



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# CoV2-Detect-Net: Design of COVID-19 prediction model based on hybrid DE-PSO with SVM using chest X-ray images

Abhishek Dixit<sup>a</sup>, Ashish Mani<sup>b</sup>, Rohit Bansal<sup>c</sup>

<sup>a</sup> Department of Computer Science, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

<sup>b</sup> Department of EEE, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

<sup>c</sup> Department of Management Studies, Rajiv Gandhi Institute of Petroleum Technology, Rae Bareilly, Uttar Pradesh, India

## ARTICLE INFO

### Article history:

Received 29 May 2020

Received in revised form 21 March 2021

Accepted 29 March 2021

Available online 6 April 2021

### Keywords:

SARS-CoV-2

COVID-19

Differential evolution algorithm

Particle swarm optimization

Support Vector machine

## ABSTRACT

For Covid-19 suspected cases, it is critical to diagnose them accurately and rapidly so that they can be isolated and provided with required medical care. A self-learning automation model will be helpful to diagnose the COVID-19 suspected individual using chest X-rays. AI based designs, which utilizes chest X-rays, have been recently proposed for the detection of COVID-19. However, these approaches are either using non-public database or having a complex design. In this study we have proposed a novel framework for real time detection of coronavirus patients without manual intervention. In our framework, we have introduced a 3-step process in which initially K-means clustering, and feature extraction is performed as a data pre-processing step. In the second step, the selected features are optimized by a novel feature optimization approach based on hybrid differential evolution algorithm and particle swarm optimization. The optimized features are then feed forwarded to SVM classifier. Empirical results show that our proposed model is able to achieve 99.34% accuracy. This shows that our model is robust and sustainable in diagnosis of COVID-19 infected individual.

© 2021 Published by Elsevier Inc.

## 1. Introduction

In Dec-2019, coronavirus infection occurred in Wuhan, China and since then this virus has impacted millions of people across the world [1]. WHO on Jan-30-2020 declared a global health emergency and coronavirus as a pandemic [2]. Several cases from across the world which includes USA, Europe, Thailand, Australia were reported in the Feb-2020. This coronavirus is named as SARS-CoV-2, a new family of SARS-COV by International Committee on Taxonomy of Viruses [3].

This coronavirus disease transmits from human to human causing rapid growth in infected patients. Due to this COVID-19 infected patients had increased across all parts of the world and the patient infected from this virus suffers from acute respiratory syndrome that includes fever, cough, chest pain and several other respiratory problems [4,5]. However, there are cases reported in which people have shown no symptoms but were infected by COVID-19 [6]. To detect coronavirus, there is Reverse Transcription Polymerase Chain Reaction (RTPCR) test available. In these tests, specimens from lower and upper respiratory organs like nasal, sputum or nasal aspirate are collected from the person suspected to be infected by COVID-19 by health specialist. But this test has limitations like the sample collection process is manual and depends on cooperation of patient and skills of specialists, which may impact the accurate diagnosis of the virus. Also, this testing process is very time

E-mail addresses: [abhishekdxit@gmail.com](mailto:abhishekdxit@gmail.com) (A. Dixit), [amani@amity.edu](mailto:amani@amity.edu) (A. Mani), [rbansal@rgipt.ac.in](mailto:rbansal@rgipt.ac.in) (R. Bansal)

consuming and rate of detection is also very low. Due to these issues, repeated test are required to be performed for accurate diagnosis [7,8].

The present testing methodology has limitations of accuracy and processing time, therefore, there is a need for rapid detection of individuals suffering from COVID-19 so that infected patients can receive quick care and treatment. Studies have shown that radiological images like X-ray or CT images can be used for the detection of COVID-19 [9]. These researches have shown the benefits of using radiological images for detection of COVID-19. However, in such case, there will be need for health specialists to read these images to identify the patients. The diagnosis of disease through radiological images can be augmented with the help of machine learning models, which can help health specialist reduce subjectivity and time taken in diagnosis. Therefore, diagnosis through machine learning models of high accuracy is required so that the authorities feel encouraged to use these models in order to get the accurate and rapid results [10].

Many researchers have proposed various AI based models for rapid detection of COVID-19 infected patients by either using X-rays or CT images. Wang et al. [11] proposed a deep learning based AI diagnostic model that uses CT images and obtained accuracy of 79.3%. Another model proposed by Xu et al. [12] that classifies healthy, COVID-19, Influenza-A viral-based pneumonia cases using CT images and claims accuracy of 86.7%. These studies use non-public datasets for testing their AI based model. Wang et al. [13] proposed an approach where public dataset of SARS-CoV-2 X-ray images are used and achieved an accuracy of 92.4%. Few more research studies have been done for the detection of COVID-19 using X-ray images [14–16 17]. In all these studies pre-processed dataset is used. Another approach proposed by Ucar et al. [18] where deep learning model is designed with Bayes optimization and public dataset with pre augmentation is used to reduce the data imbalance. Ismael et al. [19] proposed a CNN model to detect the COVID-19 cases using chest X-ray images. Hussain et al. [20] proposed a deep learning model for COVID-19 detection using chest X-ray images.

The major contribution of this study is to develop a self-learning model for rapid and real time detection of coronavirus infected individuals. This model will work as an automated tool that can be used to assist medical professionals in improving the COVID-19 diagnosis accuracy. In this framework, X-ray images are pre-processed in which step in which input X-ray images are prepared into an acceptable size and format required by the model. This includes reading an image, resizing, and augmentation. Augmentation is done for imbalanced dataset where images are artificially created by creating different combinations of images by rotation, flipping, shear and adding noise. After pre-processing we have applied segmentation of images by clustering them into similar groups. We have used K-mean clustering approach which is more common and utilized for real world data sets [21]. In the next step feature extraction is performed on the clustered dataset collected from k-means clustering. It is the process of capturing the relevant informative data from an image and build the feature set-in coarse-grained manner. Feature extraction process will generate different features by describing the characteristics of the chest x-ray image processed. These features may contain the dominant and non-dominant characteristics of an image and these irrelevant and redundant features might impact the classification accuracy and computation time of the classifier. Therefore, there is a need for a feature selection approach which will only choose the relevant and optimized features for classification and this make the entire system more efficient. We have used a novel feature selection approach that uses differential evolution algorithm(DE) [22] hybridized with particle swarm optimization (PSO) [23]. The optimized feature set is used in support vector machine (SVM) classifier. Our aim is to achieve higher accuracy in diagnosis of coronavirus infected individuals using X-ray images in real time, so optimized features and SVM helps in achieving higher accuracy, even with limited size of data set. SARS-CoV-2 is a global pandemic which spreads rapidly from human to human and so data set from x-ray images with large number of features are available, but this creates difficulty in classifying the data using standalone classifier i.e. without proper features. Therefore, there is a need for feature optimization approach that could remove the irrelevant features and, also help in dimensionality reduction by filtering the less informative features, which would lead to improvement in the accuracy of classifier. The addition of feature selection process makes the model more robust and less complex during classification computation.

Recent studies show that the evolutionary approaches are amongst the popular techniques for feature selection as they are suitable for solving complex optimization problems [24]. Xue et al. [25] performed a detailed review on various evolutionary approaches used for feature selection. Recently PSO and its variant have been adopted for feature selection problems. Zhang et al. [26] proposed an approach based on multi objective PSO for optimization of cost-based feature selection. Xue et al. [27] proposed another variant of PSO for optimization of large feature set to improve the classification accuracy. Engelbrecht et al. [28] proposed a set based PSO variant for feature selection optimization. These solutions show the advantages of PSO in feature selection but still problem in finding optimal feature set occurs when data has large number of features. In recent years, differential evolution algorithm is also introduced for optimizing feature set specially in medical domain. Maza-heri et al. [29] proposed an automated tool for detecting abnormal cardiac functions. DE, PSO algorithms are used as a feature selection approach to achieve higher accuracy. Vivekananda et al. [30] proposed a variant of DE for optimization of feature selected for cardiovascular disease. Zhang et al. [31] proposed a variant of DE for solving the multi-objective feature selection problem. However, feature selection is still a problem specially with datasets having large number of features and very less work has been done for optimizing feature set obtained from X-ray images. In order to optimize the feature selection in our framework, we have introduced a novel approach by hybridizing Differential evolution algorithm with Particle swarm optimization. In this approach, PSO is used to enhance the convergence speed and DE has been used to improve the population diversity so that optimal feature set can be selected with precision. The procedure starts with  $DE/rand/2$  mutation strategy with archival approach and therefore each generation will be able to search in larger space and avoid the premature

convergence. Lately, PSO strategy is used in the evolution process to speed up the convergence. In addition to this we have also adopted the parameters tuning during the evolution process. The coordination of changing mutation strategy and parameters tracking and tuning during evolution process help in improving performance of our proposed approach. Further the optimized feature set is then used in SVM [32] classifier for accurately detecting the COVID-19 positive and negative cases.

SVM has recently received greater consideration in designing machine learning model as it possesses the generalization ability even with relatively smaller data sets, which is not the case with other types of classifiers. The objective of ANN model is to reduce the sum of square error among target label and therefore the performance of ANN is largely dependent on size and quality the training set. However, in SVM the margin among different classes is maximized and therefore its performance is not impacted by the rare training data set [33]. Also, SVM uses kernel function that helps SVM to classify the non-linear objects.

The proposed framework model is different from previous designs in three ways: First, most of the previous models are using non-public data set or with small set of public data set whereas our model is tested with relatively large public data set available [13]. Moreover, we have applied data augmentation in order to solve the class imbalance problem with dataset. Second difference is in classifier approach. All the previous model uses CNN classifier that uses supervised learning algorithm and therefore performance is dependent on training data. There is a need to artificially augment the data in order to develop the accurate model. Our framework is developed on three step process in which initially data pre-processing is done and data is clustered using K-means clustering [21] and then feature extraction is done. This K-means clustering is unsupervised learning algorithm and there is no need to supervise the model. K-means clustering subgroups the data based on the Euclidean distance without any prior training. It is efficient yet simple approach. This is an important step to understand the underlying patterns in an image by subdividing them into clusters as per the characteristics. Third difference is the addition of feature selection step. This is an important step as the processed data from pre-processing step have large feature set and it is important to optimize the feature set so that higher classification accuracy is achieved. To achieve that we are proposing a novel hybridization approach based on differential evolution algorithm and particle swarm optimization for feature selection. In this hybridization approach we have updated the DE mutation strategy with PSO and then in crossover approach the best particles are chosen with sigmoid function. This collaboration between mutation and crossover strategy provides optimized feature subset to SVM classifier. Utilizing this framework, we have achieved higher accuracy using chest X-ray images and detect individuals suffering from coronavirus in real time.

The proposed framework is tested on large public data set of chest X-ray images and compared with other state of art approaches. The propose approach is evaluated on seven parameters as number of features optimized, accuracy, correctness, completeness, F-score, training time and classification time. We have applied k-fold cross-validation method in order to ensure the reliability of our classifier and strengthen the simulation results. These results are compared with classical SVM classifier. We have also compared the results with DE-SVM and PSO-SVM. Performance of our proposed approach is also compared with eight recently proposed models as Wang et al. [13], Li et al. [14], Afshar et al. [15], Farooq et al. [16], Chowdhury et al. [17], Ucar et al. [18], Ismael et al. [19], Hussain et al. [20]. Experimental results show that our proposed model has superior performance than other approaches. We have also performed non-parametric Friedman's test for evaluating the statistical significance of superiority of our proposed model. The test shows that our proposed model achieves highest ranking among all other algorithms. Therefore, statistical test also confirms the superiority of our proposed model.

This paper is further organized into four sections. Preliminary approaches are discussed in Section 2; Proposed framework is discussed in Section 3. Section 4 showcase the experimental results and comparative analysis. Conclusions and future work are discussed in Section 5.

## 2. Related work

This section discusses the preliminary algorithms that are used in our framework. We have used Support vector machine (SVM), Differential evolution (DE) and Particle swarm optimization (PSO) algorithms.

### 2.1. Support vector Machine (SVM)

A supervised classification algorithm SVM [32] is better suited for solving classification problems. In SVM, the main objective is to look for a hyperplane in  $n$ dimensional space in order to classify the data set. A non-linear function is set as  $\theta(\beta)$  where  $\beta$  is the training set and  $\mu * \theta(\beta) + \beta = 0$  represents the classifier hyperplane. A decision function is defined as:  $f(\beta) = \text{sign}(\mu * \theta(\beta) + \sigma)$ . In this equation,  $\mu$  represents the orientation and  $\sigma$  represents the related displacement position. Equation for hyperplane classification can be represented as:

$$\min_{\mu, \sigma, \varepsilon} \frac{1}{2} \mu^T \mu + C \sum_{i=1}^n \varepsilon_i$$

$$\text{Such that } h_i(\mu^T * \theta(\beta_i) + \sigma) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, i = 1, 2, \dots, n \quad (1)$$

where  $h_i$  and  $\beta_i$  are the training label and training data respectively. For each  $i = 1, 2, \dots, n$  where  $n$  represents the training set size,  $\beta_i \in R^n$  and  $h_i \in \{+1, -1\}$ .  $c$  is capacity constant,  $\varepsilon_i$  is a non-negative variable also known as slack variable. This slack variable is an important attribute that demonstrate the violations of margin by sample  $\beta_i$ . The characteristics of linear and non-linear classifiers are almost same but with a difference of the Lagrange formulation in non-linear. So now the formulation of the Eq. (1) re-represented as:

$$\begin{aligned} \min & \frac{1}{\sigma} \sum_{i=1}^n \sum_{j=1}^n h_i h_j \sigma_i \sigma_j \gamma(\beta_i, \beta_j) - \sum_{i=1}^n \sigma_i, \\ \text{s.t.} & \sum_{j=1}^n \sigma_j h_j = 0, 0 \leq \sigma_i \leq c, i = 1, 2, \dots, n \end{aligned} \tag{2}$$

where  $\alpha_i$  is the support value, if  $0 \leq \sigma_i < c$ , then  $e_i$  is represented as the support vector (SV),  $\gamma(\beta_i, \beta_j) = \theta(\beta_i) * \theta(\beta_j)$  is the kernel function [34].

### 2.2. Differential evolution algorithm (DE)

DE [22] is a population-based metaheuristic approach which is stochastic in nature. DE has been applied in many domains for solving different optimization problems which are complex in nature like medical image processing [35,36], Feature selection [37,31]. DE begins by population initialization as  $X_i = \{x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{iD}\}$  where  $X_i$  is the candidate solution. This process is followed by mutation, crossover and selection operation as shown below:

#### 2.2.1. Mutation

Mutation strategy in DE named as "DE/ - /n" can be represented as:

$$V_{i,G+1} = X_{r1,G} + F \cdot (X_{r2,G} - X_{r3,G}) \tag{3}$$

Where  $F$  is the scaling factor,  $r_1, r_2$  and  $r_3$  are mutually exclusive random values in range  $[1, NP]$ .  $NP$  is the population size and number of generations is represented as  $G$ .

#### 2.2.2. Crossover

Crossover operation is performed in order to improve the population diversity for target vector. Crossover rate  $CR$  is applied in order to generate new trail vector as  $U_{i,G+1}$ . This operation is described as below.

$$U_{i,G+1} = \begin{cases} V_{i,G+1} & \text{if rand} \leq CR \\ X_{i,G+1} & \text{otherwise} \end{cases} \tag{4}$$

where crossover probability is represented as  $CR \in [0, 1]$ .  $rand$  is the random number and the value is in range  $[0,1]$ .

#### 2.2.3. Selection

In this operation, fitness value of each individual vector is compared, and best optimized values are selected based on greedy selection. Thus, the target vector is defined as.

$$X_{i,G+1} = \begin{cases} U_{i,G+1} & \text{iff}(U_{i,G+1}) \leq f(X_{i,G}) \\ X_{i,G} & \text{otherwise} \end{cases} \tag{5}$$

Where objective function  $f(X)$  with  $k = 1, 2, \dots, NP$ .  $NP$  is the population size. This process is repeated until stopping criteria is met.

### 2.3. Particle swarm optimization

Kennedy and Eberhart [23] in the year 1995 introduced a novel technique based on swarm behaviour termed as Particle Swarm Optimization (PSO). In this algorithm  $X_i = (x_{i1}, x_{i2}, x_{i3} \dots x_{iD})$  shows the position and  $v_i = v_{i1}, v_{i2}, \dots, v_{iD}$  is the velocity for the particle  $i$  in search space of dimension  $D$ . Every particle in search space maintains its local and global best position as  $l_{best}$  and  $g_{best}$  respectively. PSO update its position ( $X_i$ ) and velocity ( $v_i$ ) for each iteration.

Initially PSO was introduced for solving continuous optimization problems but later as there were multiple discrete optimization problems; like feature selection, PSO was augmented to binary PSO (BPSO) [38]. In BPSO position ( $X_i$ ), and global and local best ( $g_{best}$  and  $l_{best}$ ) can only take 0 and 1 values. Therefore, for BPSO the velocity calculation is updated as:

$$v_{i,G+1} = Iw * v_{i,G} + a_1 * b_1 * (l_{best(i,D)} - X_i) + a_2 * b_2 * (g_{best(i,D)} - X_i) \tag{6}$$

where  $a_1, a_2, b_1$  and  $b_2$  are acceleration constants. In a generation  $G$ , For Particle  $i$ ,  $v_i$  represents the velocity in dimension  $D$ .  $Iw$  represents the Inertia weight that specifies the effect of previous velocity.  $l_{best(i,D)}$  and  $g_{best(i,D)}$  represents the local and global best values.

### 3. The proposed algorithm

The proposed framework for detecting COVID-19 is discussed in this section. In this framework, data is pre-processed in which input images are prepared into an acceptable size and format required by the model. This process is feature extraction in which image features like shape, size etc. are extracted followed by feature selection process for feature optimization. Finally, the optimized feature subset is forwarded to classifier for detecting the COVID-19 impacted patient.

A novel Hybrid Differential evolution algorithm based on PSO with SVM is proposed to detect the coronavirus infected individuals by classifying their chest X-ray images. The architecture framework of our proposed model is shown in Fig. 1. This proposed model is trained end to end and Chest X-ray image are required as input to diagnose the individual as Covid-19 positive or negative in real time without any manual intervention. Our framework follows the 3-step process.

1. Data pre-processing step in which input X-ray image data is prepared into an acceptable size and format required by the model. This includes reading an image, resizing, and augmentation. In Section 3.3, we have discussed about the datasets and pre-processing approach used in this study. After the pre-processing image segmentation is required to identify the similar grouping of objects known as clustering. Clustering is a crucial step, as it is an important forerunner for the feature extraction. Clustering is performed on unsupervised learning having no class information. In this process data with similar attributes are grouped together. We have used K-means clustering technique as it is generally used in overlapping clusters which is common case in real datasets. The detailed discussion on clustering analysis is done in Section 3.4. Further, feature extraction process is applied to find out the characteristics of objects within an image. These characteristics or objects like shape and colour can be used to define an object. Section 3.1 discusses the feature extraction process used in this study.
2. Feature extraction process will generate different features by describing the characteristics of the chest x-ray image processed. These features may contain the dominant and non-dominant characteristics of an image and these irrelevant and redundant features might impact the classification accuracy and computation time of the classifier. Therefore, there is a need for a feature selection approach which will only choose the relevant and optimized features for classification and this make the entire system more efficient. We have used a novel approach for feature selection-based hybrid differential evolution algorithm with PSO for feature optimization.
3. Finally, Feature subset obtained from step 2 will be used by SVM classifier for classification.

The step 1 is a data pre-processing step. The training and testing of the proposed model for detecting COVID-19 has been performed using image dataset as mentioned in [13]. The images during training phase are clustered based on “Euclidian” distance using K-means clustering approach with  $k = 2$ . The pseudo code for K-means clustering used in our model is shown in Table 1.

#### 3.1. Feature extraction

In the next step feature extraction is performed on the segmented dataset collected from K-means clustering. Feature extraction is the process of capturing the relevant informative data from an image and build the feature set-in

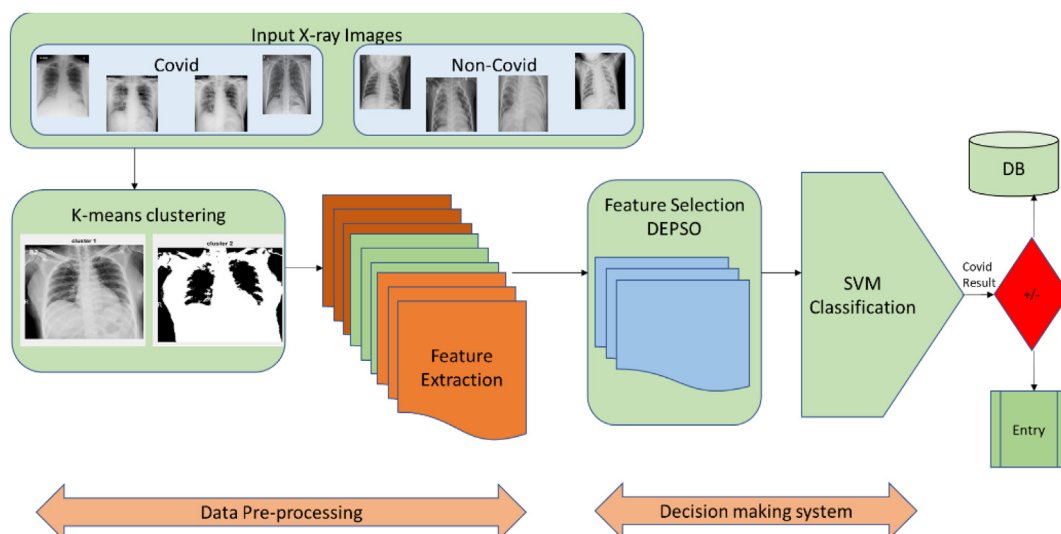


Fig. 1. Block Diagram CoV2-Detect-Net.

**Table 1**  
K-Means clustering.

1. Input: image data set for clustering and feature extraction
2. Select the number of clusters (k)
3. Select the centroids  $C_1, C_2, \dots, C_k$  randomly
4. Repeat
5. Calculate the Euclidean distance among each data and cluster.
6. Store clustered image into the region with minimum Euclidean distance.
7. Until coverage criteria

coarse-grained manner. It is like a screening process for identifying useful features. Feature extraction is useful for large dataset as it reduces the data dimensionality by obtaining the new features and remove repeated or redundant data. All the details around these feature types are stated in [39]. Table 2 showcase the procedure for feature extraction. In our model we have extracted following features.

### 3.1.1. Color features

Color features are most important features calculated from an image. Color features add more information and helps in getting content information from an image. There are various methods, but we are discussing few.

**Mean:** Mean value of image is calculated as:

$$mean = \left(\frac{1}{p \times q}\right) \sum_{x=0}^{p-1} \sum_{y=0}^{q-1} f(x, y) \quad (7)$$

**Standard Deviation:** SD is the calculation of inhomogeneity in an image.

$$SD = \sqrt{\left(\frac{1}{p \times q}\right) \sum_{x=0}^{p-1} \sum_{y=0}^{q-1} (f(x, y) - mean)^2} \quad (8)$$

### 3.1.2. Texture features

Texture features are also extracted and for that we have used a well-known grey level co-occurrence matrix (GLCM) [40]. Contrast, Correlation, Energy, Homogeneity etc. are the features extracted using GLCM approach.

**Energy:** Sum of square elements is captured in GLCM and is the uniformity in image regions. If the image pixel value is high the value of energy is also high.

**Table 2**  
Feature extraction of clustered images.

```

Input: load Clustered image set from K-means algorithm.
Calculate total length of data set.
for kk = 1: total_image
% check the size of image for grey/color image identification
[rows, columns, numberOfColorChannels] = size(image);
if numberOfColorChannels > 1
    grayImage = rgb2gray(image);
else
    grayImage = image;
end
image = double(image);
color_features.sum = sum(sum(sum(image)));
color_features.mode = mode(mode(mode(image)));
color_features.mean = mean(mean(mean(image)));
color_features.std = std(std(std(image)));
color_features.var = var(var(var(image)));
% calculate the texture feature by GLCM approach
GLCM2 = graycomatrix(grayImage,'Offset',[1 0;0 1]);
stats = graycoprops(GLCM2);
texture_features.contrast = stats.Contrast;
texture_features.correlation = stats.Correlation;
texture_features.energy = stats.Energy;
texture_features.homogeneity = stats.Homogeneity;
features = color_features; texture_features;
end
save features features

```

$$energy = \sum_{ij}^{n-1} (Q_{ij})^2 \tag{9}$$

**Entropy:** Entropy is the measure of randomness in an image.

$$entropy = \sum_{ij=0}^{n-1} -\ln(Q_{ij})Q_{ij} \tag{10}$$

**Contrast:** Measured for an image as the intensity measurement among pixels

$$contrast = \sum_{ij=0}^{n-1} Q_{ij}(i - j)^2 \tag{11}$$

**Correlation:** Measurement of correlation among the pixels.

$$correlation = \sum_{ij=0}^{n-1} Q_{ij} \frac{(i - u)(j - u)}{\alpha^2} \tag{12}$$

**Homogeneity:** calculate the similarity among pixels.

$$Homogeneity = \sum_{ij=0}^{n-1} \frac{Q_{ij}}{1 + (i - j)^2} \tag{13}$$

where,  $Q_{ij}$  =  $(i, j)$  the element in GLCM matrix,  $u = \sum_{i,j=0}^{n-1} iQ_{ij}$  is the mean of GLCM.  $n$  is the gray level count in an image.  $\alpha^2 = \sum_{i,j=0}^{n-1} Q_{ij}(i - u)^2$  is the calculation of variance of intensities of all pixels.

The next step is feature selection. In this step all the feature set collected after data pre-processing is optimized based on our hybrid differential evolution algorithm based on PSO. This is an important step as it reduces the complexity of the classifier by optimizing the feature, selecting only relevant features and thus improving the classifier's accuracy. The implementation process of our feature selection and classification is shown in Table 3. In this approach, we have used PSO for population initialization, then velocity and position of the particle is updated from its current position. In the next step, DE is applied for local search. This balances the exploitation and exploration capabilities of both the algorithms and thus improves the efficiency and prevents the problem of trapping into local optima. Further the optimized feature set are sent for classification using SVM classifier.

In our hybrid DE and PSO, classical DE with *DE/rand/2* mutation strategy is used and updated it with PSO. Further, the crossover strategy that uses a sigmoid function for comparing the best position. We have used this sigmoid function to accelerate the velocity by updating it within the range [0,1]. The position is updated as:

$$v_{i,G+1} = Iw * v_{i,G} + a_1 * b_1 * (l_{best(i,D)} - X_i) + a_2 * b_2 * (g_{best(i,D)} - X_i)$$

**Table 3**  
Feature selection and Classification with DEPSOSVM.

---

```

Input: dataset (samples, classes), Population (NP), Generation
Fmin and Fmax, CRmax and CRmin, a1, a2,
repeat
for g = 1: NP
//Calculate random integers such that r1≠r2≠r3
//For each particle i, Calculate the velocity as:
vi,G+1 = Iw * vi,G + a1 * b1 * (lbest(i,D) - Xi) + a2 * b2 * (gbest(i,D) - Xi)
//Evaluate the new position and select the best position as:
xi =  $\begin{cases} 1 & \text{if rand}() < f(v_{iD}) \\ 0 & \text{otherwise} \end{cases}$ ,
Where  $f(v_{iD}) = \frac{1}{1 + e^{-v_{iD}}}$ 
// apply DE mutation as:
DE/rand/2 Xi,G+1 = Xr1,G + F.(Xr2,G - Xr3,G) + F.(Xr4,G - Xr5,G)
// Calculate the new fitness value. f(γi) from above as:
γi = Xi,G+1 + vi,G+1
//Apply crossover as
Ci,G+1 =  $\begin{cases} \gamma_i & \text{if rand}(0, 1) \leq CR \\ X_i & \text{otherwise} \end{cases}$ 
// Pass the optimized feature set for Classification
MultisVM(samples, classes, temp) and compare with previous feature set. Choose best.
end
End

```

---



$$x_{n,i} = \begin{cases} 1 & \text{ifrand}() < f(v_{iD}) \\ 0 & \text{otherwise} \end{cases}, \text{ where } f(v_{iD}) = \frac{1}{1 + e^{-v_{iD}}} \tag{14}$$

where  $a_1, a_2, b_1$  and  $b_2$  are acceleration constants. In a generation  $G$ , For Particle  $i$ ,  $v_i$  represents the velocity in dimension  $D$ .  $Iw$  represents the Inertia weight that specifies the effect of previous velocity.  $l_{best(i,D)}$  and  $g_{best(i,D)}$  represents the local and global best values.  $f(v_{iD})$  is the sigmoid function.

As shown in Table 3, decision making system starts with hybrid DEPSO algorithm for feature selection. The algorithm starts with initialization of population where features set obtained from COVID-19 image data set of Chest X-rays are set. The hybrid DE-PSO algorithm is applied to optimize the feature set by removing the irrelevant features. A new feature threshold constraint is added to validate the redundancy relevance, duplicity and performance of fitness function. The fitness function is defined as.

$$F_{obj} = \alpha * \left( \frac{nnz(temp)}{lenth(temp)} \right) + (1 - \alpha) * (1 - perf) \tag{15}$$

where,  $\alpha$  is the threshold constraint in range  $[0, 1]$ . In the above equation  $\alpha$  is the relative importance of features and  $1 - \alpha$  is the relative importance of error rate in classification. Since number of features are always important than the classification accuracy, the value of  $\alpha$  should be  $\alpha \leq 1 - \alpha$ .

### 3.2. Computational complexity

In our proposed feature selection approach based on hybrid DEPSO, the computation complexity lies in the selection of mutation and crossover strategy using sigmoid function. Also, on the parameter self-adjusting strategy in each iteration, the computational complexity of selection strategy is represented as  $O((N \times D + N \times D + N) \times t_{max})$ . The self-adjusting strategy has the complexity of  $O(N \times M \times \log N \times t_{max})$  where,  $N$  is the population sizes,  $D$  is dimension and  $t_{max}$  is the maximum iteration and  $M$  is the count of objectives. Therefore, the computational complexity of our feature selection algorithm (DEPSO) is  $O((N \times D) \times t_{max}) + O(N \times M)$ .

As we have used SVM as a classifier and classifier’s accuracy is dependent on training data and feature set. Therefore, the complexity of SVM classifier during training is  $O(n^2f + n^3)$  and for prediction  $O(n_s f)$ , where  $n$  is the training sample,  $f$  is the number of features,  $n_s$  is the support vector count. This is due to the fact that during the training, classical SVM classifier needs to evaluate the kernel matrix  $M$  represented as  $M(x_i, y_j)$ .

### 3.3. Dataset details

Currently RT-PCR test is performed to detect the COVID-19 infected individual. Lower and upper respiratory specimens like nasal, sputum or nasal aspirate are collected from the person suspected to be infected by COVID-19 by health specialist. But this test has limitations like the sample might malfunction by mistake and this impact the accurate diagnosis. Also, the test process is very time consuming and rate of detection is also very low. Radiological images like chest X-ray images can be used as an alternative for detection of COVID-19 infected individual. Therefore, in this study chest X-ray images available as public dataset are used in our model. We have collected the COVID-19 chest X-ray images from [13] and [41]. Pneumonia infected patient’s X-ray images [42] are also used so that our model can be effectively prepared and tested. Total 13,975 X-ray images are available as public dataset. This data has 8066 normal, 5551 Pneumonia and 358 COVID-19 chest X-ray images. The sample X-ray images are shown in Fig. 2.

#### 3.3.1. Class imbalance strategies

Feature selection and classification from imbalance data impact the performance of any machine learning model. There are different strategies for handling the imbalance data [43]. This study showcases the impact of imbalance data on the per-

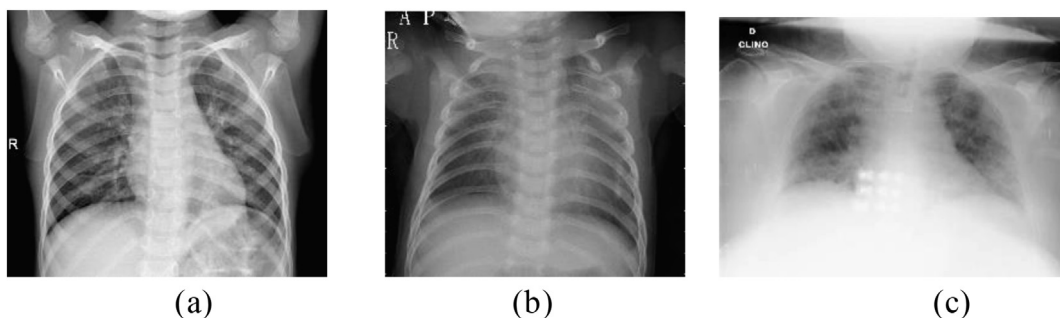


Fig. 2. Chest X-ray images (a) Normal (b) Pneumonia (C) COVID-19.

formance of classifier. Inspired from these studies we have conducted a systematic approach by resizing all the images to  $227 \times 227$  pixels. Next is to get the mirror image by flipping each image. In the next step performing shearing, adding noise, decreasing, and increasing the brightness by  $\pm 30$  pixels on both original and flipped image. Since we have limited Covid-19 x-ray images, we perform this augmentation process on Covid-19 x-ray images. Fig. 3 shows that sample operation performed, and different images obtained after augmentation process. By doing this, we obtain 24 variants of same image. After this pre-processing of image dataset, we get an acceptable number of Covid-19 x-ray images.

### 3.4. Cluster analysis

In an unsupervised learning, clustering plays a very important role. The aim of clustering is to find the interesting pattern in an unlabeled data and group them together. Various clustering techniques are suggested and proposed by researchers [44]. As there are no absolute criteria for choosing the best clustering technique available and the clustering technique mostly dependent of the problem. From the literature we see that there is a manual approach in which clusters are defined for the dataset. Other approach is known as Overlapping clustering, in this technique data is clustered on fuzzy sets based on some membership value. The other approach is hierarchical clustering technique. This algorithm starts by creating clusters of similar objects and finally each object is grouped into the distinct clusters and final clusters will have similar objects. For this study we have used 2 most widely used clustering techniques as K-means and expectation–maximization (EM) with Gaussian mixture model (GMM-EM) [45] clustering technique. Table 4 shows the accuracy comparison results of K-means and GMM-EM approach. The results show that K-means clustering is able to correctly classify 87.42% of features whereas GMM-EM is able to achieve 85.4532% accuracy. The results show that K-means clustering is showing better results in comparison to GMM-EM and therefore, we have opted k-mean algorithm as a clustering approach in our proposed model.

For calculation of “K” in K-means clustering we have used Elbow method. The basic idea in calculating the value of “K” using elbow method is to run the clustering algorithm on various values from the dataset and then calculate the sum of square error. This square error is the square of the distance between each value. “Euclidian distance” is used to calculate this distance. Every observation of the data values fit to nearest mean value cluster. All these centers choose the clusters as per the density and centroid. As shown in Fig. 4 for value of  $K = 2$  there is an “elbow point”, therefore we have used this value for clustering the image data. Image pre-processing is an important step for better clustering of images, we have applied Watershed transform of gradient magnitude as shown in Fig. 5. This helps in separating the touching objects from image.

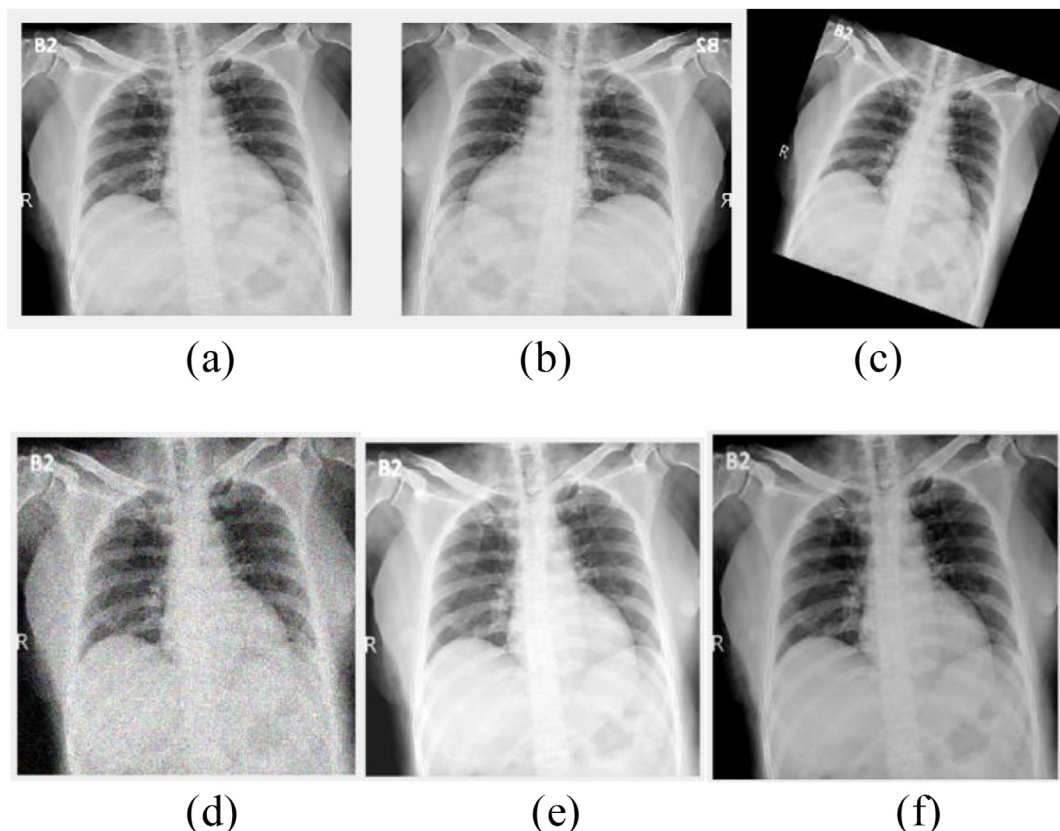
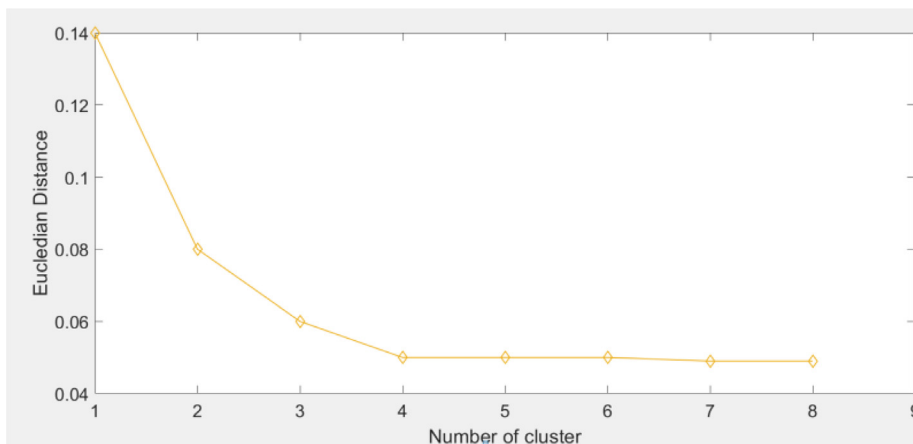


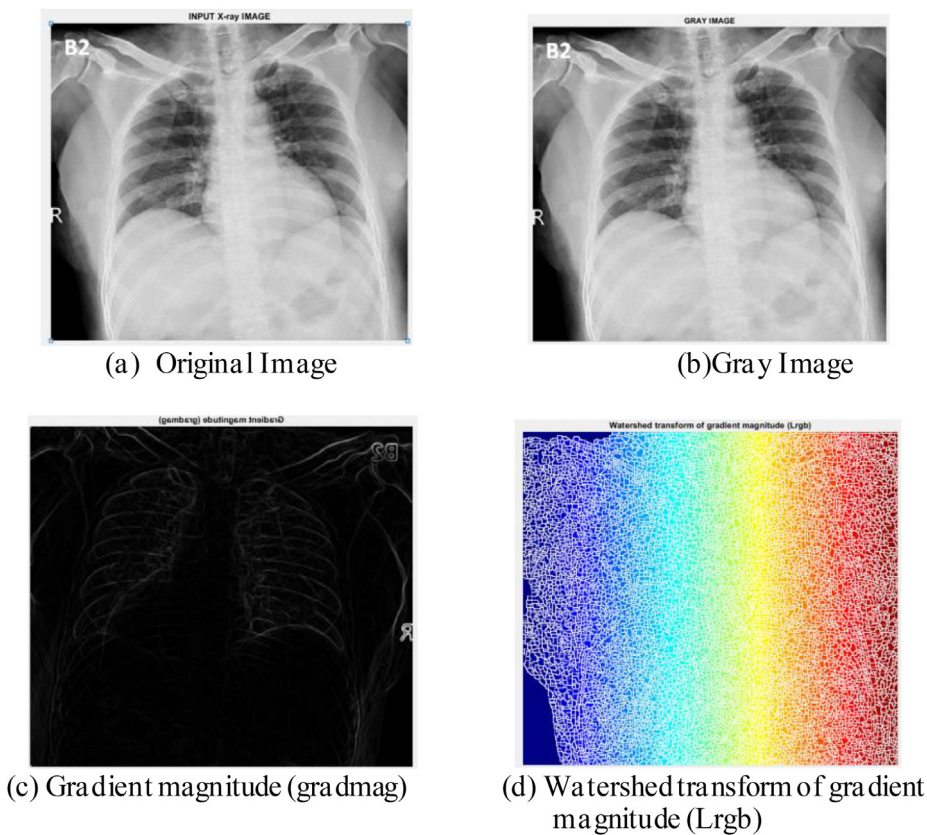
Fig. 3. Image augmentation: (a) Original Image (b) Flipped Image (c) Shear Image (d) Noise (e) increased brightness (f) decreased brightness.

**Table 4**  
Experimental results applying K-means.

Instance classification	K-means (%)	(GMM-EM) %
Correctly classified instances	87.4296	85.4532
Incorrectly classified instances	12.5704	14.5468



**Fig. 4.** Elbow curve of K-means clustering.



**Fig. 5.** Image pre-processing for K-means clustering.

### 3.5. Performance Evaluation:

We have applied k-fold cross-validation method [46] in order to ensure the reliability of our classifier and strengthen the simulation results. In this approach we have used 10-fold cross-validation by randomly dividing the 10 set of training and testing samples. Each combination is executed independently for calculating the accuracy and other parameters as mentioned below.

To measure the performance of our model, we have considered following parameters.

- Computation time: Total time spent in calculation.
- Classification accuracy:

$$C_{Acc} = \frac{TP + TN}{TP + FP + Tn + FN} \tag{16}$$

- $F_{score}$ :  $F_{score}$  is calculated to measure the test’s accuracy.

$$F_{score} = 2 \times CR \times \frac{CM}{CR + CM} \tag{17}$$

where  $TP, TN, FP, FN$  are the correctly classified class, opposite class classified, incorrectly classified and misclassified class.  $CR$  and  $CM$  are correctness and completeness.

- Friedman’s test: Non-parametric test for statistical analysis of the results.

### 3.6. Experimental results

For evaluation the performance of our proposed model, the experiment is performed using MATLAB with system configuration as 2.11 GHz, Intel<sup>®</sup> Core™ i7-8650U and 16 GB RAM. The parameters used for DEPSO algorithm is mentioned in Table 5. In the evolution stage setting up the right parameters is very important as performance is dependent on these control parameters, however choosing right parameters is a difficult process. In our approach values are adjusted in real time by monitoring the evolutionary process of each individual. During the process of evolution if any individual stagnates then the values need to be readjusted. In literature there is no guideline to set the values of control parameters, instead the random selection of values by trial and error is considered as the better option [47]. Since we have use DE and PSO algorithm in our model for feature selection and these algorithms are stochastic in nature, so accuracy is calculated by running our program over 100 iteration for 30 runs. To evaluate the performance of our proposed model we have also implemented SVM, DE with SVM (DE-SVM), PSO with SVM (PSO-SVM) classifier. We have also compared our proposed model’s efficiency with 8 recently proposed approaches (detailed mentioned in Table 6) in literature for COVID-19 detection.

In the preprocessing step data is split into 3 set as training, validation, and testing. This formation of split data is packaged as training data:80%, validation and testing data as 10% each. This validation data is required to better make the error prone objective function. Now with the separate test data set, all the images are resized to 227 × 227 pixel and the input data is reshuffled in order to reduce the overfitting’s negative impact. Table 5 shows the distribution of data and classes. All the images are pre-processed as in step 1 and 230 features are extracted from each image. These feature set are put for feature selection process using our hybrid DEPSO algorithm. Finally, the optimized feature set are sent to classifier for classification.

Experimental results shown in Table 7 showcase the proposed algorithm performance with different values of  $\alpha$  in fitness function. From the table we can see that in all the cases there is a reduction in features but with  $\alpha = 0.15$ , our proposed algorithm captures the smaller number of features with better classification accuracy as compared to other values. Our proposed approach obtains 99.35 of average accuracy in 40 runs and best accuracy of 100%. These results also show the impact of relevance and redundancy on the performance of classier. Therefore, it is important to choose the correct value of  $\alpha$ .

We have applied k-fold cross-validation method in order to ensure the reliability of our classifier and strengthen the simulation results. For our experiment we have used k as 10 and so 10 sets of equal sized training and test data samples are

**Table 5**  
Parameter settings.

S. No	Parameter Name	Value
1	Iteration	100
2	Default population	50
3	$lw$	0.9
5	$a_1, a_2$	2.05, 2.05
6	$F_{min}$ and $F_{max}$	0.5, 1.0
7	$CR_{max}$ and $CR_{min}$	1.0 and 0.8
8	$\alpha$	0.15

**Table 6**  
Dataset class.

No	Class label	Train	Validation	Test
	COVID	1229	154	153
	Pneumonia	1229	154	153
	Normal	1229	154	153
Total		3687	462	459

**Table 7**  
Proposed algorithm with different values of  $\alpha$ 

$\alpha$	Avg size	Avg accuracy	Best accuracy
All	234	78.5	
0.10	117	98.26	99.21
0.15	106	99.35	100
0.20	112	97.65	98.33
0.25	114	96.45	97.47

created. Further 9 sets are used for training and 1 is used for testing the model. We have repeated this process 10 times by applying data set on both training and testing stage. The average testing results are calculated with training and classification time and shown in Table 8. Training time is the time spent in training a classifier on 10-fold cross-validation. Classification time is the time spent in predicting the Covid-19 positive and negative cases. The results show that DESVM is able to optimize the feature set and selects 116 features, PSOSVM selects the 112 optimized feature set and our proposed algorithm is able to optimize maximum and select 106 feature set. These optimized feature results also show the impact on classifier's accuracy. We have also calculated the computational complexity of our algorithm. As we can see from the table the total time for training is 11 sec which is less as compared with DESVM and PSOSVM classifiers but more than SVM classifier. Similarly, the classification time complexity is also <1 s. The results show that our proposed hybrid approach working fast as compared to other evolutionary variants but slow in comparison to SVM classifier. This is due to feature selection step added in our framework and features subset selection from larger extracted feature set takes time but will results in fine tuning of classifier by reducing the complexity. Results also indicate that our feature selection approach is able to get more optimized feature set. This shows that adding the feature selection approach only improves the computational complexity of our classifier. These results show that our proposed algorithm is giving better performance as compared to other classifier in prediction of COVID-19 positive and negative cases.

We have also compared our proposed approach performance with 8 recently proposed models. Experimental results are shown in Table 9. From the results, we can see that the model proposed by Chowdhury et al. [17] is giving better results in terms of accuracy and correctness in comparison to other state of art algorithms. Almost similar performance is shown by

**Table 8**  
Result of classification using 10-fold cross validation method.

Evaluation parameters	SVM	DESVM	PSOSVM	DEPSOSVM
Number of Features	234	116	112	106
Accuracy	95.65	98.551	98.636	99.346
Correctness		0.985	1.000	1.000
Completeness		0.985	0.986	0.993
F-score		0.985	0.986	0.993
Training Time	2 sec	12 sec	8 sec	11 sec
Classification Time	<1 sec	<1 sec	<1 sec	<1 sec

**Table 9**  
Comparing efficiency results with other state of art algorithms.

Methodology	Accuracy (%)	Correctness (CR)	Completeness (CM)	F-score
Wang et al. [13]	92.3	0.913	0.0887	0.900
Li et al. [14]	88.9	–	–	–
Afshar et al. [15]	95.7	–	0.900	–
Farooq et al. [16]	96.2	0.969	0.969	0.969
Chowdhury et al. [17]	98.3	1.000	0.967	0.983
Ucar et al. [18]	98.3	0.983	0.983	0.983
Ismael et al. [19]	90.53	88.00	93.33	90.72
Hussain et al. [20]	94.2	94.2	96.2	91.3
DEPSOSVM (Proposed Method)	<b>99.346</b>	<b>1.000</b>	<b>0.993</b>	<b>0.993</b>

the diagnosis model proposed by Ucar et al. [18]. If we compare both these models with our proposed approach, the proposed model with DEPSOSVM is showing better results with 99.34% of accuracy as compared with all the other approaches. These results show our proposed model is giving more accurate results as compared to other models.

Experimental results show that our proposed model is showing superior performance than other approaches. There are two reasons for the same. In our framework the initial step is to pre-process the images by applying K-means unsupervised learning algorithm which is best suited for creating a self-learning model. The second reason is the feature selection step. We have added a novel feature selection approach in which differential evolution algorithm is hybridized with PSO. This hybridization helps swarm to acquire population diversity and avoiding local optima convergence. This is due to our mutation strategy that helps algorithm to maintain the population diversity and to improve the local and global search ability by utilizing DE mutation strategy at initial and PSO strategy at the later stage. DE helps in adding population diversity and PSO enhances the convergence and accuracy. Following this process, optimized feature set are collected which are then put into SVM classifier to gaining better classification accuracy particularly on the larger population and complex solution set. In our approach, DE has been used to maintain the constant parameter value for every evolution, but individual parameter can dynamically get the optimized performance in each iteration. Every dataset is different and complex so need a dynamic strategy to choose the parameter for each iteration and this will improve the overall efficiency of the algorithm. Therefore, in our approach, parameters are self-adjusted based of the status. For example, if any individual parameter becomes stationary in consecutive generations, our model will readjust them by random values as:

$$F_{i,G+1} = \begin{cases} F_{i,G} & \text{if } ES_i < ES_{max} \\ F_{min} + (F_{max} - F_{min}) * rand(1, D) & \text{otherwise} \end{cases} \quad (18)$$

$$CR_{i,G+1} = \begin{cases} CR_{i,G} & \text{if } ES_i < ES_{max} \\ CR_{min} + (CR_{max} - CR_{min}) * rand(1, D) & \text{otherwise} \end{cases} \quad (19)$$

where  $F_{i,G}$ ,  $F_{min}$  and  $F_{max}$  are the scaling factor and min and max constant values of scaling factor respectively.  $CR_{i,G}$ ,  $CR_{max}$  and  $CR_{min}$  represents the crossover rate, max and min constant values of crossover rate for  $X_{i,G+1}$  in generation  $G$ .  $rand(1, D)$  is the random values in range  $[1, D]$  and  $D$  is the size (samples, 2). The min and max values for scaling factor are 0.8 and 1.0 and crossover rate are 0.8 and 1.0. The impact of these control parameters is crucial and therefore the values are chosen by trial and error.

For handling the population diversity and premature convergence problem, we have used a novel mutation operation to produce child generation followed by crossover with particles. These generated offspring have their position updated. Throughout the iteration these new offspring update their own best position in comparison to other generated offspring.

As shown in Fig. 6, we have created the confusion matrix of our proposed framework in testing. In the figure, 1 represent the COVID data, 2 for the normal and 3 is for the Pneumonia data. Our proposed model is able to detect all the COVID-19 infected chest X-ray images of the patients. Out of 459 samples, there is only 3 cases where misclassification happened. If we check for the error rate, it is 0.007% for normal x-ray images and 0.01% for pneumonia cases. This shows the robustness and sustainability of our model. We have also captured the results using box plot as shown in Fig. 7 for good understanding.

We have also performed non-parametric Friedman’s test for evaluating the statistical performance of our proposed model. The null hypothesis is set as: For the same dataset, results  $Q$  is tested for the provided  $p$  value ( $Z_{Q-1}^2 > T$ ), where  $T$  is test statistics [48]. For testing performance metrics is applied on accuracy, fitness and  $F_{score}$ . Test statistics results and ranks for various approaches are shown in Tables 10 and 11 respectively. For this experiment, confidence level ( $\alpha$ ) is taken as 0.5. From Table 10 we can observe the difference between various approaches are significant and so this null hypothesis is rejected. Ranking order of various approaches are given in Table 11. Lower the value, higher is the ranking and so our proposed model achieves highest ranking among all other algorithms. Therefore, statistical results also confirm the superiority of our proposed model.

Friedman’s non-parametric test results shows that there is a significant difference between the comparative algorithm, but we are not able to establish the interrelation between the algorithms in comparison. García et al. [49] and Derrac et al. [50] has suggested a post-hoc analysis for  $N \times N$  comparison of multiple algorithms. Their studies suggested a non-parametric test should be conducted to determine whether there is any significant difference among the algorithms in comparison. Once this is established that there is a significant difference among the algorithms then post-hoc analysis is performed to detect which pair of algorithms are significantly different.

Post-hoc test are usually performed to control the type-1 error also known as family wise error rate (FWER) [51]. In a pairwise multiple comparison tests, FWER refers to the probability of rejecting at least one null hypothesis. With the Friedman test, the Bonferroni test is usually a suggested approach which adjust the  $p$ -values and controls the FWER [52]. But this approach is very conservative and there are modifications made to the original process [53,54]. We have used Holm’s test to determine which of these algorithms are different in  $N \times N$  comparison. This Holm’s test is also known as modified Bonferroni process and works on sequentially rejective step down approach for calculating the adjusted  $p$ -value [53]. This post-hoc analysis is based on the significant value of  $\alpha$ . For this experiment we have used  $\alpha = 0.05$ .

We have applied post-hoc analysis to further strengthen the statistical analysis and to showcase that there is a significant difference between the proposed algorithm and rest of the algorithms in comparison. We have used Bonferroni-Holm pro-

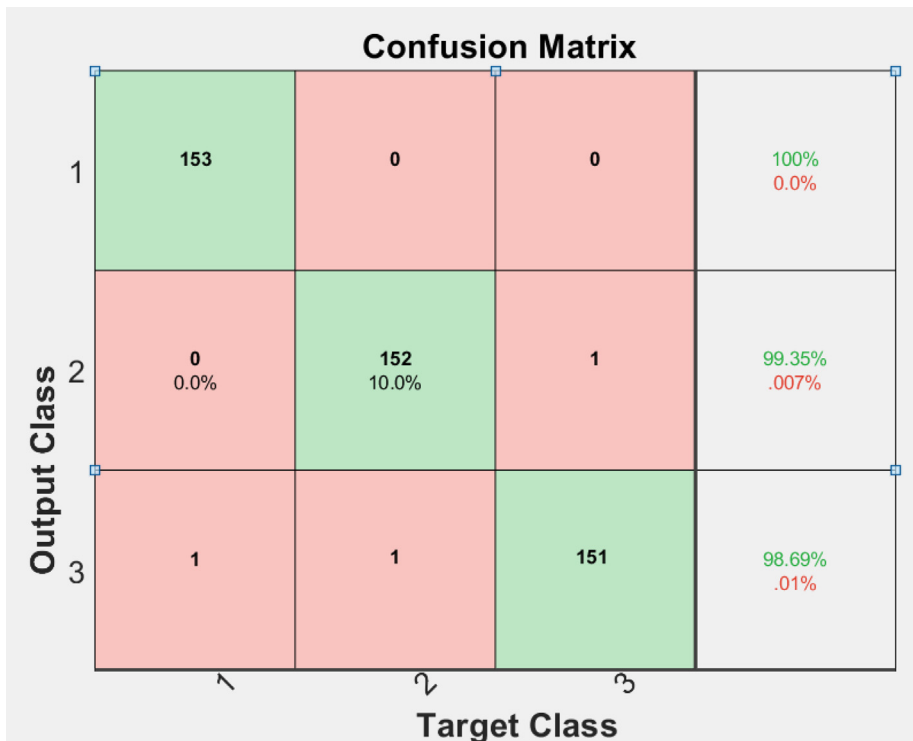


Fig. 6. Confusion Matrix of our proposed Framework.

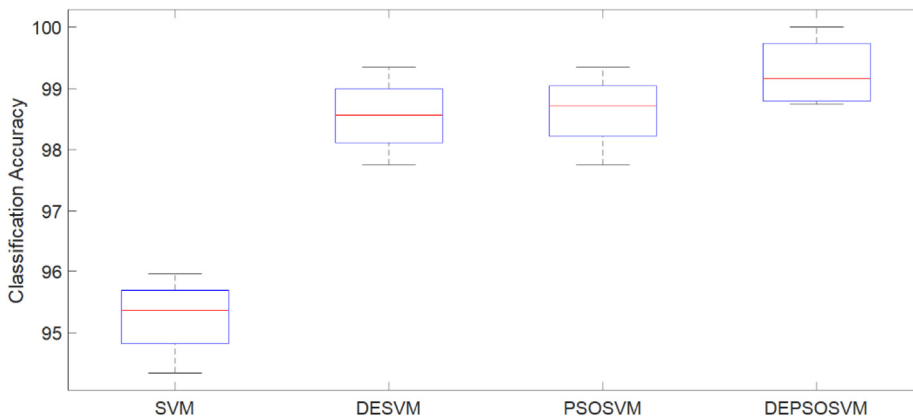


Fig. 7. Box plot of SVM and DESVM, PSOSVM and DEPSOSVM classification accuracies.

**Table 10**  
Friedman's test data.

Method	Chi sq	Asymptotic significance	Hypothesis
Accuracy	46.035	3.2097E-06	Rejected
Fitness	59.70	2.7683E-08	Rejected
F Score	39.72	4.0034E-05	Rejected

cess for this analysis. Table 12 shows the results from post-hoc analysis that includes Z-score, p-value (unadjusted), Holm's coefficient adjustment and adjusted p-value. The results are compared with proposed model and other 4 algorithms. From the results it is evident that proposed model shows the significant difference in performance in comparison to other algorithms. Therefore, post-hoc analysis results also confirms the superiority of our model.

**Table 11**  
Different algorithms rank.

Methodologies	Rank values
SVM	4.4
DESVM	2.4
PSOSVM	1.9
DEPSOSVM	1.4

**Table 12**  
Post-hoc analysis.

Evaluation Criteria	DEPSOSVM vs	Z-score	p-value	Coefficient adjustment of Holm	Adjusted p -value
Accuracy	SVM	2.6458	0.008150	3	0.024449
	PSOSVM	2.4568	0.014018	2	0.028036
	DESVM	2.2678	0.023341	1	0.023341
Fitness	SVM	4.1576	0.000032	4	0.000388
	PSOSVM	2.6458	0.008150	3	0.032598
	DESVM	2.4568	0.014018	1	0.028036
F-Score	SVM	2.3434	0.019109	2	0.038218
	PSOSVM	2.2678	0.023341	1	0.023341
	DESVM	2.6458	0.008150	4	0.032598

#### 4. Conclusion

SARS-CoV-2 also known as COVID-19 is a global pandemic. This virus transmits from human to human and spreads very rapidly. Therefore, there is a need for rapid detection of individuals suffering from COVID-19 so that infected patients should receive quick care and treatment. The current testing approach is full of limitations and require manual intervention from health specialist. Therefore, an alternative approach is required which should be quick, real time in response, giving accurate results and minimal manual intervention should be required. In this study, we have proposed a new framework which is a self-learning model, giving accurate results without manual intervention.

This model can be used for real time detection of individuals suffering from COVID-19. This framework works on 3 step process in which 1st step is data pre-processing. In this step, chest X-ray image is first clustered by using unsupervised K-means algorithm and then features are extracted. In the 2nd step, feature selection is performed using a novel differential evolution algorithm based on PSO. The large feature set obtained after feature extraction are optimized by removing irrelevant features. The optimized feature set is then put into SVM classifier to fine tune the classification process and get the better accuracy in terms of identification of COVID-19 positive and negative cases.

The proposed framework is tested on large public data set of chest X-ray images and compared with other state of art approaches. Our approach is performing superior as compared to other approaches and giving better accuracy, correctness, completeness and  $F_{score}$ . We have also compared the computation time with SVM and with classical DE and PSO algorithm. Our proposed approach is giving better optimized feature set in less time as compared with classical DE and PSO. We have also performed Friedman's non-parametric test for statistical analysis and our model is performing better as compared to other approaches. This shows that our model is robust and sustainable learning model and has a great potential to be used as an automation tool for detecting the COVID-19 infected individuals in real time. The proposed model can be used public places like shopping mall, Metro, school/colleges, Airports, Railways entry points to detect the individuals infected from COVID-19.

In future we want to extend this work and test our model on larger set of data and evaluate other evolutionary approaches.

#### CRedit authorship contribution statement

**Abhishek Dixit:** Conceptualization, Methodology, Software, Writing - original draft. **Ashish Mani:** Visualization, Investigation, Software, Validation, Writing - review & editing. **Rohit Bansal:** Supervision, Validation, Writing - review & editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



## References

- [1] WHO, "Coronavirus disease 2019 info web site –<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>," 2020.
- [2] WHO, "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)," [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)), 2020.
- [3] T. T.-Y. Lam, M. H.-H. Shum, H.-C. Zhu, Y.-G. Tong, X.-B. Ni, Y.-S. Liao, W. Wei, W. Y.-M. Cheung, W.-J. Li, L.-F. Li, G. M. Leung, E. C. Holmes, Y.-L. Hu and Y. Guan, "Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China," *medRxiv and bioRxiv*, 2020.
- [4] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, P.-R. Hsueh, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges, *Int. J. Antimicrob. Agents* 55 (3) (2020) 105924.
- [5] C. Lin, Z. Chen, B. Xie, Z. Sun, Y. Ding, X. Li, M. Niu, S. Guo, J. Lei, COVID-19 pneumonia patient without clear epidemiological history outside Wuhan: An analysis of the radiographic and clinical features, *Clin. Imaging* 65 (2020) 82–84.
- [6] C. Lin, Y. Ding, B. Xie, Z. Sun, X. Li, Z. Chen, M. Niu, Asymptomatic novel coronavirus pneumonia patient outside Wuhan: The value of CT images in the course of the disease, *Clin. Imaging* 63 (2020) 7–9.
- [7] R. Liu, H. Han, F. Liu, Z. Lv, K. Wu, Y. Liu, Y. Feng, C. Zhu, Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020, *Clin. Chim. Acta* 505 (2020) 172–175.
- [8] C. DKW1, P. Y. C. SMS, H. KPY, K. P. L. Y. N. DYM, W. CKK, Y. P. W. Q. P. M and P. LLM, "Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia," *Clinical Chemistry*, vol. 66, no. 4, p. 549–555, 2020.
- [9] D. Wang, B. Hu, C. Hu, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China, *JAMA Network* 323 (11) (2020) 1061–1069.
- [10] K. DS, G. M, C. W, L. H. Valentim CCS and e. a. Baxter SL, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [11] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng and B. Xu, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *medRxiv and bioRxiv*, 2020.
- [12] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Y. Chen, J. Su, G. Lang, Y. Li, H. Zhao, K. Xu, L. Ruan and W. Wu, "Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia," *arXiv:2002.09334v1*, 2020.
- [13] L. Wang and A. Wong, "Images, COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray," *arXiv:2003.09871*, 2020.
- [14] X. Li, C. Li and D. Zhu, "COVID-Xpert: An AI Powered Population Screening of COVID-19 Cases Using Chest Radiography Images," *arXiv:2004.03042*, 2020.
- [15] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Platanotis and A. Mohammadi, "COVID-CAPS: A Capsule Network-based Framework for Identification of COVID-19 cases from X-ray Images," *arXiv:2004.02696*, 2020.
- [16] M. Farooq and A. Hafeez, "COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs," *arXiv:2003.14395*, 2020.
- [17] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi and M. B. I. Reaz, "Can AI help in screening Viral and COVID-19 pneumonia?," *arXiv:2003.13145*, 2020.
- [18] F. Ucar, D. Korkmaz, COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the, *Med. Hypotheses* 140 (2020) (2020) 109761.
- [19] A. M. Ismael and A. k. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert Systems with Applications*, vol. 164, p. 114054, 2021.
- [20] E. Hussain, M. Hasan, M.A. Rahman, I. Lee, CoroDet: A Deep Learning Based Classification for COVID-19 Detection using Chest X-ray Images, *Chaos, Solitons Fractals* (2020) 110495.
- [21] Anil K. Jain, Data clustering: 50 years beyond K-means, *Pattern Recogn. Lett.* 31 (8) (2010) 651–666.
- [22] R. Storn, K. Price, Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, *J. Global Optim.* 11 (4) (1997) 341–359.
- [23] J. Kennedy, R. Eberhart, Particle swarm optimization, in *IEEE international conference on neural networks*, 1995.
- [24] Y. Zhou, W. Zhang, J. Kang, X. Zhang, X. Wang, A problem-specific non-dominated sorting genetic algorithm for supervised feature selection, *Inf. Sci.* 547 (2021) 841–859.
- [25] B. Xue, M. Zhang, W.N. Browne, X. Yao, A survey on evolutionary computation approaches to feature selection, *IEEE Trans. Evol. Comput.* 20 (4) (2016) 606–626.
- [26] Y. Zhang, D.-w. Gong, J. Cheng, Multi-Objective Particle Swarm Optimization Approach for Cost-Based Feature Selection in Classification, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 14 (1) (2017) 64–75.
- [27] Y. Xue, B. Xue and M. Zhang, "Self-Adaptive Particle Swarm Optimization for Large-Scale Feature Selection in Classification," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 5, 2019.
- [28] A.P. Engelbrecht, J. Grobler, J. Langeveld, Set based particle swarm optimization for the feature selection problem, *Eng. Appl. Artif. Intell.* 85 (2019) 324–336.
- [29] V. Mazaheri, H. Khodadadi, Heart arrhythmia diagnosis based on the combination of morphological, frequency and nonlinear features of ECG signals and metaheuristic feature selection algorithm, *Expert Syst. Appl.* 161 (2020) 113697.
- [30] T. Vivekanandan, N. Ch Sriman Narayana Iyengar, Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease, *Comput. Biol. Med.* 90 (2017) 125–136.
- [31] Y. Zhang, D.-w. Gong, X.-z. Gao, T. Tian, X.-y. Sun, Binary differential evolution with self-learning for multi-objective feature selection, *Inf. Sci.* 507 (2020) 67–85.
- [32] Vladimir N. Vapnik (Ed.), *The Nature of Statistical Learning Theory*, Springer New York, New York, NY, 1995.
- [33] Abe and Shigeo, *Support Vector Machines for Pattern Classification*, Springer, New York, 2005.
- [34] C.J. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, *Data Min. Knowl. Disc.* 2 (2) (1998) 121–167.
- [35] O. Tarkhaneh, H. Shen, An adaptive differential evolution algorithm to optimal multi-level thresholding for MRI brain image segmentation, *Expert Syst. Appl.* 138 (2019) 112820.
- [36] S. Mitra, B. Uma Shankar, Medical image analysis for cancer management in natural computing framework, *Inf. Sci.* 306 (2015) 111–131.
- [37] M. B. Subramanya, J. Virmani and Kriti, "A DEFS Based System for Differential Diagnosis Between Severe Fatty Liver and Cirrhotic Liver Using Ultrasound Images," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*, Academic Press, 2019, pp. 53–72.
- [38] J. Kennedy, R. Eberhart, "A discrete binary version of the particle swarm algorithm," in *IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation*, Orlando, FL, USA, USA, 1997.
- [39] G.-H. Liu, J.-Y. Yang, Content-based image retrieval using color difference histogram, *Pattern Recogn.* 46 (1) (2013) 188–198.
- [40] H. M and V. R. S, "GLCM Architecture For Image Extraction," *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol. 3, no. 1, pp. 75–82, 2014.
- [41] J. P. Cohen and L. D. Paul Morrison, "COVID-19 Image Data Collection," *arXiv:2003.11597v1*, 2020.
- [42] D. Kermany, K. Zhang and M. Goldbaum, "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images," *Mendeley Data*, V2, 2018.
- [43] A. de Haro-García, G. Cerruela-García, N. García-Pedrajas, Ensembles of feature selectors for dealing with class-imbalanced datasets: A proposal and comparative study, *Inf. Sci.* 540 (2020) 89–116.
- [44] M.C. Thrun, Q. Stier, Fundamental clustering algorithms suite, *SoftwareX* 13 (2020) 100642.

- [45] S. Kermani, N. Samadzadehghdam, M.E. Tavakol, Automatic color segmentation of breast infrared images using a Gaussian mixture model, *Optik* 126 (21) (2015) 3288–3294.
- [46] P. Pławiak, Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals, *Swarm Evol. Comput.* 39 (2018) 192–208.
- [47] N. Nasimul, B. Danushka, I. Hitoshi, “An adaptive differential evolution algorithm,” in *IEEE Transaction on, Evol. Comput.* (2006).
- [48] D.G. Pereira, A. Afonso, F.M. Medeiros, Overview of Friedman’s Test and Post-hoc Analysis, *Commun. Stat. – Simul. Comput.* 44 (10) (2015) 2636–2653.
- [49] S. García, A. Fernández, J. Luengo, F. Herrera, Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power, *Inf. Sci.* 180 (10) (2010) 2044–2064.
- [50] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, *Swarm Evol. Comput.* 1 (1) (2011) 3–18.
- [51] J.H. Zar, *Biostatistical Analysis*, 5th Edition., Northern Illinois University, Pearson, 1999.
- [52] O.J. Dunn, Multiple Comparisons Among Means, *J. Am. Stat. Assoc.* 56 (293) (1961) 52–64.
- [53] S. Holm, A Simple Sequentially Rejective Multiple Test Procedure, *Scand. J. Stat.* 6 (2) (1979) 65–70.
- [54] Y. Hochberg, A sharper Bonferroni procedure for multiple tests of significance, *Biometrika* 75 (4) (1988) 800–802.