ISSN: 0976-3104

SPECIAL ISSUE: Emerging Technologies in Networking and Security (ETNS)
Sountharrajan et al

ARTICLE OPEN ACCESS

PARTICLE SWARM OPTIMIZED FEATURE SELECTION FOR ALZHEIMER CLASSIFICATION

S Sountharrajan*, P Thangaraj, E Suganya
Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu. INDIA

ABSTRACT

Alzheimer's disease (AD) refers to a neuro-degenerative chaos that is a general kind of dementia which leads to memory loss, and lack of cognitive functioning and so on. Magnetic Resonance Imaging (MRI) is popularly utilized for human body imagings. MRI is a fundamentally non-invasive method giving high degree clarity on the soft tissue inside the brain better than conventional Computed Tomography (CT), ultrasound, Positron Emission Tomography (PET), etc. SVM is a kind of ANN (artificial neural network) which has got training from supervised learning methods and had showed the benefits of decreasing the training-testing error and hence producing greater recognition precision. This paper investigates empirically PSO's (Particle Swarm Optimization's) effectiveness towards selection of features.

Published on: 2nd -December-2016

KEY WORDS

Alzheimer's disease (AD), Magnetic Resonance Imaging (MRI), Particle Swarm Optimization (PSO)

***Corresponding author: Email:** thangarajp.cse@rediffmail.com

INTRODUCTION

Alzheimer's disease (AD) is a typically occurring kind of dementia. The clinical symptoms are featured as cognitive decline along with degradation of everyday life as well as other symptoms related to neuropsychiatric or behavioural change. It is very difficult to diagnose AD in its early stage because there is no existing known biomarker. AD has a stealthy onset that can be of either genetic or environment oriented reasons. Distinguishing various kinds of dementia is complex. The transitional phase named as Mild Cognitive Impairment (MCI) occurs between usual and demented behavior. The disorder is diagnosed with symptoms of high decline in cognition than predicted for the particular age but at the same time no disruption is found in the routine activities.

The key symptom of MCI is featured by the decline in memory as well as the impairment of cognition. Researchers report the association of AD and MCI that MCI provides the threat of 10 to 64 percent in leading to AD [1]. AD is defined as continuous neurodegenerative illness and it is different from MCI by its symptom of progressive decline of routine activities. The occurrence of AD is found to increase drastically at 65 years of age and it is also studied that around 26 million of people in the globe are affected by AD, the number is also expected to increase rapidly by as much as four times by 2050. In order to treat or prevent AD, many researches have been undertaken to diagnose it in early stage itself. Hence, by detecting the change in the tissue of brain that reflects the pathological process of MCI can assist in preventing or postponing further progression or its conversion in to AD. This early detection and efficient intervention could possibly decrease further impairment.

The major clinical factor to identify and predict the advancement of AD is the universally accepted cognition test. The research with the objective to improve early detection is done with a tool kit which includes brain images, biomarker of Cerebrospinal Fluid (CSF) as well as biomarker of plasma.

The main reason to select features is to decide the quantity of feature subset which is necessary and accurate for the classifier to differentiate the link or action into usual or invasive. It is also found that the feature subset p is low when compared to subset m. FS investigates the combination space to obtain the best combination of features.

COMPUTER SCIENCE

Normally FS is classified into three categories: First is the conventional exponential algorithm, where large quantities of subsets are assessed and the quantity grows in an exponential manner with dimensional growth. Here, exhaustive search is the classical algorithm used. The second category being sequential technique which adds or removes the feature sequentially, but the limitation is that the algorithm has the challenge of being trapped into local minima. The final category is the meta-heuristic algorithm that includes random search to ensure escape from local minima.

Particle Swarm Optimization (PSO) is a populace oriented stochastic technique to solve optimization problem. In PSO, the particle wanders in the search area to find optimal solution. A candidate solution is obtained from the present location of the particle. Every particle goes in search of optimal position by altering the velocity based on rules simulating the behavior of birds. PSO is a part of the category of swarm intelligent methods which is chosen for solving problem of optimization [2].

RELATED WORK

Jongkreangkrai et al [3] enhances the AD classification performance by merging hippocampus and amygdale volume and thickness of entorhinal cortex. Its aim is to examine the helpful feature got through MRI to classify the AD affected patients with the help of Support Vector Machine (SVM). The samples of MR brain images weighed as T1 of 100 normal and AD patients were treated with the software Free Surfer to evaluate hippocampus and amygdala volume and thickness of entorhinal cortex of both the hemispheres of brain. Comparative volume of hippocampus and amygdale are computed to correct variations in each head size. SVM is deployed with different fusions of feature: (H: hippocampus relative volume, A: amygdala relative volume, E: entorhinal cortex thickness, HA: hippocampus and amygdala comparative volume and ALL: all features). Receiver operating characteristic (ROC) analysis is applied to assess the technique. AUC range of five fusions are 0.8575 (H), 0.8374 (A), 0.8422 (E), 0.8631 (HA) and 0.8906 (ALL) respectively. Even though “ALL” gave the maximum AUC, no numerically important difference had been noticed except in the case of “A”. Result proved the feasibility of recommended features towards AD patient via computer aided classification.

Moradi et al [4] presents a new Magnetic Resonance Imaging (MRI) oriented technique to predict the conversion of MCI to AD within 1 to 3 years prior to medical diagnoses. Primarily, a new MRI bio-marker for MCI to AD progressions is designed with the help of semi supervised learning and then combined with age factor and cognitive metric of the subject with a supervised learning algorithm that results in the cumulative biomarker. Added value of this new feature to predict the MCI to AD change is formulated on the received data from Alzheimer's Disease Neuroimaging Initiative (ADNI) databases. With this ADNI data, the MRI biomarker attained a 10 fold cross validated region within the ROC curve (AUC) of 0.7661 in differentiating progressive MCI patient (pMCI) from stable MCI patient (sMCI). Cumulative biomarker according to MRI data altogether along with base cognitive metrics and patient’s age reached a 10 fold cross validated AUC value of 0.9020 in differentiating MCI from sMCI. The output presents the demonstration of the probability of proposed technique to diagnose AD early and the significant part of MRI in predicting the conversion of MCI to AD. Moreover, the result showed that merging of MRI data with cognition test result enhances the precision of predicting the conversion of MCI to AD.

Zhang et al [5] suggested a technique that primarily deployed digital wavelet transform for extracting attributes and later applied Principal Component Analysis (PCA) for reducing the space of the feature. Then, kernel support vector machine (KSVM) is built along RBF kernel, utilizing Particle Swarm Optimization (PSO) for optimizing the attributes C and σ. Over fitting is avoided by the use of five-fold cross validation. The experiment used a data set of 90 brain images downloaded from the web site of Harvard Medical School. The fivefold cross-authentication classification output proved the classification precision as 97.78% which is greater than 86.22% by BP-NN and 91.33% by RBF-NN. The parameters are selected in comparison to PSO and other random selection technique. The result proved that PSO is more efficient to construct best KSVM.

Alzheimer’s disease (AD) is a well-known precursor to dementia. The treating of AD is more effective if it is detected at the early stage. Being the pre-cursory phase of AD, Mild cognitive impairment (MCI) is considered as better focus point for early diagnosis, but its diagnosis is much difficult because of cognitive deterioration’s subtlety. Lee et al [6] develops a technique for automation of the detection of AD and MCI with SVM as well as Diffusion Tensor Imaging (DTI) data. The method applied two SVM models: first is to classify AD and MCI and next is to classify AD and Normal Control (NC). Both the model uses the Fractional Anisotropy (FA) and the

THE IIOAB JOURNAL ISSN: 0976-3104

mode of anisotropy (MO) values of DTI were utilized as attributes. In the two models, MO values gave the result of optimal performance than FA values. In an individual assessment the AD-MCI classifier shows sensitivity of 69.2 %, specificity of 100 % and precision of 89.7 %, and the AD-NC classifier shows sensitivity of 84.6 %, specificity of 90.9 % and precision of 87.5 %. The output proves satisfactory and recommends classification of DTI on the basis of SVM is most strong while detecting MCI and AD in earlier stage.

Yang et al [7] proposes an MRI-based classification structure to differentiate AD and MCI patient from healthy subjects with the use of several attributes as well as various classifiers. As the first stage, the features are extracted (volume and shape) from the given MRI data with sequential image processing steps followed by the application of principal component analysis (PCA) to transform the collection of attributes of probably co-related parameters into a small set of value of linearly non-correlated parameters, minimizing the dimensions of features space. In the end, a new data mining structure merging with SVM, PSO to classify AD/MCI is designed. The hybrid technique is compared with conventional classifiers, namely SVM and SOM (Self Organizing Map) were trained for the classification of patients. It is noted that for the suggested structure the precision of classification had enhanced to 82.35% and 77.78% among the patient with AD as well as MCI. The results attained up to 94.12% and 88.89% in AD and MCI through fusing of features of volume and shape with the use of PCA. Current result suggests that new multivariate techniques of pattern matching attain a clinically relevant precision for a priori prediction of the development from MCI to AD.

MATERIALS AND METHOD

[Figure- 1] depicts the flowchart of the proposed method. The technique is given as follows:

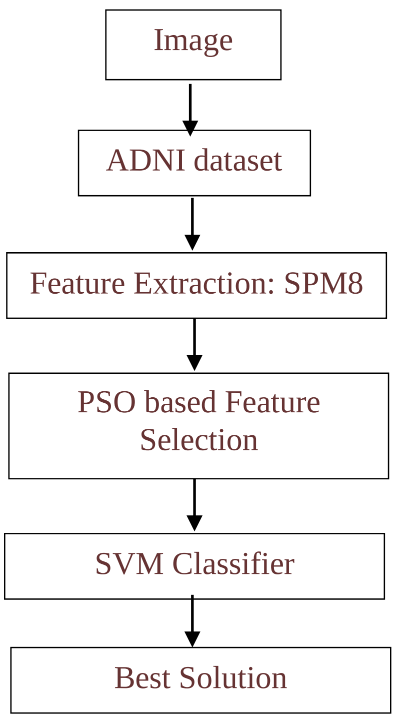

Fig: 1. Flowchart for Proposed Methodology

....

ALZHEIMER'S DISEASE NEURO-IMAGING INITIATIVE (ADNI) DATABASE

This study is done with the data got from Alzheimer's Disease Neuro-imaging Initiative (ADNI) database (http://adni.loni.usc.edu/). The main objective of ADNI is to evaluate whether MRI, PET, and other existing biomarkers, clinical and as well as neuropsychological evaluation can be merged to calculate the progression of MCI and early AD. Determining responsive and particular biomarkers of early AD development is needed to help the researcher and physiologists to design novel treatment and measure the efficiency along with decrease in time and costs of trials.

FEATURE EXTRACTION

Statistical parametric mapping means building and evaluating the extensive numerical procedures for testing the hypotheses on functional imaging of data. These notions were included in the SPM software which is designed to analyze the sequences of brain image data that could be the series of images from various cohorts or time-series from one subject. A huge benefit of SPM is its uni-variant system unlike other multivariate methods where huge quantity of observation is required. Even though wide research had been undertaken in the domain of features extraction, no investigation was performed in the domain of AD CAD since the image represents huge volume of data and imaging study has restricted subjects [8]. To classify AD, feature vector is of great dimension and methods to minimize the input space dimensionality are necessary to improve the classification precision in total.

PARTICLE SWARM OPTIMIZATION (PSO)

PSO is a populace-based method that was built by Eberhat and Kennedy. PSO is a successful and well known global search method. This is the most appropriate protocol to handle the issues related to features selection because of the given factors: assesses the fitness function of the complete swarm in every step. The position of the given two best particles is used to update the v_i of ith particle: (1) the best position travelled by a particle (p_B) and (2) the best position of the neighbors of ith particle that had been travelled till then (n_B). If the complete swarm is considered as neighbourhood region, then it becomes the global best and is called as " g_B ". From the given information, the enhanced equation may be deduced as: Here, rand() is an arbitrary number generator whose value is between [0, 1]. The rand() is performed while they occur. c is called the "inertia weight". While c is lower than 1, the particle helps exploitations over explorations; else if c is greater than 1, the particle helps explorations over exploitations. The parameter a_1 and a_2 are non-negative constants termed as "acceleration coefficient" [9]. The position of the particle swarm is modified based on formula (3) and (4). They will get near to one another from several directions. The PSO executes the process repeatedly till the stopping factor is reached. It is seen that vmax (the maximum velocity) must be decided prior, with the objective to keep the optimizer in a considerable range.

FEATURE SELECTION BASED PSO

The idea of PSO for best features selection [10] issue is taken into consideration. Take a huge feature space complete with features subset. Every features subset may be taken as a location in such space. Consider that there exists N features, then it is found that there will be 2N types of subsets. The subset with minimum distance and high quality classification is known as the best location. Then, a particle swam is inserted into the features space and each particle takes one location. With the aim to reach the optimal position, the particles start flying in the space. In due course of time, they alter the positions, interact among themselves and go in search of the local as well as global optimum positions. In time, they ought to converge at a good probably optimum position. This capacity for exploration ensures the better performance of feature selection and hence the discovery of best possible subset.

The position of the particle is given as binary bits of string of length N, wherein N is the complete quantity of entities. Each bit representing a feature, the value `1' refers to the respective feature being chosen and '0' if not chosen. Every position is a feature subset. The velocity of every particle is given as positive integer, ranging within 1 and Vmax. It states the count of particle bits that need to be modified at a specific time to be similar to the global optimum position. The count of various bits among 2 particles is related to the variation in their positions respectively.

Next to the velocity updation, the position of the particles is modified by the novel velocity. Consider that the novel velocity is V, and the quantity of various bits between the present particle and gbest is g x. Two cases are found to occur while the position is being updated:

1) V<= g x . Here, the velocity of the particle is less or equal to the difference in position between the particle and gbest. Distinct from the gbest, V bits of the particles are modified in random. Instead of being similar to gbest, the particles move towards the global optimum with continuous exploration of the search area.

2) V> g x . Here, the velocities of the particles overrun the location difference between the particle and gbest. Also, the particles are modified to be similar to gbest, it is further changed in random ('random' means 'exploration ability') modification (V- g x) of bits outside of the various bits between the particle and gbest. Hence even after reaching the best position, the particle moves further in other directions, for more searches.

The high velocity Vmax acts as a limitation to monitor the global exploring ability of a particle swarm. A greater Vmax helps global explorations, at the same time small Vmax stimulates the local exploitations. If Vmax is very less, then the particle faces mode challenge to escape from the local optima whereas with high Vmax, the particle may fly rapidly away from a good solution.

SUPPORT VECTOR MACHINE (SVM)

SVM is a kind of ANN, which is trained by the use of supervised learning, shows the benefit on minimizing errors on training as well as testing, which results in getting better recognition precision. [11]. But, some features data is linearly nonseparable. Also, in few circumstances, the features will not be properly separable, specifically in the case of border between the classifications. In order to permit flexibility to some extent in dividing the groups, SVM uses a costs parameter, represented as C, to monitor the tradeoff between permitting training defects and forced rigid margin. Costs function with C is given as follows, wherein ς_i denotes a slack variable,

$$\cos t = C\sum_{i=1}^{N}(\varsigma_i) \tag{1}$$

Mapping the patterns in a high dimensional features space was developed by fusing features to generate a kernel matrix. The kernel matrix is normally built with a kernel function that has 2 patterns as argument and gives output value. This research employs a radial basis function (RBF) kernel. One- against-rest is used to assemble the classifier which distinguishes individual class. This technique contains of building 1 SVM in each class and training is given to differentiate sample of 1 class from the others. Generally, unknown pattern's classification is performed based on the highest output of all the SVM.

$$k(x_i, \mathrm{y}_i) = \mathrm{e}^{-\gamma\|x_i - y_i\|\mathrm{Fit}_p}, i = j = 1, 2, ..., n \tag{2}$$

wherein x_i represents the input vector, y_i represents the jth prototype vector, and $Fit_p correctly-classified / total$ number of testing data. Lastly, the best solution can be given by usage of Lagrange technique,

$$L_p \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\varsigma_i - \sum_{i=1}^{m}\alpha_i\{y_i(w.x_i + b) - 1 + \varsigma_i\},$$
$$L_D \equiv \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\alpha_i\alpha_j y_i y_j^k(x_i, y_j) \tag{3}$$

where $\|w\|$ is the Euclidean norm of w, αi which denotes Lagrange multipliers, L_p represents the Lagrange function, and L_D represents the dual solution of L_p. C as well as γ are utilized to monitor the tradeoff between training error as well as generalization capability in SVM with RBF kernel. Hence PSO is used to obtain the best fusion of C and γ.

The usage of an SVM to classify the image is an instance for linear discrimination. In the base model, it is a binary classifier, that implies that it splits the space into where the MR image is distributed into 2 classes through identification of a dividing hyper plane. The inspiration in implementing SVM is that it applies the principles of structural risk minimization that has the objective to obtain a hyper plane that increases the length between training classes.

RESULTS

[Table- 1] shows the summary of results. **[Figure- 2 to 4]** shows the classification accuracy, sensitivity and specificity respectively.

Table: 1. Summary of Results

	C=1, gamma =0.01	C=10, gamma =0.01	C=100, gamma =0.01	C=1, gamma =0.001	C=10, gamma =0.001	C=100, gamma =0.001
Classification Accuracy	84.67	77.33	86.67	86	81.33	86.75
AD- Sensitivity	0.8667	0.7111	0.8889	0.9111	0.7556	0.8696
MCI - Sensitivity	0.8333	0.8333	0.8667	0.8667	0.8667	0.9167
CS-Sensitivity	0.8444	0.7556	0.8444	0.8	0.8	0.8
AD- Specificity	0.8889	0.875	0.9	0.9072	0.898	0.9192
MCI - Specificity	0.9167	0.8462	0.9286	0.9167	0.875	0.9268
CS- Specificity	0.9468	0.8913	0.9583	0.949	0.9149	0.9406

COMPUTER SCIENCE

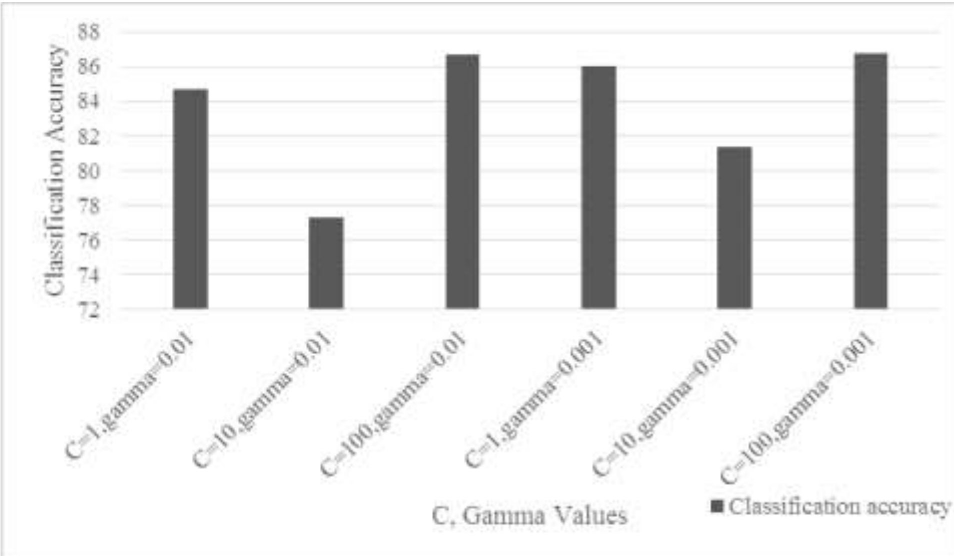

Fig: 2. Classification Accuracy

[Table- 1] and [Figure- 2] shows that the classification accuracy of C=100, Gamma=0.001 performs better by 2.43% than C=1, Gamma=0.01, by 11.48% than C=10, Gamma=0.01, by 0.09% than C=100, Gamma=0.01, by 0.87% than C=1, Gamma=0.001 and by 6.45% than C=10, Gamma=0.001.

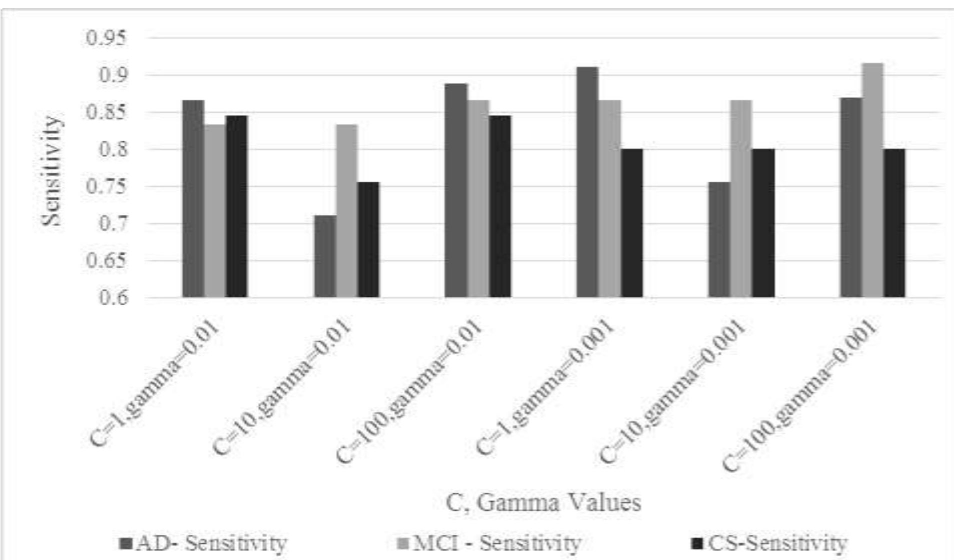

Fig: 3. Sensitivity

[Table- 1] and [Figure- 3] shows that the sensitivity of AD sensitivity performs better than MCI and CS for C=100, Gamma=0.01. Results show that the average values for C=100, Gamma=0.01 performs better by 2.16% than C=1, Gamma=0.01, by 12.24% than C=10, Gamma=0.01, by 0.86% than C=1, Gamma=0.001, by 7.08% than C=10, Gamma=0.001 and by 0.53% than C=100, Gamma=0.001.

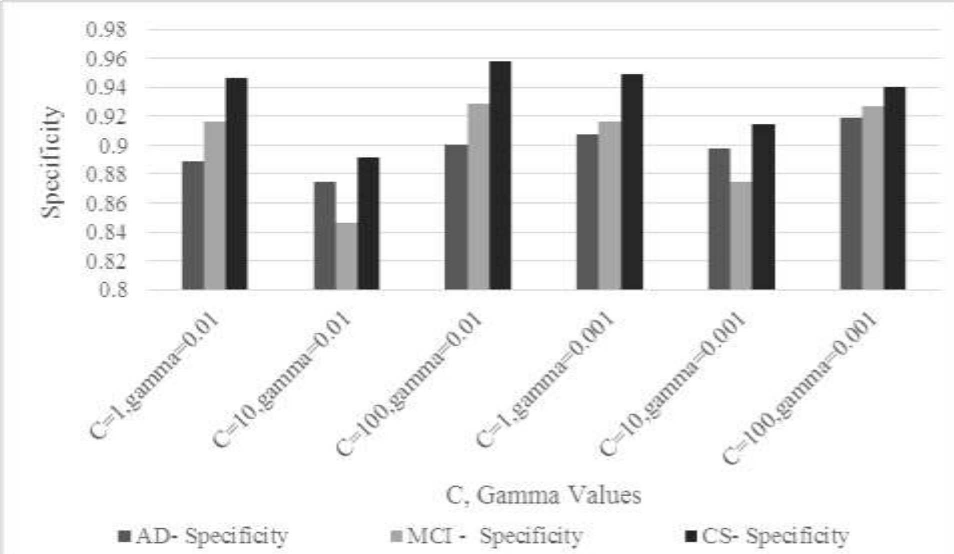

Fig: 4. Specificity

[Table- 1] and [Figure- 4] shows that the specificity of AD specificity performs better than MCI and CS for C=100, Gamma=0.01. Results show that the average values for C=100, Gamma=0.01 performs better by 1.25% than C=1, Gamma=0.01, by 6.5% than C=10, Gamma=0.01, by 0.5% than C=1, Gamma=0.001, by 3.62% than C=10, Gamma=0.001 and by 0.01% than C=100, Gamma=0.001.

CONCLUSION

Alzheimer's Disease (AD), a general kind of dementia, is noticed very often in the people of old age. In order to detect AD, one of the best imaging methods is MRI that permits the quantitative estimation of features of brain to assess AD through a non-invasive method. PSO conducts search through a particle swarm that updates from each round of search. In order to find the best solution, every particle flies in the direction of previously known best position (pbest) and the best global position in the swarm (gbest). To assess the performance of every obtained solution, SVM is constructed. To experiment AD, MCI and CS specificity and Sensitivity are utilized Result shows that the classification accuracy of C=100, Gamma=0.001 performs better by 2.43% than C=1, Gamma=0.01, by 11.48% than C=10, Gamma=0.01, by 0.09% than C=100, Gamma=0.01, by 0.87% than C=1, Gamma=0.001 and by 6.45% than C=10, Gamma=0.001.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

The authors report no financial interests or potential conflicts of interest.

REFERENCES

[1] Alzheimer's disease facts and figures, [2012] http://www.alz .org/alzheimers disease facts figures.asp? type =homepage.

[2] Wang S, Phillips P, Yang J, Sun P, Zhang Y. [2016] Magnetic resonance brain classification by a novel binary particle swarm optimization with mutation and time-varying acceleration coefficients. *Biomedical Engineering/Biomedizinische Technik*.

[3] Jongkreangkrai C, Vichianin,Y, Tocharoenchai C, Arimura H. [2016, March]. Alzheimer's Disease Neuroimaging Initiative. Computer-aided classification of Alzheimer's disease based on support vector machine with combination of cerebral image features in MRI. In *Journal of*

Physics: Conference Series 694(1): 012036). IOP Publishing.

[4] Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Alzheimer's Disease Neuroimaging Initiative. [2015] Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects.*NeuroImage* 104:398-412.

[5] Zhang Y, Wang S, Ji G, Dong Z. [2013] An MR brain images classifier system via particle swarm optimization and kernel support vector machine. *The Scientific World Journal*.

[6] Lee W, Park B, Han K. [2015] SVM-Based Classification of Diffusion Tensor Imaging Data for Diagnosing Alzheimer's Disease and Mild Cognitive Impairment. In International Conference on Intelligent Computing(pp. 489-499). Springer International Publishing.

[7] Yang ST, Lee JD, Chang TC, Huang C H, Wang J. J, Hsu WC, Li KY. [2013]. Discrimination between Alzheimer's disease and mild cognitive impairment using SOM and PSO-SVM. *Computational and mathematical methods in medicine*.

[8] Ramírez J, Górriz JM, Salas-Gonzalez D, Romero A, López M, Álvarez I, Gómez-Río M. [2013] Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features. *Information Sciences* 237:59-72.

[9] Zhang Y, Wang S, Ji, G. [2015] A comprehensive survey on particle swarm optimization algorithm and its applications. Mathematical Problems in Engineering 38.

[10] Kennedy J. [1997, April]. The particle swarm: social adaptation of knowledge. In *Evolutionary Computation, 1997. IEEE International Conference on* (pp. 303-308). IEEE.

[11] Cortes C, Vapnik V. [1995] Support-vector networks. *Machine learning* 20(3): 273-297.