Optimization Letters
https://doi.org/10.1007/s11590-021-01816-y

ORIGINAL PAPER

A VNS based framework for early diagnosis of the Alzheimer's disease converted from mild cognitive impairment

Pinar Karadayi-Ataş[1] · Aise Zulal Sevkli[2] · Kadir Tufan[3]

Received: 12 April 2020 / Accepted: 16 October 2021

Abstract

Mild cognitive impairment (MCI) is an intermediate stage between age-related cognitive decline. Alzheimer's disease (AD) is a more serious decline in dementia. Early identification of mild cognitive impairment with a high risk of Alzheimer's disease is very important for increasing the success rate of the treatment. In this study, we present a Variable Neighborhood Search (VNS) based framework that uses Magnetic Resonance Imaging (MRI) data to diagnose early conversion from MCI to AD. The proposed framework has been built in three main phases: preparing dataset, feature selection, and classification. After preparing the dataset, a VNS algorithm selects the most predictive MRI features for classification. Then, a Linear Support Vector Machine is utilized to classify the selected features. All data in this study are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database with 860 subjects, eight different monthly periods, and 286 features in each period. The results obtained from the framework outperform those of previous research in terms of accuracy, sensitivity, and specificity values. The results of this study demonstrate that our framework has a huge potential for early prediction and detection of mild cognitive impairment to Alzheimer's disease conversion.

✉ Aise Zulal Sevkli
sevkliaz@miamioh.edu

Pinar Karadayi-Ataş
pinaratas@arel.edu.tr

Kadir Tufan
ktufan33@umt.edu.al

1 Department of Computer Engineering, Faculty of Engineering and Architecture, Istanbul Arel University, 34537 Istanbul, Turkey

2 Department of Computer Science and Software Engineering, Miami University, Oxford, OH, USA

3 Metropolitan University of Tirana, Tirana, Albania

Published online: 15 November 2021

Keywords Alzheimer's disease · Mild cognitive impairment · Variable neighborhood search · ADNI · Linear-support vector machine · Symmetrical uncertainty

1 Introduction

Alzheimer's disease (AD) refers to a neurological irregularity in which the death of brain cells causes memory problems and cognitive decline. The disease starts mild and gets progressively worse with time. Both the nerve and the nerve's connection are damaged when time passes. The cause of AD is unknown, but generally, age, family health history, previous severe head injuries, and lifestyle triggers the risk of developing this condition. AD symptoms develop gradually, and the ailments aggravate from bad to worse in several years; thus, it is a progressive condition. The stages of AD include the following: preclinical Alzheimer's disease, mild Alzheimer's disease, moderate Alzheimer's disease, and severe Alzheimer's disease. A patient who is in the preclinical stage may look like a normal control (NC) after a physical examination and mental status testing. In a period of 10 to 20 years, specific regions of the brain are affected. A patient who is in mild Alzheimer's stage suffers from memory loss. In the stage of moderate Alzheimer's disease, patients have a problem remembering family members and friends. In addition, the memory loss problem increases with time. The patient becomes completely dependent on others for care in a severe Alzheimer's stage [1].

Clinical measures for AD are some standard measures such as the Mini-Mental Score Exam (MMSE) and Clinical Dementia Rating (CDR) [9]. With these two measures, the second and third stages of AD can be detectable; as a result, these measures are intelligible but inadequate for early detection of AD. Brain Magnetic Resonance Images (MRI) are a legitimate method for determining the degeneration amount of the brain structure, which consists of the hippocampus, entorhinal cortex, cerebral cortex, volume, shape, and thickness. Nowadays, 33.9 million people have AD, and the number of Alzheimer's patients is expected to triple over the next 40 years. The rapid rise of AD indicates the importance of early diagnosis and treatment [2].

Mild Cognitive Impairment (MCI) is defined as a syndrome in which the cognitive decline is greater than the expected limit for an individual's age and education level but does not interfere notably with activities of daily life [18]. For the effective treatment of AD, it would be important to identify MCI patients at high risk for conversion to AD. For the early detection of AD, the MCI stage plays an important role [28]. Several groups have proposed the use of many techniques for analyzing regions of MRI data for predicting the future conversion of MCI to AD [17, 28, 32].

In this study, we propose an application-specific generic software called framework that uses MRI data to diagnose early conversion from MCI to AD. The framework employs the Variable Neighborhood Search (VNS) method for selecting the best discriminative features from a very large feature list of MRI data. Selected features are utilized to classify the real MCI subjects and the MCI subjects that will be converted into AD in the future.

The remainder of this paper is organized into four sections: Sect. 2 contains an explanation of previous studies on AD. Section 3 gives details of the proposed framework.

Experimental results are provided in Sect. 4. Finally, Sect. 5 contains the discussion and concluding remarks.

2 Literature review

In the literature, researchers used various pattern recognition techniques for diagnosing early AD. In one of the previous studies, researchers used the Spatial Pattern of Abnormalities for Recognition of Early AD (SPARE-AD) technique, which is derived from high-dimensional pattern classification algorithms. The aim of the researchers is to predict MCI to AD conversion with good sensitivity, and their method's accuracy was very low (62%) [13]. To reduce the high dimensionality of the image data that is yielding a low dimensional embedding, some researchers used the non-linear manifold learning techniques. After feature reduction, they used a semi-supervised classifier, which utilizes both labeled and unlabeled images to boost performance. With the Laplacian Support Vector Machine Classifier (LapSVM), their accuracy result was 78%, and that of specificity was 78% [38].

In another study, researchers evaluated the performance of ten approaches, which are three methods based on cortical thickness, five voxel-based methods, and two methods based on the hippocampus. Their accuracy result was 67% [12]. Some researchers compared the classification accuracy achieved with Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). Their accuracy result was 68% [32]. Another study aims to compare and combine MRI data from the two-study cohorts using an automated image analysis pipeline and a multivariate data analysis approach. Researchers combined those two cohorts. They used multivariate analysis (orthogonal partial least squares to latent structures-OPLS). Their accuracy was 59%, and sensitivity was 74% [31]. Some researchers proposed a different method that consists of three key components which are a domain transfer feature selection component, a domain transfer sample selection component and a domain transfer Support Vector Machine classification component. Their accuracy result was 69.4%, and that of specificity and sensitivity was 64% and 73%, respectively [8].

Some researchers used logistic regression with stability selection for the integration and selection of potential predictors for MCI to AD conversion. After the selection of features from ADNI, they applied Support Vector Machines (SVM) to build the classifier. They used their method to determine the area under the receiver operating characteristic curve (AUC), which is the standard method to assess the accuracy of predictive distribution models. Its AUC value is 86% [33]. In the feature selection step, the ROIs (Region of Interest) were used for the prediction of anatomical regions, which were involved at different times prior to the progression from MCI to AD. After the feature selection step, they used linear discriminant analysis (LDA) techniques for classification. Their AUC values for progressive MCI (pMCI) vs. stable MCI (sMCI) was 81%, pMCI (12 month) vs. sMCI was 76%, pMCI (24 month) vs. sMCI was 71% and pMCI (36 month) vs. sMCI was 64% [14].

Another technique used kernel regression methods. With this method, they estimated the brain age by using normal brain-aging patterns. Their AUC value is 78%

[17]. By using logistic regression models, some researchers aimed to predict conversion from MCI to AD. Their accuracy result was 72% [30]. In a recent study [39], they applied statistical analysis based on receiver operating characteristic curves periodically to detect conversion to probable Alzheimer's disease. The accuracy value of this work is 88.7%. Furthermore, in the study of [40], feature ranking according to their respective t-test scores was applied at the feature selection step, and they applied a genetic algorithm and SVM. Their best accuracy result is 75.0%. Another study [41] developed a morphological factor method. They achieved 72.3% accuracy from MRI data. The study in [7] investigated a multivariate data analysis method using multimodality data (i.e., MRI and CSF).

Recently, Moradi et al. [27] developed a machine learning framework by using regularized logistic regression and random forest classifier for Alzheimer's conversion prediction in MCI subjects. Their accuracy was 82%, sensitivity was 87%, and specificity was 74%).

In this study, the proposed framework is aimed at designing and implementing an effective feature selection based on VNS and developing accurate classification modeling with SVM. To the best of our knowledge, VNS has not been implemented for ADNI before. Furthermore, the existing literature for predicting the early diagnosis of Alzheimer's disease has produced less than the desired prediction. Therefore, there is a need for robust models with high predictive power. The objective of this paper is to validate the large set of features derived from patients' MRI (as they were reported to ADNI) based on VNS and to develop accurate classification modeling. The novelty of this paper is based on these objectives and the new insights extended from the findings.

3 A VNS-based framework

In this study, we propose a VNS based framework that uses MRI data to diagnose early conversion from MCI to AD. Our framework consists of three fundamental phases. The first phase is preparing the dataset. The second one is a feature selection by using VNS. The third fundamental phase is the classification by using Linear-Support Vector Machine. The overall structure of the proposed framework is illustrated in Fig. 1. The following four subsections explain these phases and the framework in detail.

3.1 Phase 1: preparing the dataset

The purpose of this study is to predict early diagnosis of Alzheimer's disease by using Magnetic Resonance Imaging (MRI) data. MRI data tables were chosen because, in many studies, MRI data are used for detecting conversion from MCI to AD [3, 4, 19, 25]. MRI is performed in longitudinal and cross-sectional studies. We used cross-sectional measures in MRI data because cross-sectional measures are described as the most accurate predictors of conversion in the literature [5]. Cross-sectional MRI values, calculated by FreeSurfer [15] version 4.3, are combined by the diagnostic summary table, which also has labels of subjects. These labels are NC, MCI, and AD. Those two tables are merged according to their intersection of columns whose name

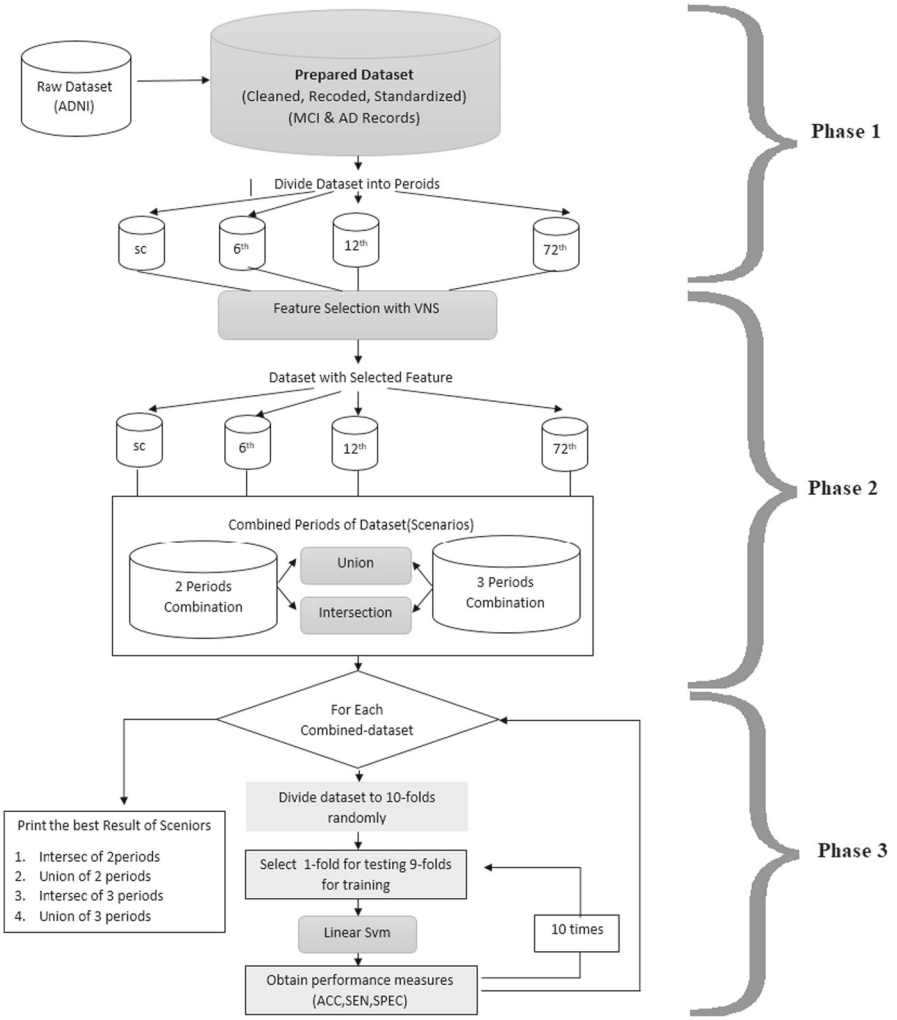

Fig. 1 The schema of the proposed framework

is Roister Id (RID). Furthermore, the dataset is prepared considering null values in the columns and the rows of the table.

3.2 Phase 2: feature selection with VNS

The high dimensional feature vectors of MRI data impose a high dimensional cost as well as the risk of overfitting during classification. In general, not all the features are equally useful for classification purposes. Therefore, removing some of them may improve the classification. In the general context, the objective of feature selection is to find a subset of features for minimizing the classification error rate.

There are different types of feature selection algorithms. One of them is the filter-type feature selection method, in which feature selection is made by using a quality measure function `Fitness(S)` independent of learning (classification) algorithms. We can formulate the feature selection problem in our framework as below.

Given D a set of d subjects characterized by the pair (x_i, y_i), where each $x_i \epsilon F$ *is* an instance described by a vector of l MRI features repeated for p periods $F = (F_1, F_2, \ldots, F_{lp})$ and $y_i \epsilon Y$ is the known class label of x_i, the aim of the feature selection problem for our framework is to

$$max\left\{Fitness(S) : S2^F\right\}, \tag{1}$$

where $S \subset F$ is any subset of MRI features, $2^F = \{S : S \subset F\}$ is the set of all subsets of F. In this study, we used the fitness function of the Correlation-based Feature Selection algorithm [16, 20] to evaluate candidate solutions.

$$Fitness(S) = \frac{m.S\overline{U}(S, Y)}{\sqrt{m + m(m-1).S\overline{U}(S, S)}} \tag{2}$$

where *Fitness(S)* is the heuristic quality of S containing m features, $\underline{SU}$ (S, Y) is the mean feature-class correlation and, $\underline{SU}(S, S)$ is the average feature-feature intercorrelation.

$$S\overline{U}(S, Y) = \frac{1}{m}\sum\nolimits_{X_i \in S} SU(X_i, Y) \tag{3}$$

$$S\overline{U}(S, S) = \frac{2}{m(m-1)} \cdot \sum\nolimits_{\substack{X_i \in S \\ X_i \neq X_j}} SU(X_i, Y_j) \tag{4}$$

The numerator of *Fitness(S)* can be thought of as providing an indication of how the feature subset S is powerful to predict the class. The denominator of *Fitness(S)* shows how much redundancy there is among the features. Therefore, the BVNS tries to maximize the fitness value as much as possible to find an optimal feature subset.

To measure the correlation between features and between features and class, Symmetrical Uncertainty (SU) is used [20]. It is defined as follows:

$$SU(X, Y) = 2\left[\frac{IG(X|Y)}{H(X) + H(Y)}\right] \tag{5}$$

where IG refers to the mutual information between two features or between a feature X and a class label Y. H(X) is defined as:

$$\mathrm{H}(X) = -\sum\nolimits_i P(x_i) \cdot \left(\log_2\right) \log_2(P(x_i)), \tag{6}$$

where H(X) is a monotonic function of the probability $P(x_i)$ and expresses the information content. The mutual information between two features or between a feature X

and a class label Y is:

$$IG(X|Y) = H(X) - H(X|Y), \tag{7}$$

Finally, the formula of IG can be obtained as follows:

$$IG(X|Y) = -\sum_j P(y_j) \sum_i P(x_i|y_j) \log_2(P(x_i|y_j)), \tag{8}$$

The primary advantage of filter methods is their speed and their ability to scale large datasets [30]. To use of this advantage in our framework, we used a filter-type feature selection method.

For eliminating irrelevant and redundant features and selecting the best discriminative feature subset S_{best} from a large feature list of MRI data (*lp* features), we utilized a Variable Neighborhood Search (BVNS) based feature selection method before classification. Some other filter methods are also used for eliminating redundant features from the ADNI database in previous studies [7, 11, 22, 23]. However, VNS, to the best of our knowledge, has not been implemented for ADNI before. Moreover, comparing VNS to other filter-based popular techniques shows that VNS is the competitive strategy, which is capable of finding a small size of the feature with similar predictive power than the other algorithms [16].

VNS proposed by Hansen and Mladenovic [21] is a metaheuristic algorithm for solving combinatorial and global optimization problems based on the principle of systematic changes of neighborhood structure within the search. BVNS follows the basic version of VNS, which has three phases: Shake, Local Search, and Neighborhood Change. Shake selects a random solution among the solutions of the current neighborhood search space. When the Local Search reaches the local minimum, the algorithm decides whether to move its current solution to the local minimum in the Neighborhood Change phase. If the local minimum of the local search is better than that of the current solution, the algorithm starts again with the improved solution from the first neighborhood; otherwise, the algorithm continues from the Shaking phase with the next neighborhood structure until all neighborhood structures are exhausted. This process repeats until the maximum number of iterations is met. The pseudocode of BVNS is given in Algorithm 1 where N_k ($k = 1,2,3\ldots,k_{max}$) represents kth neighborhood structure, and $N_k(S)$ denotes the set of feature subsets in the kth neighborhood of the feature subset $S \subset F$. First, the algorithm starts to generate the initial subset S and initialize S_{best} as S. Then, the Shaking method selects a subset S' randomly from $N_{\text{k}}(S)$. VNS continues to apply a Local search method to S' to obtain the improved subset S''. If the quality of S'' is better than the quality of S, S'' is selected as a new starting point for searching ($S \leftarrow S''$) with the first neighborhood ($k \leftarrow 1$). The function *UpdateBest* is called to update S_{best} if it is needed. If S'' is not better than the current subset S, then the neighborhood is changed to the next one ($k \leftarrow k + 1$). If there is no improvement on S in the last neighborhood ($N_{k_{max}}$), the algorithm starts all over again with $k = 1$ until the stopping criteria are met.

```
Procedure BVNS (F: set of all MRI features, Y: set of known class labels)
   Initialize the set of neighborhood structures Nk←k-exchange, k = 1,2, 3..., kmax
   S, R ← Generate_Initial_Solution(F,Y) // R is the set of relevant features
   Sbest ← S
   repeat
       k ← 1
       repeat
          S′, R′ ← ShakeMethod(S, Nk, R)
          S′′, R′′ ← LocalSearch(S′, R′)
          if Fitness(S″) ≥ Fitness(S) then
              S ← S″, R ← R″, k ← 1
              UpdateBest(Sbest, S)
          else
              k ← k + 1
       until (k== kmax)
   until (StoppingCriterion)
   return Sbest
End Procedure
```

Algorithm 1. Pseudocode of BVNS

The implementation details of BVNS, i.e., initial solution, the methods used in shaking, and local search phases, are explained below.

Solution representation: A candidate solution S to the feature selection problem is formed by using a variable-length integer set. Each element of the set represents a feature index. In our dataset D, each period contains at most l features of MRI data. We index these features from 1 to l. If a candidate solution S equals {21, 56, 78, 94, 102, 148}, this means that the solution S consists of only six features located at 21st, 56th, 78th, 94th, 102nd, and 148th indexes in the set F that contains all MRI features.

Initial Solution: In general, one of the two methods is used to generate initial solutions: randomly or systematically. To start the search from a convenient location on the search space, we prefer to use a systematic way to create an initial solution. The pseudocode of the initial solution algorithm is given in Algorithm 2. The initial solution is generated by removing irrelevant features, which contain no information about the class, and redundant features, which have correlated information about the class.

```
Procedure Generate_Initial_Solution (F: set of all MRI features, Y: set of known class labels)
 Initialize initial solution S=F
   -- Remove irrelevant features--
   for each feature Fi : F
      If SU(Fi,Y) equal 0 then
          Remove Fi from S
   --Find predominant features and their relevant features--
   Select a feature Xi in S where its SU(Xi,Y) is highest
   Do Until all features labeled as either predominant feature or relevant feature
      If Xi is not labeled as predominant feature then
          Label Xi as predominant feature
          for each feature sj that SU(Xj,Y)< SU(Xi,Y) :S
                if Xj is not a predominant feature and its SU(Xj,Y) is next-highest then
                        If SU(Xi,Xj) > SU(Xj,Y) then
                                Label Xj as relevant feature of Xi

   --Remove redundant features--
   for each feature Xi: S
       If Xi is not labeled as predominant then
           Remove Xi from S
           Add Xi to relevant-set R
   return S, R
End procedure
```

Algorithm 2. Pseudocode of Initial Solution Generation

SU (X, Y) takes value in [0, 1]. A value of 0 means that X and Y are irrelevant; a value of 1 means that X and Y are highly correlated. While generating the initial solution, all features with *SU* measure equal to 0; in other words, all uncorrelated features are removed. Feature grouping is another method to hold all relevant features together in a group. The feature has the most predictive power; in other words, the biggest correlation between a feature and its class is selected as a predominant feature in the group. Because of this, all relevant features other than the predominant feature can be accepted as redundant features, and they are removed from the solution as well.

In order to identify predominant groups and features, we follow the same strategy used in [16]. First, we sort the features in the ascending order of $SU(X_i, Y)$. The last feature, which has the highest *SU* value, is selected as the first predominant feature. Then any of the features in the list is checked by comparing the correlation between the features X_i and X_j and the *SU* value of X_i and the class label set Y. If the correlation between such features is larger, then X_j is labeled as a relevant feature of X_i. This process is repeated until no relevant feature is found. The second predominant feature, which has the second-largest $SU(X_i, Y)$ value among the non-relevant features and its relevant features are found. This process is repeated until no predominant feature is found. The algorithm returns predominant features as an initial solution S and their relevant features set R, which are used to generate a neighborhood solution in the Shake phase of BVNS.

Neighborhood Structures in the Shake Method: The neighborhood structures (NS) are the key elements of VNS, and the performance depends on both the choice and the order of the neighborhood structures. In this study, a *k-exchange* neighborhood structure is performed. First, randomly k features are selected in solution S, then they are exchanged with randomly selected k features in the relevant-set R, which is constructed while generating an initial solution. A value of k determines the number of NS that is employed in the Shake method. In other words, the Shake phase of the BVNS uses a *k-exchange* neighborhood structure, starting from $k = 1$ *to* k_{max}, to generate a new neighborhood solution. In the implementation, k_{max} is set as 1% of the number of features in the dataset D. If k_{max} is less than 5, we set it as 5.

Local Search Method: The solution returned by the Shake function is improved by the local search. In this study, we implement two local search methods: Forward Selection (FS) and Backward Elimination (BE). FS starts with the current solution and greedily adds a relevant feature ($r'i$) from R' one at a time based on the *SU* value between $r'i$ and the class Y. FS keeps trying to add a feature until it finds the first improvement in the fitness of the solution. After that, the improved solution is sent to BE to get a better feature subset in the local search. BE takes the improved solution of FS as an initial solution, and then it moves a feature from the solution list to the relevant-set greedily. The greedy choice of BE is the value of the *SU* function, which measures the correlation between the solution's features. BE removes one of the features randomly from the feature pair that has the highest correlation between them. The process of BE continues to remove features as long as the fitness of the solution is improved. In the end, the local search returns the improved solution (S'') and the updated relevant-set (R'').

3.3 Phase 3: classification with support vector machine

After finding S_{best} (the best discriminative features subset) by BVNS, the classification phase starts for the prediction of conversion from MCI to AD. For this purpose, the Linear-Support Vector Machine (SVM) is used in our framework. SVM is one of the most effective and popular supervised learning models for classification, which is first introduced in [10]. It draws an optimal hyperplane between classes during the training phase. Optimal Hyperplane can be expressed as follows:

$$y = \langle w, x \rangle + b \tag{9}$$

where x is an input vector, w represents the weight vector for the decision surface, y is the output, b refers to bias term, and $\langle , \rangle$ is the dot product. Max-margin hyperplane, which is also known as the support vector, can be obtained by following optimization formula:

$$w^2 + C \sum_{i=1}^{l} \varepsilon_i, \tag{10}$$

$$s.t. : y_i \left(w^T x_i + b \right) \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0, i = 1, \ldots, l$$

where C represents the cost that is also referred to as regularization constant, ε_i is the penalty term for avoiding overfitting. The important thing of the problem is to give the value of C neither big nor small value since large C results in a small margin while small C results in a large margin for l training points.

In this study, some parameters are utilized for linear-SVM. A trial-and-error method was applied to set the optimal C value. For linear function, the kernel function was chosen because of its simplicity. Moreover, the linear kernel function yields highly satisfactory results in similar problems [24, 26].

4 Experimental results

All MRI data in this study are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. ADNI is designed to develop clinical, imaging, biomedical, and genetic biomarkers for early detection of AD. It is accessible via http://adni.loni.usc.edu/ with previous authorization. The dataset has MRI data of subjects in 6-year periods which are starting month (SC), 6th month, 12th month, 24th month, 36th month, 48th month, 64th month, and 72nd month (p = 8), and its labels are NC, MCI, and AD with 286 features ($l = 286$). In this study, columns (features), which have a minimum of 80% null values, are removed from the dataset. Moreover, subjects whose features with null values are removed from the data table. After applying phase 1 of the framework, the prepared dataset has 860 subjects ($d = 860$); they are grouped as NC (213 subjects), MCI (total 397 subjects, 20 of them converted from NC), and AD (total 250 subjects, 183 of them converted from MCI). In total, there are

2288 features (*lp*), which consist of cross-section measures of the brain comprising the hippocampus, parahippocampus, and entorhinal cortex.

The proposed framework is implemented using C# programming language and Matlab library, and the experiments are performed on a machine with an Intel Xeon CPU E5-2660 v3 @ 2.60 GHz with Microsoft Windows 10.

4.1 Experimental results of feature selection with BVNS (Phase 2)

The clean dataset is divided into periods (SC, 6th, 12th, 24th, 36th, 48th, 64th, and 72nd), and BVNS is run to each period six times with different seed values. This run-repetition is necessary for all algorithms if they contain random functions. Running with the different seed values shows the robustness of the algorithm against the randomness. The list of the best feature subset for each repetition is given in Table 1. If a feature is found at least three times through the repetitions, it is added to the result list. For instance, feature 77 is found three times (at run 4, 5, and 6) by BVNS in the SC period, so that feature 77 is in the final solution; on the other hand, feature 173 found two times, it was not added to the final solution. As a result, redundant and relevant features are eliminated.

Note that BVNS did not find any predominant feature for 36th, 48th, 64th, and 72nd months. Consequently, the initial solution of BVNS is not generated for those periods. For this reason, we removed these periods from the dataset. In other words, this result indicates that periods of 36th, 48th, 64th, and 72nd seem not useful for predicting the conversation from MCI to AD. The reason can be the brain MR image is remaining nearly the same after the 36th month, as observed in [6]. There is no significant change in brain MRI between the 36th month- 72nd month.

After eliminating redundant and irrelevant features for each period, we need to decide how we can use these remaining features for the classification task. The straightforward solution can be working with all remaining dominant features for classification. On the other hand, there are some common features among the periods. While some of them disappear after some periods, some others appear only for a specific period. To observe the effect of the dominant features, belong to a specific period to the classification process, the following four scenarios are designed in the proposed framework:

Scenario 1: Intersection of selected features of two-month periods: In this scenario, first, BVNS is applied individually for each period. Then, all possible pairs of months are listed as given in Table 2. For each pair of months, the intersection of selected features is found and used to construct the classification feature vector. For example, to construct the classification feature vector of (6th, 12th) month pair, first, common features between 6 and 12th are found. Next, the common feature list of 6th is combined with the common feature list of 12th. Finally, combined feature vectors are classified by using Linear-SVM; also, their performances are recorded.

Scenario 2: Union of selected features of two-month periods: This scenario is very similar to the previous one. The only difference is that instead of the intersection, the union operator is used for constructing the classification feature vector.

Table 1 Feature Selection Solutions for each repetition of BVNS for each period

	1st run seed = 10	2nd run seed = 20	3rd run seed = 30	4th run seed = 100	5th run seed = 1000	6th run seed = 12,323	Final Solution
SC Total:845 subjects	{7, 49, 80, 88, 95, 116, 118, 129,131, 138, 164}	{7, 49,79, 80, 88, 95, 116,118,129, 131, 138, 164, 173, 177}	{7, 49, 80, 88, 95, 116, 118, 129, 131, 138, 164}	{7, 49, 69, 77, 80, 88, 95, 116, 118, 129, 131, 138, 164, 173, 177}	{7, 49, 77, 80, 116, 118, 129, 131, 138, 164}	{7, 49, 77, 80, 88, 95, 116, 118, 129, 131, 138, 164, 177}	{7, 49, 77, 80, 88, 95, 116, 118, 129, 131, 138, 164, 177}
6th Month Total:740 subjects	{5, 7, 33, 77, 79, 80, 88, 95, 116, 118, 129, 131, 134, 138, 143, 155, 159, 164, 173, 204, 205}	{5, 7, 33, 79, 80, 88, 95, 116, 118, 129, 131, 134, 138, 159, 164, 173, 204, 205}	{5, 7, 33, 77, 79, 80, 88, 93, 95, 116, 118, 129, 131, 134, 138, 143, 159, 164, 173, 185, 204, 205}	{5, 7, 33, 77, 79, 80, 88, 93, 95, 116, 118, 129, 131, 134, 138, 145, 155, 164, 173, 185}	{5, 7, 33, 77, 79, 80, 88, 93, 116, 118, 129, 131, 134, 138, 145, 155, 159, 164, 173, 185, 204, 205}	{5, 7, 33, 77, 79, 80, 88, 95, 116, 118, 129, 131, 134, 138, 143, 164, 173, 205}	{5, 7, 33, 77, 79, 80, 88, 93, 95, 116, 118, 129, 131, 134, 138, 143, 155, 159, 164, 173, 185, 204, 205}
12th Month Total:689 subjects	{5, 7, 53, 77, 79, 80, 89, 96, 117, 119, 130, 132, 137, 139, 143, 164}	{5, 7, 53, 77, 79, 80, 89, 96, 117, 130, 132, 137, 139, 155, 164}	{7, 89, 96, 119, 130, 132, 139, 164}	{5, 7, 53, 77, 79, 80, 89, 96, 117, 119, 130, 132, 137, 139, 155, 164}	{5, 7, 53, 77, 79, 80, 89, 96, 117, 119, 130, 132, 136, 139, 164}	{7, 53, 77, 79, 80, 96, 117, 119, 130, 132, 137, 139, 164, 171}	{5, 7, 53, 77, 79, 80, 89, 96, 117, 119, 130, 132, 137, 139, 164}
24th Month Total:556 subjects	{79, 80, 89, 117, 119, 130, 132, 136, 139, 165, 174}	{5, 79, 80, 89, 117, 119, 130, 132, 136, 139, 165, 174}	{7, 79, 80, 89, 117, 119, 130, 132, 136, 139, 165, 174}	{7, 79, 80, 89, 117, 119, 130, 132, 136, 139, 165}	{7, 79, 80, 89, 117, 119, 130, 132, 136, 139, 165, 174}	{5, 79, 80, 89, 117, 119, 130, 132, 136, 139, 165}	{5, 79, 80, 89, 117, 119, 130, 132, 136, 139, 165,174}

Table 2 Two months and three months variations

Two months variations	
(SC, 6th)	(6th, 12th)
(SC, 12)	(6th, 24th)
(SC, 24th)	(12th, 24th)
Three months variations	
(SC, 6th, 12th)	(SC, 12th, 24th)
(SC, 6th, 24th)	(6th, 12th, 24th)

Scenario 3: Intersection of selected features of three-month periods: In this scenario, common features selected by BVNS for all possible 3-months combinations (listed in Table 2) are used for classification.

Scenario 4: Union of selected features of three-month periods: This scenario is very similar to the previous one. The only difference is that instead of the intersection, the union operator is used for constructing the classification feature vector.

4.2 Experimental results of classification with SVM (Phase 3)

In the Linear-SVM classification step, the dataset is divided as training and test sets. For a reliable measurement of the performance of the SVM, the *n-fold cross-validation* technique is used [29]. The data set is divided into *n* subsets, and the holdout method is repeated *n* times. In each iteration, one of the *n* subsets is used as the test set, and the other *n–1* subsets are included as the training set. Then, the average error across all *n* trials is computed. As it is widely used in the literature, tenfold-cross-validation was chosen in this work.

As for the calculation of effectiveness and efficiency of the proposed classification scheme, we calculated several terms, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). They include the following:

- True Positive: A patient who has AD, correctly identified as he/she, has AD.
- False Positive: A Patient whose stage of the disease is MCI, incorrectly identified as he/she, has AD.
- True Negative: A Patient whose stage of the disease is MCI, correctly identified as his/her stage of the disease, is MCI.
- False Negative: A patient who has AD, incorrectly identified as his/her stage of the disease, is MCI.

We use two classification functions: sensitivity and specificity.

- Sensitivity: It is the true positive rate, and it measures the proportion of positive that is correctly identified. It can be expressed as follows:

$$Sensitivity = \frac{TP}{TP + FN} * 100(\%) \tag{11}$$

Table 3 The best combination of each scenario and its average results for tenfold cross-validation

Scenario # (best combination)	Avg. Training accuracy (%)	Avg. Testing accuracy (%)	Avg. Testing sensitivity (%)	Avg. Testing specificity (%)
1 (6th, 24th)	99.53	99.96	99.93	100.00
2 (6th, 24th)	96.13	96.72	88.00	100.00
3 (6th, 12th,, 24th)	98.57	99.44	98.00	100.00
4 (SC, 6th, 24th)	94.83	96.72	88.00	100.00

- Specificity: It is the true negative rate, and it measures the proportion of negatives that are correctly identified. It can be expressed as follows:

$$Specificity = \frac{TN}{TN + FP} * 100(\%) \tag{12}$$

- Accuracy: It is measured by specificity and sensitivity. It shows the effectiveness of the classifier.

$$Accuracy = \frac{TP + TN}{TN + FP + TP + FN} * 100(\%) \tag{13}$$

We use BVNS results for the classification. Since there are not enough selected features for 36th months or later, features from SC, 6th, 12th, and 24th month are used as intersection and union forms for all possible combinations. In other words, intersections and unions of all possible 2 and 3 months are applied. The best combinations of the scenarios and their average results for tenfold cross-validation are listed in Table 3. Each fold-result for each scenario can be accessed from the appendix link.

Scenario 1: Intersection of selected features of two-month periods
Among all possible combinations for this scenario, the best performance is achieved by the intersection of combination is 6th and 24th-month combination. Totally 305 patients were used. BVNS selects eight features from the intersection of combination (6th and 24th month) for the classification. Furthermore, user-defined constants C was taken as 0.001. The linear-SVM method classified AD from MCI with very high specificity 100% and high sensitivity: 99.93%. Testing and training errors were also very small; both errors were 0.001. Training and testing accuracies are 99.53% and 99.96%, respectively.
Moreover, the effect of the varying user-defined parameter C on the performance of the proposed framework is a critical problem. If the performance of the suggested system is highly dependent upon C, then the generalization capacity of the framework is arguable. In other words, the proposed framework works effectively with the given data and gives poor results when the data change. For this reason, for each configuration, the performance parameters of the suggested method with respect to the varying C are tabulated. Performance result changes according to user-defined constants C showed in Table 4. Changing C on a large scale has a minimal effect on the performance of the

Table 4 Changing of user-defined value C on the best result of scenario 1 (6th–24th month)

User-defined constants (C)	Training accuracy (%)	Testing accuracy (%)	Testing sensitivity (%)	Testing specificity (%)
1	98.35	97.69	91.66	100.00
0.1	97.86	97.69	95.13	100.00
0.01	97.78	97.69	91.66	100.00
0.001	99.53	99.96	99.93	100.00
0.0001	97.99	97.69	95.69	100.00
0.00001	99.20	96.69	96.11	100.00
0.000001	98.32	97.69	91.76	100.00

proposed framework. Thus, the suggested framework is robust by changing parameter C.

Scenario 2: Union of selected features of two-month periods

After trying all possible combinations, the best performance is achieved by the union of the 6th and 24th-month combination. Totally 258 patients were used. BVNS selects 32 features. User-defined constant C was taken as 0.1. The linear-SVM method classified AD from MCI with very high specificity, 100%. Training and testing accuracies are 96.13% and 96.72%, respectively, and its sensitivity is 88.00%.

Scenario 3: Intersection of selected features of three-month periods

The feature vector, which is constructed from the intersection of selected features of 6th, 12th, and 24th months, gives the best performance for this scenario. Totally 309 patients were used. In this case, the number of selected features is 8. User-defined constant C was taken as 0.001. The linear-SVM method classified AD from MCI with a specificity of 100% and a sensitivity of 98%. Training accuracy is 98.57%, and the testing accuracy is 99.44%, nearly equal.

Scenario 4: Union of selected features of three-month periods

The union of (SC, 6th, 24th) produces the best result. Totally 198 patients were used. The number of features selected for this classification is 39. User-defined constant C is taken as 0.001. The proposed framework gives a good classification performance where training and testing accuracies are 94.83% and 96.72%, respectively. Moreover, its classification specificity and sensitivity values are 100% and 88%.

In sum, different variations of intersection and union of the selected features are used to observe the performances of different scenarios. Among them, scenario 1 achieves the best performance: the intersection of selected features of two-month periods (6th-24th month). The best performance is achieved in the early periods, which is clinically most desired. In other words, early detection is very critical for increasing the success rate of the therapy, and the proposed method provides this very early. The results also show that scenarios of unions are less successful than scenarios of intersections. More features provided by the union operator may result in worse classification performance.

Table 5 Results of the classification algorithms for the feature list of the intersection of (6th and 24th)

Classification algorithms	Accuracy (%)
BVNS + LDA	87.24
BVNS + Decision Tree	85.30
BVNS + Naive Bayes	84.32
BVNS + KNN	84.75
BVNS + Linear SVM	99.96

4.3 Comparison with the other classification algorithms

In this study, SVM was used as a base classifier for the proposed framework. In order to show the power of our framework, we perform four other popular classification algorithms on the intersection of the 6th and 24th months-features-list selected by BVNS, which gives the best accuracy results for SVM. The classification methods used for comparison are explained below. All four classification methods are implemented by using the Statistics and Machine Learning Toolbox in Matlab.

- **Linear Discriminant Analysis (LDA):** LDA is utilized in many Machine Learning applications to reduce the number of features to eliminate the phenomenon called the curse of dimensionality [34]. In this study, regularized LDA is applied by using the fit discriminant analysis classifier method of Matlab®.
- **Decision Tree:** Decision Tree Analysis is one of the most widely used and practical methods for supervised learning. It is non-parametric and can be used for both regression and classification problems. It splits the dataset into different branches based on different conditions [35].
- **Naive Bayes:** A Naive Bayes classifier model is a member of the probabilistic learning model. It is based on the famous Bayesian theorem. This method is applicable when the number of input features is high because it is not much affected by the curse of dimensionality [36]. In this work, we applied the Matlab fitcnb method.
- **K Nearest Neighbors (KNN):** KNN is a non-parametric statistical estimation technique. It stores all available cases, and any new case is classified into these cases based on the similarity measure defined [37]. It is trained as an 8-nearest neighbor classifier.

The average of tenfold cross-validation results of four methods based on accuracy is given in Table 5. It is clear that the highest value for average accuracy is depicted in the BVNS + LinearSVM. In other words, the BVNS + LinearSVM approach is significantly superior to the other four approaches in predicting early conversion from MCI to AD**.**

4.4 Comparison with the literature results

Results of our proposed framework are compared with the results of some similar frameworks that use the same ADNI database for early diagnosis of the AD converted from MCI. The results, which are taken from original papers, are listed in Table 6.

Table 6 Performance comparison of similar methods that use the ADNI database

Author	Data	Result	Conversation time	Methods
Davatzikos et al. [13]	MRI and CSF	AUC = 73% Max ACC = 62%	0–36 months	SPARE-AD
Zhang and Shen [38]	MRI, PET and cognitive scores	AUC = 77% ACC = 78% SEN = 79% SPE = 78%	0–24 months	Non-linear manifold learning techniques and semi-supervised classifier
Cuingnet et al. [12]	MRI data	ACC = 67% SEN = 62% SPE = 69%	0–18 months	Methods based on cortical thickness, five voxel-based methods, and two methods based on the hippocampus
Wolz et al. [32]	Combination of different MRI-based features	ACC = 68% SEN = 67%	0–48 months	Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM)
Westman et al. [31]	MRI data	ACC = 59% SEN = 74%	0–12 months	Multivariate analysis
Cheng et al. [8]	MRI, PET, CSF	ACC = 69.4% SEN = 64.3% SPE = 73.5%		Domain transfer feature selection component
Ye et al. [33]	Basic measures and MRI data	AUC = 86%	0–48 months	Logistic regression with stability selection
Shaffer et al. [30]	MRI, PET, CSF and basic measurements	ACC = 72%	0–48 months	Logistic regression models
Gaser et al.[17]	Age and MRI data	AUC = 78%	0–36 months	Kernel regression methods
Moradi et al. [27]	MRI, age and cognitive measures	AUC = 90% ACC = 82% SEN = 87% SPE = 74%	0–36 months	regularized logistic regression and random forest classifier
Our VNS-based Framework	MRI data	ACC = 99.96% SPE = 100% SEN = 99.93%	0–24 months	BVNS & Linear SVM

AUC: area under the receiver operating characteristic curve, ACC: accuracy, SEN: sensitivity, SPE: specificity

All the studies on this table work on the same dataset with different methods, they are using the same evaluation metrics and same time periods. It is clearly seen that the proposed framework outperforms the others. ADNI dataset has an enormous number of features, and extracting the discriminative feature is very crucial. BVNS has made feature selection successfully. Hence, we can get rid of the curse of the dimensionality problem. The performance of the method for different union and intersection scenarios also outperforms the previous methods in Table 6. This means that the proposed framework is consistent. Furthermore, each scenario's accuracy is generally still better than that of the performance methods suggested in the literature.

The best accuracy of the proposed framework is 99.96%, whereas the second-best performance is less than 90%. There is a very clear gap between these performances. As stated earlier, this performance has almost remained unchanged while the user-defined parameter C is changing. This means that the generalization power of the suggested method is very high. Different scenarios to construct the feature vector are tested in the proposed framework, and almost all of them have better performance than that of the literature's best.

Compared to the other studies using the same ADNI database, the method suggested in this study has some other advantages adding to its performance. Those are as follows:

- The diagnosis of conversation from MCI to AD is estimated to take a shorter time than others in the literature. In the literature, diagnosis is generally covered between 0–36th months. However, this study suggests that the highest performance is obtained within 0–24th months (after the subject applies to the doctor). 99.96% accuracy in a short time after the patient applies to the clinic is very critical because the main aim of this research is to prognosis which patient will convert from MCI to AD in his or her future life as early as possible.
- Achieving the best performance in the early months (6th, 24th) is also convenient with clinical findings because in the first stage of AD, brain activity is healthier, and observation of change in the brain is more specific than feature stages of the disease.
- To the best of our knowledge, in this study, BVNS was used successively for the first time in feature selection of MCI and AD subjects from the ADNI database. This method can be used to select discriminative features for other problems where the curse of dimensionality is a challenge.
- The short-processing time makes it possible for the proposed framework to be merged into real-time systems used in clinics.

5 Conclusion

It is important to detect preclinical AD as early as possible for maximal treatment effect. Some studies have addressed this issue in the literature, but their accuracies are far from being satisfactory. In this study, an early diagnostic framework for the MCI subjects that will convert to AD in the future is proposed. Test and training data were obtained from ADNI, which is a high-dimensional dataset. To select the best descriptive features, a VNS based feature selection method is implemented. Correlation-based

Symmetrical Uncertainty is used to evaluate the solutions. After the feature selection step, four feature combination scenarios are modeled and tested with the Linear—SVM method to classify AD and MCI on the dataset. Combining VNS for feature selection with SVM for classification gives high accuracy, high sensitivity, and high specificity values compared to those of previous studies conducted on the same dataset.

Developing a VNS-based framework as a diagnosis aid tool for clinical usage would be a desirable purpose for future research, and the same approach can be applied when problems with a curse of dimensionality are a major issue.

Acknowledgements We wish to thank M. Garcia-Torres for providing us with his implementation.

References

1. Alzheimer's Association: Alzheimer's disease facts and figures. Alzheimer's Dement, **4**(2), 110–33 (2008)
2. Barnes, D.E., Yaffe, K.: The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol. **10**(9), 819–828 (2011)
3. Beheshti, I., Demirel, H., Neuroimaging, A.D.: Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease. Comput. Biol. Med. **64**, 208–216 (2015)
4. Beheshti, I., Demirel, H., Neuroimaging, A.D.: Feature-ranking-based Alzheimer's disease classification from structural MRI. Magn. Reson Imaging **34**(3), 252–263 (2016)
5. Brunton, S., Ginestet, C., Muehlboeck, S., Westman, E., Simmons, A.: Cross-sectional, longitudinal and laterality measures of hippocampus, parahippocampus and entorhinal cortex in mild cognitive impairment. Alzheimers Dement. **9**(4), P430 (2013)
6. Guo, H., et al.: Evaluation of common structural brain changes in aging and Alzheimer disease with the use of an MRI-based brain atrophy and lesion index: a comparison between T1WI and T2WI at 1.5 T and 3T. Am. J. Neuroradiol. **35**(3), 504–512 (2014)
7. Chaves, R., Ramirez, J., Gorriz, J.M., Puntonet, C.G.: Association rule-based feature selection method for Alzheimer's disease diagnosis. Expert Syst. Appl. **39**(14), 11766–11774 (2012)
8. Cheng, B., Zhang, D. and Shen, D.: Domain transfer learning for MCI conversion prediction. In: Medical Image Computing and Computer-Assisted Intervention– MICCAI 2012 (p. 82–90). Springer, Berlin (2012)
9. Choi, S.H., et al.: Estimating the validity of the Korean version of expanded clinical dementia rating (CDR) scale. J. Korean Neurol. Assoc. **19**(6), 585–591 (2001)
10. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
11. Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., et al.: Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PLoS ONE **6**(7), e21896 (2011). https://doi.org/10.1371/journal.pone.0021896
12. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., et al.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage **56**(2), 766–781 (2011)
13. Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q.: Prediction of MCI to AD conversion, via MRI, CSF biomarkers, pattern classification. Neurobiol. Aging **32**(12), 2322.e19-2322.e27 (2011). https://doi.org/10.1016/j.neurobiolaging.2010.05.023
14. Dubois, B., Feldman, H.H., Jacova, C., et al.: Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. Lancet Neurol. **6**(8), 734–746 (2007)
15. Fischl, B.: FreeSurfer. Neuroimage **62**(2), 774–781 (2012)
16. García-Torres, M., Gómez-Vela, F., Melián-Batista, B., Moreno-Vega, J.M.: High-dimensional feature selection via feature grouping: A Variable Neighborhood Search approach. Inf. Sci. **326**, 102118 (2016)
17. Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H.: BrainAGE in mild cognitive impaired patients: Predicting the conversion to Alzheimer's disease. Alzheimer's disease neuroimaging initiative. PLoS ONE **8**(6), e67346 (2013). https://doi.org/10.1371/journal.pone.0067346
18. Gauthier, S., et al.: Mild cognitive impairment. Lancet **367**(9518), 12621270 (2006)

19. Gorji, H.T., Haddadnia, J.: A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. Neuroscience **305**, 361–371 (2015)
20. Hall, M.A.: Correlation-based feature selection for machine learning. The University of Waikato, (1999)
21. Hansen, P., Mladenovic, N.: Variable neighborhood search. Comput. Oper. Res. **24**, 1097–1100 (1997)
22. Hanzel, D.K., Trojanowski, J.Q., Johnston, R.F., Loring, J.F.: High-throughput quantitative histological analysis of Alzheimer's disease pathology using a confocal digital microscanner. Nat. Biotechnol. **17**(1), 53–57 (1999)
23. Latchoumane, C., Vialatte, F.B., Solé-Casals, J., Maurice, M., Wimalaratna, S.R., et al.: Multiway array decomposition analysis of EEGs in Alzheimer's disease. J. Neurosci. Methods **207**(1), 41–50 (2012)
24. Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., Yan, Z.: A review on the forecasting of wind speed and generated power. Renew. Sustain. Energy Rev. **13**(4), 915–920 (2009)
25. Liu, L., Fu, L., Zhang, X., Zhang, J., Zhang, X., et al.: Combination of dynamic (11)C-PIB PET and structural MRI improves diagnosis of Alzheimer's disease. Psychiatry Res **233**(2), 131–140 (2015)
26. Maji, S., Berg, A.C., and Malik, J.: Classification using intersection kernel support vector machines is efficient. In IEEE Computer Vision and Pattern Recognition (2008)
27. Moradi, E., et al.: Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage **104**, 398–412 (2015)
28. Prestia, A., et al.: Prediction of AD dementia by biomarkers following the NIA-AA and IWG diagnostic criteria in MCI patients from three European memory clinics. Alzheimer's Dement **11**(10), 1191–1201 (2015)
29. Refaeilzadeh, P, Tang, L and Liu, H (2009) Cross-validation. In Encyclopedia of database systems. Springer. p. 532–538.
30. Shaffer, J.L., et al.: Predicting cognitive decline in subjects at risk for Alzheimer's disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. Radiology **266**(2), 583–591 (2013)
31. Westman, E., et al.: AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. Neuroimage **58**(3), 818–828 (2011)
32. Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., et al.: Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS ONE **6**(10), e25446 (2011). https://doi.org/10.1371/journal.pone.0025446
33. Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., et al.: Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. BMC Neurol. **12**, 46 (2012). https://doi.org/10.1186/1471-2377-12-46
34. Modarress, D., Elghobashi, S., Tan, H.: Two-component LDA measurement in a two-phase turbulent jet. AIAA J. **22**(5), 624–630 (1984)
35. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
36. Patil, T.R., Sherekar, S.S.: Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int. J. Comput. Sci. Appl. **6**(2), 256–261 (2013)
37. Fukunage, K., Narendra, P.M.: A branch and bound algorithm for computing k-nearest neighbors. IEEE Trans. Comput. **7**, 750–753 (1975)
38. Zhang, D. and Shen, D.: Semi-supervised multimodal classification of Alzheimer's disease. In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. IEEE (2011)
39. Modrego, P.J., Fayed, N., Pina, M.A.: Conversion from mild cognitive impairment to probable Alzheimer's disease predicted by brain magnetic resonance spectroscopy. Am. J. Psychiatry **162**(4), 667–675 (2005)
40. Beheshti, I. et al.: 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. **83**, 109–119 (2017)
41. Westman, E., Muehlboeck, J.S., Simmons, A.: Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. Neuroimage **62**, 229–238 (2012). https://doi.org/10.1016/j.neuroimage.2012.04.056