

Classification of Breast Cancer using Differential Evolution and Least Squares Support Vector Machine

¹Omar S.Soliman, ²Eman AboElHamd

^{1,2}Faculty of Computers and Information, Cairo University, Egypt

ABSTRACT: *Breast Cancer is becoming a leading cause of death among women in the whole world; meanwhile, it is confirmed that the early detection and accurate diagnosis of this disease can ensure a long survival of the patients. This paper proposes a hybrid classification algorithm using Differential Evolution (DE) and Least Squares Support Vector Machine (LS-SVM). LS-SVM technique is used for classification. Since LS-SVM classifier is so sensitive to the changes of its parameter values, DE algorithm is used as an optimization technique for LS-SVM parameters. This will guarantee the effectiveness of the hybrid algorithm by searching for the optimal values of the classifier. The proposed algorithm is implemented and evaluated on Wisconsin Breast Cancer Dataset (WBCD) from UCI Repository of Machine Learning Databases. The proposed algorithm was compared with different classifier algorithms which were applied on the same database. The experimental results showed the superiority of the proposed algorithm that could achieve a classification accuracy of 99.75%.*

Keywords: Breast Cancer (BC), Differential Evolution (DE), Particle Swarm Optimization (PSO), Least Squares Support Vector Machine (LS-SVM).

1. INTRODUCTION:

Breast Cancer (BC) is a type of cancer originating from breast tissue. BC tumors are classified as non-invasive or invasive. The most common type of BC is ductal carcinoma, which begins in the lining of the milk ducts. Another type of BC is lobular carcinoma, which begins in the lobules of the breast. Invasive breast cancer is breast cancer that spreads from where it began in the breast ducts or lobules to surrounding normal tissue. BC occurs in both men and women, although male BC is rare. Tumors are classified as being either benign or malignant, depending on whether or not they can spread by invasion and metastasis. Benign tumors are tumors that can't spread by invasion or metastasis; hence, they only grow locally. Malignant tumors are tumors that are capable of spreading by invasion and metastasis. By definition, the term "Cancer" applies only to malignant tumors.

According to The International Agency for Research on Cancer (IARC) [21], global cancer burden rises to 14.1 million new cases and 8.2 million cancer deaths in 2012. The most commonly diagnosed cancers worldwide were those of the lung (1.8 million, 13.0% of the total) then breast (1.7 million, 11.9%). According to the American Cancer Society's estimates for BC in the United States for 2014, the chance that breast cancer will be responsible for a woman's death is about 3%. About 12% women in the

US will develop invasive BC during their lifetime. The chance of a woman having invasive BC during her life is about 1 in 8. The chance of dying from BC is about 1 in 36. BC death rates have been going down. This is probably the result of finding the cancer earlier and better treatment [24].

The aim of this paper is to help physicians in the early diagnosis for BC Patients. As it is sometimes found after symptoms appear, but many women with early BC have no symptoms. This is why getting the recommended screening tests before any symptoms develop is so important. Screening refers to tests and exams used to find a disease, like cancer, in people who do not have any symptoms. The goal of screening exams, such as mammograms, is to find cancers before they start to cause symptoms. BC that are found because they can be felt tend to be larger and are more likely to have already spread beyond the breast. In contrast, Breast Cancers found during screening exams are more likely to be small and still confined to the breast. The size of a BC and how far it has spread are important factors in predicting the prognosis for a woman with this disease. Most doctors feel that early detection tests for breast cancer save many thousands of lives each year, and that many more lives could be saved if even more women and their health care providers took advantage of these tests [1]. The rest of the paper is organized as follows; section 2 mentions the problem background and related work. The proposed algorithm is introduced in section 3. Experimental results and discussion are presented in section 4. The last section is devoted to conclusions and further research.

2. BACKGROUND AND RELATED WORK

Breast Cancer is a worldwide problem because of the severity of the disease. Diagnosis and treatment of the disease become a challenging research problem for many researchers. In [1] an Artificial Neural Network (ANN) approach based on the pareto-differential evolution (PDE) algorithm augmented with local search for the prediction of BC was presented. ANN algorithm was also used to increase the accuracy and objectivity of BC diagnosis in [16]. A comparison between three algorithms (Evolutionary, Hierarchical and k-medoids) was proposed in [2]. BC diagnosis based on a SVM method combined with feature selection had been proposed in [3]. A study in [7] compared 3 data mining methods (Decision Tree (DT) C5.0, SVM and Logistic Regression) which attempted to identify the determining factors of BC. The

results showed that DT outperformed both SVM and logistic regression. A hybrid technique using (PSO-SVM) was proposed for BC diagnosis [8]. In [11] a bias-corrected $100(1 - \sigma)\%$ simultaneous confidence bands for LS-SVM classifiers based on a regression framework was presented. In [10] SVM was trained using linear, polynomial and radial basis function (RBF) kernels and applying PSO to each kernels for different datasets to get better accuracy for BC, Lung cancer and Heart diseases. In [18] a model that did a simultaneous optimization for SVM kernel parameters and feature subset without degrading the SVM classification accuracy was introduced. In [12] a study compared PSO based ANN, Adaptive Neuro Fuzzy Inference System (ANFIS) and a case-based reasoning (CBR) classifier with a logistic regression model and DT model.

In [14] a classification model was proposed to predict ALN status by means of selecting BC patients' basic clinical and histological features that can be obtained in each hospital. Their model composed of back propagation Multi-Layer Perceptron (MLP), Logistic Regression (LR) and Genetic Algorithm (GA) based MLP models. A decisions support system for early BC detection that employed a Correlation Feature Selection procedure and an Artificial Immune Recognition System (AIRS) was presented in [15]. In [6] MLP was applied for BC patients' classification. In [31] BC diagnosis was conducted using LS-SVM classifier algorithm. A Hybrid technique using SVM based classifier along with Novel PSO and Binary PSO was presented in [18]. A hybrid PSO and Tabu Search (HPSOTS) approach for gene selection for tumor classification was developed in [9]. An algorithm for rule based reasoning which was developed for mammographic findings to provide support for the clinical decision to perform biopsy of the breast [29]. In [25] a study that was built to investigate the use of fuzzy methodologies to create an easy to interpret set of classification rules, capable of placing the large majority of patients into one of the specified groups. PSO, Quantum-behave Particle Swarm for training SVM as well as LSSVM and active set strategy were introduced [27]. In [30] the asymptotic behaviors of SVM are fused with genetic algorithm (GA) and the feature chromosomes were generated. In [31] diagnose BC based on the extracted tumor features. Feature extraction and selection are critical to the quality of classifiers founded through a hybrid of K means and SVM algorithms.

2.1 Differential Evolution

Differential Evolution (DE) is a stochastic, population based, real-valued algorithm. It was introduced by Storn and Price in 1996. DE is an efficient, effective and powerful population-based stochastic search technique for solving optimization problems over continuous space, which has been widely applied in diverse fields especially in scientific and engineering ones [26]. DE has a lot of advantages such that its conceptual simplicity and ease of use, finding the true global minimum regardless of the initial parameter values, fast convergence, and using few

control parameters besides its sustainability for parallelization [13]. The Main ad-vantage of DE is that it internally works on continuous space and only switches to the integer during the evaluation of the objective function. This characteristic supports higher accuracy compared to some other well-known algorithms. While DE shares similarities with other Evolutionary Algorithms (EA), it differs significantly in the sense that distance and direction information from the current population is used to guide the search process. In DE, a target vector is mutated using a difference vector (obtained as a weighted difference between randomly selected individuals) to obtain a trial vector. This trial vector is then stochastically recombined with the parent to produce an off-spring, which is only accepted if the fitness of the offspring is better than that of the parent. There are plenty of DE strategies that differ in the way that the target vector is selected, the number of difference vectors used, and the way in which recombination is applied. DE algorithm composed of 4 main phases, Initialization, Mutation, Crossover and finally selection of the coming generation [5].

A. Initialization Phase

DE algorithm aims at evolving a population of size N with D -dimensional parameter vector, called (Individuals). These individuals encode the candidate solution. i.e. $X_{i,G} = \{x_{i,G}^1, \dots, x_{i,G}^D\}$, $i = 1, \dots, N$ towards the global optimum, where i index denotes the population in the generation G . The initial population should better cover the entire search space as much as possible by uniformly randomizing individuals within the range $[0, 1]$.

B. Mutation Operation

After initialization, DE employs the mutation operation to produce a mutant vector $V_{i,G}$ for each individual $X_{i,G}$ (Target Vector) in the current population. For each target vector $X_{i,G}$ at the generation G , its associated mutant vector $V_{i,G} = \{v_{i,G}^1, v_{i,G}^2, \dots, v_{i,G}^D\}$ can be generated via certain mutation strategy.

There are lots of different mutation strategies. In the proposed algorithm [DE/best/1] mutation strategy was used, as in (1).

$$V_{i,G} = X_{best,G} + F(X_{r_1,G} - X_{r_2,G}) \quad (1)$$

Where r_1, r_2 are mutually exclusive integers and normally generated within the range $[1, N]$. These indices are randomly generated once for each mutant vector and they are different from the index i . F is the scaling factor which is a positive control parameter for scaling the difference vector. $X_{best,G}$ is the best individual vector with the best fitness value in the population at generation G .

C. Crossover Operation

This operation is a posterior step after the mutation phase. It is applied to each pair of the target vector $X_{i,G}$ and its corresponding mutation vector $V_{i,G}$ to generate a trial vector $U_{i,G} = \{U_{i,G}^1, U_{i,G}^2, \dots, U_{i,G}^D\}$. Basically, DE employs the binomial (uniform) crossover defined in (2):

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{if } (\text{rand}[0,1] \leq CR) \text{ or } (j = j_{rand}) \\ x_{i,j}, & \text{otherwise} \end{cases}, j = 1, 2, \dots, D \quad (2)$$

In (2) CR is a user defined constant within the range [0,1], which controls the fraction of parameter values copied from the mutant vector. j_{rand} is a randomly chosen integer in the range [1, D]. The binomial crossover operator copies the j th parameter of the mutant vector $V_{i,G}$ to the corresponding element in the trial vector $U_{i,G}$ if $\text{rand}_j[0,1] \leq CR$ or $j = j_{rand}$. Otherwise, it is copied from the corresponding target vector $X_{i,G}$. The remaining parameters of the trial vector $U_{i,G}$ are copied from the corresponding target vector $X_{i,G}$. The condition $j = j_{rand}$ is introduced to ensure that the trial vector $U_{i,G}$ will differ from its corresponding target vector $X_{i,G}$ by at least one parameter.

D. Selection Operation

If the values of some parameters of a newly generated trial vector exceed the corresponding upper and lower bounds, they are randomly and uniformly reinitialized within the pre-specified range. The objective function values of all trial vectors are evaluated; then, a selection operation is performed. The objective function value of each trial vector $f(U_{i,G})$ is compared to that of its corresponding target vector $f(X_{i,G})$ in the current population. If the trial vector has less or equal objective function value than the corresponding target vector, the trial vector will replace the target vector and enter the population of the next generation. Otherwise, the target vector will remain in the population for the next generation. The selection operation can be expressed in (3)

$$X_{i,G+1} = \begin{cases} U_{i,G}, & \text{if } f(U_{i,G}) \leq f(X_{i,G}) \\ X_{i,G}, & \text{otherwise} \end{cases}, j = 1, 2, \dots, D \quad (3)$$

DE main steps are stated in Algorithm-1 [5]

Algorithm-1: Differential Evolution

- Step 1:** Randomly generate a population of N vectors, each of D dimensions.
- Step 2:** Calculate the objective function value $f(X_i)$ for all target vectors X_i
- Step 3:** Select 3 points from the population and generate mutant individual V_i using (1)
- Step 4:** Apply Crossover operation on each target vector x_i with mutant individual (generated in step 3) to generate a trail vector U_i using (2)
- Step 5:** Calculate the objective function value for vector U_i
- Step 6:** Choose better of the two (function value at target and trail point) using (3)
- Step 7:** Check whether a convergence criterion is met, if yes then stop; otherwise go to step 3.

2.2 Least Squares Support Vector Machine

LS-SVM classifiers are one particular sample of Support Vector Machine (SVM). LS-SVM is used for finding a hyper plane, which separates various classes. It obtains this optimal hyper-plane by using maximum Euclidean

distance to the nearest point. The LS-SVM classifier maps the input vectors into a high dimensional feature space for non-separable data. Then, the LS-SVM classifier finds an optimal separating hyper-plane in this higher dimensional space [17].

Given a training dataset of N points $\{x_k, y_k\}_{k=1}^N$ with input data $x_k \in \mathbb{R}^n$ and output $y_k \in \mathbb{R}$, we consider the following optimization problem in primal weight space:

$$\min_{w,b,e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \quad (4)$$

Such that

$$y_k - (w^T \phi(x_k) + b) = e_k, k = 1, 2, \dots, N \quad (5)$$

Where γ is a regularization factor, e_k the difference between the desired output y_k and the actual output, and $\phi(\cdot)$ is a nonlinear function mapping the data points into a high dimensional Hilbert space; in addition, the dot product in the high-dimensional space is equivalent to a positive definite kernel function $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$.

In primal weight space, a linear classifier in the new space is calculated using (6); where w is the weight vector and $b \in \mathbb{R}$ which called as the bias term.

$$y(x) = \text{sgn}(w \cdot \phi(x) + b) \quad (6)$$

The dual space of this primal space was found by solving the Lagrangian function in (7)

$$L(w, e, \alpha) = J(w, e) - \sum_{k=1}^N \alpha_k (w^T \phi(x_k) + e_k - y_k) \quad (7)$$

Where α_k Lagrangian multipliers and are called Support Vectors. The optimal solution for objective function using (7) must satisfy the following Karush-Kuhn Tucker (KKT) conditions [23].

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N \alpha_k y_k \phi(x_k) \quad (8)$$

$$\frac{\partial L}{\partial e} = 0 \rightarrow \alpha_k = \gamma e_k, k = 1, \dots, N$$

$$\frac{\partial L}{\partial w} = 0 \rightarrow w^T \phi(x_k) + e_k - y_k = 0, k = 1, \dots, N$$

The linear system in, (9) will results after elimination of w and e which generates the Support Vector α'_k

$$\left(K + \frac{1}{\gamma} \right) \alpha = y \quad (9)$$

Where $y = [y_1, y_2, \dots, y_N]^T$, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ and $K \in \mathbb{R}^{N \times N}$ is the kernel matrix. The resulting LS-SVM model for function estimation is as in (10), where $K(\dots)$ is the kernel function.

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) \quad (10)$$

LS-SVM (Algorithm-2) was implemented using Ra-dial Basis Function (RBF) in (11).

$$K(x, x_k) = \exp\left(-\frac{\|x-x_k\|^2}{\sigma^2}\right) \quad (11)$$

LS-SVM main steps are stated in Algorithm-2

Algorithm-2: LS-SVM:

- Step 1:** Load the training data set of n data points, $\{x_i, y_i\}_{i=1}^n$ where x_i is the i^{th} input vector and $y_i \in \mathbb{R}$ is the corresponding i^{th} target with values $\{2, 4\}$.
- Step 2:** Generate random weights for each input data point.
- Step 3:** Determine the value of the bias term b and initialize the error e for each point randomly.

- Step 4:** Initialize γ and σ using random values.
- Step 5:** Search for values of e , w and b that minimize the objective function (4) and (5).
- Step 6:** Construct the Lagrangian function in (7) with the solution that must satisfy the KKT conditions in the set of equations in (8)
- Step 7:** Calculate number of support vectors (σ^*) using (9)
- Step 8:** Training data for LS-SVM model could be classified using (10) with RBD kernel function in (11)
- Step 9:** Classify any new point in (6) using RBF kernel function in (11)
- Step 10:** Loop until stopping criteria is met, usually until reach the maximum number of iterations.

2.3 Overview on Wisconsin Breast Cancer Database (WBCD)

Cancer is one of the major causes of mortality all over the world. Research into cancer diagnosis and treatment has become an important issue for the scientific community. Breast Cancer is the 2nd severe type of cancer especially between women. BC is a malignant tumor that is developed from breast cells. There are some risk factors that increase the chance of developing breast cancer in women.

The proposed Algorithm worked on the Wisconsin Breast Cancer Dataset (WBCD) taken from UCI repository of machine learning databases. The dataset contains 699 instances taken from needle aspirates from patients breasts, of which 458 cases belong to benign class and the remaining 241 cases belong to malignant class. It should be noted that there are 16 instances which have missing values, in this study all the missing values are replaced by the mean of its surrounding values. Each record in the database has 9 attributes in addition to the class attribute which was represented as 2 for benign and 4 for malignant cases. The 9 attributes listed in Table-1 and are graded (1-10) with their mean and standard deviation values.

Table-1: WBCD description of attributes

Attribute Description	Attribute Values	Mean	Standard Deviation
1. Clump thickness	1-10	4.44	2.83
2. Uniformity of cell size	1-10	3.15	3.07
3. Uniformity of cell shape	1-10	3.22	2.99
4. Marginal adhesion	1-10	2.83	2.86
5. Single epithelial cell size	1-10	2.23	2.22
6. Bare nuclei	1-10	3.54	3.64
7. Bland chromatin	1-10	2.45	2.45
8. Normal nucleoli	1-10	2.87	3.05
9. Mitoses	1-10	1.60	1.73

3. PROPOSED ALGORITHM

Reference to the previous literature review and background, a hybrid classification algorithm that integrates DE algorithm and LS-SVM is proposed. The proposed algorithm for BC diagnosis and treatment is composed of two main phases, Parameters Optimization and Classification. DE algorithm was used as a parameters optimization technique aiming to improve the sitting of the parameter values of LS-SVM; Hence, overcoming its sensitivity to the changes in the values of the parameters. Classification phase using LS-SVM technique consists of two main phases, Training phase followed by a phase of testing the algorithm. A block diagram of the proposed algorithm is depicted in Figure (1).

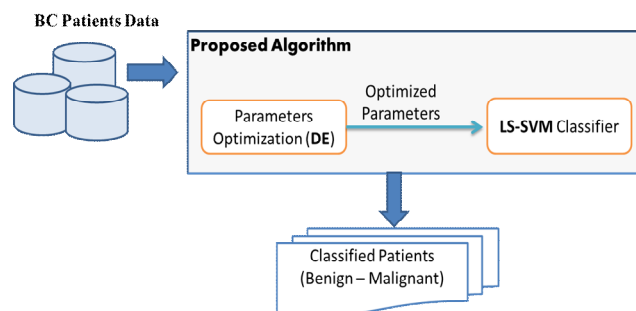


Figure 1: Block Diagram of the Proposed Algorithm

The aim of parameters optimization phase using DE is to find the optimal values for the parameters of the LS-SVM classifier (The regularization factor (γ) and Gaussian Kernel function (σ)). The second phase utilizes LS-SVM to classify the BC patients into one of two classes (Benign/ Malignant) using the optimized parameters.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the results obtained in the proposed algorithm trying to highlight its effectiveness. The experiments were done on WBC Database taken from UCI repository of Machine Learning Databases. The proposed algorithm is composed of 2 main phases. The first phase is the parameters' optimization which was conducted using DE algorithm. The second phase is the classification of BC patients using LS-SVM method. LS-SVM is a parameter dependent algorithm which means that it's so sensitive to the changes of its parameter values. Hence, DE algorithm seeks to find the best combination of the parameter values for LS-SVM that helps in maximizing the classification accuracy. In this study, the optimized parameters in LS-SVM were (σ, γ). BC patients were classified into one of two classes (Benign/Malignant).

The classification phase was divided into 2 main sub-phases, training followed by a testing phase. The proposed algorithm was applied on total of 669 records in WBC dataset. Classification accuracy was calculated as a performance measurement using (12)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

Where TP and TN are True Positive and True Negative respectively, which are the pro-portion of positive and negative cases that were correctly identified. Positive cases are the records with Benign label and negative ones are with Malignant label. FP and FN stand for False Positive and False Negative which are the proportion of negative cases that were incorrectly classified as positive and the proportion of positive cases that were incorrectly classified as negative respectively [19]. Three types of Cross Validation (CV) technique were conducted in this study, (80-20, 50-50 and 70-30). The classification accuracy resulted from each type was recorded, as shown in Figure(2)

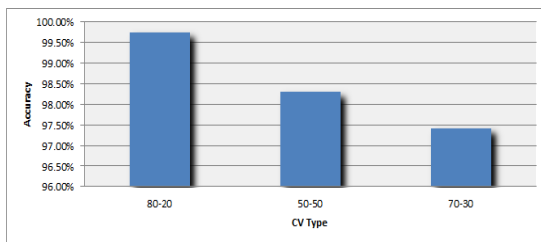


Fig.2 Classification Accuracy obtained from DE optimization technique

It's obvious from Figure(2) that the best classification accuracy was obtained from (80-20) CV which is 99.75%. To assert the contribution of DE algorithm and its effectiveness in the classification process; the proposed algorithm was compared with some of the previous results reported by earlier methods. Table-2 shows the comparison between the proposed algorithm and different classifier algorithms which were applied on the same database.

Table-2: Classification accuracy comparison

CI Technique	Accuracy
PSO-SVM [10]	93.55%
PSO-RBF Kernel [8]	99.3%
K-Means, SVM [31]	97.38%
Proposed Algorithm	99.75%

As shown in Table-2, the proposed algorithm outperformed the other CI techniques by obtaining the highest classification accuracy. It's obvious from Table-2 that PSO algorithm was used in many previous works. This is may be because PSO is a global search algorithm which has great benefits as an optimization technique. Although both PSO and DE are global search algorithms, DE could outperform PSO as proven in this study. This might be because PSO is a parametric method which depends on many parameters (inertia weight (ω), Acceleration Rates ((C_1, C_2))). Wrong choice of these parameter values may lead to divergent or cyclic trajectories. In contrast, DE doesn't have this problem.

Besides that PSO may lose its exploration capability [42]. On the other hand, K-means algorithm has different drawbacks that might prevent it from obtaining higher classification accuracy [32]. One of these disadvantages is that it is not work well in the global clusters. In contract, DE doesn't have this problem as it is a global search, population based and real-valued algorithm.

5. CONCLUSION AND FUTURE WORK

This paper proposed a hybrid classification algorithm for BC patients which integrates DE and LS-SVM algorithms. The proposed algorithm was composed of two main phases. The first phase is Parameters Optimization, followed by a Classification phase. Classification had two main phases, Training and Testing phases. The input parameters for LS-SVM were optimized using DE algorithm. The LS-SVM algorithm was used to classify BC patients into one of two classes (Benign/Malignant). DE algorithm could guarantee the robustness of the hybrid algorithm by searching for the optimal values for LS-SVM parameters. Optimizing the parameters could guarantee the effectiveness of the proposed algorithm by increasing the classification accuracy. The proposed algorithm was implemented on WBCD from UCI repository of Machine Learning Databases. Different types of CV were conducted and the results showed that 80-20 CV outperformed the other CV types by obtaining the highest classification accuracy of 99.75%. The proposed algorithm was compared with different classifier algorithms. The experimental results showed the effectiveness of the proposed algorithm. DE algorithm might obtain better results than what was mentioned in this study either by changing the schema or changing the parameter settings. Hence, as a future work, other schemas for DE (i.e. DE/best/2) might be used or DE parameters might be optimized to choose the best value for F and CR.

References

- [1] Hussein A Abbass. An evolutionary artificial neural networks approach for breast cancer diagnosis, Artificial Intelligence in Medicine, 25(3):265{281, 2002.
- [2] Pedro Henriques Abreu, Hugo Amaro, Daniel Castro Silva, Penousal Machado, and Miguel Henriques Abreu, Personalizing breast cancer patients with heterogeneous data, In The International Conference on Health Informatics, pages 39-42. Springer,2014.
- [3] Mehmet Fatih Akay. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 36(2):3240-3247, 2009.
- [4] Mafaz Mohsin Al-Anezi, Marwah Jasim Mohammed, and Dhufir Sami Hammadi. Artificial immunity and features reduction for effective breast cancer diagnosis and prognosis. International Journal of Computer Science Issues (IJCSI), 10(3), 2013.
- [5] Musrrat Ali, Millie Pant, and Ajith Abraham. Simplex differential evolution. Acta Polytechnica Hungarica, 6(5):95-115, 2009.

- [6] Y Benchaib, Alexis Marcano-Cedeño, Santiago Torres-Alegre, and Diego Andina. Application of artificial metaplasticity neural networks to cardiac arrhythmias classification. In *Natural and Artificial Models in Computation and Biology*, pages 181-190. Springer, 2013.
- [7] Cheng-Min Chao, Yao-Lung Kuo, and Bor-Wen Cheng. Three artificial intelligence techniques for finding the key factors in breast cancer, *Journal of Statistics and Management Systems*, 15(2-3):389-404, 2012.
- [8] Hui-Ling Chen, Bo Yang, Gang Wang, Su-Jing Wang, Jie Liu, and Da-You Liu. Support vector machine based diagnostic system for breast cancer using swarm intelligence. *Journal of medical systems*, 36(4):2505-2519, 2012.
- [9] Kun-Huang Chen, Kung-Jeng Wang, Min-Lung Tsai, Kung-Min Wang, Angelia Melani Adrian, Wei-Chung Cheng, Tzu-Sen Yang, Nai-Chia Teng, Kuo-Pin Tan, and Ku-Shang Chang. Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm. *BMC bioinformatics*, 15(1):49, 2014.
- [10] Smruti Rekha Das, Pradeepta Kumar Panigrahi, Kaberi Das, and Debahuti Mishra. Improving rbf kernel function of support vector machine using particle swarm optimization. *International Journal*, 2012.
- [11] Kris De Brabanter, Peter Karsmakers, Jos De Brabanter, Johan AK Suykens, and Bart De Moor. Confidence bands for least squares support vector machine classifiers: A regression approach. *Pattern Recognition*, 45(6):2280-2287, 2012.
- [12] Mei-Ling Huang, Yung-Hsiang Hung, Wen-Ming Lee, RK Li, and Tzu-Hao Wang. Usage of case-based reasoning, neural network and adaptive neuro-fuzzy inference system classification techniques in breast cancer dataset classification diagnosis. *Journal of medical systems*, 36(2):407-414, 2012.
- [13] Dervis Karaboga and Selçukökdem. A simple and global optimization algorithm for engineering problems: Differential evolution algorithm. *Turkish Journal of Electrical Engineering & Computer Sciences*, 12(1), 2004.
- [14] R Karaiş, M Tez, YA Kiliç, Y Kuru, and İGuler. A genetic algorithm model based on artificial neural network for prediction of the axillary lymph node status in breastcancer. *Engineering Applications of Artificial Intelligence*, 26(3):945-950, 2013.
- [15] CD Katsis, I Gkogkou, CA Papadopoulos, Y Goletsis, PV Boufounou, and G Stylios. Using artificial immune recognition systems in order to detect early breast cancer. *International Journal of Intelligent Systems & Applications*, 5(2), 2013.
- [16] Javed Khan, Jun S Wei, Markus Ringner, Lao H Saal, Marc Ladanyi, Frank Westermann, Frank Berthold, Manfred Schwab, Cristina R Antonescu, Carsten Peterson, et al, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, *Nature medicine*, 7(6):673-679, 2001.
- [17] Masayuki Kurosaki, Yasuhito Tanaka, Nao Nishida, Naoya Sakamoto, Nobuyuki Enomoto, Masao Honda, Masaya Sugiyama, Kentaro Matsuura, Fuminaka Sugauchi, Yasuhiro Asahina, et al. Pre-treatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis c using genetic polymorphism in il28b and viral factors. *Journal of hepatology*, 54(3):439-448, 2011.
- [18] Monalisa Mandal, Anirban Mukhopadhyay, and Ujjwal Maulik. Fuzzy rule-based classifier for microarray gene expression data by using a multiobjective pso-based approach. In *Fuzzy Systems (FUZZ)*, 2013 IEEE International Conference on, pages 1-7. IEEE, 2013.
- [19] Javad Haddadnia Hadi Varharam Mohammad Fiuzy, Azam Qarehkhani. Introduction of a method to diabetes diagnosis according to optimum rules in fuzzy systems based on combination of data mining algorithm (d-t), evolutionary algorithms (aco) and artificial neural networks (nn). *The Journal of Mathematics and Computer Science (JMCS)*, 6(4):272-285, 2013.
- [20] Mahamed GH Omran, Andries P Engelbrecht, and Ayed Salman. Differential evolution based particle swarm optimization. In *Swarm Intelligence Symposium, 2007. SIS 2007*. IEEE, pages 112-119. IEEE, 2007.
- [21] D Max Parkin, Paola Pisani, and J Ferlay. Global cancer statistics. *CA: A cancer journal for clinicians*, 49(1):33-64, 1999.
- [22] A Kai Qin, Vicky Ling Huang, and Ponnuthurai N Suganthan. Differential evolution algorithm with strategy adaptation for global numerical optimization. *Evolutionary Computation, IEEE Transactions on*, 13(2):398-417, 2009.
- [23] Xigao Shao, Kun Wu, and Bifeng Liao. Single directional smo algorithm for least squares support vector machines. *Computational intelligence and neuroscience*, 2013.
- [24] Rebecca Siegel, Deepa Naishadham, and Ahmedin Jemal. *Cancer statistics, 2013*. CA: a cancer journal for clinicians, 63(1):11-30, 2013
- [25] Daniele Soria, Jonathan M Garibaldi, Andrew R Green, Desmond G Powe, Christopher C Nolan, Christophe Lemetre, Graham R Ball, and Ian O Ellis. A quantifier-based fuzzy classification system for breast cancer patients. *Artificial intelligence in medicine*, 58(3):175-184, 2013.
- [26] Rainer Storn. On the usage of differential evolution for function optimization. In *Fuzzy Information Processing Society, 1996. NAFIPS, 1996 Biennial Conference of the North American*, pages 519-523. IEEE, 1996.
- [27] Tao Sun, Jingjing Wang, Xia Li, Pingxin Lv, Fen Liu, Yanxia Luo, Qi Gao, Huiping Zhu, and Xiuhua Guo. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in ct based on a multi-dimensional data set. *Computer methods and programs in biomedicine*, 111(2):519-524, 2013.
- [28] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293-300, 1999.
- [29] Wei-zh Tu. The design of multi-level decision making system based on the adaptive neural fuzzy inference system{case studies on health care and financial service. *2(12):7491-7499*, 2013.
- [30] Mingyuan Zhao, Chong Fu, Luping Ji, Ke Tang, and Mingtian Zhou. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications*, 38(5):5197-5204, 2011.
- [31] Bichen Zheng, Sang Won Yoon, and Sarah S Lam. Breast cancer diagnosis based on feature extraction using a hybrid of k-means and support vector machine algorithms. *Expert Systems with Applications*, 41(4):1476-1482, 2014.
- [32] Ray, Siddheswar and Turi, Rose H, Determination of number of clusters in k-means clustering and application in colour image segmentation, *Proceedings of the 4th*

international conference on advances in pattern recognition and digital techniques, 137—143, 1999

AUTHORS



Omar Soliman Associate Professor with the department of Operations Research and Decision Support, Faculty of Computers and Information, Cairo University, Giza, Egypt. Currently he is a lecturer in the Faculty of Computers and Information, Cairo University. Mainly his research areas of interest include Artificial Intelligence, Machine Learning, Optimization, Decision Support Systems, Data mining and knowledge discovery.



Eman AboElHamd Currently pursuing the master's degree with the department of Operations research and Decision Support in Faculty of Computers and Information, Cairo University, Giza, Egypt. She is also a teacher assistant at Faculty of Computers and Information, Cairo University. Her current research interests include Artificial Intelligence, Machine learning and Decision Support Systems.