternary classification considers Alzheimer's disease versus mild cognitive impairment versus normal control in a single run. A two-stage sparse representation model is used to design the proposed classifier. To update the weight coefficients, we suggest a regularized Levenberg–Marquardt learning. It allows selecting a subset of significant training samples. To determine the appropriate subset size, we investigate an objective function in terms of classification accuracy. For optimization, we suggest a hybrid particle swarm optimization–squirrel search technique. The experiment conducted on the Alzheimer's Disease Neuroimaging Initiative database shows our method outperforms other state-of-the-art methods in terms of computation time and accuracy. The use of different training–testing partition ratios makes the proposed method immune to biased results, overfitting, and underfitting difficulties. Moreover, results are obtained from 100 iterations to confirm its stability. The suggested model may be helpful for further research in medical image analysis.

KEYWORDS
disease classification, hybrid particle swarm optimization-squirrel search algorithm, optimal sparse representation, ternary classifier

1 | INTRODUCTION

Alzheimer's Disease (AD) is a degenerative brain disease. It is a common form of dementia and a progressing irreversible disease that mainly occurs in adults. According to the Alzheimer's Association, around 50 million people worldwide suffer from AD and associated dementia.[1] Mild cognitive impairment (MCI) is a stage of memory loss or other cognitive ability declines in people who can still conduct most activities of daily living independently. It can develop for a variety of causes. Some people with MCI may go on to acquire dementia, while others may not. If the hallmark alterations in the brain are present, MCI can be an early stage in the illness continuum for neurodegenerative disorders such as Alzheimer's. MCI can recover to normal cognition or remain stable in some people. Thus, people who are suffering cognitive changes should seek help as soon as possible for proper diagnosis and treatment options. There exists a high probability that subjects with MCI may develop AD. Despite its severity, there is no report on treatment planning to cure AD or to stop its progression. However, early detection could be a preventive measure to diagnose the development of AD and its prodromal stage MCI. Researchers around the globe are working to develop computerized techniques for the early diagnosis of AD.[2–12] The primary motivation behind the development of these techniques are helping experts interpret the disease, reducing workload, reducing false treatment planning due to exhaustion, minimizing intra- and inter-expert variation, and so on. Furthermore, experts may use them to provide a second opinion.

A large and growing body of literature has investigated varieties of methods to solve the AD classification task. A weighted multi-modality sparse representation classification method (wmSRC) is studied in Reference 5. It utilized a weighted combination scheme to extract the complementary information (features) from different modalities like magnetic resonance imaging (MRI), fluorodeoxyglucose-positron emission tomography (FDG-PET), and florbetapir PET. In this approach, 340 subjects (113 AD, 110 MCI, and 117 normal control [NC]) are used for experimentation. The method's performance is measured using two binary classification problems: AD versus NC and MCI versus NC. The authors employed performance metrics such as accuracy, sensitivity, specificity, and area under receiver operating characteristics (AUC) for validation. At the same time, every subject may not go through multimodality tests for diagnosis. Multitemplate-based methods are also utilized to perform AD/MCI classification from NC.[6,7] Unlike traditional template-based methods, these schemes use multiple templates during the registration to avoid the biasing problem. In addition, they employ these templates to take out numerous sets of feature representations. A sparse feature selection algorithm is applied to reduce the dimension of the feature representation derived from the multitemplate method in.[6] Finally, an ensemble-based classifier using multiple support vector machine (SVM) outputs performs the classification. The multitemplate learning method utilized 459 subjects (97 AD, 234 MCI, and 128 NC). To assess its performance, the researchers used a 10-fold cross-validation technique. They supplied average values of their classification results.

An inherent structure-based multi-view learning (ISML) method is reported in Reference 7. It extracts the multiview feature representations of each individual using the selected template. It is tested on 459 subjects (97 AD, 234 MCI, and 128 NC). Three binary classification tasks are employed to evaluate the ISML method: AD versus NC, progressive MCI versus NC, and progressive MCI versus stable MCI. For the validation, it applied a 10-fold cross-validation approach. The average result of the 10-folds in the cross-validation is utilized to assess the ISML method's performance. However, it does not consider a similarity among the individual subjects during the feature selection. Consideration of this could probably improve the feature selection strategy. Sometimes, the nonlinear registration methods in AD classification may increase the computational complexity and response time. Taking this into consideration, a landmark-based model that avoids the registration, as well as the tissue segmentation, is suggested.[8] The landmark-based technique used 358 AD, 831 MCI, and 430 NC subjects. The authors presented results of two binary classification problems: AD versus NC and MCI versus NC. The authors employed a two-fold cross-validation technique to validate their approach. Nevertheless, its performance relies on the number of training images. Insufficient training images may decrease the accuracy.

More recent attention is focused on the machine learning techniques with the feature selection strategy in AD diagnosis. They allow us to reject irrelevant features to avoid high-dimensionality problems. The feature selection is also referred to as the detection of effective biomarkers. A relational regularized discriminative least square regression (R2DLSR) algorithm is proposed for the diagnosis of AD or MCI using multimodal images.[9] In this technique, the feature selection is achieved by using a sparse learning method that contains multi-relation regularization. The algorithm utilizes 805 subjects (226 AD, 393 MCI, and 186 NC) in the experiment. To evaluate the classification performance 10-fold cross-validation method is employed. However, it integrates a few relational information for the feature selection. Inclusion of the more effective relationships could probably boost the performance. An approach incorporating voxel-based morphometry (VBM) followed by the feature ranking is used for the early diagnosis of AD.[10] Genetic algorithm (GA) is utilized to capture an optimal feature subset. Then classification is performed using the SVM. The approach is tested on 458 subjects (160 AD, 136 MCI, and 162 NC). It is validated by utilizing the 10-fold cross-validation strategy. It is evaluated using performance metrics like accuracy, sensitivity, specificity, and AUC. The approach shows satisfactory results using the T1-weighted (T1-w) brain MRI. Thin-plate spline metric learning for the SVM (TML-SVM)——a deep learning (DL) method, is studied for AD as well as MCI classification.[11] The model uses a nonlinear metric learning algorithm, which transforms the features into a linearly separable space for the SVM. It experiments on 338 subjects (94 AD, 121 MCI, and 123 NC). Two binary classification problems: AD versus NC and MCI versus NC are used to assess the approach. The authors divided their samples into five-folds: three-folds for training, one-fold for hyper-parameter validation, and one-fold for testing. They repeated the experiment 10 times with several random five-splits to confirm generalizability. The technique is evaluated utilizing three assessment metrics: accuracy, sensitivity, and specificity. It shows successful results using the T1-w brain MRI. However, it is suggested that fine-tuning could further improve the performance.

Recent studies during the last 5 years also demonstrate that techniques combining information from multi-modal medical images are proven to give accurate results. For instance, nonlinear graph fusion (NGF) uses the complementary information from multiple biomarkers: MRI, FDG-PET, and cerebrospinal fluid.[12] The NGF technique employed a total of 147 subjects (37 AD, 75 MCI, and 35 NC) in the study. The technique is investigated in a variety of settings, including AD versus NC, MCI versus NC, and AD versus MCI versus NC with multiclass classification. The authors used a 25%-out cross-validation method. Seventy-five percent of the total samples are randomly picked for training in each validation, while the remaining 25% are used for testing. To eliminate the sample bias, the validation is done over 100 times. Over 100 iterations, the average classification accuracy and average AUC are calculated. However, the technique is validated with few medical images only. The performance of all the models discussed above is listed in Table 5. In view of all the literature that is mentioned so far, most of the baseline methods use the binary classifier. They need three runs to solve the ternary classification problem—AD versus MCI versus NC, that is, AD versus MCI, AD versus NC, and MCI versus NC, which requires more computational time. This has motivated us to develop a single classifier to perform the ternary classification task in a single run.

Despite the advancements in the last few decades, early detection of AD and its prognosis remains a difficult task and requires further research. Another reason for our motivation is to develop a better substitute for the existing techniques. In this context, a sparse representation-based classifier is a better substitute to solve the problem at hand. It uses sparse coding to represent a test sample as a weighted linear mixture of the training samples. It is effectively used in a variety of pattern recognition and medical image processing tasks. The benefits of using the sparse representation-based classifier are manifold. It transforms the input data to the sparse domain, resulting in several zeros in the data matrix. This transformation helps to keep the information in a compressed form. It is also cheaper to store. Sparse representation-based classifiers have excellent properties of self-learning, adaptivity, generalizing tasks, nonlinearity, and so on. The goal is to classify the unknown test sample to its relevant class. If we transform the original representation into a sparse representation, then it aids in deciding the contribution of relevant and irrelevant training examples for accurate classification.[13–15]

Typically, such approaches rely on sparsity. It is evaluated using l_0-norm or l_1-norm technique. As noted by Zhang et al., instead of evaluating sparsity utilizing l_0-norm or l_1-norm, the collaborative representation in sparse-based classifiers is mainly responsible for improving the classification accuracy.[16] The method implements a l_2-norm-based regularized least square approach to evaluate the sparsity in the representation. It performs the classification with less computational complexity as well as response time. The l_2-norm-based sparse representation approaches have achieved much attention in the face recognition application.[16–18] In most of the sparse representation-based classifiers, the sparsity is evaluated for all the training samples. Recently, a new subset selection strategy is implemented in the sparse representation-based classifier for face recognition.[18] The approach implements a similarity measure such as Euclidean distance to extract a group of important training samples neighbors to the unknown test sample. This strategy allows rejecting some of the undesired training samples. Their inclusion in the representation may increase the misclassification rate. The classification is performed by evaluating the sparsity in the selected subset using the l_2-norm minimization. We have used the sparse representation concept for breast cancer classification using a Gauss–Newton representation-based algorithm with numerical data only.[19] However, these approaches suffer from drawbacks like (1) ignore weights during the evaluation of similarity measure and (2) lacks in deciding the optimal size of the subset (number of significant training samples) during its selection, which ultimately affects the classification accuracy. Traditional approaches decide the size of the subset on a trial and error basis. These deficiencies motivated us to create a better alternative to the existing sparse representation-based classifiers.

In this paper, to solve the ternary classification problem, that is, AD versus MCI versus NC in a single run, we suggest a novel optimal sparse representation (OSR)-based classifier. The proposed classifier combines sparse representation with the RLM and optimization to obtain improved classification results. The necessities of combining these methods are (1) the suggested OSR updates the weights to neglect some insignificant training samples in the feature space. (2) To avoid computational complexity, the RLM is used instead of the l_0- or l_1-norm to get the updated weights. (3) To overcome computational complexity which may arise due to the large training dataset, we suggest a PSO–SSA optimization to obtain an optimal subset size for the subset selection instead of applying a trial and error approach. Thus, a novel combination of these methods is quite essential. Nevertheless, it is a better substitute for the existing techniques. The results show promising improvement in the classification performance compared to the baseline methods. We highlight our simulation analysis on the ADNI database. In the simulation, we use different ratios of training–testing partitions to avoid biased results, overfitting, and underfitting problems. The reported results are obtained over 100 iterations, which confirm its stability. The following section provides a detailed explanation of the proposed scheme.

The overall structure of this paper takes the form of four sections, including this introductory section. Section 2 is concerned with the methodology employed in this study. The simulation results, implementation, evaluation, and comparison with the state-of-the-art are presented in Section 3, which includes a discussion about the database used for simulation and the implications of the suggested method. Finally, Section 4 is the conclusion, which gives a summary and critique of the findings.

2 | PROPOSED OSR-BASED CLASSIFIER

In this work, we propose an OSR-based classifier. The classifier is applied to the ternary classification problem—AD versus MCI versus NC. Most of the previously reported sparse-based classifiers make use of the input image in the representation. The novel idea of performing classification in a low dimensional feature space improves the computational efficiency of the proposed classifier. Unlike previous sparse representation-based classifiers, our proposed classifier performs the classification in two different stages. The first stage considers the weights associated with all the training samples in the feature space for the subset selection. Here, RLM learning is incorporated to update the weight coefficients in both the stages with less computational complexity. In addition, to tackle the exhaustive search for deciding the size of the subset, we suggest a hybrid PSO–SSA optimization method. The optimum value of K (size of the subset) is obtained by maximizing a new objective function based on the classification accuracy. In this context, a first-hand objective function is proposed. Hence, the first stage allows selecting an optimum group of significant training instances considering that it affects the final classification decision. In the second stage, the unknown test sample is expressed as the linear weighted sum of the determined K training feature vectors. It uses the representation results to classify the test sample using a maximum class contribution criterion.

Before classification, the standard multiview approach is utilized to extract the features from the input images.[20–22] To be precise, having a feature extraction as a preprocessing step (instead of using the original image) ahead of the classifier stage is an advantage because it reduces the dimensional complexity. Figure 1 shows the block diagram of the proposed OSR-based classifier. It is clear from the figure that features from the training images are extracted only once. Every time a new test image is classified, features are extracted from the test image. The training instances, randomly initialized weights, and the test instances are input to the proposed classifier. These inputs are utilized to form a sparse representation, $y = f(X, W)$, where X indicates the training feature vector, W denotes the random weight vector, and y is the test feature vector. In the first stage, updated weights (ϕ) are calculated using an RLM technique to maintain consistency in the representation $y = f(X, W)$. Now, after obtaining the updated weights $y = f(X, W)$ is represented

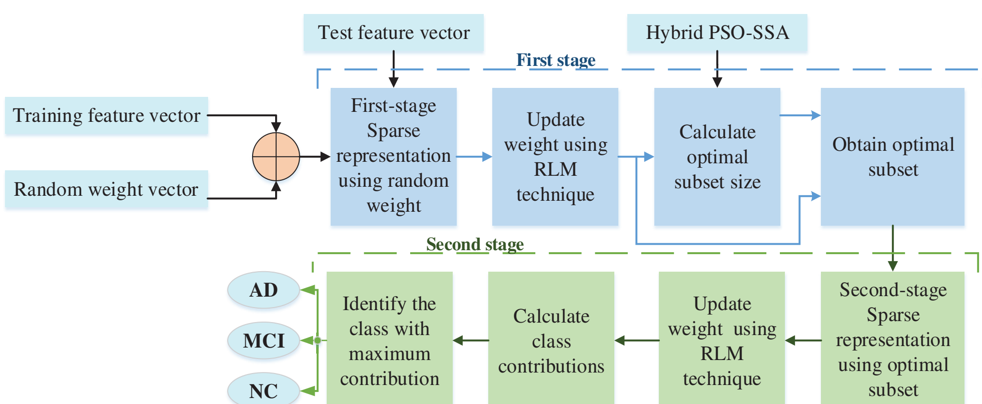

FIGURE 1 Block diagram of the proposed method

as $y = f(X, \phi)$. Secondly, a group of (size K) significant training instances is extracted. A hybrid PSO–SSA optimization technique is introduced to find the optimal size of the subset (K). This subset is used to form a new sparse representation of the test feature vector, that is, $y = f(U, \phi)$. Thirdly, the weights are further updated (ψ). Finally, each class contribution towards the test instance is computed. A detailed explanation of the classifier is given below.

In the first stage, the feature vector of a test image is represented as a weighted linear sum of the training features. Let $X = [x_1, x_2, ..., x_N]$ is the set of training feature vectors, where N is the total number of training images. Consider y be the feature vector of an unknown test image. The proposed scheme assumes the feature vector y as:

$$y = w_1 x_1 + w_2 x_2 + ... + w_N x_N. \tag{1}$$

Each x_i is a training feature vector, and w_i is a random weight component associated with each x_i. The contribution of the ith vector is $w_i x_i$. The random values of w_i results in incorrect contribution of each x_i. So an updated $W = [w_1, w_2, ..., w_N]$ is required to maintain the consistency. Literature study suggests that it can be obtained by minimizing the sum of square error, as discussed in References 19,23, and 24. In this context, we propose the RLM technique to generate the updated W. It is a modified version of Levenberg–Marquardt. A regularization parameter (λ) is introduced in this technique to make the least square solution stable. It also imposes a weaker sparsity constraint on the solution. Additionally, it avoids the singularity problem in the data matrix during the classification process. The strategy to obtain the updated W is given below:

$$W_{\text{next}} = W_{\text{now}} + \Delta w, \tag{2}$$

where $\Delta w = (X^{\mathrm{T}} X + \lambda I)^{-1} X^{\mathrm{T}} \Delta f$. Note that λ is a regularization term, I represents the identity matrix, and Δf denotes the deviation of the actual output from the desired output. Here, $\lambda = 0.1$, after testing with different values of λ in the range [0, 0.5].

Here, all the x_i in Equation (1) are from different classes. However, in Equation (1), some unsuitable training feature vectors give an adverse influence in representing y belonging to a specific class. Thus, if all the N feature vectors in Equation (1) are used for the classification, this may increase the misclassification rate.

In this paper, we suggest finding a group of K significant training instances. Since the K useful training feature vectors may be from different classes, we must exploit all of the training feature vectors from each class to obtain an OSR. Furthermore, we must efficiently explore the search space. To put it in another way, the two most important conditions for searching a collection of K relevant training feature vectors are exploitation and exploration. That is why a hybrid PSO–SSA is proposed. It combines the advantages of exploring and exploiting the training feature vectors to find an optimal K value. At the same time, since there are fewer parameters to tune and constraints to consider, it can achieve a better global convergence. The suggested solution may be helpful in future research to solve problems of this nature. The value of K is obtained by optimizing a new objective function, based on the classification accuracy, using the suggested hybrid PSO–SSA. It has the characteristics of PSO[25,26] and SSA.[27] It can reach the global minima compared to individual optimization techniques like GA, PSO, and SSA because of its superior exploration and exploitation capabilities. The suggested hybrid technique integrates the intrinsic social quality of PSO with the foraging behavior of squirrels. It uses a co-evolutionary algorithm that can simultaneously update the position of a particle. Like other optimization techniques, the population is initialized within the exploration domain. Each particle in the population is called a candidate solution. The position of each particle is updated:

$$P(t+1)_{\text{PSO-SSA}} = P(t)_{\text{PSO-SSA}} + \left\{ V(t+1)_{\text{PSO}} + P(t+1)_{\text{SSA}} \right\}. \tag{3}$$

The velocity in PSO is updated as

$$\begin{aligned} V_i(t+1)_{\text{PSO}} = b(t) V_i(t) &+ c_1 r_1 (p\text{best}_i - P_i(t)) \\ &+ c_2 r_2 (\text{gbest}_i - P_i(t)), \end{aligned} \tag{4}$$

where $V_i(t)$ is the velocity of ith particle at tth iteration, $P_i(t)$ is the current location of the ith particle. The value of t starts from 1. So, $P_i(0)$ signifies the initial position of the particles with 0 iteration, c_1 and c_2 are acceleration coefficients, r_1 and r_2 are random values within [0, 1], $b(t) = \text{rand} \times \frac{t}{t_{\max}} (b_{\max} - b_{\min}) + b_{\min}$ is the inertia weight function in which $b_{\max}$ and $b_{\min}$ are the range of inertia weight and $t_{\max}$ is the total iteration. The gliding of flying squirrel in SSA is given as:

$$P(t+1)_{\text{SSA}} = \begin{cases} P_{\text{at}}(t) + d_{\text{g}} G_{\text{c}} (P_{\text{ht}}(t) - P_{\text{at}}(t)) R_1 & \geq \delta \\ P_{\text{nt}}(t) + d_{\text{g}} G_{\text{c}} (P_{\text{at}}(t) - P_{\text{nt}}(t)) R_2 & \geq \delta \\ P_{\text{nt}}(t) + d_{\text{g}} G_{\text{c}} (P_{\text{ht}}(t) - P_{\text{nt}}(t)) R_3 & \geq \delta \\ \text{rand}, \text{otherwise} & \end{cases}, \tag{5}$$

where d_{g} is the random gliding distance, $P_{\text{ht}}(t)$ is the position of flying squirrel that reached hickory nut tree, $P_{\text{at}}(t)$

is the position of flying squirrel on acorn nut tree, $P_{nt}(t)$ is the position of flying squirrel on normal tree, G_c indicates gliding constant, δ denotes predator presence probability, and R_1, R_2, and R_3 are random numbers. The objective function used in the hybrid PSO–SSA to find the local/global best particle is given as

$$F(P) = \operatorname{argmax}(\mathrm{CA}), \tag{6}$$

where $\mathrm{CA} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$, TP is true positive, TN is true negative, FP is false positive, and FN is false negative.[28] The focus of this work is to get an optimal K for minimizing the classification error. Any comparison among different optimization techniques is beyond the scope of this paper. The optimal K obtained from the hybrid PSO–SSA is used to choose an optimal subset using Euclidean distance measure, given as[29]:

$$D_i = \|y - \phi_i x_i\|_2; \quad i = 1,2,\dots,N, \tag{7}$$

where $\phi = [\phi_1, \phi_2, \dots, \phi_N]$ is the updated set of weight vector W. An optimal subset, that is, $\widehat{U} = [\widehat{u}_1, \widehat{u}_2, \dots, \widehat{u}_K]$ is selected from X based on K smallest D_i value. Every time a new test sample is presented, the proposed classifier needs to run the whole optimization process to get new updated ϕ. However, the strategy is efficient even if a large training dataset is used. This is because, during the optimization, the whole training dataset is not used. Only the optimum number of subsets is used to get the updated ϕ.

In the second stage, y is expressed as a linear weighted sum of $\widehat{U}$, given as:

$$y = \phi_1 \widehat{u}_1 + \phi_2 \widehat{u}_2 + \dots + \phi_K \widehat{u}_K, \tag{8}$$

where $\widehat{u}_i$ for ($i = 1, 2, \dots K$) are the identified nearest training feature vectors and ϕ_i are the updated weight values from the first stage. The new representation in Equation (8) now contains the subset, due to which the impact of the training instances is altered. Moreover, the weights are no longer optimum. So, we have further updated the weights using the process as discussed in the first stage.

Consequently, the role of each Lth class in the subset towards y is computed:

$$\widehat{U}_L = \sum_{i=1}^{n_L} \psi_i \widehat{u}_i, \quad L = 1,2,\dots,C, \tag{9}$$

where $\psi = [\psi_1, \psi_2, \dots, \psi_K]$ is the set of updated weight components of ϕ obtained using the learning technique discussed earlier. Note that n_L represents the number of training instances in the Lth class. For instance, in a two-class problem, Equation (9) results in two vectors: U_1 and U_2. These two vectors indicate the contribution of the two classes towards the test feature vector. This idea is extended here to generate more than two vectors to create a multiclass space. This idea of the proposed classifier makes it a multiclass classifier, which helps to output the contributions of the multiple classes in a single run. It is the main contribution of the work. The class of the test feature vector is calculated as

$$C = \min_L \left\{ \left\| y - \widehat{U}_L \right\|_2 \right\}. \tag{10}$$

In Equation (10), a smaller distance between y and $\widehat{U}_L$ indicates a maximum contribution of the Lth class. Hence, it is classified to the class L. The algorithm of the proposed method is given in Algorithm 1.

3 | RESULTS AND DISCUSSIONS

This section contains a thorough examination and explanation of the findings. Before that, the database used,

Algorithm 1

Proposed OSR-based classifier

Step 1: Initialize the random weight vector $W = [w_1, w_2, \dots, w_N]$.
Step 2: Update weight vector W, that is, $\phi = [\phi_1, \phi_2, \dots, \phi_N]$ using RLM technique in Equation (2).
Step 3: Initialize the parameters of PSO–SSA (c_1, c_2, r_1, r_2, $b_{\max}$, $b_{\min}$, d_g, G_c, δ, R_1, R_2, R_3 and $t = 1$).
Step 4: Initialize a swarm of S particles, velocity, local best particle (Lbest) and global best particle (Gbest) for the swarm.
Step 5: Evaluate the fitness value of each particle using Equation (6).
Step 6: Update the position using Equation (3).
Step 7: Update Lbest and Gbest.
Step 8: If not converged ($t < 100$) go to Step 5.
Step 9: Find the optimal subset $\widehat{U} = [\widehat{u}_1, \widehat{u}_2, \dots, \widehat{u}_K]$ using Equation (8).
Step 10: Represent y in terms of $\widehat{U}$ using Equation (8).
Step 11: Again update weights $\psi = [\psi_1, \psi_2, \dots, \psi_K]$ of y in Step 10 using RLM technique.
Step 12: Calculate the contribution of each class $\left(\widehat{U}_L\right)$ using Equation (9).
Step 13: Identify the class of y using Equation (10).

simulation setting, assessment approach, and some details about the suggested method implementation are covered.

3.1 | Database

Data used in the preparation of this article are obtained from the ADNI database (http://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 goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

In this paper, we have used the brain MRI samples of 200 cases, which include 76 NC subjects, 72 MCI subjects, and 52 AD patients. The parameters used to select these subjects from the whole database are illustrated in Table 1.

Representative images (ADNI database) are displayed in Figure 2.

3.2 | Implementation

We have implemented the suggested technique in MATLAB on a MAC platform with Intel core i5. The parameter sets used in the proposed method are shown in Table 2.

We have employed different ratios of training–testing partition splits to validate the robustness of the suggested technique, as shown in Figure 3. Here, the entire dataset containing 200 brain MRI samples is partitioned into 10 equal-sized blocks. Each block contains $200/10 = 20$ brain MRI samples. To avoid overfitting, we employ the partition blocks in nine different training–testing ratios, such as 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, and 9:1. In these ratios, the first number is the number of blocks used for training (green color blocks). Moreover, the second number is the number of blocks used for testing (orange color blocks). For instance, in the 1:9 ratio, one block (20 brain MRI samples) is used for training, and nine blocks ($9 \times 20 = 180$ brain MRI samples) are used for testing. These test samples are used to generate the test score (shown as test score #1 in Figure 3), which measures the model's performance on the 1:9 ratio. It is to be noted that the test score represents the different performance metric values obtained using the test samples. The performance metrics utilized to evaluate the model are discussed in the following subsection. Likewise, test results (or scores) are obtained for the other ratios. The final test score is the average of the nine test scores. It shows the proposed model's overall performance in one iteration. It is to be noted that, to demonstrate the stability of the proposed model, we have reported the average mean, and average standard deviation of the test scores obtained over 100 iterations.

3.3 | Evaluation

The suggested technique is validated using different performance indices such as classification accuracy (CA), sensitivity (SEN), specificity (SPE), confusion matrix (CM), and AUC are utilized.[28,30] In the context of a

TABLE 1 Subjects demographic information from the ADNI database

Parameter		Diagnostic group: AD	MCI	NC
Age (75–85)	Mean	80.15	79.69	79.03
	SD	3.1239	2.7759	2.8549
Gender	Male	27	49	41
	Female	25	23	35
Mini-mental state examination (20–30)	Mean	24.83	24.69	25.26
	SD	3.3159	3.1644	2.8281
Clinical dementia rating (0–1)	Mean	0.5301	0.4955	0
	SD	0.3931	0.4035	0
Research group	Patient			
Slice thickness	3–5 mm			
View	Axial			
Scanner	1.5 Tesla MRI			
Modality	T2-weighted (T2) MRI			

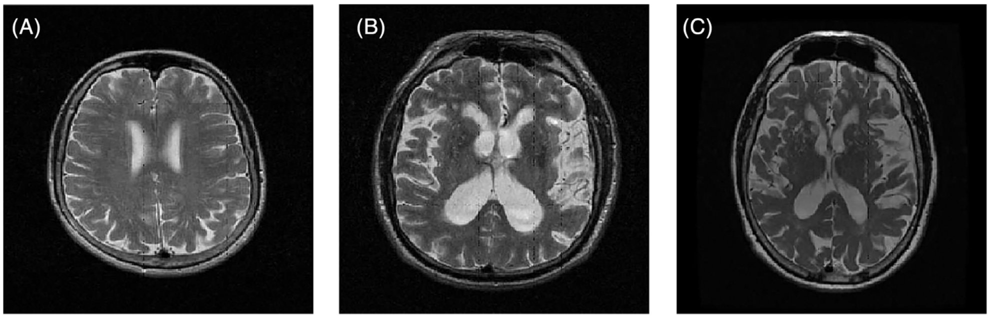

FIGURE 2 Examples of brain MR images. (A) NC; (B) MCI; (C) AD

TABLE 2 Parameter setting for the proposed classifier

Parameter	Value
Particle dimension	1
Population size	30
Acceleration co-efficient (c_1, c_2)	2
Inertia weight ($b_{\max}$)	0.9
Inertia weight ($b_{\min}$)	0.4
Gliding constant (G_c)	1.9
Predator presence probability (δ)	0.1
Number of iteration in hybrid PSO–SSA (t)	100
Termination criteria of learning technique	100 iterations
Regularization parameter, λ	0.1

ternary classification problem, sensitivity and specificity are defined as follows:

$$\mathrm{SEN}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{11}$$

$$\mathrm{SPE}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \tag{12}$$

where, TP $\rightarrow$ true positives, TN $\rightarrow$ true negatives, FP $\rightarrow$ false positives, and FN $\rightarrow$ false negatives. A CM is used to determine the TP, TN, FP, and FN for each class in a ternary classification problem. Table 3 illustrates one of the best performing results of the CM for the AD versus MCI versus NC, obtained using the proposed technique. It is a 3×3 matrix (ternary classification problem). The diagonal elements represent the correct classification. For instance, 327, 322, and 219 indicate the correct classification of normal subjects, MCI subjects, and AD patients, respectively. The value off the diagonal denotes misclassification, such as a value of 4 (3rd element in the 2nd column) indicates misclassification of 4 MCI subjects as normal. It is to be noted that the CM in Table 3 is obtained by summing the CMs of all training–testing ratios, that is, summation of nine CMs in one iteration. A greater summation value of the elements along the diagonal indicates better accuracy.

If we consider class AD, the TP, TN, FP, and FN values are calculated as follows:

- TP: All AD instances that are classed as AD, that is, 219, are included in the TP of AD.
- TN: All non-AD occurrences that are not identified as AD are included in the TN of AD, that is, 322 + 327 + 5 + 4 = 658.
- FP: All non-AD cases that are classed as AD, that is, 3 + 7 = 10, are included in the FP of AD.
- FN: All AD instances that are not classed as AD, that is, 6 + 7 = 13, are included in the FN of AD.

By replacing AD with MCI or NC, the four metrics TP, TN, FP, and FN of the MCI or NC class are obtained.

3.4 | Simulation results

In this section, quantitative evaluation results are presented for the ternary classification task at hand. Simulations are performed using different training–testing partition ratios, as discussed above. Table 4 shows the classification performance of the proposed OSR method in terms of accuracy, sensitivity, specificity, and AUC obtained by averaging over 100 iterations. To calculate these validation measures, we have employed the standard approach used in ternary classification problems. The accuracy, sensitivity, specificity, and AUC are calculated for each class in one iteration. Then, these measures are computed for 100 iterations to demonstrate the stability of the proposed model, as shown in Table 4. Finally, the overall validation measure is evaluated by averaging over each class. For instance, overall average CA = (97.85% + 98.15% + 92.06%)/3 = 96.02%.

3.5 | Comparison with the state-of-the-art

We compared the suggested OSR-based classifier with state-of-the-art methods in Table 5. For comparison, the baseline methods used for AD classification are selected

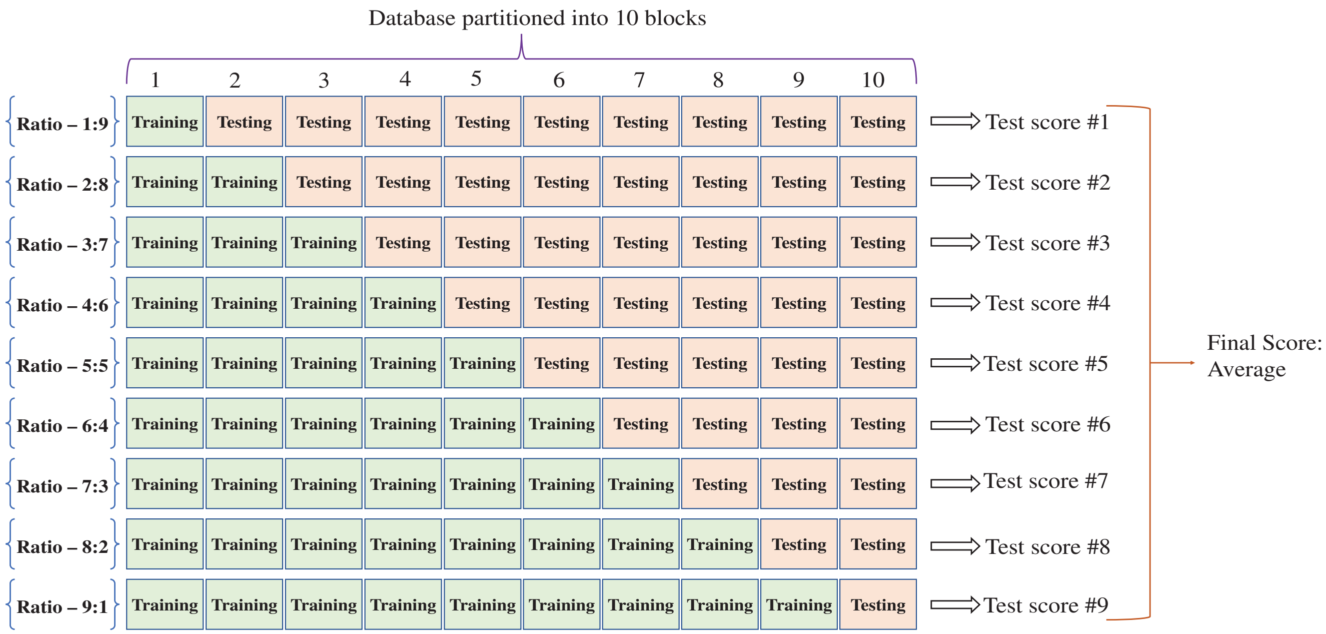

FIGURE 3 Training–testing partition splits utilized in the suggested method

TABLE 3 Confusion matrix for the ternary classification problem

	True class		
Predicted class	**AD**	**MCI**	**NC**
AD	219	3	7
MCI	6	322	5
NC	7	4	327

from the recent literature published during the last 5 years. In addition to different baseline methods (discussed in the introduction section), Table 5 also includes DL models like convolutional neural network (CNN) and the Interpretable DL model. A total of 1409 subjects (294 AD, 763 MCI, and 352 NC) are used in Reference 31. The technique uses various binary classification problems: AD versus NC, AD versus MCI, MCI versus NC, and so on. To validate their technique, the authors performed 10-fold cross-validation. In the interpretable DL model, 417 subjects (188 AD and 229 NC) are utilized for investigation. The researchers have solved a single binary classification problem, namely AD versus NC. The approach is validated on accuracy, sensitivity, and specificity.

From Table 5, it is observed that the suggested OSR-based method outperforms all other methods. The accuracy, sensitivity, and specificity of CNN reported for AD versus NC are 99.2%, 0.9890, and 0.9950, respectively in Reference 31. It is more than our suggested method. However, the authors have solved three binary classification problems to achieve the task at hand. Moreover, the values of the performance metrics of CNN for AD versus MCI and MCI versus NC are much low compared to the suggested method. Furthermore, if we take the average of the three binary classifications, then its performance is much lower than our suggested approach. Unlike other methods, the suggested model is effective in solving the ternary classification problem—AD versus MCI versus NC in a single run, which is usually solved as AD versus NC and MCI versus NC. Finally, we can observe that it also outperforms the NGF method in the AD versus MCI versus NC classification case.

In this work, the Friedman test is performed to show the statistical significance of the proposed OSR method.[33] Table 6 shows the p values from this test. It is observed that the proposed method gives a p value less than 0.05 for all the methods. Hence, it indicates that our approach is significantly different from other methods. It provides improved results, as confirmed from the above tables.

In this study, the central contribution is the subset selection stage in the sparse representation. For demonstrating the significance of the subset selection stage, the classification performance of using only the first stage sparse representation is reported and compared with the suggested OSR in Table 7 using the ADNI database. It is witnessed that the subset selection stage in the proposed method yields improved classification results.

TABLE 4 Classification performance of the proposed OSR over 100 iterations

Class	CA (%)	SEN	SPE	AUC
AD				
Average value	97.85	0.9825	0.9528	0.9937
SD	0.1882	0.6053	0.0011	0.0013
MCI				
Average value	98.15	0.9906	0.9954	0.9847
SD	0.1614	0.6086	0.0013	0.0020
NC				
Average value	92.06	0.9891	0.9894	0.9941
SD	0.1925	0.6125	0.0015	0.0012
Overall average value	96.02	0.9874	0.9792	0.9908

TABLE 6 Statistical analysis result using the Friedman test

Method	*p* value
R2DLSR[9]	0.0016
Multiple template learning method[6]	0.0027
ISML[7]	0.0027
Landmark-based method[8]	0.0016
wmSRC[5]	0.0016
NGF[12]	0.0016
TML-SVM[11]	0.0016
VBM + feature ranking+GA[10]	0.0016
CNN[31]	0.0027
Interpretable DL[32]	0.0016

TABLE 5 Comparison of CA, SEN and SPE using the ADNI database

Method	CA (%)	SEN	SPE
VBM + Feature ranking+GA[10]			
AD vs. NC	93.01	0.8913	0.9680
TML-SVM[11]			
AD vs. NC	91.95	0.8949	0.9382
MCI vs. NC	83.72	0.8474	0.8272
NGF[12]			
AD vs. NC	91.80	0.8890	0.9470
MCI vs. NC	79.50	0.8510	0.6710
AD vs. MCI vs. NC	60.20	–	–
wmSRC[5]			
AD vs. NC	94.80	0.9560	0.9400
MCI vs. NC	74.50	0.6640	0.8210
Landmark-based method[8]			
AD vs. NC	83.1	0.8050	0.8510
MCI vs. NC	73.6	0.7530	0.6970
ISML[7]			
AD vs. NC	93.83	0.9278	0.9569
Multiple template learning method[6]			
AD vs. NC	93.06	0.9485	0.9071
R2DLSR[9]			
AD vs. NC	94.68	0.9790	0.9138
MCI vs. NC	80.32	0.6435	0.8667
CNN[31]			
AD vs. MCI	85.90	0.8360	0.8830
MCI vs. NC	76.10	0.7510	0.7710
Interpretable DL[32]			
AD vs. NC	83.40	0.7670	0.8890
Proposed OSR			
AD vs. MCI vs. NC	96.02	0.9874	0.9792

TABLE 7 Comparison of one-stage and two-stage sparse representation methods

	Indices		
Method	CA (%)	SEN	SPE
One-stage	92.15	0.8930	0.8997
Two-stage	96.02	0.9874	0.9792

Table 8 presents the computation complexity (based on the number of executions) of the different baseline methods and the proposed method for AD versus MCI versus NC classification. From the table, it is observed that all the baseline methods except NGF use the binary classifier. Thus, it is implicit that they need to execute the algorithm three times to solve the ternary disease classification problem at hand. Like the proposed OSR technique, the NGF method needs only a single run to perform the same task. However, its performance in terms of the evaluation metric values (reported above) is low compared to the proposed OSR method. Hence, the suggested OSR scheme is a better substitute in terms of the computational complexity to solve the multiclass disease classification problem.

3.6 | Discussions

In this paper, we have investigated a novel OSR-based scheme for the AD classification task. A standard database involving three different classes is used for the classification task. The overall performance is measured by evaluating different performance indices (CA, SEN, SPE, CM, and AUC). The classification is achieved in two stages. In the first stage, the main objective is to choose an optimal group containing the significant training instances for accurate classification. The approach used in this stage for the subset selection provides two advantages: (1) adds self-learning capability to the sparsity in the representation and (2) obtain the optimal value of K (size of the subset) using a new objective function based on CA. The objective function based on CA used for obtaining the optimal value of the subset size is very encouraging. An optimum value of K is obtained by maximizing the CA. This objective function may help to design such classifiers in the future.

In the second stage, the RLM technique is used for the calculation of the updated weight vector in the subset. It facilitates us to measure the sparsity in the selected subset. The ultimate purpose of this stage is the assessment of the class contribution in expressing the test feature. Notably, the results demonstrate that the proposed OSR-based classifier yields better performance compare to the other relevant methods.

TABLE 8 Number of executions for AD versus MCI versus NC classification

Method	No. of execution
wmSRC[5]	3 runs
R2DLSR[9]	3 runs
VBM + feature ranking+GA[10]	3 runs
NGF[12]	1 run
CNN[31]	3 runs
Interpretable DL[32]	3 runs
Proposed OSR	1 run

The main reason for its improved performance is maybe the optimal subset selection (using sparse representation at two stages) in representing the test feature vector. Since the suggested method takes advantage of this strategy, it is able to discard some unsuitable training feature vectors. Furthermore, it retains the suitable training feature vectors in the representation. The inclusion of these suitable training feature vectors eases in evaluating the influence of the class to which a test instance is likely to belong. Moreover, RLM learning is used to evaluate the optimum sparsity at both stages. Due to this optimum sparsity evaluation, better optimal subset selection is achieved in the first stage. Additionally, the optimum class contribution ability is calculated in the second stage, which improves the classification performance. Researchers have suggested methods using multi-modal images.[5,9,10,12] However, their methods need the subjects to undergo multiple scanning, which is not desirable. This is a limitation of their methods. Our method prevents subjects from undergoing multiple scanning. The time complexity (number of runs) of our method is discussed in Table 8. Although the time complexity of our method is higher than the binary classifier for its increased number of stages, there is a merit of achieving reduced overall time complexity while dealing with multiclass problems. Since the methods reported are binary in nature, they fail to solve the AD versus MCI versus NC problem in a single run.[5,9,10,12] Alternatively, our method is quite intriguing and helps us to solve this problem in a single run. The reason behind our success is due to the improvements in sparsity utilizing the second stage. Furthermore, the suggested objective function inherently includes the mechanism for maximizing CA.

3.7 | Technical novelty

In this section, we highlight the technical novelties of this research work. The proposed technique makes several noteworthy contributions: (1) using features to create a single feature vector makes the proposed OSR-based classifier more robust for classification. In this way, it adds a synergistic behavior to the classifier. (2) We introduced a hybrid PSO–SSA for optimum subset selection that helps us to achieve improved classification results. (3) Due to the incorporation of the optimization technique, the subset selection is optimal when a new test instance is found. (4) The use of different proportions for validation ensures that there is no overfitting problem. (5) the reported results are obtained over 100 iterations, which confirms its stability, and (6) the novelty of the objective function reported in this work lies in the fact that it may generate curiosity among the researchers for its worthy application in the related field. This proposal gives a new direction to solve the problem like AD versus MCI versus NC in a single run. Our classifier can identify whether AD versus MCI versus NC from the brain MR image in a single run. Furthermore, the computational complexity is reduced by using a regularized term instead of using an l_1 or l_2 norm for the weight update.

4 | CONCLUSION

Unlike the conventional classifiers based on the sparse representation methods, our proposed classifier based on the OSR works on two stages to perform coarse-to-fine classification of an unknown test sample. We have obtained an improved classification accuracy because an optimal subset is used that includes the most appropriate training feature vectors only. The new objective function reported yields an optimum subset size due to its inbuilt ability (in terms of *CA*). The individual weight components are updated in both stages, which leads to the improvement in the representation. Usually, the AD versus MCI versus NC classification is solved as an individual binary classification problem. On the other hand, the suggested research work is best suited to solve ternary classification tasks in a single run. The introduction of two stages in sparse representation shows its uniqueness in solving such problems. The classification rate is significantly improved compared to the one-stage sparse representation method. The simulation results explicitly reveal that our scheme achieved the best performance. Unlike the existing classifiers, which depend on the whole training dataset through the entire classification process making the task inefficient, our proposed classifier uses an optimal subset only that makes it efficient. It is believed that the results of *CM* using the ADNI database presented in this paper may attract researchers for their future references.

The current study has examined only one database. The future scope of the work is the assessment of its performance using various modalities. In addition, the idea can be extended to design a computer-aided diagnosis (CAD) system. A graphical user interface can be developed for the CAD system.

ACKNOWLEDGMENTS

Data collection and sharing for the ADNI project were supported by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health grant U01 AG024904) and DOD ADNI (Department of Defense award number W81X-WH-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; EuroImmum; 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 & Co., 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.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ADNI, "Alzheimer disease neuroimaging initiative," http://www.loni.usc.edu/ADNIl.32

ORCID

Sanjay Agrawal https://orcid.org/0000-0002-7966-1412
Rutuparna Panda https://orcid.org/0000-0002-8676-0144

REFERENCES

1. “Alzheimer association.” http://www.alz.org/. Accessed January 2021.
2. Thiyaneswaran B, Anguraj K, Kumarganesh S, Thangaraj K. Early detection of melanoma images using gray level

co-occurrence matrix features and machine learning techniques for effective clinical diagnosis. *Int J Imag Syst Technol*. 2021;31(2):682-694.
3. Jeevanantham V, MohanBabu G. Detection and diagnosis of brain tumors-framework using extreme machine learning and CANFIS classification algorithms. *Int J Imag Syst Technol*. 2021;31(2):540-547.
4. Mathiyalagan G, Devaraj D. A machine learning classification approach based glioma brain tumor detection. *Int J Imag Syst Technol*. 2021;31(3):1424-1436.
5. Xu L, Wu X, Chen K, Yao L. Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment. *Comput Methods Prog Biomed*. 2015;122(2):182-190.
6. Liu M, Zhang D, Shen D. Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. *IEEE Trans Med Imag*. 2016;35(6):1463-1474.
7. Liu M, Zhang D, Adeli E, Shen D. Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis. *IEEE Trans Biomed Eng*. 2016;63(7):1473-1482.
8. Zhang J, Gao Y, Gao Y, Munsell BC, Shen D. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. *IEEE Trans Med Imag*. 2016;35(12):2524-2533.
9. Lei B, Yang P, Wang T, Chen S, Ni D. Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis. *IEEE Trans Cybern*. 2017;47(4):1102-1113.
10. Beheshti I, Demirel H, Matsuda H, Alzheimer's Disease Neuroimaging Initiative. 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*. 2017;83:109-119.
11. Shi B, Chen Y, Zhang P, Smith CD, Liu J, Initiative ADN. Nonlinear feature transformation and deep fusion for Alzheimer's disease staging analysis. *Pattern Recogn*. 2017;63:487-498.
12. Tong T, Gray K, Gao Q, Chen L, Rueckert D, Alzheimer's Disease Neuroimaging Initiative. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recogn*. 2017;63:171-181.
13. Zhang X, Ma S, Wang S, Zhang J, Sun H, Gao W. Divisively normalized sparse coding: toward perceptual visual signal representation. *IEEE Trans Cybern*. 2021;51(8):4237-4250.
14. Yang P, Zhou F, Ni D, et al. Fused sparse network learning for longitudinal analysis of mild cognitive impairment. *IEEE Trans Cybern*. 2021;51(1):233-246.
15. Aishwarya N, Bennila Thangammal C. A novel multimodal medical image fusion using sparse representation and modified spatial frequency. *Int J Imag Syst Technol*. 2018;28(3):175-185.
16. Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: which helps face recognition? *Computer Vision (ICCV), 2011 IEEE International Conference on IEEE*; IEEE; 2011:471-478.
17. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell*. 2009;31(2):210-227.
18. Xu Y, Fang X, Li X, et al. Data uncertainty in face recognition. *IEEE Trans Cybern*. 2014;44(10):1950-1961.
19. Dora L, Agrawal S, Panda R, Abraham A. Optimal breast cancer classification using gauss–newton representation based algorithm. *Expert Syst Appl*. 2017;85:134-145.
20. Wang Q, Guo Y, Wang J, Luo X, Kong X. Multi-view analysis dictionary learning for image classification. *IEEE Access*. 2018;6:20174-20183.
21. Jiang Y, Deng Z, Chung F-L, et al. Recognition of epileptic eeg signals using a novel multiview tsk fuzzy system. *IEEE Trans Fuzzy Syst*. 2017;25(1):3-20.
22. Han L, Jing X-Y, Wu F. Multi-view local discrimination and canonical correlation analysis for image classification. *Neurocomputing*. 2018;275:1087-1098.
23. J. S. R. Jang, C. T. Sun, and E. Mizutani, Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice Hall, 1997.
24. Gill PE, Murray W, Wright MH. *Practical Optimization*. University of Michigan, Academic Press; 1981.
25. Thangaraj R, Pant M, Abraham A, Bouvry P. Particle swarm optimization: hybridization perspectives and experimental illustrations. *Appl Math Comput*. 2011;217(12):5208-5226.
26. Kennedy J, Eberhart R. Particle swarm optimization (PSO). *Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia*; IEEE; 1995:1942-1948.
27. Jain M, Singh V, Rani A. A novel nature-inspired algorithm for optimization: squirrel search algorithm. *Swarm Evol Comput*. 2019;44:148-175.
28. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. 2009;45(4):427-437.
29. Zhang T. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE Trans Inf Theory*. 2011;57(7):4689-4708.
30. Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett*. 2006;27(8):861-874.
31. Basaia S, Agosta F, Wagner L, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks. *NeuroImage Clin*. 2019;21:101645.
32. Qiu S, Joshi PS, Miller MI, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*. 2020;143(6):1920-1933.
33. Demšar J. Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res*. 2006;7:1-30.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Dora L, Agrawal S, Panda R, Abraham A. An efficient multiclass classifier for classification of Alzheimer's disease/mild cognitive impairment/Normal subjects. *Int J Imaging Syst Technol*. 2021;1-13. doi:10.1002/ima.22656