A genetic algorithm-based deep RBF neural network for medical classification

Roguia Siouda
Department of computer sciences
University of 8 Mai 1945
Guelma Algeria
siouda.roguia@univ-guelma.dz

Mohamed Nemissi
Department of Electronics and telecommunications
University of 8 Mai 1945
Guelma Algeria
nemissi.mohamed@univ-guelma.dz

Hamid Seridi
Department of computer sciences
University of 8 Mai 1945
Guelma Algeria
seridi.hamid@univ-guelma.dz

ABSTRACT

In this paper, we introduce a deep RBF neural network for medical classification. The proposed classifier consists of two parts: an auto-encoder and an RBF neural network. The auto-encoder is used to decrease the number of the characteristics of the presented samples. Then, the obtained new features are presented to the RBF neural network. The design of the RBF neural networks is performed in two stages. First, the subtractive clustering method is used to define the centers of RBFs. Second, the genetic algorithm is used to optimize the widths of RBFs. To assess the proposed classifier, we perform tests over three medical datasets from the UCI machine-learning repository and we compare its performances with other methods.

KEYWORDS

RBF neural network, Auto-encoder, Genetic algorithm, Subtractive clustering method, Classification

ACM Reference format:

R. Siouda, M. Nemissi and H. Seridi. 2020. A genetic algorithm-based deep RBF neural network for medical classification. In *Proceedings of ACM ISPR conference, Hammamet, Tunisia, 16-18 October 2020. (ISPR'2020)*, 6 pages. https://doi.org/10.1145/3432867.3432868

ISPR '20, October 16–18, 2020, Virtual Event, Tunisia

ACM ISBN 978-1-4503-7502-3/20/10…$15.00
https://doi.org/10.1145/3432867.3432868

1 INTRODUCTION

With the current technological evolution, the collected biomedical data have been interestingly growing. They include a large variety of types, like blood tests, imaging data, biomedical signals, patients' records …etc. This constitutes an important means for the development of the medical systems, but it requires more computational efforts. Therefore, dealing with such amount of data impose using more powerful technics. Among these technics, neural networks have been successfully used because of their learning capabilities. In this work, we use neural networks for both data reduction and classification. More precisely, we use an auto-encoder for dimensionally reduction and RBF neural network for the classification task.

Radial Basis Function neural networks (RBFNNs) constitute an important type of neural networks. They are inspired form some biological neurons, which have local response. Consequently, they include a special type of hidden neurons: RBFs. Training RBFNNs can be done in one, two or three stages [1].

Recently, many approaches based on bio-inspired have been proposed for designing RBFNNs. For example, in [2], the authors proposed a design method based on combining hierarchical fuzzy clustering and PSO. In this method, the centers are defined using fuzzy clustering and the widths and the output weights are estimated using PSO. In [3], the authors proposed a polynomial RBFNN (p-RBF NNs) based on K-means clustering and optimized using Particle Swarm Optimization (PSO) and Differential Evolution (DE). In this system, the weights of RBFs are performed by four types of polynomials, which their coefficients are estimated using the Weighted Least Square Estimation (WLSE) method. PSO and DE are used to optimize the centers and their widths respectively. In [4], the authors proposed designing RBFNN using a cooperative particle swarm optimization (CPSO). It consists of two distinct swarms. The first swarm calculates the network structure and the centers using a non-symmetric fuzzy means algorithm, while the second calculates the widths. In this system, the two swarms cooperate by exchanging information. In [5], the authors proposed designing an RBFNN classifier in three stages. First, K-means clustering method is used for defining the centers of RBFs. Then, the PSO is used for optimizing the widths

of RBFs. Finally, the back-propagation algorithm is used for fine-tuning all parameters, i.e. centers, widths and output weights. In [6] the authors proposed a hybrid technique, in which the artificial bee colony (ABC) algorithm is used for the feature selection and k-nearest neighbor (k-NN) with genetic algorithm (GA) is used for the classification. The authors in [7] proposed an expert system

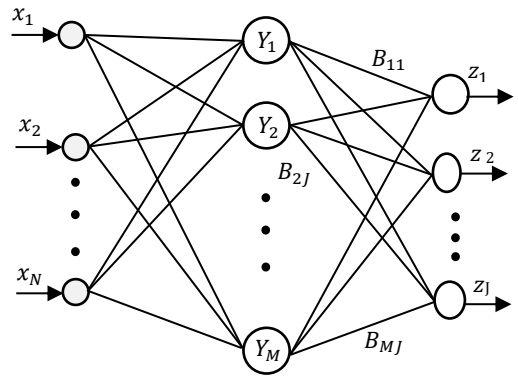

Figure 1: An example of RBF neural network with N-M-J architecture.

based on the genetic algorithm, simulated annealing and SVM for the diagnosis of hepatitis disease. SVM is used for classification and the hybrid of GA-SA for the selection of the most significant feature subset of the dataset and the optimization of the kernel parameters of the SVM as well.

Reducing the input dimensionality in the classification problems permits using less features and consequently it reduces the size of the neural network used as classifier. The training process becomes faster as the number of weights is less. In the case of RBFNNs, it also permits reducing the number of hidden neurons [8]. Furthermore, by reducing the input, the irrelevant data can be removed, which improve the overall performances. In the domain of data reduction, deep learning has been introduced as an effective method that perform unsupervised feature learning [9]. In this work, we use sparse auto-encoder (SAE), which has been widely employed. Therefore, the auto-encoder is first used to reduce input features. Afterward, the obtained features are presented to the RBFNN. The subtractive clustering method is then used to define the centers and the GA to define the widths of RBFs.

The rest of this paper is organized as follows. Section 2 gives some backgrounds about the RBFNNs, Auto-encoder and GA. Section 3 describes the proposed classifier. Section 4 presents and analyses the experimental results conducted on three biomedical datasets taken from the UCI machine-learning repository. Finally, section 5 concludes this paper.

2 THEORETICAL BACKGROUND

2.1 Radial Basis Function Neuronal Networks (RBFNNs)

An RBF network is a special type of single hidden feed-forward neural networks [10]. It includes radial basis functions (RBFs) as activation function for hidden neurons. Therefore, this type of neural networks differs from the Multi Layered Perceptron (MLP), which is based on sigmoid functions.

Fig. 1 illustrates an example of RBFNNs. In this example, the network has N input neurons, M hidden neurons and J output neurons. In the classification problems, the number of input neurons is equal to input dimension. In this work, we use only one output neuron as the medical problems are generally bi-classes.

We use the Gaussian functions, which are the most commonly used as activation functions. The outputs of the hidden layer are then given by:

$$Y_m(x) = exp\left(-\frac{\|X-V_m\|^2}{2\sigma_m^2}\right) \quad (1)$$

Where $X = (x_1, x_2, \ldots, x_N)$ is the input vector, $V_m = (v_{1m}, v_{2m}, \ldots, v_{Nm})$ and σ_m are, respectively, the center vector and width corresponding the m^{th} RBF.

The network outputs are linear weighted combinations of the hidden outputs. The j^{th} output, Z_j, is given by:

$$Z_j = \sum_{m=1}^{M} B_{mj} Y_m(x) \quad (2)$$

Where, M is the number of hidden neurons, Y_m is the output of the m^{th} hidden neuron and B_{mj} is the weight connecting the m^{th} hidden neuron and the j^{th} output.

2.2 Sparse auto-encoder (SAE)

An auto-encoder is a neural network trained using an unsupervised learning algorithm to give the target values equal to the inputs ($y(i) = x(i)$). It aims at learning a function $h(x) \approx x$; so, the output $\hat{x}$ is similar to the input x. Consequently, for a limited number of hidden neurons, interesting information can be found. Fig. 2 gives an example of auto-encoder neural network.

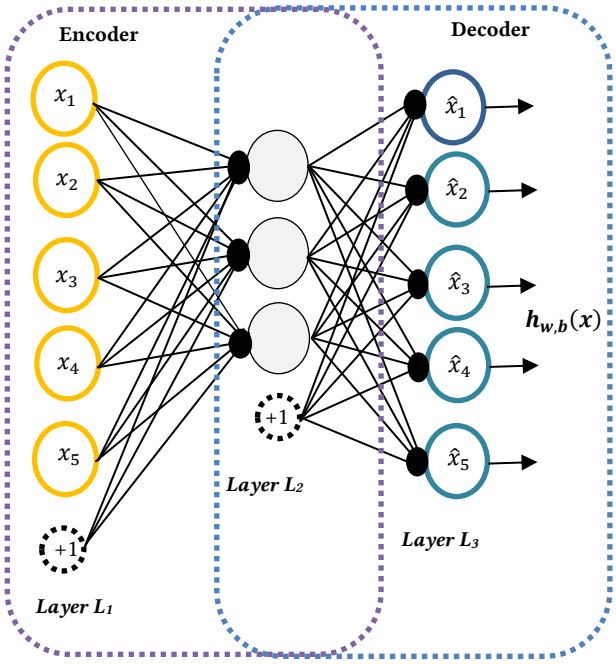

Figure 2: An example of auto-encoder with 5 inputs and 3 hidden neurons.

Sparse Auto-Encoder (SAE) aims at learning sparse features by adding a sparse penalty term inspired by the sparse coding [11]. This term is added to the cost function so that the learned features are not a simple repetition of the inputs. The sparse penalty aims at minimizing the number of "active" hidden neurons. Suppose $\alpha_j(x)$ denotes the activation of the j^{th} hidden neuron. Therefore, the average activation of this neuron is given by:

$$\rho_j = \frac{1}{n}\sum_{i=1}^{n}\left[a_j(x(i))\right] \quad (3)$$

Where, n is the dimension of the feature space.
The sparsity can be performed by adding a regularization term that indicates the difference between the average activation value, $\hat{\rho}_j$, and a target value, ρ. This term can be done by the Kullback-Leibler divergence as follows:

$$\Omega_{spar} = KL(\rho\|\rho_j) = \rho log\frac{\rho}{\rho_j} + (1-\rho)log\frac{1-\rho}{1-\rho_j} \quad (4)$$

The cost function can then be given as follows:

$$f = MSE\left(X - \hat{X}\right) + \alpha\Omega_{spar} + \beta\Omega_w \quad (5)$$

Where, $MSE\left(X - \hat{X}\right)$ is the mean squared error and Ω_w is the sum squared of all the networks weights.

2.3 Genetic Algorithm (GA)

The GA was initially developed by Holland and his colleagues in 1975. It borrows the mechanisms of natural evolution to find out the optimum solutions to the defined problems [12].
It starts with a randomly generated initial population covering the search space and then uses the processes of reproduction that combine the biologically-analogical selection, crossover, and mutation operators to create a new and fitter population of possible solutions. The reproduction process is iterated by simulating the natural selection and survival paradigms until the most appropriate solution is obtained.

- Selection

The selection process concept consists in selecting the most appropriate individuals and let them transfer their genes to the next generation.

- Crossover

In a genetic algorithm, crossover is the most significant step. A crossover point is randomly chosen in the genes chain for each parent's pair to be mated.

- Mutation

Some genes may be subject to a mutation with a low random probability in certain new offspring. This means that some bits can be flipped.

Every phase or component in GA plays a significant role in producing an optimal solution.

3 THE PROPOSED APPROACH

The aim of this work is to introduce a classification system based on a deep auto-encoder and an RBF neural network. The role of the auto-encoder is to reduce the dimensionally of the features, and consequently to decrease the networks size and simplify the optimization process. The design of the proposed classifier consists of three stages:
First, the auto-encoder is used to reduce the dimensionally of the features. The obtained new features constitute the input of the RBFNN.
Second, the Subtractive Clustering method is used to define the number of RBFs and their centers. This method was introduced by Chiu. et All in 1994 [13]. It is a simple method that permits determining the number of clusters automatically. In this work, we use it to define both the number of RBFs and their centers. In general, the clustering methods are unsupervised algorithms applied on unlabeled samples. In this work, we apply the clustering on each class separately in order to have more accurate RBFs [14].
Third, the GA is used to define the widths of the RBFs. Every individual constitutes a solution; it contains the widths corresponding to all RBFs. In this process, the output weights of RBFNN are calculated using the pseudo-inverse method which considerably simplify the training process [15, 16]. The used objective function is:

$$Fitness = \sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(Z_q - T_q\right)^2} \quad (6)$$

The complete process of the proposed method is illustrated in Fig. 3.

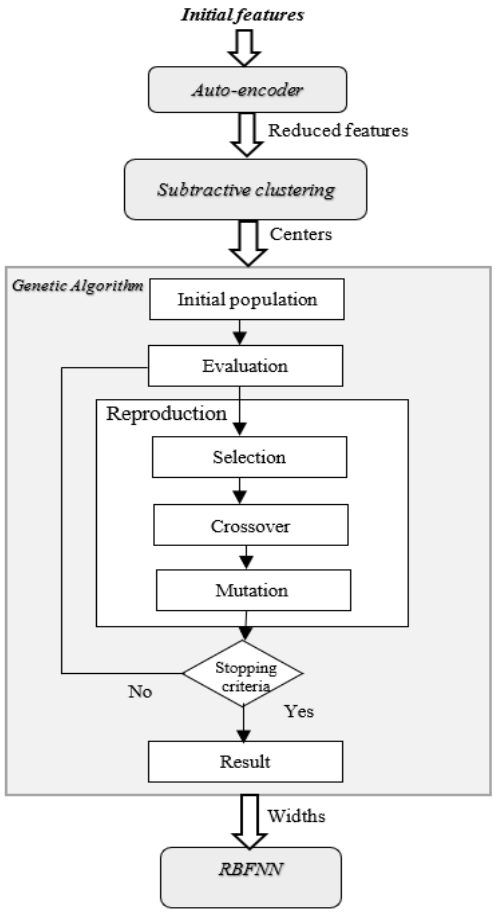

Figure 3: Principal scheme of the proposed method.

4 TESTS AND EXPREMENTS

To evaluate the performance of our classifier, we used three benchmark datasets from the UCI machine-learning repository [17]: Wisconsin Diagnostic Breast Cancer (WDBC); Parkinson; Hepatitis.

- The Wisconsin Diagnostic Breast Cancer (WDBC) dataset consists of 569 observations with 357 benign and 212 malignant. In this dataset, each observation has 32 attributes. The first two attributes correspond to the unique identification number and the diagnosis status or classes.
- The Parkinson disease dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado. This dataset contains 195 records and 2 classes: the presence and absence of Parkinson disease.
- The Hepatitis dataset is obtained from the Carnegie-Mellon University and it contains 155 instances belonging to two classes: live or die.

In order to evaluate the generalization performance, we used a 10-fold cross validation. According to this strategy, the dataset is randomly split into10 subsets, 9 of them are utilized as training data and the rest as test set. This procedure is repeated 10 times and consequently, all samples appear once in a test set.

Fig.4 displays the average accuracy classification results of the WDBC dataset using different features and numbers of RBFs.

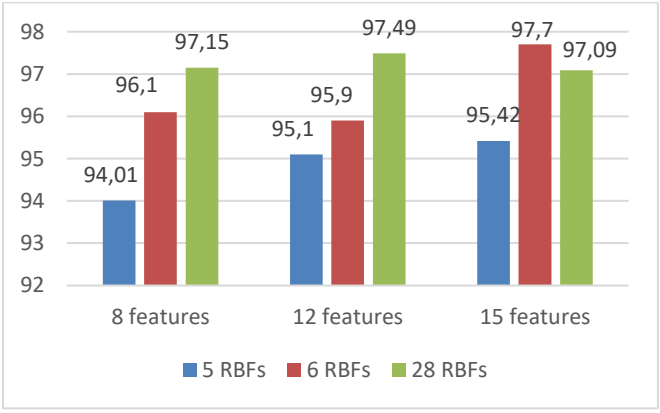

Figure 4: The classification results of the WDBC dataset.

Table 1. compares the proposed method with other works applied on the same dataset (WDBC). These works include a variety of machine learning techniques, i.e. fuzzy logic, neural networks, SVM, PCA, ...etc. We note that our classifier outperforms some of these works expect one which consists in an unsupervised feature extraction algorithm based on Deep Learning.

Table 1: Comparison of Wisconsin diagnostic breast cancer (WDBC) dataset.

Authors, Year	Methods	Accuracy (%) [Cross Validation]
Gouda I. Salama, (2012) [18]	SMO	97.71 [10-CV]
Bichen Zheng, (2014) [19]	K-means+SVM	97.38 [10-CV]
G. Naga Ramadevi, (2015) [20]	PCA+SVM	96.84 [10-CV]
Mehrbakhsh Nilashi, (2017) [21]	EM+PCA+Fuzzy-RuleBased	93.2 [10-CV]
Yawen Xiao, (2018) [22]	**SAE+SVM**	**98.25 [10-CV]**
Proposed	Auto-encoder+RBFNN	97.70 [10-CV]

Fig.5 displays the average accuracy classification results of the Parkinson dataset using different features and numbers of RBFs.

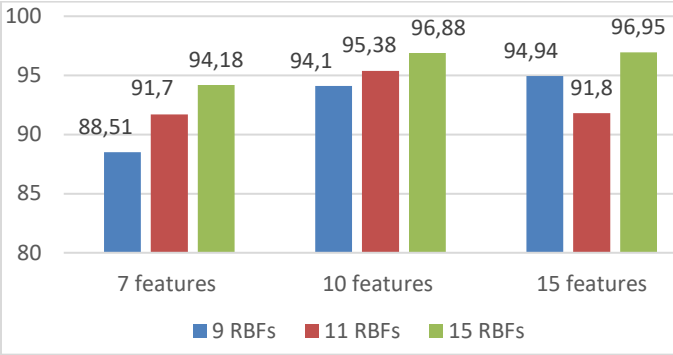

Figure 5: The classification results of the Parkinson dataset.

Table 2. compares the proposed classifier with other works applied on the same dataset (Parkinson). These works include neural networks, random forest, ELM ...etc. We note that our classifier outperforms all these works.

Table 2: Comparison of Parkinson dataset.

Authors, Year	Methods	Accuracy (%) [Cross Validation]
S. Anto (2015) [7]	GASA-SVM	87.00
Hui-Ling Chen, (2015) [23]	mRMR-KELM	95.97 [10-CV]
Prerna Sharma1, (2018)[24]	MGWO + random forest	93.87 [70-30]
Deepak Gupta, (2018)[25]	OCFA	92.19 [70-30]
Proposed	**Auto-encoder+ RBFNN**	**96.95 [10-CV]**

Fig. 6 displays the classification results obtained in the Hepatitis dataset using different numbers of features and RBFs.

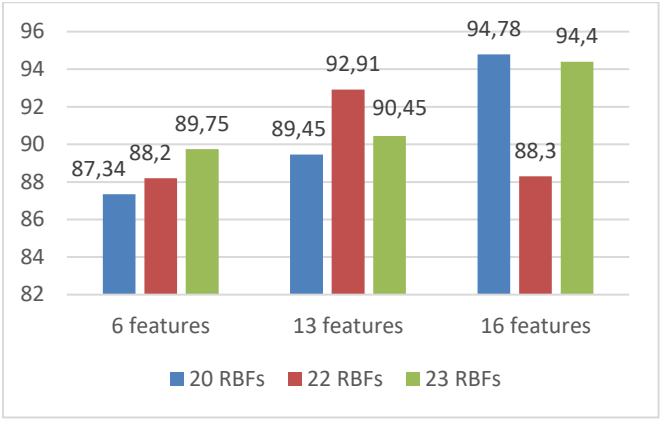

Figure 6: The classification results of the Hepatitis dataset.

Table 3. compares the proposed classifier with some other state-of-the-art works applied on the same dataset (Hepatitis). These works include neural networks, PCA, ELM, RBFNN, PSO ...etc. We note that our classifier outperforms all these works.

Table 3: Comparison of Hepatitis dataset.

Authors, Year	Methods	Accuracy (%) [Cross Validation]
Sultan Noman Qasem, (2011) [26]	RBFNN+TVMOPSO	82.26 [50-50]
Nursuci Putri Husain, (2017) [27]	LSSVM+ IACA algorithm	93.7 [80-20]
Ibrahim Aljarah, (2018) [28]	RBFNN+BBO	84.72 [67-33]
Proposed	**Auto-encoder+ RBFNN**	**94.78 [10-CV]**

5 CONCLUSIONS

A deep RBF neural network based on GA have been proposed. In this system, the feature space is reduced using an auto-encoder in order to decrease the RBF network size and simplify the optimization process. The RBFNN training is performed by defining the centers of RBFS using subtractive clustering method and optimizing the widths using GA. The proposed classifier was evaluated over three medical datasets and compared with some state-of-the-art methods. The obtained results showed good generalization performances and they are promising for more enhancement.

REFERENCES

[1] F. SCHWENKER, H. A. KESTLER and G.PALM. 2001. Three learning phases for radial-basis-function networks. Neural networks, 14(4-5), 439-458.

[2] A. D. Niros, G. E. Tsekouras, D. Tsolakis, A. Manousakis-Kokorakis and D. Kyriazis. 2015. Hierarchical fuzzy clustering in conjunction with particle swarm optimization to efficiently design RBF neural networks. Journal of Intelligent & Robotic Systems, 78(1), 105-125.

[3] OH. S. K. Kim, W. D. Pedrycz, W. and S. C. Joo. 2012. Design of K-means clustering-based polynomial radial basis function neural networks (pRBF NNs) realized with the aid of particle swarm optimization and differential evolution. Neurocomputing, 78(1), 121-132.

[4] A. Alexandridis, E. Chondrodima and H. Sarimveis. 2016. Cooperative learning for radial basis function networks using particle swarm optimization. Applied Soft Computing, 49, 485-497.

[5] R. Siouda and M. Nemissi. 2018. An Optimized RBF-Neural Network for Breast Cancer Classification. International Journal of Informatics and Applied Mathematics, 1(1), 24-34.

[6] S. Kumar and G. Sahoo. 2017."Medical Dataset Classification Using k-NN and Genetic Algorithm." Computational Intelligence in Data Mining. Springer, Singapore, 813-823.

[7] S. Anto and S. Chandramathi. 2015. "An expert system based on SVM and hybrid GA-SA optimization for hepatitis diagnosis." International Journal of Computer Engineering in Research Trends 2.7, 437-443.

[8] X. Fu and L. Wang. 2003. Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 33(3), 399-409.

[9] G. E. Hinton and R. R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.

[10] S. N. Qasem and S. M. Shamsuddin. 2010. Generalization improvement of radial basis function network based on multi-objective particle swarm optimization. J. Artif. Intell, 3(1), 1-16.

[11] B. A. Olshausen and D. J. Field. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381(6583), 607.

[12] J. H. Holland. 1975. Adaptation in Natural and Artificial Systems, University of Michigan, Ann Arbor, Mich, USA.

[13] S. L. Chiu. 1994. Fuzzy model identification based on cluster estimation. Journal of Intelligent & fuzzy systems, 2(3), 267-278.

[14] H. Salah, M. Nemissi, H. Seridi and H. Akdag. 2019. Subtractive Clustering and Particle Swarm Optimization Based Fuzzy Classifier. International Journal of Fuzzy System Applications (IJFSA), 8(3), 108-122.

[15] M. NEMISSI, H. SALAH and H. SERIDI. 2018. Breast cancer diagnosis using an enhanced Extreme Learning Machine based-Neural Network. In 2018 International Conference on Signal, Image, Vision and their Applications (SIVA) (pp. 1-4). IEEE.

[16] M. Nemissi, H. Seridi and H. Akdag. 2014. One-against-all and one-against-one based neuro-fuzzy classifiers. Journal of Intelligent & Fuzzy Systems, 26(6), 2661-2670.

[17] K. Bache and M. Lichman. 2013. UCI Machine Learning Repository [http://archive. ics. uci. edu/ml]. Irvine, CA: University of California. School of information and computer science, 28.

[18] G. I. Salama, M. Abdelhalim and M. A. E. Zeid. 2012. Breast cancer diagnosis on three different datasets using multi-classifiers. Breast Cancer (WDBC), 32(569), 2.

[19] B. Zheng, S. W. Yoon and S. S. Lam. 2014. 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.

[20] G. N. Ramadevi, K. U. Rani and D. Lavanya. 2015. Importance of feature extraction for classification of breast cancer datasets-a study. International Journal of Scientific and Innovative Mathematical Research, 3(2), 763-368.

[21] M. Nilashi, O. bin Ibrahim, H. Ahmadi and L. Shahmoradi. 2017. An analytical method for diseases prediction using machine learning techniques. Computers & Chemical Engineering, 106, 212-223.

[22] Y. Xiao, J. Wu, Z. Lin and X. Zhao. 2018. Breast cancer diagnosis using an unsupervised feature extraction algorithm based on deep learning. In 2018 37th Chinese Control Conference (CCC) (pp. 9428-9433). IEEE.

[23] H.L.Chen, G.Wang, C.Ma, Z.N.Cai, W.B.Liu and S.J.Wang. 2016. An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson′ s disease. Neurocomputing, 184, 131-144.

[24] P. Sharma, S. Sundaram, M. Sharma, A. Sharma and D. Gupta. 2019. Diagnosis of Parkinson's disease using modified grey wolf optimization. Cognitive Systems Research, 54, 100-115.

[25] D. Gupta, A. Julka, S. Jain, T. Aggarwal, A. Khanna, N. Arunkumar and V. H. C. de Albuquerque. 2018. Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. Cognitive systems research, 52, 36-48.

[26] S. N. Qasem and S. M. Shamsuddin. 2011. Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis. Applied Soft Computing, 11(1), 1427-1438.

[27] N. P. Husain, N. N. Arisa, P. N. Rahayu, A. Z. Arifin and D. Herumurti. 2017. Least Squares Support Vector Machines Parameter Optimization Based on Improved Ant Colony Algorithm For Hepatitis Diagnosis. Jurnal Ilmu Komputer dan Informasi, 10(1), 43-49.

[28] I. Aljarah, H. Faris, S. Mirjalili and N. Al-Madi. 2018. Training radial basis function networks using biogeography-based optimizer. Neural Computing and Applications, 29(7), 529-553.