JCSE Journal of Computing Science and Engineering

Regular Paper

Journal of Computing Science and Engineering,
Vol. 10, No. 4, December 2016, pp. 118-127

An Improved Sample Balanced Genetic Algorithm and Extreme Learning Machine for Accurate Alzheimer Disease Diagnosis

Vasily Sachnev*
School of Information, Communication and Electronics Engineering, The Catholic University of Korea, Bucheon, Korea
bassvasys@hotmail.com

Sundaram Suresh
School of Computer Engineering, Nanyang Technological University, Singapore
ssundaram@ntu.edu.sg

Abstract
An improved sample balanced genetic algorithm and Extreme Learning Machine (iSBGA-ELM) was designed for accurate diagnosis of Alzheimer disease (AD) and identification of biomarkers associated with AD in this paper. The proposed AD diagnosis approach uses a set of magnetic resonance imaging scans in Open Access Series of Imaging Studies (OASIS) public database to build an efficient AD classifier. The approach contains two steps: "voxels selection" based on an iSBGA and "AD classification" based on the ELM. In the first step, the proposed iSBGA searches for a robust subset of voxels with promising properties for further AD diagnosis. The robust subset of voxels chosen by iSBGA is then used to build an AD classifier based on the ELM. A robust subset of voxels keeps a high generalization performance of AD classification in various scenarios and highlights the importance of the chosen voxels for AD research. The AD classifier with maximum classification accuracy is created using an optimal subset of robust voxels. It represents the final AD diagnosis approach. Experiments with the proposed iSBGA-ELM using OASIS data set showed an average testing accuracy of 87%. Experiments clearly indicated the proposed iSBGA-ELM was efficient for AD diagnosis. It showed improvements over existing techniques.

Category: Smart and intelligent computing

Keywords: Alzheimer disease; OASIS; Improved samples balanced genetic algorithm; Extreme Learning Machine

I. INTRODUCTION

Alzheimer disease (AD) is an example of dementia or neurological disorder with various symptoms, including memory loss, confusion, and learning difficulties. AD is most common in elderly people. AD diagnosis, especially in the early stages, is one of the most challenging problems in the field of medicine. AD diagnosis using simple and progressive methods can give an early signal to start treatment which may slow down the disease [1]. Hence, new and more advanced techniques are needed for AD diagnosis at early stages.

Brain imaging is one of the most powerful tools for AD diagnosis and AD research. Brain imaging investigates the human brain by visualizing brain tissues using a progression of different methods. Several brain imaging

Open Access http://dx.doi.org/10.5626/JCSE.2016.10.4.118 http://jcse.kiise.org

Received 08 November 2016; **Accepted** 06 December 2016
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

techniques have been developed for diagnosis and research, including the following: computed tomography (CT), single-photon emission computed tomography (SPECT), positron emission tomography (PET), and magnetic resonance imaging (MRI). According to recent studies for AD diagnosis [2, 3], CT has a poor accuracy in detecting AD. Both SPECT and PET use radioactive isotopes for testing [4, 5]. They are harmful to patients after several tests. In addtrion, data collected from SPECT and PET might not be sufficient for proper AD diagnosis as SPECT has low spatial resolution which causes lower accuracy for AD diagnosis. Compared to CT [2, 3], SPECT and PET [4, 5], MRI [6-8] is an efficient and fast brain imaging tool with spatial resolution of the brain shape and volume high enough for accurate AD diagnosis. MRI is also able to give a dynamic analysis of the brain shape and volume. Dynamic analysis can detect fast modifications in brain activity which is extremely important for AD diagnosis [6, 7].

MRI is widely used for accurate AD diagnosis in the early stages. Analysis of MRIs taken from AD patients in the early stages shows modifications in the hippocampus and entorhinal cortex brain areas [9-11]. However, authors who presented methods to differentiate AD patients and normal persons by examining the volume of the hippocampus and entorhinal cortex in the brain areas chosen manually. Such manual method is not always accurate. It depends on researchers which might lead to mistakes. Hence, an automatic approach is highly desirable.

Automatic AD diagnosis approaches use features extracted from MRI and traditional machine learning techniques for accurate AD classification [6-8, 11]. AD diagnosis is possible due to artificial changes in the brain's volume captured by MRI. MRI analysis based on regions-of-interest (ROI) can identify brain areas responsible for AD and numerically rank the discovered brain areas for further analysis. Lao et al. [11] and Chupin et al. [12] have presented an ROI based approach for AD diagnosis which can automatically segment the hippocampus areas using probabilistic and anatomical methods. ROI based AD diagnosis is simple in nature. It may efficiently detect AD patients when tissue loss in the brain is significant. However, ROI based AD diagnosis can fail when tissue loss in the brain is small. Hence, a new brain evaluation methodology different from ROI needs to be developed.

A suitable evaluation methodology based on MRI has been invented by Ashburner and Friston [13]. They proposed whole brain morphometric evaluation for accurate AD detection. It is efficient even when tissues loss is relatively small. Their brain morphometric evaluation can identify and measure modification of tissue volume in individual brains or between brains of AD patient and normal person. Brain morphometric evaluation will create a set of relevant morphometric features for accurate AD diagnosis. The extracted morphometric features or voxels represent probabilities of the gray and white matters as well as cerebrospinal fluid tissues. Kloppel et al. [6] and Davatzikos et al. [7] have trained Support Vector Machine (SVM) classifiers for AD diagnosis using a set of morphometric features. The proposed iSBGA-ELM (improved sample balanced genetic algorithm and Extreme Learning Machine) of this study utilizes a set of morphometric features to build an AD classifier and process experiments.

Mahanand et al. [14] have utilized MRI from the public Open Access Series of Imaging Studies (OASIS) [15] database to build a set of 5,788 features extracted using the voxel-based morphometry (VBM) approach. They reduced the set of features using Principal Component Analysis (PCA). The reduced set of features was used to build an AD classifier using a self-adaptive resource allocation network (SRAN). Experiments with 30 normal persons and 30 AD patients from the OASIS database clearly indicate that discovery of a reduced set of features is sufficient for accurate AD diagnosis [14].

Although SVM is a popular and efficient machine learning tool for solving various classification problems, including the AD classification problem [6, 7], the training phase of SVM is computationally extensive. It usually needs significant time to build a classifier with high classification accuracy. Hence, a new machine learning technique with acceptable classification accuracy with a fast training phase is needed.

Most of the set of morphometric features extracted from MRI of OASIS data are redundant for analysis. Thus, direct use of the complete set of extracted morphometric features to build an AD classifier does not guarantee high classification performance. Hence, a search for a reduced set of voxels (or features) to create an AD classifier with better classification accuracy is needed.

The search for a reduced set of features with promising properties is one of the most common optimization problems in many research areas of science. Saraswathi et al. [16] have searched for a reduced set of genes from popular Global Cancer Map (GCM) to build a cancer diagnosis approach using integer coded genetic algorithm and particle swarm optimization coupled with Extreme Learning Machine (ICGA-PSO-ELM). This method has a serious drawback; the number of genes has to be assigned manually. The cited cancer classifier detects 14 types of cancer with high classification accuracy [16]. In comparison, Sachnev et al. [17] have used a binary coded genetic algorithm to search for an optimal set of genes from the GCM database. They used about 52 discovered biomarkers from a set of 92 chosen genes to build a cancer classifier.

An efficient AD diagnosis approach based on the iSBGA-ELM is proposed in this paper. The proposed iSBGA is a completely automatic approach for searching for a robust set of voxels and building a classifier for accurate AD diagnosis. The iSBGA uses two crossovers designed specifically for data from the OASIS database

(namely a *regular sample balanced crossover* and an *irregular sample balanced crossover*) and sample balanced mutation to create a basis for iSBGA. A reduced set of robust voxels is then used to train an AD classifier based on the ELM. The ELM classifier with maximum classification accuracy is our final AD diagnosis approach. The reduced set of robust voxels which creates the ELM classifier with the highest accuracy is then used to discover biomarkers responsible for AD.

This paper is organized as follows. The OASIS database is presented in Section II. The framework of the proposed iSBGA-ELM is presented in Section III in detail. Experimental results are presented in Section IV. Section V concludes the paper.

II. OASIS

The publicly available OASIS database [15] is a famous database of MRI scans for AD research. OASIS contains MRI from two sets of data (218 people from 18 to 59 years old, and 198 people from 60 to 96 years old). The second set of data (198 people) was used in this paper for analysis. Of the 198 people, 98 are normal people without AD. The Clinical Dementia Rating (CDR) of the 98 patients is 0; whereas the remaining 100 are AD patients, including 70 with very mild AD (CDR = 0.5) and 28 with mild AD (CDR = 1).

The MRIs in the OASIS database were created using a Siemens 1.5-T vision scanner in a single imaging session. T1-weighted 3D MPRAGE (magnetization-prepared rapid acquisition gradient echo) datasets of the whole brain were acquired. The acquired volumes had 128 sagittal 1.25 mm-thick slices without gaps and a pixel resolution of 256×256 (1×1 mm). Finally, brain morphometric evaluation was used to extract a set of 19,879 voxels (features) from the MRIs of the OASIS database.

The AD classification problem considered here was a binary classification problem with 19,879 features and 198 samples (100 AD patients and 98 normal people). The proposed iSBGA-ELM processed the presented OASIS data and built an efficient AD classifier.

Detailed explanation of the proposed iSBGA-ELM is presented below.

III. PROPOSED iSBGA-ELM APPROACH FOR AD DIAGNOSIS

The proposed iSBGA-ELM utilizes features extracted from MRIs of the OASIS database to create an accurate AD diagnosis approach. The proposed iSBGA-ELM contained two major steps: Voxels' selection and AD classification (see Fig. 1).

The iSBGA-ELM started by processing the MRIs (see Voxels' selection procedure in Fig. 1). Each MRI was converted into a set of 19,879 voxels using brain morphometric evaluation. The Voxels' selection procedure initiated a search for a reduced set of voxels based on the proposed iSBGA. Each reduced set of voxels was then used to build an AD classifier using the ELM.

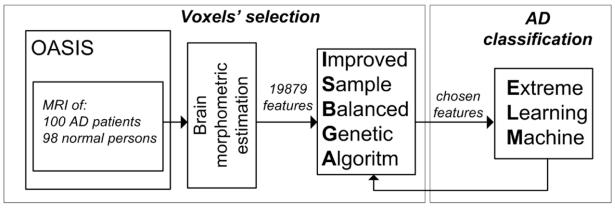

Fig. 1. Framework of the proposed Alzheimer disease (AD) diagnosis based on iSBGA-ELM.

In the proposed method, combining the iSBGA and the ELM created an efficient unified framework. The efficiency of the created ELM classifier was mostly dependent on the set of reduced voxels chosen by iSBGA. Recent AD research has focused on discovering sets of voxels (or biomarkers) which probably have some responsibility for AD. It was discovered that the subsets of voxels which created AD classifiers with accuracies close to the maximum were slightly different from the set of known biomarkers. Thus, slight modifications in the set of reduced voxels chosen by iSBGA may or may not improve the accuracy of AD classification. Sets of voxels can be modified to find improvements. They can be modified again and again until no further improvement could be reached. Such an iterative strategy is the basis of the genetic algorithm.

A. Improved Sample Balanced Genetic Algorithm

The proposed iSBGA is a modification of well-known genetic algorithm (GA) adapted for searching for the reduced set of voxels most suitable to create an AD classifier with maximum classification accuracy.

The GA is a famous optimization tool for solving complex optimization problems in many areas of science and engineering. GA exploits self-adaptive gene recombination mechanisms from nature. In GA, each optimization problem is specified by the set of chromosomes. Manipulations with chromosomes extracted from an optimization problem helps GA to search for an optimal solution. The iSBGA proposed here uses a string of binary coefficients as a set of meaningful chromosomes.

String of binary coefficients: A set of voxels from OASIS database has to be transformed into a set of relevant chromosomes. Each chromosome is a critical value for a given optimization problem. The set of extracted chromosomes builds a solution. A fitness function then uses those parameters to calculate a numerical measure or fitness value which is then used to evaluate each solution.

In the proposed AD diagnosis approach, each voxel

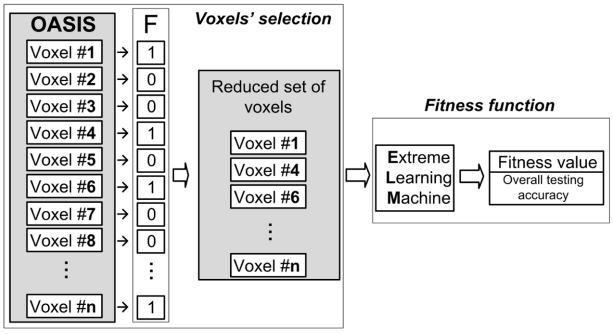

Fig. 2. Binary solution for Alzheimer disease (AD) diagnosis based on iSBGA-ELM (improved sample balanced genetic algorithm and Extreme Learning Machine).

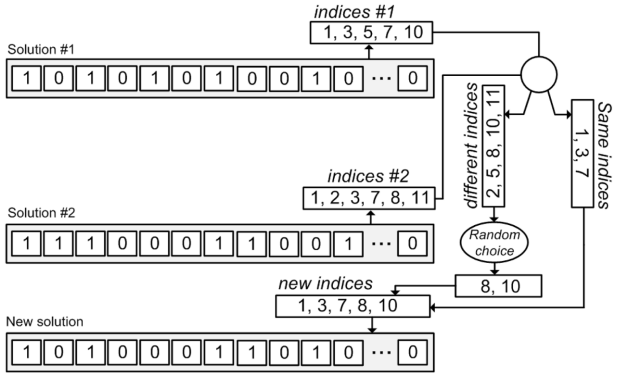

Fig. 3. Regular samples balanced crossover.

from the OASIS database represents a chromosome in the proposed GA framework. The OASIS database contains 19,879 voxels. Thus, each solution for the AD classification problem contains 19,879 chromosomes. The proposed iSBGA searches for a reduced set of voxels/chromosomes from the 19,879 voxels/chromosomes available. In this research, each chromosome determines the appearance status of a corresponding voxel. According to the chromosome value, each voxel can be either picked or not picked in the reduced set of voxels. Thus, the solution from the proposed iSBGA is a set of values: pick/NOT-pick, or "true"/"false", or in binary "1"/"0" (see Fig. 2). In the proposed method, binary "1" links to chosen voxels while binary "0" links to skipped voxels. Finally, the string of 19,879 binary coefficients (the binary solution) builds a reduced set of voxels for further analysis (see Fig. 2).

The proposed iSBGA creates new binary solutions by using three genetic operators: *regular sample balanced crossover*, *irregular sample balanced crossover*, and *sample balanced mutation*.

1) Genetic Operators

GA uses crossover and mutation to create new solutions (see Fig. 3). Genetic operators manipulate chromosomes of given optimization problem in a similar way to chromosome exchange mechanism from nature. Crossover creates a new genome by exchanging genetic materials from two input sources (genomes). The created genome may contain properties from both input sources. Such genetic recombination is chaotic and unpredictable in nature. Crossover achieves an exchange of properties between the inputs and outputs. Such properties may be either enhanced or degraded. Mutation modifies genes randomly and usually causes significant property degradation. Sometimes mutations can create new properties not existing in the input genomes.

Genetic operators in the GA process chromosome exchange between input sources similar to crossover and mutation found in nature. GA crossover recombines chromosomes from two randomly chosen solutions and builds a new solution. GA crossover always follows a fixed strategy to build new solutions.

The efficiency of the GA depends on problem specification, the efficiency of the chosen genetic operators, and GA settings. Problem specification deals with finding a proper way to transform a given optimization problem into the set of relevant chromosomes. A wrong choice of problem specification can lead to incorrect work and failure of the whole GA. The efficiency of the chosen genetic operators mostly depends on the given optimization problem and data. Different optimization problems may need different crossover and mutation methods or their combination. The GA settings affect the main procedures of GA. They may cause various problems. Incorrect settings can damage the convergence of the GA, create in balance between populations, and finally cause the GA to fail.

The proper choice of the crossover for GA is a big challenge. Wrong choice significantly damages the generalization ability of the GA. The concept of hybrid crossovers may solve such problem. The hybrid crossover provides a choice from several different crossovers in a pool. A pool may contain a few well-known crossovers or a few crossovers with novel design. It may even contain special crossovers. The hybrid crossover randomly picks one crossover from the pool and processes it to generate a new solution. Thus, the final set of solutions is obtained by using all crossovers listed in the pool. Hybrid crossovers are efficient if the optimization problem is relatively new and the proper crossover is difficult to choose. Sometimes a hybrid crossover is useful when a single crossover does not guarantee efficient search. Therefor, a combination of a few crossovers is needed.

Search for an optimal set of voxels for the AD classification problem using GA based on several existing crossovers including hybrid crossovers failed. Hence, a new crossover or set of crossovers designed specifically for the AD classification problem using the GA framework is needed.

In this paper, two new crossovers and one mutation designed specifically for GA focused on solving the AD classification problem are presented. *Regular sample balanced crossover*, *irregular sample balanced crossover*, and *sample balanced mutation* are presented in this paper. GA based on well-known crossovers failed due to problem specification based on the binary string with 19,879 coefficients for each given AD classification problem. Each binary solution contains 19,879 binary coefficients. However, the number of binary coefficients equal to "1" is always relatively small. Each binary solution contains on average 20-200 binary coefficients equal to "1", while the rest of the coefficients are "0". According to the problem specification, only binary coefficients "1" are meaningful in defining the efficiency of GA. All examined crossovers significantly modified the number of binary coefficients "1" in the new solution compared to the number of coefficients "1" in input solutions, which caused significant performance loss for AD classification. In the worst case scenario, crossovers might create solutions with all zeros, which means there are no chosen voxels to build a classifier. This case is unacceptable. Hence, the new *regular sample balanced crossover*, *irregular sample balanced crossover*, and *sample balanced mutation* are proposed to handle this challenge.

Regular samples balanced crossover manages the challenge by controlling a number of binary coefficients "1" in a new solution compared to the number of binary coefficients "1" in the input solutions. Regular samples balanced crossover is displayed in Fig. 3.

The proposed *regular samples balanced crossover* exchanges chromosomes from "solution #1" and "solution #2" to build a "new solution" (see Fig. 3). *Regular samples balanced crossover* keeps binary coefficients "1" located in both input solutions and randomly picks binary coefficients "1" different in both input solutions. The proposed *regular samples balanced crossover* collects locations of the binary coefficients "1" in both input solutions (see "indices #1" and "indices #2" in Fig. 3). Then the extracted indices are divided into the set of "different indices" and "same indices". "Same indices" are directly moved to the set of "new indices" which creates a "New solution". "Different indices" are divided randomly by "Random choice" according a random parameter s_r. Selected indices are then unified with "same indices" to build the "new solution". Random parameter s_r in "Random choice" is chosen in the range of 0.3–0.7. The random parameter is randomly generated every time GA calls for a crossover.

Fig. 3 presents an example of the proposed *regular samples balanced crossover*. In the given example, two input solutions ("solution #1" and "solution #2") are given. The random parameter s_r is equal to 0.6. Binary coefficients "1" are located in the positions of {1, 3, 5, 7, 10} and {1, 2, 7, 8, 10} of "solution #1" and "solution #2", respectively (see "index #1" and "index #2"). Then, extracted indices are divided into "same indices" {1, 3, 7} and "different indices" {2, 5, 8, 10, 11}. "Random choice" randomly picks $\lceil L \cdot s_r \rceil$ indices from "different indices", where L is the number of indices in the set "different indices". Then, $\lceil L \cdot s_r \rceil = \lceil 5 \cdot 0.6 \rceil = 3$. "Random choice" picks indices {2, 8, 10}. Then, "new indices" {1, 2, 7, 8, 10} unify the "same indices" {1, 3, 7} and the indices chosen by "Random choice" {2, 8, 10}. Finally, "new indices" create "new solution" (see Fig. 3).

The proposed *regular sample balanced crossover* keeps the balance of the binary coefficients "1" between input and output solutions and efficiently processes any binary solutions for the AD classification problem. If the number of common binary coefficients "1" from input solutions is relatively small, then the size of the set "same indices" is small or even 0 (when there are no same indices in the input solutions). The new solution is created mostly from "different indices" picked randomly. In this case, the new solution and input solutions have the maximum possible difference. Such solutions may either significantly degrade or improve the properties. If the number of common binary coefficients from input solutions is significant, then most of the indices from both input solutions belong to the "same indices". In this case, new solution is mostly created from the "same indices". Thus, the proposed *regular samples balanced crossover* balances GA in various scenarios. In the beginning, GA manages solutions generated randomly and input solutions have a small number of binary "1" in common. New solutions are then created from "different indices", similar to a chaotic search. Near the end, GA should have already identified a set of common locations of binary "1". Thus, most of the indices from input solutions will belong to the "same indices" set. New solutions are only slightly modified compared to input solutions, similar to a search for an optimal solution.

Irregular samples balanced crossover is a modified version of the *regular sample balanced crossover* presented above. *Irregular samples balanced crossover* cre-

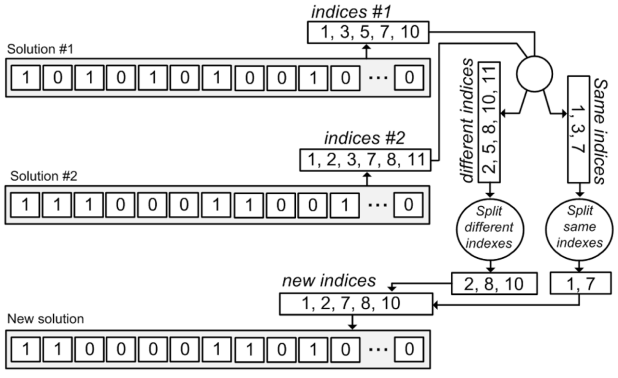

Fig. 4. Irregular samples balanced crossover.

ates new solution from randomly picked "different indices" and "same indices" (see Fig. 4). The split parameter for "Split different indices" is s_{dif} in the range of 0.3–0.7. The split parameter "Split same indices" is s_{same} in the range of 0.8–1.0. Split parameters s_{dif} and s_{same} are randomly generated every time when GA calls for a crossover.

In the example presented in Fig. 4, s_{dif} and s_{same} are 0.6 and 0.8, respectively. The *irregular samples balanced crossover* presented in Fig. 4 randomly picks $\lceil L_{dif} \cdot s_{dif} \rceil = \lceil 5 \cdot 0.6 \rceil = 3$ from "different indices" and $\lceil L_{same} \cdot s_{same} \rceil = \lceil 3 \cdot 0.8 \rceil = 2$ from "same indices". Thus, "new index" {1, 2, 7, 8, 10} unifies 3 indices from "different indices" {2, 8, 10} and 2 indices from "same indices" {1, 7}.

Irregular samples balanced crossover is needed when GA has almost finished the search. A significant portion of the generated solutions contain the same indices. In a case when "solution #1" and "solution #2" are the same, *regular sample balanced crossover* generates new solution exactly the same as "solution #1" and "solution #2". *Irregular sample balanced crossover* does not have this drawback. In the previous scenario, *irregular sample balanced crossover* generates new solution different from "solution #1" and "solution #2".

In this research, the concept of hybrid crossover is implemented. The hybrid crossover randomly picks either *regular sample balanced crossover* or *irregular sample balanced crossover. Regular sample balanced crossover* is efficient except when "solution #1" and "solution #2" are the same. *Irregular sample balanced crossover* efficiently handles cases when "solution #1" and "solution #2" are the same. It does not degrade GA performance significantly otherwise.

The proposed *samples balanced mutation* randomly modifies binary coefficients from input solution so that the number of binary "1" in new solution and the input solution stays the same.

Fitness function is a special procedure to evaluate solutions in GA. The fitness function calculates a fitness value which gives a numerical estimate of the importance of each solution in the GA framework. Solutions with maximum or minimum fitness value (depending on what kind of optimal solution is the target for the given optimization problem) are optimal or suboptimal solutions for the given problem. Fitness values are used to rank the examined solutions in the GA framework. In GA, all important solutions with fitness values closer to the maxima (or minima) are used to generate slightly different solutions with better fitness values. Lower ranked solutions are discarded.

In the proposed iSBGA-ELM framework, the fitness function is an AD classifier based on the ELM. The AD classifier is trained using a reduced set of voxels specified by the current string of binary coefficients or solution. The given set of voxels is used to train 10 ELM classifiers using random parameters. The average testing accuracy of the 10 created ELM classifiers is a fitness value.

The selection procedure sorts given solutions according to fitness values and selects solutions with promising fitness values for further processing in GA framework. The selection procedure assigns a selection probability to each solution according to its fitness value. Thus, a solution with a higher rank has a better chance to survive and generate another solution in GA. Solutions with lower ranks are mostly ignored.

In the proposed iSBGA-ELM, the geometric ranking method [19] is used as a selection procedure. The method sorts all given solutions in descending order according to fitness value and assigns a probability to each solution P_j as follows:

$$P_j = q'(1-q)^{r_j - 1} \tag{1}$$

where

$$q' = \frac{q}{1-(1-q)^N}$$

q' is the selection parameter, r_j is a rank of j-th solution in the partially ordered set, and N is the population size. The detailed explanation of the geometric ranking method is given previously [18]. In this research, the parameter $q = 10^{-3}$ is chosen.

Termination criteria: The GA stops when no better result can be produced during the last 50 generations.

iSBGA-ELM framework: iSBGA-ELM processes the initialization step and several generations (see Fig. 5). Each generation contains selection procedure, a set of genetic operators, and a fitness function. The proposed iSBGA starts from the initialization step. A total of 200 binary solutions are generated randomly. Each solution is processed by the fitness function and its fitness value is then calculated. Each solution is a set of 19,879 binary coefficients. The number of binary "1" in each initial solution is limited to the range of 20–200. Each initial solution then builds a reduced set of voxels from the OASIS database (see Fig. 2). The reduced set of voxels is used to train 10 AD classifiers based on the ELM. The created ELM classifiers are fitness functions and the average overall testing accuracy is the fitness value. Thus, a combination of 200 initial binary solutions $\{F_1^0, F_2^0, F_3^0, ..., F_{200}^0\}$ and corresponding fitness values $\{f_1^0, f_2^0, f_3^0, ..., f_{200}^0\}$ build the initial population of the GA,

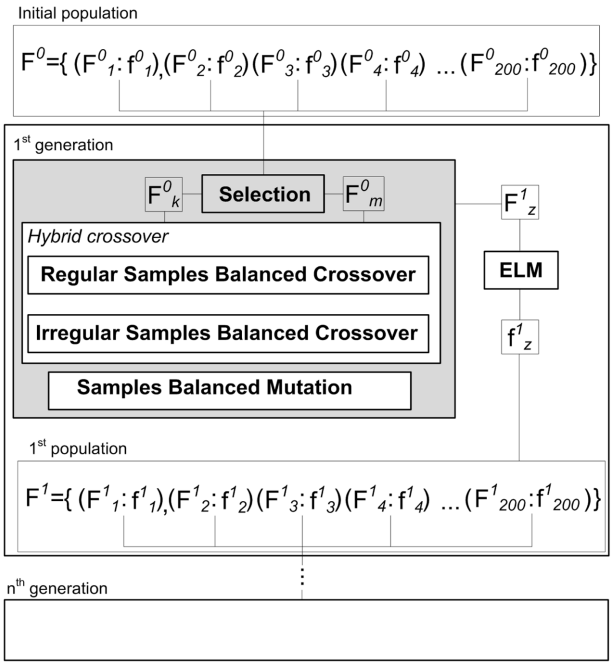

Fig. 5. Framework of the proposed iSBGA-ELM (improved sample balanced genetic algorithm and Extreme Learning Machine).

F^0 (see Fig. 5). Each *n*-th generation in iSBGA starts from the selection procedure based on the geometric ranking method. The selection procedure picks solutions from the previous population F^{n-1} according to the assigned probabilities. Pairs of selected binary solutions are processed by genetic operators. The hybrid crossover based on *regular sample balanced crossover* and *irregular sample balanced crossover* together with the *sample balanced mutation* then creates a set of new binary solutions, $\{F^n_1, F^n_2, F^n_3, ..., F^n_{200}\}$. Similar to the initialization step, each binary solution F^n_i is evaluated by a fitness function and its fitness value f^n_i is then calculated. The combination of all binary solutions $\{F^n_1, F^n_2, F^n_3, ..., F^n_{200}\}$ and corresponding fitness values $\{f^n_1, f^n_2, f^n_3, ..., f^n_{200}\}$ builds *n*-th population F^n. In each generation, crossover creates 70% or 140 new solutions, and mutation creates the remaining 30% or 60 new solutions. GA processes generations until new generations no longer produce better results during the last 50 generations.

B. Extreme Learning Machine

The ELM is a machine learning technique with extremely fast learning phase and good generalization performance. Technically, ELM is a single hidden layer feed-forward neural network where input weights and bias of the hidden neurons are randomly assigned while output weights are estimated analytically [19].

In the proposed AD diagnosis approach, an ELM is used to solve the AD classification problem. The AD classification problem is a binary classification problem with 198 samples (100 AD patients and 98 normal persons) and 19,879-dimensional feature space. In the proposed AD diagnosis approach, iSBGA significantly reduces the feature space and simplifies the AD classifiers based on the ELM.

The framework to build an ELM classifier controlled by Gaussian hidden neurons for the AD classification problem is presented below.

Data: Training and testing sets are created using given binary solutions and the OASIS database. First, 198 samples (100 AD, 98 normal) from OASIS are randomly divided into a training set (70 AD, 68 normal) and testing set (30 AD, 30 normal). The 19,879 features from the OASIS database are reduced by iSBGA to a set of *m* features. Then, the training set contains 138 samples, i.e., $\{(X^1_{tra}, c^1_{tra}), ..., (X^t_{tra}, c^t_{tra}), ..., (X^{138}_{tra}, c^{138}_{tra})\}$; testing set contains 60 samples, i.e., $\{(X^1_{test}, c^1_{test}), ..., (X^t_{test}, c^t_{test}), ..., (X^{60}_{test}, c^{60}_{test})\}$, where X_{test} and X_{tra} are *m*-dimensional feature vector and $c^t \in \{1, 2\}$ or {AD patient, normal person} in a class label. The coded class label y^t is calculated as follows:

$$y^t = \begin{cases} 1, & \text{if AD patient} \\ -1, & \text{if normal person} \end{cases} \quad (2)$$

Therefore, $y_{tra} = \{y^1_{tra}, y^2_{tra}, ..., y^{138}_{tra}\}$ and $y_{test} = \{y^1_{test}, y^2_{test}, ..., y^{60}_{test}\}$. They are vectors with coded class labels for testing and training, respectively.

The framework of the ELM is summarized as follows:
Training phase:
1) Assign the number of hidden neurons L.
2) Generate sets of input weights $A_{m\times L}$ and width (bias) $b_{L\times 1}$ of hidden Gaussian neurons randomly.
3) Compute the hidden layer output matrix G:

$$G = \begin{pmatrix} g^1_1 & \cdots & g^N_1 \\ \vdots & \ddots & \vdots \\ g^1_L & \cdots & g^N_L \end{pmatrix} \quad (3)$$

where N is a number of samples (N = 138 for training and N = 60 for testing), g^t_j is a response of *j*-th hidden neuron for *t*-th sample calculated as follows:

$$g^t_j = \exp\left(-\left(\frac{(x^t_{tra} - A_j)^T \cdot (x^t_{tra} - A_j)}{2 \cdot b^2_j}\right)\right) \quad (4)$$

4) Compute output weights β:

$$\beta = y_{tra} \cdot G^{\dagger} \quad (5)$$

where † is a Moore-Penrose generalized inverse.
5) Compute predicted coded class labels $\hat{y}_{tra}$:

$$\hat{y}_{tra} = \beta \cdot G \quad (6)$$

6) Define the predicted class label as follows:

$$\hat{c}_{tra}^{k} = \begin{cases} 1, & \text{if } \hat{y}_{tra}^{k} \geq 0 \\ 2, & \text{if } \hat{y}_{tra}^{k} < 0 \end{cases}$$

Testing phase:

1) Compute the hidden layer output matrix G for testing (N = 60) using Eq. (3), where

$$g_j^t = \exp\left(-\left(\frac{(x_{test}^t - A_j)^T \cdot (x_{test}^t - A_j)}{2 \cdot b_j^2}\right)\right)$$

2) Compute predicted coded class labels $\hat{y}_{test}$:

$$\hat{y}_{test} = \beta \cdot G$$

3) Define predicted class label for testing as follows:

$$\hat{c}_{test}^{k} = \begin{cases} 1, & \text{if } \hat{y}_{test}^{k} \geq 0 \\ 2, & \text{if } \hat{y}_{test}^{k} < 0 \end{cases}$$

Accuracy of any ELM classifier mostly depends on the randomly chosen weights and hidden neuron bias. In the proposed iSBGA-ELM, each binary solution is used to create 10 ELM classifiers with randomly generated weights and bias. The fitness value is an average of the 10 overall testing accuracies of the ELM classifiers. This 10-fold validation balances the generalization performance of the ELM but neglects the effects of randomness.

IV. EXPERIMENTAL RESULTS

Experiments with the proposed iSBGA-ELM included training the AD classifier using the ELM on a set of 138 samples (70 AD patients, 68 normal persons) and testing using 60 samples (30 Ad patients, 30 normal persons). Experiments also included searching for a reduced set of voxels from the OASIS database with maximum fitness value, i.e., maximum overall testing accuracy of the created ELM classifier.

The proposed iSBGA-ELM processed 122 generations until the termination criteria were satisfied, i.e., no improvement since the 72th generation. The iSBGA-ELM found a set of 38 voxels which were used to create the ELM classifier for AD diagnosis. The best ELM classifier from the set of 10 created ELM classifiers from the 10-fold validation showed 93% overall training accuracy, while the average showed 87% of overall testing accuracy.

A. Comparison with Existing Methods

Comparison with existing techniques should cover completely automatic AD diagnosis approachs based on machine learning techniques and the OASIS database. Direct comparison is possible if the examined methods also use a complete set of samples (198 patients) from the OASIS database and the complete set of 19,879 voxels. Unfortunately, researchers mostly used a manually reduced set of voxels and samples. Thus, direct comparison was impossible.

Kloppel et al. [6] have used 4 sets of patients with AD and healthy controls for cross database analysis. Each set could unify MRI that was created using one scanner with fixed settings. They used the SVM to create an AD classifier and found that the classification accuracy was about 81%–96% udner different scenarios.

Davatzikos et al. [7] have collected MRI from 30 AD patients and 20 normal persons, extracted morphometric features, and trained an AD classifier based on the SVM classifier. They reported a classification accuracy of about 90%.

Mahanand et al. [14] have used a set of MRI from 30 AD patients and 30 normal from OASIS database and extracted a set of 5,788 features using the VBM approach. A reduced set of 20 features was used to build an AD classifier using a SRAN. They reported that the overall testing accuracy was about 91%.

As mentioned earlier, direct comparison was impossible for all examined existing techniques. Besides, the iSBGA-ELM proposed here used a complete set of samples and voxels for accurate AD diagnosis. Experiments with a complete set of samples and voxels are more challenging. It increases the complexity of the classifiers and reduces the accuracy. Thus, the proposed AD diagnosis approach based on iSBGA-ELM was used to solve the more complex classification problem. It showed acceptable classification accuracy.

V. CONCLUSION

In this paper, an iSBGA-ELM was used to create an efficient AD diagnosis approach. Hybrid crossover based on the proposed *regular sample balanced crossover* and *irregular sample balanced crossover* maintained the convergence of the proposed iSBGA-ELM and obtained a reduced set of voxels with the highest fitness value.

Our experimental results clearly indicated that the proposed approach based on iSBGA-ELM was efficient for AD diagnosis. The proposed method showed an overall testing accuracy of 87% in tests of all samples available in the OASIS database.

AKNOWLEDGMENTS

This work was supported by Catholic University of Korea research funds 2016.

REFERENCES

1. P. M. Thompson, K. M. Hayashi, R. A. Dutton, M. C. Chiang, A. D. Leow, E. R. Sowell, G. De Zubicaray, J. T. Becker, O. L. Lopez, H. J. Aizenstein, et al., "Tracking Alzheimer's disease. *Annals of the New York Academy of Sciences*, vol. 1097, no. 1, pp. 183-214, 2007.
2. O. L. Lopez, J. T. Becker, C. A. Jungreis, D. Rezek, C. Estol, F. Boiler, and S. T. DeKosky, "Computed tomography—but not magnetic resonance imaging—identified periventricular white-matter lesions predict symptomatic cerebrovascular disease in probable Alzheimer's disease," *Archives of Neurology*, vol. 52, no. 7, pp. 659-664, 1995.
3. K. A. Jobst, L. P. Barnetson, and B. J. Shepstone, "Accurate prediction of histologically confirmed Alzheimer's disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and DSM-III-R criteria, SPECT, X-ray CT, and Apo E4 in medial temporal lobe dementias," *International Psychogeriatrics*, vol. 10, no. 3, pp. 271-302, 1998.
4. J. Ramírez, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, and C. G. Puntonet, "Early detection of the alzheimer disease combining feature selection and kernel machines," In *International Conference on Neural Information Processing*, Heidelberg: Springer, 2008, pp. 410-417.
5. M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, and M. Gomez-Rio, "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease," *Neurocomputing*, vol. 74, no. 8, pp. 1260-1271, 2011.
6. S. Kloppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681-689, 2008.
7. C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging," *Neurobiology of Aging*, vol. 29, no. 4, pp. 514-523, 2008.
8. C. R. Jack, R. C. Petersen, Y. C. Xu, P. C. O'Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen, "Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397-1397, 1999.
9. R. J. Killiany, B. T. Hyman, T. A. Gomez-Isla, M. B. Moss, R. Kikinis, F. Jolesz, R. Tanzi, K. Jones, and M. S. Albert, "MRI measures of entorhinal cortex vs hippocampus in preclinical AD," *Neurology*, vol. 58, no. 8, pp. 1188-1196, 2002.
10. G. B. Frisoni, M. P. Laakso, A. Beltramello, C. Geroldi, A. Bianchetti, H. Soininen, and M. Trabucchi, "Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease," *Neurology*, vol. 52, no. 1, pp. 91-91, 1999.
11. Z. Lao, D. Shen, Z. Xue, B. Karacali, S.M. Resnick and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *NeuroImage*, vol. 21, no. 1, pp. 46-57, 2004.
12. M. Chupin, E. Gerardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehericy, H. Benali, L. Garnero, and O. Colliot, "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus*, vol. 19, no. 6, pp. 579-587, 2009.
13. J. Ashburner and K. J. Friston, "Voxel-based morphometry: the methods," Neuroimage, vol. 11, no. 6, pp. 805-821, 2000.
14. B. S. Mahanand, S. Suresh, N. Sundararajan, and M. A. Kumar, "Alzheimer's disease detection using a self-adaptive resource allocation network classifier," in *Proceedings of the 2011 International Joint Conference on Neural Networks (IJCNN),* San Jose, CA, 2011, pp. 1930-1934.
15. D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498-1507, 2007.
16. S. Saraswathi, S. Sundaram, N. Sundararajan, M. Zimmermann, and M. Nilsen-Hamilton, "ICGA-PSO-ELM approach for accurate multiclass cancer classification resulting in reduced gene sets in which genes encoding secreted proteins are highly represented," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 2, 452-463, 2011.
17. V. Sachnev, S. Saraswathi, R. Niaz, A. Kloczkowski, and S. Suresh, "Multi-class BCGA-ELM based classifier that identifies biomarkers associated with hallmarks of cancer," *BMC Bioinformatics*, vol. 16, article no. 166, 2015.
18. S. Suresh, S. N. Omkar, V. Mani, and T. G. Prakash, "Lift coefficient prediction at high angle of attack using recurrent neural network," *Aerospace Science and Technology*, vol. 7, no. 8, pp. 595-602, 2003.
19. G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489-501, 2006.

Vasily Sachnev

Vasily Sachnev received his B.S. and M.S. degrees in Electrical Engineering from the Komsomolsk-na-Amure State Technical University, Russia, in 2002 and 2004, respectively. He received Ph.D. degree in Multimedia Security at the Center of Information Security and Technology (CIST), Graduate School of Information Management and Security, Korea University, Seoul, Korea in 2009. He joined the department of Information, Communication and Electronics Engineering of the Catholic University of Korea in 2010, where he is currently working as an assistant professor. His research interests include multimedia security, steganography, steganalysis, machine learning and bioinformatics.

Suresh Sundaram

Suresh Sundaram received his B.E. degree in electrical and electronics engineering from Bharathiyar University in 1999, and M.E. (2001) and Ph.D. (2005) degrees in Aerospace Engineering from Indian Institute of Science Bangalore, India. He was post-doctoral researcher in School of Electrical Engineering, Nanyang Technological University, Singapore, from 2005–2007. Subsequently, he was selected as an ERCIM research fellow for the period of 2007–2008 and he spent valuable time in the project team PULSAR at INRIA Sophia-Antipolis, France. For a short period, he was working as Faculty at Industrial Engineering, Korea University, Seoul. Later, he was with Indian Institute of Technology at Delhi, as an Assistant Professor in Electrical Engineering from 2008–2009. From 2010, he is working as Associate Professor in School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include computational cognitive system, neural networks, intelligent control, medical image processing, mathematical optimization, and game theory.