Automatic System for COVID-19 Diagnosis

Seyyid Ahmed Medjahed[1], Mohammed Ouali[2]

[1] Ahmed Zabana University Center,
Algeria

[2] Thales Canada Inc.,
Canada

seyyidahmed.medjahed@cu-relizane.dz, mohammed.ouali@usherbroorke.ca

Abstract. During the last months, the virus COVID 19 spread globally, quickly and affected many people. This last, is an infection caused by severe acute respiratory. Unfortunately, the number of cases increases significantly and early diagnosis of this disease can help to save the health of patient and his entourage by stopping contamination. In this paper, we propose a process of COVID 19 diagnosis in Chest X-rays. This process is composed of three main steps. The first one is the feature extraction using four approaches. The second one is the feature selection phase using a new feature selection approach. The last phase is the classification. The classifier used in this approach is composed of four supervised classification approaches. The proposed work has been tested COVID-19 in X-ray images obtained by PyImageSearch.

Keywords. COVID-19 diagnosis, feature extraction, feature selection, classification, multi-verses optimizer.

1 Introduction

Since the last months of 2019, a new disease has been appeared called COVID 19 (Coronavirus). This virus has been appeared firstly in Wuhan capital of Hubei, China. COVID 19 is spread during close to contaminated surface of when people cough or sneeze. The symptoms are fever, cough, fatigue, myalgia. Unfortunately, on March $11, 2020$, the world health organization declared a pandemic. Until today we record 659544 cases and 30630 deaths.

Intelligence Artificial and precisely Machine learning has been a very active research field in medical diagnosis by analyzing X-ray images. Generally, the process of use of machine learning is based into building a strong training model. The main keys of this process are feature extraction and feature selection. The first one is the representation of an image as feature vector, among of feature extraction methods, we cite, Histogram of oriented gradients, Local binary patterns, Color histograms, Fourier, Gabor, Discrete cosine transform, etc.

The second one is the feature selection which allows to select the relevant and the optimal subset of feature by removing the non-informative and redundant features. Classification output depends largely of the features used to build the training model. The features are not all relevant, several features are considered as noise and reduce the accuracy rate.

Feature selection is a process of preprocessing data that attempts to select the optimal subset of features considered as the relevant and informative one before classification. Feature selection approaches are divided into three categorize: filter, wrapper and embedded. Filter approach. Filter approach uses the general characteristics of features. It classifies the feature according to certain measure such as Fisher Score [5], Pearson correlation [20], mutual information [19], etc.

Wrapper approach is based on the generation of features subset by using classifier. The score using to evaluate the candidate subset is generally the classification accuracy rate. The wrapper approach provides a good results compared to filter approach, but, it is time consuming because it called several times at each iteration the classifier, and the results can highly be related to that classifier. Wrapper approach uses generally a meta heuristic such as Genetic Algorithm [7], Particle Swarm Optimization [8], Gravitational Search Algorithm [16], Binary Bat Algorithm [14], etc.

The last category is embedded approach which integrates the feature selection during the classification process such as SVM-RFE [15].

Many works have been done on medical diagnosis based on feature selection. In [1], the authors attempt to diagnosed the autism spectrum disorder using electroencephalogram. The authors use a feature selection approach based on mutual information, information gain, minimum redundancy maximum relevancy and genetic algorithm. The classifies are K nearest neighbor and support vector machine. Xiaoke et al. [6], present multi-modal neuroimaging feature selection using for diagnosis Alzheimer's disease.

The authors propose a new multi-modal neuroimaging feature selection based on consistent metric constraint for AD analysis. The multi-kernel support vector machine is using as classifier. In [3], the authors propose to incorporate the feature selection approaches for neonatal seizure diagnosis. The feature selection is based on decision support system using the electroencephalography. They use ten different feature selection algorithms to select the optimal subset of feature. In [12], the authors trait the problem of neurological disorder diagnosis for autism. They proposed to use feature selectin approach to reduce the high-dimensionality of connectome data. The authors proposed a new feature selection approach called brain network atlas guided feature selection to disentangle the healthy from the disordered connectome.

In this paper, we propose a complete process of COVID 19 diagnosis. This process is composed of three phases.

The first one is the feature extraction based on four approaches. The second phase is feature selection, we propose a new approach based on Multi-Verse Optimizer and a new objective function. The last phase is the classification analysis. This work will be tested on COVID 19 Chest X-ray images.

The rest of paper is organized as follows: In section 2, we detailed the proposed approach. In section 3, we present and discuss the experimental results. In section 4, the conclusions and some future work are presented.

2 Proposed Approach

The approach proposed in this work contains three phases: feature extraction, feature selection and classification.

2.1 Feature Extraction

The first step consists of the extraction of features. The process is to convert the image to feature vector. In this study, we propose to combine four feature extraction approach which are:

— Pyramid Histogram of Orientation Gradients consists to the gradient orientation in the image used generally for object detection. (dalal trigs 2005) it consist of counting the occurrence of gradient orientation. The image is divided to sub regions at different resolutions [4].

— Fourier features is a very used approach in image processing. It divided the image into sine and cosine components. The number of frequencies is the number of pixels in image [18].

— Gabor feature attempts to extract characteristics of scale, orientation and spatial locality which are combined to recognize a region [9].

— Discrete cosine transform is member of the class of sinusoidal unitary transforms. It is a feature extraction method that divide the image into sub blocks of differing importance related to the visual quality [17].

The basic of combining all these feature extraction approaches is to use all the advantage of each one, and, get a sufficient number of features.

2.2 Feature Selection

In this section, we present the main step which is feature selection. This last represents a primordial step in classification process. It allows to select the optimal sub set of features considering as the relevant and informative subset. Improving this step allows to increase the quality of classification and by consequence the accuracy rate.

In this study, we cast the feature selection problem as a combinatorial optimization problem defined as follows:

Let's supposing $F = \{F_1, \ldots, F_n\}$ entire set of all features provided by the first step (feature extraction). We define a binary variable $X = \{X_1, \ldots, X_n\}$ which is decision variable that can be 0 or 1. 0 means that the feature is selected and it will be used to build the training model and 0 otherwise.

To evaluate the quality of candidate feature subset, a certain measure must be defined. The objective function computes the score of selecting the candidate feature subset or not. The objective function $J(X)$ proposed in this study is composed of two terms, the classification accuracy rate $J_1(X)$ and number of selected bands $J_2(X)$:

$$J(X) = \alpha \times J_1(X) + (1 - \alpha) \times J_1(X).$$

The main goal of this objective function is to reduces the classification error rate and the number of selected features together.

The classification accuracy rate is obtained by five classifiers: Support Vector Machine using Gaussian Kernel, K Nearest Neighbor, Naï¿½ve Bayes, Discriminant Analyses Classifier, Decision Tree. for each new instance, to select the correct class, we compute the number of occurring classes in between the five classifiers: Let's Z a new instance to be classified, the proposed approach worked as follows:

$$Class(Z) = Max(occuringNumber(C_1(Z), C_2(Z), \\ C_3(Z), C_4(Z), C_5(Z))),$$

where $C_1(Z), C_2(Z), C_3(Z), C_4(Z), C_5(Z)$ are the class of Z using Support Vector Machine, K Nearest Neighbor, Naive Bayes, Discriminant Analysis Classifier, Decision Tree. $J_1(X)$ has the following form:

$$J1(X) = CAR(\text{Using selected features}(X_i \times F_i)),$$

where CAR is classification accuracy rate.

The second term of objective function is the number of selected band which attempts to minimize the number of selected features:

$$J2(X) = \frac{Number\_Sel.\_Features(SUM(X_i = 1))}{n},$$

where n is the total number of features.

To minimize the objective function, we propose to use the Multi-Verses Optimizer.

2.2.1 Multi-Verses Optimizer

Multi-Verses Optimizer is naturel inspired optimization algorithm based on Multi-verse theory. Our universe was created by a big explosion called big bang. The universe is on expansion through space which is caused by the eternal inflation. Inflation is the main source of forming planets, starts, black hole, etc. [13, 2].

Muti-Verses theory admets that it exists other universes with different physical laws. In cosmology, three concepts exists: White hole, black hole, worm hole. These three concepts are the main keys of multi-verses theory.

The Multi-verse assumes that there are many universes also created by big bang [13, 2].

Multi-Verses Optimizer is based on the following rules [13, 2]:

— The higher inflation rate, the higher probability of having while hole,

— The higher inflation rate, the lower probability of having black hole,

— Universes with higher inflation rate tend to send objects thrrough white hole,

— Universe with lower inflation rate tend to receive more objects through black holes,

— The objects in universe may face random movement towards the best universe via worm holes regardless of the inflation rate,

— Each solution is a universe and each variable in the solution is an object in the universe. The concept of white and black holes is used for exploration and the concept of wormhole is used for exploitation [13, 2].

The mathematical model is defined as follows [13, 2]:

Let's U a universe with:

$$
U = \begin{bmatrix}
x_1^1 & x_1^1 & . & . & . & x_1^1 \\
x_1^1 & x_1^1 & . & . & . & x_1^1 \\
. & . & & . & & . \\
. & . & & . & & . \\
. & . & & . & & . \\
x_1^1 & x_1^1 & . & . & . & x_1^1
\end{bmatrix}
$$

d is the number of variables and n is the number of universes (candidate solution):

$$
x_i^j = \begin{cases}
x_k^j & r_1 < NI(U_i), \\
x_i^j & r_1 \geq NI(U_i),
\end{cases}
$$

where x_i^j is the jth variable of ith universe. U_i is the ith universe. $NI(U_i)$ is the normalized inflation rate ot the ith universe. x_i^j is the jth variable ok kth universe selected by roulette wheel selection mechanism. r_1 is random number between 0 and 1 [13, 2].

The pseudocodes for this part are as is presented in Algorithm 1 [13, 2].

In order to provide local changes for each universe and have high probability of improving the inflation rate using wormholes, we assume that wormhole tunnels are always established between a universe and the best universe formed so far. The formulation of this mechanism is as follows:

$$
x_i^j = \begin{cases}
\begin{cases}
X_j + TDR \times ((ub_j - lb_j) \\
\qquad \times r_4 + lb_j) & r_3 < 0.5, \\
X_j - TDR \times ((ub_j - lb_j) \\
\qquad \times r_4 + lb_j) & r_3 \geq 0.5, \\
\qquad\qquad x_i^j.
\end{cases}
\end{cases}
$$

x_i^j is equal to the first term of equation if $r_2 < WEP$ elsewise is equal to x_i^j.

Algorithm 1 First part

1: SU = sorted universes
2: NI = Normalize inflation rate of the universe (Compute fitness fuction)
3: **for** Each universe indexed by i **do**
4:    Black Hole Index $= i$
5:    **for** Each object indexed by j **do**
6:      r_1 = random number between 0 and 1
7:      **if** $r_1 < NI(U_i)$ **then**
8:        While Hole Index = RouletteWheelSelection(NI)
9:        U(Black Hole Index, j) = SU(While Hole Index, j)
10:     **end if**
11:   **end for**
12: **end for**

Here X_j is the jth variable of the best universe formed so far. lb_j and Ub_j is the lower and upper bound of jth variable. x_i^j is the jth variable of the ith universe r_2, r_3, r_4 are random numbers between 0 and 1.

The pseudocode of this part is defined as presented in Algorithm 2 [13, 2].

Algorithm 2 Second part

1: **for** Each universe indexed by i **do**
2:   **for** Each object indexed by j **do**
3:    r_2 = random number between 0 and 1
4:    **if** r_2¡WormholeExistanceProbability **then**
5:      r_3 = random number between 0 and 1
6:      r_4 = random number between 0 and 1
7:      **if** r_3¡0.5 **then**
8:        $U(i,j)$ = Best Universe(j) + TravellingDistanceRate x $(ub_j - lb_j)$x$r_4 + lb_j$
9:      **else**
10:       $U(i,j)$ = Best Universe(j) - TravellingDistanceRate x $(ub_j - lb_j)$x$r_4 + lb_j$
11:      **end if**
12:     **end if**
13:   **end for**
14: **end for**

As seen in the pseudocode, there are two main coefficients: WormholeExistenceProbability (WEP) and TravellingDistanceRate (TDR). The first one increase linearly over the iterations in order to

emphasize the exploitation phase. The second one TDR represents the distance rate that an object can be teleported by a wormhole around the best universe. These two coefficients is defined as follows [13, 2]:

$$WEP = \min + l \times (\frac{\max - \min}{L}),$$

$$TDR = 1 - \frac{l^{\frac{1}{p}}}{L^{\frac{1}{p}}},$$

where l current iteration, L maximum iterations, p is the exploitation accuracy over the iterations [13, 2]. MVO algorithm is defined as presented in Algorithm 3 [13, 2].

Firstly, the algorithm generates randomly a set of universes. In each iteration, by using white and black holes, the objects can move between universe with high inflation rates to universe with low inflation rate.

Each universe faces random teleportation in its objects via worm holes to the best universe [13, 2].

2.2.2 Proposed Binary Multi-Verse Optimizer

We propose a binary version of of MVO algorithm. The problem of feature selection is a binary problem where 1 means that the feature is selected and 0 otherwise.

In other terms, if $X_i = 1$, the feature F_i is selected and used to build the training model, else, if $X_i = 0$, the feature F_i is not selected [10, 11]. This is why, we use the sigmoid function as follows:

$$S(x_i^j) = \frac{1}{1 + \exp(-x_i^j)}, \qquad (1)$$

$$x_i^j = \begin{cases} 1 \ if \ S(x_i^j) \geq 0.5, \\ 0 \ if \ S(x_i^j) < 0.5. \end{cases}$$

Algorithm 3 MVO Algorithm

1: Generate random universe (U)
2: Initialize WER, TDR and Best Universe
3: SU = Sorted universes
4: NI = Normalize inflation rate of the universe (Compute fitness fuction)
5: **for** Each iteration **do**
6:    Compute the fitness function of all universes
7:    **for** Each universe indexed by i **do**
8:      Update WEP and TDR
9:      Black Hole Index $= i$
10:      **for** Each object indexed by j **do**
11:        r_1 = random number between 0 and 1
12:        **if** $r_1 < NI(U_i)$ **then**
13:          While Hole Index = RouletteWheelSelection(NI)
14:          U(Black Hole Index, j) = SU(While Hole Index, j)
15:        **end if**
16:        r_2 = random number between 0 and 1
17:        **if** $r_2 ¡$ WormholeExistanceProbability **then**
18:          r_3 = random number between 0 and 1
19:          r_4 = random number between 0 and 1
20:          **if** $r_3 ¡ 0.5$ **then**
21:            $U(i, j)$ = Best Universe(j) + TravellingDistanceRate x $(ub_j - lb_j)$x$r_4 + lb_j$
22:          **else**
23:            $U(i, j)$ = Best Universe(j) - TravellingDistanceRate x $(ub_j - lb_j)$x$r_4 + lb_j$
24:          **end if**
25:        **end if**
26:      **end for**
27:    **end for**
28: **end for**

---

3 Experimental Results

We present the results obtained by the experiments in this section. Performances demonstration are conducted in terms of classification accuracy rate, sensitivity, specificity, positive predictive value and negative predictive value. The following formula are used to compute these measures. Let us define NTP as Number of True Positives; NTN as Number of True Negatives; NFP as Number of False Positives; and NFN as Number of False Negatives. Then we can define the following measures.

Accuracy Rate	$\frac{NTP+NTN}{NTP+NTN+NFP+NFN}$,
Sensitivity	$\frac{NTP}{NTP+NFN}$,
Specificity	$\frac{NTN}{NFP+NTN}$,
Positive Predictive Value	$\frac{NTP}{NTP+NFP}$,
Negative Predictive Value	$\frac{NTN}{NTN+NFN}$.

3.1 Datasets

The images dataset used in this work is obtained by Adrian Rosebrock and available in (PyImageSearch.com). This dataset is composed of 50 X-ray images divided into two categories: 25 images represents normal chest and the remaining 25 are classed as COVID 19. The images have different size. Figure 1 illustrated some chest X-ray images.
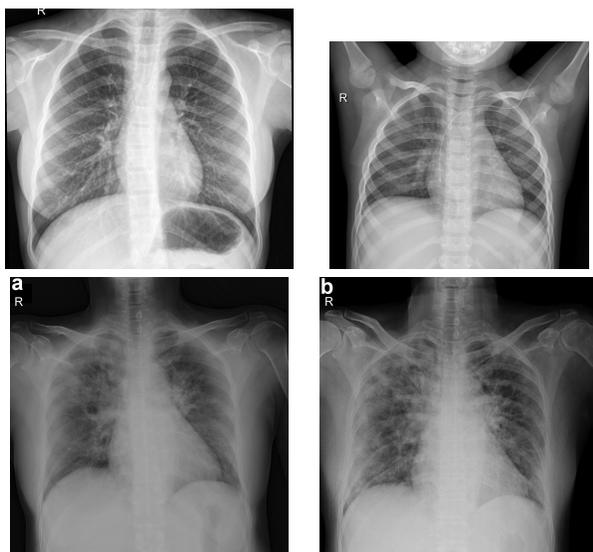

Fig. 1. Some chest X-ray Images.

3.2 Parameters Setting

In classification, is very primordial to define the training and testing sets. In this study, we propose to divide the dataset into two subset: 70% instances used for training and 30% used for test. To avoide the problem of overtraining, in each iteration of the algorithm, we split randomly the dataset.

The parameters of the proposed approach are defined as follows:

Firstly, parameters of MVO:

— Number of universes is 60,

— Number of iterations is 50,

— Coefficient p is 6,

— Value of min is 0.2,

— Value of max is 1.

As mentioned above, for feature extraction step, we have used four approach:

— Features from F_1 to F_{765} are obtained by Pyramid Histogram of Oriented Gradients,

— Features from F_{766} to F_{773} are obtained by Fourier,

— Features from F_{774} to F_{841} are obtained by Gabor,

— Features from F_{842} to F_{844} are obtained by DCT.

The total number of features is 844.

3.3 Results and Discussion

In this section, we present the experimental results obtained by the proposed approach. Table 1 presents the results.

Table 1. Comparaison between the results (%) obtained by the proposed approach and some classifier. PPV (Positive Predictive Value), NPV (Negative Predictive Value)

	SVM	KNN	CNB	DCA	DTREE	This study
Accuracy	75	75	90	85	75	95
Sensitivity	70	80	100	70	90	90
Specificity	80	70	80	100	60	100
PPV	77,77	72,72	83.34	100	69,23	100
NPV	72,72	77,78	100	76,92	85,71	90,90

Table 1 represents the classification accuracy rate, sensitivity, specificity, positive predictive value, negative predictive value obtained by the proposed approach and some classifier (classifier using all the features) Support Vector Machine (SVM), K Nearest Neighbor (KNN), Classifier Native Bayes (CNB), Discriminant Analyses Classifier (DAC), Decision Tree (DTREE).

By analyzing the results, we clearly observe that the proposed approach provides a high

classification accuracy rate that reaches 95% following by native bayes with 90% of accuracy. Discriminant analyses classifier provides 85% of accuracy. The rest of classifiers reach 75% of classification accuracy rate.

For the proposed approach the sensitivity is 90% and the specificity reaches 100%. This means that the proposed approach can return correctly a positive result for 90% of people who has the disease and a false value for the 10% of peoples. With 100% value of specificity means that the proposed approach returns correctly a negative result for 100% of people. The positive and negative predictive value are very satisfactory.

The total number of feature is 844. The proposed approach has selected 492 features which means that 58% of features has been selected.

This paper can be summarized with the following points:

1. The proposed approach is composed of three steps: feature extraction, feature selection and classification.

2. The features set is composed of features extracted by using: Pyramid Histogram of Orientation Gradients, Fourier, Gabor and Discrete cosine transform.

3. The feature selection approach is based on Muti-Verse Optimizer and a Binary version is proposed.

4. The fitness function is composed of two important terms: accuracy rate and the number of selected features. The goal is to minimize the classification error rate and also the number of selected features.

5. The classification approach used to compute the fitness function and the classification accuracy rate is based on five classifiers: Support Vector Machine using Gaussian Kernel, K Nearest Neighbor, Native Bayes and Discriminant Analyses Classifier and Decision Tree. The class affected to instance is analyzed and choosen among the five classes generated by the different classifiers

4 Conclusion

This paper proposes an automatic system for COVID 19 diagnosis. The system is composed of three main steps: feature extraction, feature selection and classification. We propose to combine four feature extraction approaches Pyramid Histogram of Orientation Gradients, Fourier, Gabor and Discrete cosine transform.

The next step is feature selection which allows to select the relevant features. For this step, a wrapper approach is proposed based on Multi-Verse Optimizer and a binary version of MVO is defined. The objective function is to minimize the number of features and to minimize the classification error rate. We combine five classifiers: Support Vector Machine using Gaussian Kernel, K Nearest Neighbor, Native Bayes and Discriminant Analyses Classifier and Decision Tree. the class affected to the instance is the class that has the maximum number of occurring between all the classes generated by the classifier. The dataset is a set of chest X-ray images available on PyImageSearch.com. Performance evaluation has been done by analyzing the classification accuracy rate, sensitivity, specificity, positive predictive value and negative predictive value. The analysis of the results indicates that the proposed approach provides satisfactory results compared to classifier without feature selection. As future work, is will be very interesting to test this approach in a big dataset contains many images.

References

1. **Abdolzadegana, D., Hossein, M., & Majid Ghoshunib, M. (2020).** A robust method for early diagnosis of autism spectrum disorder from EEG signals based on feature selection and DBSCAN method. *Biocybernetics and Biomedical Engineering*, Vol. 40, No. 1, pp. 482–493.

2. **Abualigah, L. (2020).** Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications. *Neural Computing and Applications*, Vol. https://doi.org/10.1007/s00521-020-04839-1, pp. 1–21.

3. **Acikoglu, M. & Arslan Tuncer, S. (2020).** Incorporating feature selection methods into a machine learning-based neonatal seizure diagnosis. *Medical Hypotheses*, Vol. 135, pp. 109–464.

4. **Dalal, N. & Triggs, B. (2005).** Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA*, Vol. 1, pp. 886–893.

5. **Gu, Q., Li, Z., & Han, J. (2011).** Generalized Fisher score for feature selection. *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pp. 266–273.

6. **Hao, X., Bao, Y., Guo, Y., Yu, M., Zhang, D., Risacher, S. L., Saykin, A. J., Yao, X., & Shen, L. (2020).** Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. *Medical Image Analysis*, Vol. 60, pp. 101–625.

7. **Huang, C. L. & Wang, C. J. (2006).** A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications*, Vol. 31, No. 2, pp. 231–240.

8. **Kennedy, J. & Eberhart, R. (1995).** Particle swarm optimization. *Proceedings of IEEE International Conference on Neural Networks IV*, Vol. 4, pp. 1942–1948.

9. **Kumar, A. & Pang, G. (2002).** Defect detection in textured materials using Gabor filters. *Industry Applications, IEEE Transactions on*, Vol. 38, pp. 425–440.

10. **Medjahed, S. A., Saadi, T. A., Benyettou, A., & Ouali, M. (2016).** Gray wolf optimizer for hyperspectral band selection. *Applied Soft Computing*, Vol. 40, pp. 178–186.

11. **Medjahed, S. A., Saadi, T. A., Benyettou, A., & Ouali, M. (2017).** Kernel-based learning and feature selection analysis for cancer diagnosis. *Applied Soft Computing Journal*, Vol. 51, No. February, pp. 39–48.

12. **Mhiriab, I. & Rekik, I. (2020).** Joint functional brain network atlas estimation and feature selection for neurological disorder diagnosis with application to autism. *Medical Image Analysis*, Vol. 60, pp. 101596.

13. **Mirjalili, S., Mirjalili, S. M., & Hatamlou, A. (2016).** Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Computing and Applications*, Vol. 27, No. 2, pp. 495–513.

14. **Mirjalili, S., Mirjalili, S. M., & Yang, X.-S. (2013).** Binary bat algorithm. *Neural Computing and Applications*, Vol. 25, No. 3-4, pp. 663–681.

15. **Mishra, S. & Mishra, D. (2015).** SVM-BT-RFE: An improved gene selection framework using Bayesian T-test embedded in support vector machine (recursive feature elimination) algorithm. *Karbala International Journal of Modern Science*, Vol. 1, No. 2, pp. 86–96.

16. **Rashedi, E., Nezamabadi, H., & Saryazdi, S. (2009).** GSA: A gravitational search algorithm. *Information Sciences*, Vol. 179, No. 13, pp. 2232–2248.

17. **Saeed, D., Ali, A., & Mohammad-Shahram, M. (2007).** Feature extraction using discrete cosine transform for face recognition. *9th International Symposium on Signal Processing and Its Applications, ISSPA*, pp. 1–4.

18. **Stromberg, W. D. & Farr, T. G. (1986).** A fourier-based textural feature extraction procedure. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-24, pp. 722–731.

19. **Vergara, J. R. & Estévez, P. A. (2014).** A review of feature selection methods based on mutual information. *Neural Computing and Applications*, Vol. 24, pp. 175–186.

20. **Wosiak, A. & Zakrzewska, D. (2018).** Integrating correlation-based feature selection and clustering for improved cardiovascular disease diagnosis. *Complexity*, pp. 1–11.