



Multi-objective Genetic Algorithm Based Deep Learning Model for Automated COVID-19 Detection Using Medical Image Data

S. Bansal¹ · M. Singh¹ · R. K. Dubey² · B. K. Panigrahi³

Received: 8 March 2021 / Accepted: 26 August 2021 / Published online: 1 September 2021
© Taiwanese Society of Biomedical Engineering 2021

Abstract

Purpose In early 2020, the world is amid a significant pandemic due to the novel coronavirus disease outbreak, commonly called the COVID-19. Coronavirus is a lung infection disease caused by the Severe Acute Respiratory Syndrome Coronavirus 2 virus (SARS-CoV-2). Because of its high transmission rate, it is crucial to detect cases as soon as possible to effectively control the spread of this pandemic and treat patients in the early stages. RT-PCR-based kits are the current standard kits used for COVID-19 diagnosis, but these tests take much time despite their high precision. A faster automated diagnostic tool is required for the effective screening of COVID-19.

Methods In this study, a new semi-supervised feature learning technique is proposed to screen COVID-19 patients using chest CT scans. The model proposed in this study uses a three-step architecture, consisting of a convolutional autoencoder based unsupervised feature extractor, a multi-objective genetic algorithm (MOGA) based feature selector, and a Bagging Ensemble of support vector machines based binary classifier. The proposed architecture has been designed to provide precise and robust diagnostics for binary classification (COVID vs.nonCOVID). A dataset of 1252 COVID-19 CT scan images, collected from 60 patients, has been used to train and evaluate the model.

Results The best performing classifier within 127 ms per image achieved an accuracy of 98.79%, the precision of 98.47%, area under curve of 0.998, and an F1 score of 98.85% on 497 test images. The proposed model outperforms the current state of the art COVID-19 diagnostic techniques in terms of speed and accuracy.

Conclusion The experimental results prove the superiority of the proposed methodology in comparison to existing methods. The study also comprehensively compares various feature selection techniques and highlights the importance of feature selection in medical image data problems.

Keywords Coronavirus (COVID-19) · Convolutional autoencoder · Multi-objective genetic algorithm · Feature subset selection

✉ R. K. Dubey
RahulKumar.Dubey@in.bosch.com

S. Bansal
Shrey.Bansal.cs118@cse.iitd.ac.in

M. Singh
Mukul.Singh.cs118@cse.iitd.ac.in

B. K. Panigrahi
bkpanigrahi@ee.iitd.ac.in

¹ Computer Science and Engineering Department, Indian Institute of Technology Delhi, New Delhi 110016, India

² Robert Bosch Engineering and Business Solutions Private Limited Head Office, 123, Hosur Rd, 7th Block, Koramangala, Bengaluru, Karnataka 560095, India

³ Electrical Engineering Department, Indian Institute of Technology Delhi, New Delhi 110016, India

1 Introduction

A chest infection disease affects the functioning of the lungs [1]. The common lung infections are lung cancer, Chronic Obstructive Pulmonary Disease (COPD), bronchitis, pneumonia and, asthma. Coronavirus disease (COVID-19) is a lung infection disease caused due to the novel discovered virus known as SARS-CoV-2 [2]. COVID-19 began with reports of unknown causes of pneumonia in Wuhan City, China, around December 2019. The worldwide economy was impacted by the unprecedented rise in COVID-19 cases and it has been declared a pandemic by the World Health Organization [3].

On 18 June 2020, a total of 8,379,081 patients became infected with COVID-19, and 215 countries listed 450,101

deaths [3]. The standard diagnostic test for COVID-19 is the Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) [4]. Due to PCR's high selectivity and sensitivity, it is prevalent. The limitations of the PCR technique are (1) time consuming, (2) expensive, (3) shortage of kits, and (4) long production time [5]. A faster and cheaper testing mechanism is required to tackle the alarming rates of spread of COVID-19. Radiological analysis like Chest CT (computed tomography) scans and X-Rays produce high hit-rate in COVID-19 diagnosis. Authors in [6] established a high correlation between radiological results and RT-PCR. The above reasons encouraged developing a cheaper and faster COVID-19 screening mechanism using a radiological approach [7].

From the comprehensive analysis of the COVID-19 diagnosis field, it is inferred that the best alternative for COVID-19 detection to the RT-PCR test kits is chest radiography (X-rays and CT scan) [8]. However, CT scan modality seems to be more efficient than chest X-ray for the following reasons: (1) X-rays provide only a 2D perspective whereas CT scan provides a detailed 3D view of the organ, (2) in X-rays, ribs overlap the lungs and heart, whereas, the CT scan does not. A deep-learning-based three-step model is proposed for CT-scan based screening, consisting of a convolutional autoencoder (CAE) based unsupervised feature extractor, an

evolutionary algorithm based feature subset selector, and a feature classifier.

A CNN-based dense autoencoder has been used as the feature extractor because of CNN's high representational power and the generality of unsupervised learning from it. The Autoencoder ensures an accurate and diverse feature set, while the feature selector removes all redundant and irrelevant features improving the performance. After obtaining a reduced representation of raw data as a diverse set of features, the evolutionary algorithm based feature subset selectors is used to select optimal feature subsets. Finally, the bagging ensemble of support vector machines (SVM) is trained on the subsets chosen by the various selectors, and their performance is compared.

2 Related Works

Table 1 consists of various state of the art techniques currently available in the literature of COVID-19 diagnosis. Further, a detailed analysis of the review is presented.

Works from [10, 11, 13, 14] have used pre-trained CNN models for COVID-19 diagnosis. Transfer Learning techniques are useful when data is limited, but they often fail to learn intricate features unique to the required dataset. Some

Table 1 Related work results analysis on COVID-19 screening

References	Technique	Key findings
[9]	infection Size Aware Random Forest method (iSARF)	The accuracy of 87.9%, a sensitivity of 90.7%, and a specificity of 83.3% are achieved on chest CT scan
[10]	ResNet-18 (CNN model)	The performance parameters are: specificity: 92.2%, sensitivity: 98.2% and, AUC: 0.996
[11]	Pre-trained CheXNet and DenseNet	An accuracy of 90.5%, a sensitivity of 100% is achieved using 5323 (COVID19-115, normal - 1341, and pneumonia-3867) Chest X-ray images
[12]	Joint Classification and Segmentation (JCS)	Used a dataset of 400 COVID-19 patients (144,167 images) and 350 Non-COVID patients. The model achieves a dice score of 78.3%, sensitivity of 95% and a specificity of 93% for the segmentation task
[13]	Domain Exten. Transfer Learning (DETL) with Gradient Class Activation Map (Grad-CAM)	The Data A - Binary classes disease (13 diseases) and normal. Data B - Four classes (normal, pneumonia, other diseases, and Covid19). An accuracy of 95.3% using X-ray scans
[14]	AlexNet, VGG16, VGG19, GoogleNet, and ResNet50	Pre-trained models used to train CNN on 742 chest CT scans for two binary classes (COVID and non-COVID). The highest accuracy of 82.91% is achieved with the ResNet50 pre-trained CNN model
[15]	3-Dimensional deep learning	The specificity of 92.2%, a sensitivity of 98.2%, and AUC of 0.996 is achieved by the 3-D CNN model
[16]	Detail-Oriented Capsule Nets + Peekaboo (patch crop and drop strategy)	A recall of 91.5%, accuracy of 87.6%, precision of 84.3%, and AUC of 96.1 is achieved on chest CT scan dataset for classification to binary classes(COVID-19 and Non-COVID)
[17]	Multi-Objective Differential Evolution (MODE) deep learning	The performance parameters of MODE outperforms by 1.927% of Kappa statistics, 1.68% of specificity, 1.82% of sensitivity and, 2.09% of F-measure in comparison to authentic CNN models

authors have performed fine-tuning, but retraining the last few layers might not change the basic features extracted by the CNN.

Authors in [9, 12, 15, 16] have used random forest, peeka-boo, and segmentation classification. They have not used explicit feature extractors, and since the classification uses chest CT images, a deep feature extractor architecture like CNN might perform significantly better in this case.

The authors in the literature have obtained quality results by focusing only on feature extractors and classifiers. In our work, we propose to shift the attention from feature extraction to feature selection as it is critical to remove the redundant features in an unsupervised extractor and improve the performance of any standard classifier.

The author in [17] obtained improved results using MO-DE [18] feature selector over Deep CNN models, thus showcasing the importance of proper feature selection technique in medical image classification. We extend their work further and try to analyze and compare various feature reduction and selection techniques ranging from linear dimensionality reduction (principal component analysis-PCA) to various multi-objective feature selectors. We obtained state-of-the-art results, validating their results, and obtaining an improved, robust model for COVID-19 screening.

Further, authors in [19] have found genetic selectors to outperform standard results on the Flavia dataset. Authors in [20] use a Nondominated Sorting Genetic Algorithm II (NSGA-II) based MOGA for feature selection and evaluate its performance on various datasets. Authors in [21] show the use of the GA based feature selector for network intrusion detection. Authors in [22] compare GA based feature selectors on medical datasets focusing on diagnostic radiology. Authors of [22] compare GA based feature selectors to other approaches. In the stated studies, optimization of internal parameters of the MOGA has not been explored. Further, there is no comparative analysis among MOGA and other multi-objective evolutionary techniques for feature selection on medical images. Multi-Objective Optimization using Evolutionary Algorithms has not been well explored in its use as a feature selector.

We try to improve upon the previous works by analyzing the effects of optimizing parameters of MOGA. We also studied and compared MOGA with other multi-objective evolutionary techniques for feature selection on COVID-19 CT Scan Image Dataset, not done previously by any works.

3 Theoretical Background

3.1 Autoencoder (AE) Based Feature Extractor

Autoencoders [23] are unsupervised learning methods trained to reconstruct their inputs, usually by going through

a compressed representation of lower dimensionality [24]. Structurally an AE comprises two parts, namely an Encoder and a Decoder. Figure 1 summarizes the structure of an AE.

The encoder (E) converts the input image (x) to an encoded representation (h), which reflects the features of the image due to the constraint to reduce dimensionality. An encoder deterministically maps its input to a reduced representation generally using an affine map:

$$h = E(W \cdot x + b) \quad (1)$$

here W denotes the weights for the encoder part, b represents the bias, and h represents the reduced representation. Similarly, the decoder (D) takes the reduced representation (h) and outputs the reconstructed image (y). An Autoencoder is trained to minimize the reconstruction error of its input. Hence, training of AE can be seen as a minimization of the following cost function:

$$Cost = \frac{1}{N} \sum_j \mathcal{L}_{oss}[x_i, y_i] \quad (2)$$

where N represents the number of images, x_i and y_i represent the i th input-output image pair, and \mathcal{L}_{oss} is the reconstruction error between two images. Mean squared error has been used as the reconstruction error. CAE combines convolutional operations with the architecture of an AE. The authors of [25] have shown that CAE shows high accuracy in finger vein identification. Since CNN can extract a very detailed set of feature maps from images, convolutional AE has been used as a feature extractor in this study.

3.2 Multi Objective Genetic Algorithm Based Feature Selector

3.2.1 Multi Objective Genetic Algorithm (MOGA)

Multi-Objective Optimization is the process of simultaneously optimizing more than one competing objective function. Two Objectives have been considered in this work, namely, classification accuracy and size of feature subset. These are competing objectives, and a single solution optimizing both might not exist. An alternative is to generate a set known as the Pareto Optimal set of solutions. A Pareto

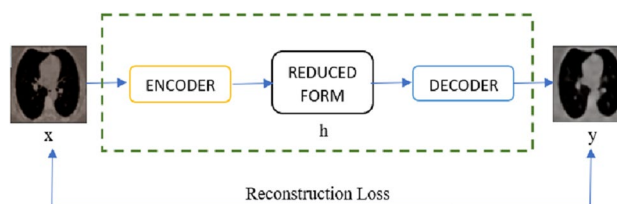


Fig. 1 Schema of basic autoencoder

set is a set of solutions where no solution is dominated by any other solution in the set. There is always a degradation in some objectives, required to improve any objective in a Pareto set of solutions.

Consider a set of M objectives that have to be minimized. $\mathcal{H} = \{h_1, h_2, \dots, h_M\}$. Consider $x_1, x_2 \in \{\text{Pareto Set}\}$, then x_1 dominates x_2 if:

$$\begin{aligned} \forall i = 1, 2, \dots, M, \{h_i(x_1) \leq h_i(x_2)\} \text{ and} \\ \exists i = 1, 2, \dots, M, \{h_i(x_1) < h_i(x_2)\} \end{aligned} \tag{3}$$

A solution is said to be Pareto Optimal if there exists no solution which dominates it. All such Pareto Optimal solutions together form the Pareto Optimal Set.

There exist various algorithms for multi-objective Genetic Optimizations. NSGA-II [26] is one such elitist principle-based algorithm much superior to classic gradient-based approaches. NSGA-II has been used to carry out the multi-objective feature subset selection in this study. Figure 2 summarizes the implementation of NSGA-II.

3.2.2 Initial Population and Encoding

Solutions in the population (a.k.a chromosomes) are represented as binary strings. The i th gene in a chromosome is one if the solution contains the i th feature of the input set. For the initial population, random binary chromosomes have been generated.

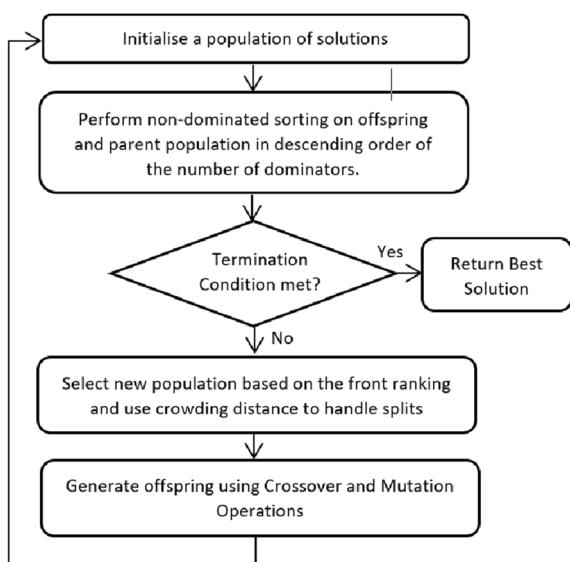


Fig. 2 The working of NSGA-II

3.2.3 Crossover and Mutation

The creation of two new offspring chromosomes using the selected parent pair is known as crossover. Single point crossover has been used in this work, where each gene is randomly selected from one of the parents. Parents are selected using tournament-based selection.

Mutation conserves population diversity. Mutation involves random modifications in the value of the chromosomes. Random bit flip has been used as the mutation operator in this study.

3.2.4 Termination

The MOGA based selector terminates when either the maximum number of generations or the stall generation limit has been reached. After termination, the selector returns the final population with objective scores and front rankings.

3.3 Ensemble SVM Based Classification

The SVM ensemble with Bagging is used in classification as SVM is a weak learner [27]. Using many small classifiers can increase robustness and produce low error. Bagging [28], uses randomized training sets for creating different models. A single classifier’s training set is randomly generated by drawing N random data points (N is the size of the original training set) from the original training set with replacement. Figure 3 illustrates the structure of the bagging ensemble-based SVM.

As described above, bootstrap builds K duplicate training datasets from the given training data set (TR)

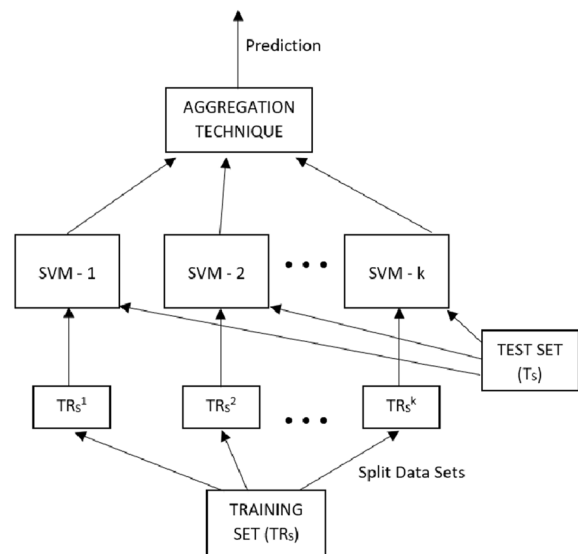


Fig. 3 A general architecture of SVM ensemble with an aggregation step

$\{TR_k | k = 1, 2, \dots, K\}$ using random re-sampling with replacement.

After training, the independently trained SVMs are aggregated. Thus, majority voting has been used in the study because it uses upper layer SVM to combine several lower layer SVMs (double layer hierarchical combining).

4 Proposed Method

A 3-step architecture is proposed for the screening of COVID-19 chest CT scans. The proposed architecture consists of a feature extractor, a feature selector, and a classifier. Flowchart summarizing the proposed architecture is depicted in Fig. 4

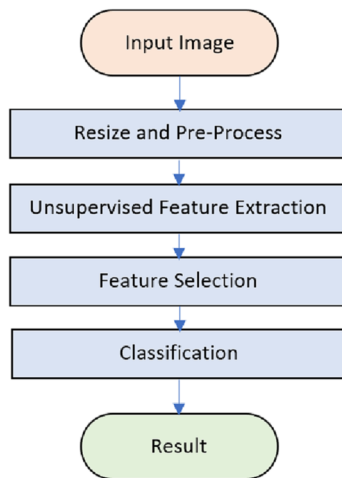


Fig. 4 Flowchart summarizing the proposed architecture

An autoencoder based unsupervised learning approach is used to generate features from the CT scan images automatically. This gives us a diverse feature set, essential for this classification.

Though diverse, the features extracted by the Autoencoder have very high dimensionality and suffer from a redundancy of features. To remove the extra features, a MOGA based feature selector is proposed to select an optimal set of features.

Finally, for classification, a bagging based ensemble of support vector machines is used to carry out the binary classification of the feature sets into COVID-19 and non-COVID classes. A brief outline of the various methods is highlighted in the subsequent study.

4.1 Auto Encoder Structure and Training

The input image of size $128 \times 128 \times 3$, is fed into the CNN, which contains convolutional layers (kernel size 3) and max-pooling layers (downscaling factor of 2). ReLu activation is applied after every convolution. The encoder layers have 32, 16, and 8 filters (output channels), respectively. A decoder follows the encoder to reconstruct the image using deconvolution and up-sampling layers. The output of the encoder has the shape $14 \times 14 \times 16$. This is flattened to generate a feature vector of length 2048 per CT Scan image. The CNN architecture has been summarized in Fig. 5.

The Auto Encoder is trained using the training set with the validation set for validation, as explained in Sect. 4.1. Adam optimizer has been used for training the AE, with Mean Squared Error (MSE) as the loss function. The AE has been trained for two hundred epochs with a batch size of 10 per epoch. Figure 6 shows a reconstruction of test set images by AE.

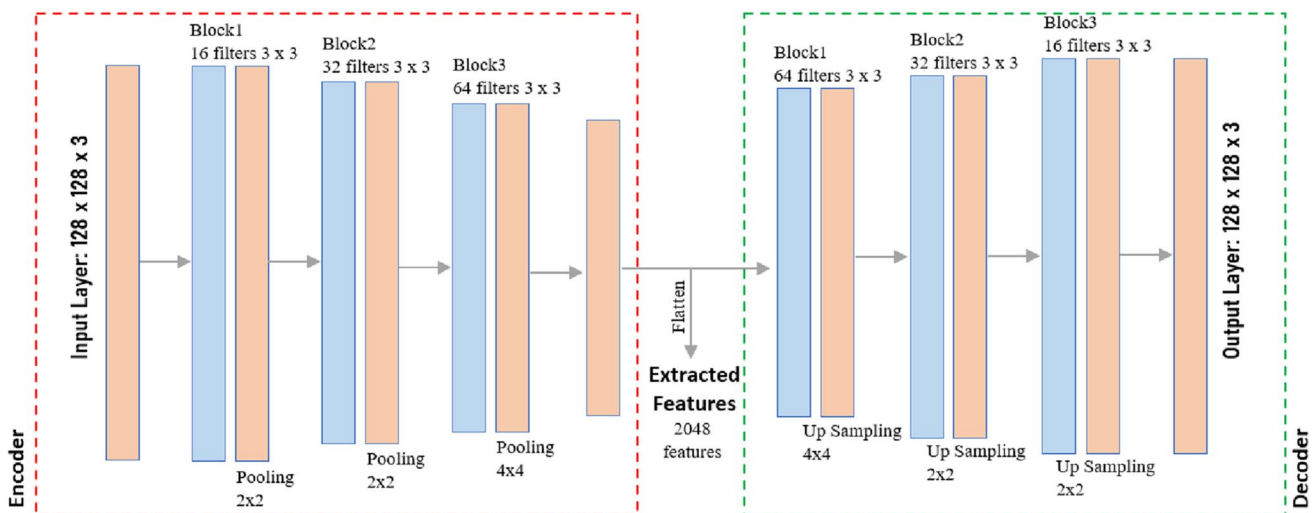


Fig. 5 Proposed architecture of Convolution Auto-encoder

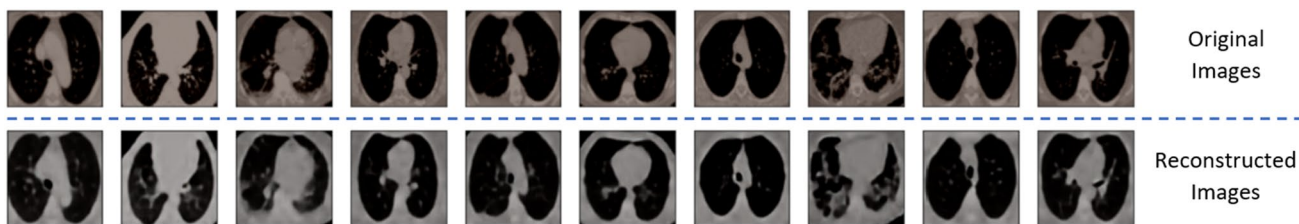


Fig. 6 Original and reconstructed (from Convolution Autoencoder) chest CT images from validation set

4.2 Feature Selector

The feature extractor extracts 2048 features from an input image of $128 \times 128 \times 3$. MOGA has been applied for selecting a superlative set from the extracted features using two fitness criteria:

$$C_1 = \frac{1}{S} \quad C_2 = Accuracy(\mathcal{F})$$

where S is the cardinality of \mathcal{F} and \mathcal{F} is the subset of features selected, and Accuracy is classification accuracy on the test set. Reducing the number of features ensures that there are no redundant or irrelevant features in the dataset. Classification accuracy is measured on the test set using an SVM.

Instead of constant Crossover and mutation rates, linear crossover and mutation rates have been used in this study. This ensures a high initial mutation rate preventing premature convergence and a low mutation rate when MOGA is close to the Pareto front. Similarly, the crossover rate is initially low to maintain diversity and gradually increases. Figure 8 shows the plot of the crossover and mutation rates against generations for the MOGA.

The summary of GA Parameters is given in Table 3. For evaluation, an average of 100 runs has been considered. The run summary of the MOGA based selector showing the min., max., avg., and std. dev. of the number features and highest accuracy for the given generation (using SVM as a classifier) is shown in Table 2. The plot of highest accuracy vs. No of features selected by MOGA is shown in Fig. 7

4.3 Feature Classifier

An ensemble of support vector machines (SVM) is used to classify the selected features. The bagging technique is used to construct the SVM ensemble. For classification, the dataset is randomly divided into ten parts, and the individual SVMs are trained independently (bootstrap techniques). These individual models are then aggregated by the deterministic averaging process to make a

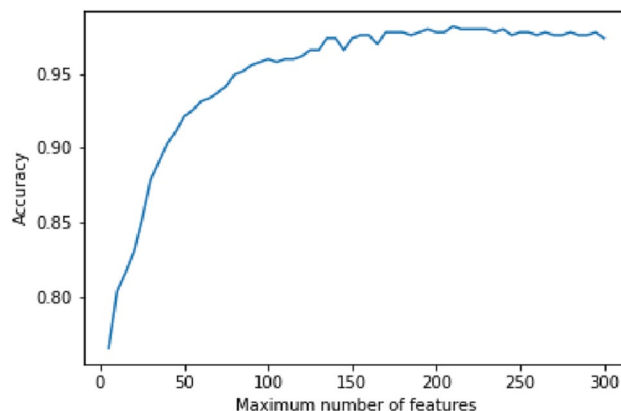


Fig. 7 The plot of highest accuracy vs the maximum number of features)

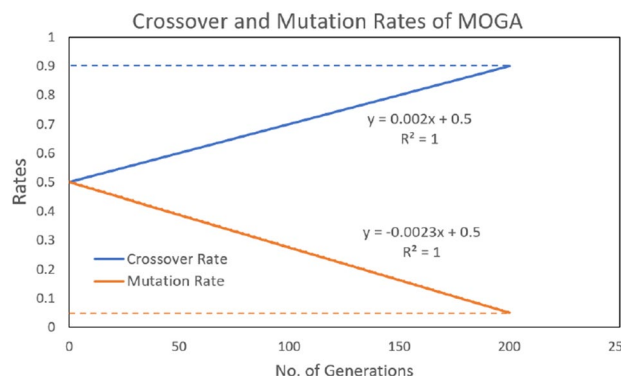


Fig. 8 Crossover and mutation rates of MOGA vs generations plot

joint decision. Each SVM has an RBF kernel with C and Gamma tuned values using the Genetic Algorithm-based Hyperparameter Optimizer. The classifier’s performance, evaluated using the test set, and the number of features is stated in Table 8.

Table 2 Summary of the run for MOGA selector

Generations	Number of evaluations	Number of features				Accuracy
		Min	Max	Avg	Std. dev.	
10	147	265	308	287.45	7.71	0.65
20	126	263	305	284.3	6.21	0.79
30	142	259	302	279.64	7.34	0.83
40	112	257	299	276.34	5.66	0.86
50	139	253	294	273.3	7.27	0.88
60	154	245	291	268.18	5.83	0.89
70	128	242	292	264.82	4.59	0.91
80	146	239	291	261.74	6.69	0.92
90	135	234	287	258.84	5.52	0.93
100	128	235	285	256.4	6.34	0.94
110	117	231	281	250.24	5.21	0.94
120	115	228	279	248.32	5.91	0.95
130	128	224	274	245.96	6.47	0.97
140	108	227	275	247.32	8.45	0.96
150	113	226	275	246.33	7.93	0.97
160	123	226	271	243.9	5.24	0.98
170	105	224	272	242.14	4.35	0.97
180	101	224	271	241.66	6.01	0.98
190	107	221	269	240.72	6.81	0.98
200	97	224	273	242.35	4.46	0.99

Table 3 Summary of the hyperparameters used for MOGA

Parameter	Value
Population size	200
No. of generations	200
Mutation probability	Linear increase from 0.5 to 0.9
Crossover probability	Linear decrease from 0.5 to 0.1
Selection procedure	Tournament based
Mating pool size	50

Table 4 The brief details of the dataset for CT scans

Set of images	Training set	Validation set	Test set	Total
COVID	756	236	260	1252
non-COVID	732	261	237	1230
Total	1488	497	497	2482

5 Data and Validation

5.1 Dataset

The images of CT Scans used in this study are taken from the public database of COVID-19 CT Scans by the name of “SARS-CoV-2 Ct-Scan Dataset” published and maintained by Soares et al. [29]. The dataset consists of 2482 images of chest CT Scans, out of which 1252 are from patients infected with COVID-19. The remaining 1230 images are from patients of other non-COVID pulmonary Diseases. The presence of other non-COVID respiratory diseases allows the model to learn COVID specific features.

The patients considered in the compilation of the dataset mentioned above are from various hospitals in Sao Paulo, Brazil. The COVID-19 CT Scan images are

collected from 60 patients (32 males and 28 females). The non-COVID CT Scan images were also collected from 60 patients (30 males and 30 females).

The dataset has been split into three sets, namely training (0.6), validation (0.2), and testing (0.2). The splitting is random, and an average of 5 splits is stated for all evaluations. The summary of the dataset after splitting is stated in Table 4.

5.2 Evaluation Metrics

The screening performance of the model was assessed by accuracy (ACC), precision (PRE), area under ROC curve (AUC), recall/sensitivity (REC), and F1 score (F1). Precision is the number of true positives over total positive predictions. Recall is defined as the number of true positives over the number of correct classifications. F1 score is simply the harmonic mean of precision and sensitivity of the model. AUC is the total area contained under a ROC Curve, and it shows the usefulness of tests on the model.

5.3 Experimentation

5.3.1 Comparison of AE Depths

Depth of any Neural Network directly affects its performance, and an optimal depth ensures an accurate and robust model. The reconstruction Structural Similarity Index (SSIM) and Mean Squared Error (MSE) has been used to compare various autoencoders. Three different autoencoders have been considered for this with 2, 3, and 4 convolution layers, respectively, in the encoder. The exact structure of the autoencoders is given below:

- 2-Layers: two convolution layers of kernel 3x3 with 32 and 64 filters, respectively. Each layer is followed by a max-pooling layer of 2×2 .
- 3-Layers (proposed) : three convolution layers of kernel 3×3 with 16, 32, and 64 filters, respectively. Each layer is followed by a max-pooling layer of 2×2 .
- 4-Layers: four convolution layers of kernel 3x3 with 8, 16, 32, and 64 filters, respectively. Each layer is followed by a max-pooling layer of 2×2 .

The analysis is summarized in Table 5. The AE has been trained on the train set and tested on the validation set for this analysis. The size of images used is 128x128, and the pixel values have been scaled to lie between 0 and 1.

5.3.2 Effect of Bagging Estimators on Performance

Bagging ensemble uses several estimators instead of a single estimator for prediction. This improves performance since a single estimator may have high test error, but it is overcome by using many small estimators. A different number of estimators are compared based on their accuracy on the validation set, and the box plot of the accuracy vs. the number of estimators is shown in Fig. 9. It can be seen that the accuracy improves till 20 estimators, then it saturates.

5.3.3 Comparing Different Population Sizes

Optimal population size is obtained by applying the proposed MOGA based selector on the validation set. For obtaining the accuracy values, multiple runs were conducted,

Table 5 Comparison of AE depth on the basis of reconstruction

Auto-encoder	SSIM score	MSE score
2-Layer	0.76	0.0014
3-Layer	0.91	0.0004
4-Layer	0.89	0.0007

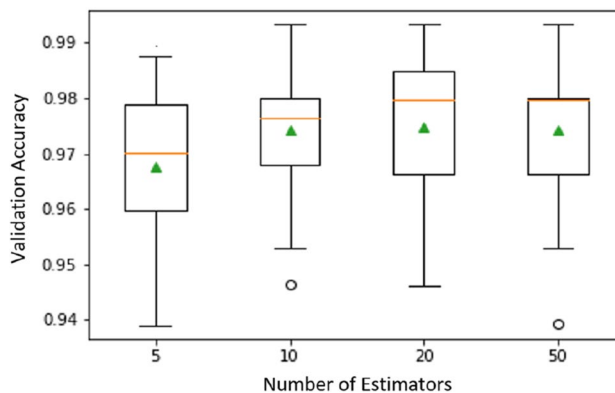


Fig. 9 Box plot of validation set accuracy vs number of bagging estimators

and an average of these was recorded. The graphs show the accuracy against the population size of MOGA, which is varied between 50 and 300 in increments of 50. The plot shows that the performance improves up to size 200, after which it stabilizes. Figure 10 shows the plot of the accuracy vs. population size.

5.3.4 Comparing Generation Size of MOGA

Improvement of Pareto fronts with generation is studied in this section. The fronts are plotted using 5 points from each generation, with the parameters for MOGA being as stated in Table 3. Y-axis represents the selected subset’s accuracy on the validation set, while the X-axis represents the inverse of the number of features selected. It can be seen that the fronts improve till 150 generations and then the front stabilizes. This is also observed in the overlap of the fronts in generation 150 and 200. Figure 11 shows the generation wise Pareto fronts.

5.3.5 Comparison with Simple GA, PCA and No-Selector

This section compares the MOGA based selector, PCA, and Simple GA. Accuracy on the validation set is taken as the comparison metric. PCA, a popular dimensionality reduction technique, is applied with a variance set to 0.95. Simple GA tries to find the optimal feature set using validation accuracy as the fitness function. Direct classification with all the extracted features without any feature selection has also been performed. The results obtained are summarized in Table 6. Directly using the features without selection results in poor performance of the model. The proposed model outperforms all the techniques in terms of accuracy. In terms of the number of features, it can be seen that MOGA selects considerably fewer features than simple GA.

Fig. 10 The plot of population size vs validation accuracy for MOGA selector

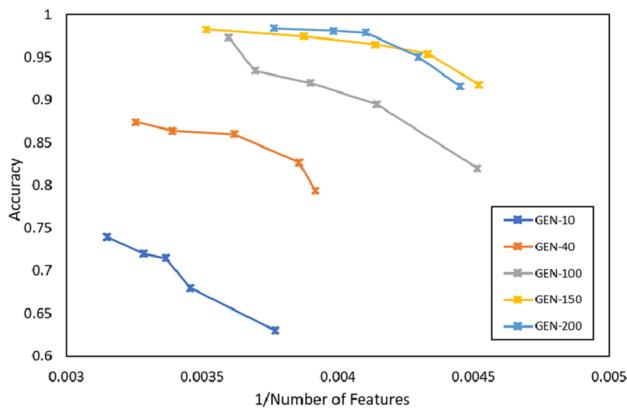
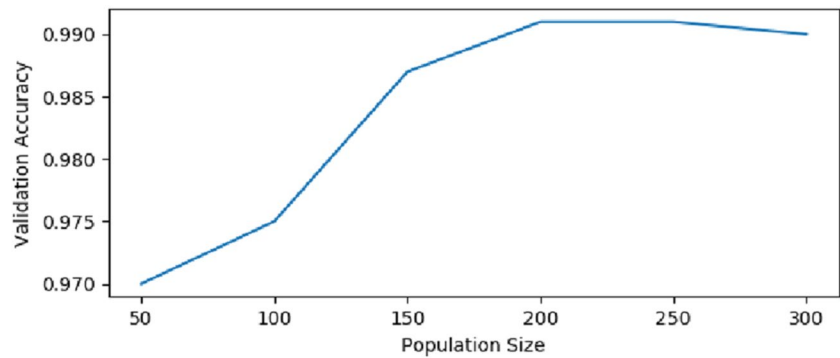


Fig. 11 The plot of various Pareto Fronts w.r.t. Generations, for the proposed MOGA. Five points are chosen from each generation for plotting this graph

Table 6 Validation accuracies with MOGA, PCA and GA as feature selector

Feature selector	Accuracy	No. of features
MOGA	0.991	244
Simple GA	0.961	316
PCA	0.947	207
No selector	0.893	2048

5.3.6 Comparison of Crossover and Mutation Rates

Crossover and mutation rates are the parameters that control the convergence of the MOGA selector. A non-constant linear crossover rate has been used in this study to improve the selector's ability to find the optimal front. The proposed selector is compared with constant crossover and mutation rate based MOGA. Accuracy on the validation set averaged over multiple runs is used for this comparison. The result of the analysis is summarized in Table 7.

Table 7 Validation accuracy of MOGA selector using different crossover and mutation rates

Crossover rate	Mutation rate	Accuracy
Linear increase from 0.5 to 0.9	Linear decrease from 0.5 to 0.1	0.991
0.9	0.1	0.973
0.8	0.2	0.977

5.3.7 Comparison of Feature Selectors

Multiple feature selection techniques, namely Multi-Objective Particle Swarm Optimization (MOPSO) [30], Multi-Objective Differential Evolution (MODE) [18] and MOGA are compared in this study. The standard implementations of these techniques (except the proposed method) are used for this analysis. Table 8 shows the evaluation results on different selectors. For evaluation, features are extracted using the proposed AE architecture, selected using different selectors, and finally classified using SVM Ensemble. For more effective comparison, the test set, which is unseen by the selectors, is used for the evaluation. The details of the train-test split are provided in Sect. 4-A. The results obtained show that the proposed model outperforms other multi-objective feature selection techniques. Figure 13 shows the confusion matrices obtained for different feature selectors on the test set.

6 Results and Analysis

For evaluation, the dataset is split according to Table 4. The proposed method has been evaluated using the test set, composed of 260 COVID-19 chest CT images and 237 non-COVID chest CT Images. The performance is measured based on the evaluation metrics discussed in Sect. IV-B. The features are extracted using the AE encoder defined in III-A and selected using MOGA as described in III B. Finally, the

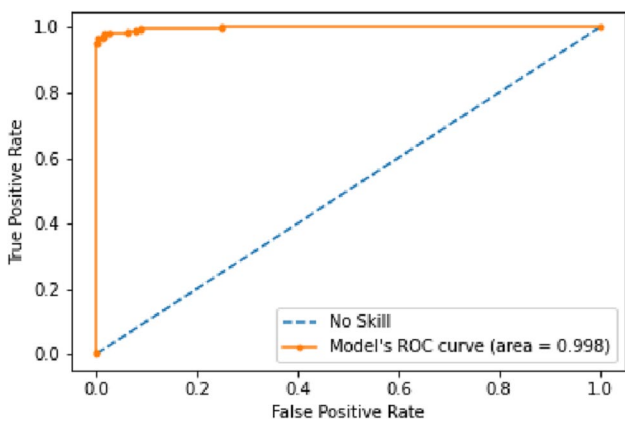


Fig. 12 ROC characteristics curve for the proposed methodology (convolutional autoencoder + MOGA + Bagging Ensemble with SVM)

The proposed architecture achieves an accuracy (ACC) of 98.79%, precision (PRE) of 98.47%, sensitivity (SEN) of 99.23%, F1 score (F1) of 98.85%, specificity (SPE) of 98.31%, net positive rate (NPV) of 99.14% and area under ROC curve (AUC) of 99.8%. The prediction time on the system used is 127 ms per image. The proposed model outperforms the current state of the art COVID-19 diagnostic techniques in terms of speed and accuracy.

The receiver operator characteristic curve on the proposed model’s test set is depicted in Fig. 12. The area obtained under the ROC Curve (AUC) is 0.998. A high value of AUC shows the robustness of the proposed model. Table 8 summarizes the evaluation results of the proposed architecture. Confusion matrices for different feature selectors on the test set are shown in Fig. 13.

features are classified using a Bagging Ensemble of SVM classifiers, described in III C.

The proposed methodology is implemented on python software, running on a CPU. The system architecture uses an Intel Core i7 processor with a 4 GB graphic card, running at 1.80 GHz, a 64-bit operating system, and 16 GB RAM.

The proposed study (MOGA) outperforms other multi-objective feature selectors. A decrease in ACC is expected in GA with an increase in the number of variables. As the number of variables for optimizing the selection are less, MOGA outperforms MODE and MOPSO.

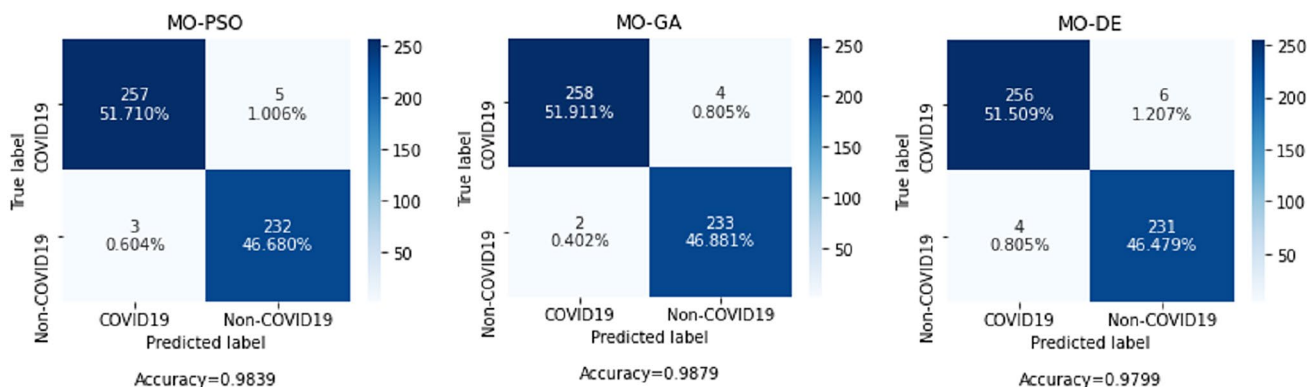


Fig. 13 Confusion matrices of the proposed methodology with different multi objective feature selectors on the test set

Table 8 Comparative assessment of various feature selectors on the test set

Feature selector	Number of features selected	Classification summary				Evaluation metric						
		TP	TN	FP	FN	ACC	PRE	SEN	F1	SPE	NPV	AOC
Multi objective Differential evolution	252	256	231	6	4	0.9798	0.9770	0.9846	0.9808	0.9746	0.9829	0.991
Multi objective Particle swarm Optimization	227	256	232	5	4	0.9839	0.9809	0.9884	0.9846	0.9789	0.9872	0.994
Multi objective Genetic algorithm (proposed study)	244	258	233	4	2	0.9879	0.9847	0.9923	0.9885	0.9831	0.9914	0.998

Bold indicates the best performance

7 Conclusion

An unsupervised learning-based approach is proposed for feature generation because of the higher feature diversity obtained from such an approach. Various evolutionary and non evolutionary feature selectors are compared in this study, and finally, a MOGA based selector is proposed. An ensemble of SVMs is used for the final classification. The bagging technique is used in the ensemble as it works well with complex feature maps.

The study further finds many insights in feature extraction, feature selection, and classification, which are listed below.

- Unsupervised learning-based feature extractors can provide detailed and accurate feature maps for medical image classification.
- Evolutionary Feature Selectors remove data redundancy better than standard techniques like PCA in terms of accuracy and number of features. Not using a feature selector results in inferior performance
- Optimizing the number of features and accuracy forces the model to learn from a smaller feature set, resulting in a more robust model since only the most productive features are retained.
- MOGA outperforms MOPSA and MODE in medical image classification because of the large number of parameters that need to be optimized for MOPSA and MODE.
- Variable Crossover and Mutation rates for MOGA can significantly improve performance in medical image classification.
- Bagging improves a classifier's performance, as a large number of classifiers produce a lower test error than a single classifier. This is because diversity compensates for bias.

The proposed model achieves better results than state-of-the-art techniques for all performance metrics. With such high-performance results and a little prediction time compared to Physical RT-PCR tests, the proposed model can be an effective and efficient COVID-19 Chest CT Scan screening Technique. Shortly, clinically verified AI-based diagnosis may be the way for rapid screening and early containment of outbreaks. With increasing structured medical data, deep learning models can be helpful for it.

Further, the study proposes that techniques like unsupervised feature extractor and evolutionary feature selector can help address the problem associated with limited COVID-19 radiology data. The study also comprehensively compares various feature selection techniques and highlights the importance of feature selection in medical

data problems. The study uses open-sourced dataset for COVID-19 screening. The technique's effectiveness is limited by the dataset available and needs to be verified on other data. Also, for clinical validation, there will be a need to localize the infection regions, map them in the images, and track the degree of infection.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. WebMD. (2020). Retrieved June 18th, 2020.
2. WHO. (2020). Retrieved June 18th, 2020.
3. Worldometer. (2020). Retrieved June 18th, 2020.
4. Tahamtana, A., & Ardebili, A. (2020). Real-time RT-PCR in Covid-19 detection: Issues affecting the results. *Expert Review of Molecular Diagnostics*, 20, 453–454.
5. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., et al. (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (covid-19) in China: A report of 1014 cases. *Radiology*.
6. Ye, Z., Zhang, Y., Wang, Y., Huang, Z., Song, B.: Chest CT manifestations of new coronavirus disease 2019 (covid-19): A pictorial review. *European Radiology*.
7. Kanne, J. P., Little, B., P., Chung, J. H., Elicker, B. M., & Ketaj, L. H. (2020). Essentials for radiologists on covid-19: An update—Radiology scientific expert panel. *Radiology*.
8. Gozes, O., Frid-Adar, M., Sagie, N., Zhang, H., Ji, W., & Greenspan, H. (2020). Coronavirus detection and analysis on chest CT with deep learning.
9. Shi, F., Xia, L., Shan, F., Wu, D., Wei, Y., Yuan, H., Jiang, H., Gao, Y., Sui, H., & Shen, D. (2020). Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification.
10. Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Chen, J., Su, Y., Lang, G., Li, Y., Zhao, H., Xu, K., Ruan, L., & Wu, W. (2020) Deep learning system to screen coronavirus disease 2019 pneumonia.
11. Mangal, A., Kalia, S., Rajgopal, H., Rangarajan, K., Nambodiri, V., Banerjee, S., & Arora, C. (2020). Covidaid: Covid-19 detection using chest X-ray.
12. Wu, Y.-H., Gao, S.-H., Mei, J., Xu, J., Fan, D.-P., Zhao, C.-W., & Cheng, M.-M. (2020). Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation.
13. Basu, S., Mitra, S., & Saha, N. (2020). Deep learning for screening covid-19 using chest X-ray images.
14. Loey, M., Smarandache, F., & Khalifa, N. E. M. (2020) A deep transfer learning model with classical data augmentation and CGAN to detect Covid-19 from chest CT radiography digital images. Preprints.
15. Gozes, O., Maayan, F.-A., Greenspan, H., Browning, P. D., Zhang, H., et al. (2020). Rapid AI development cycle for the coronavirus

- (covid-19) pandemic: Initial results for automated detection patient monitoring using deep learning CT image analysis.
16. Mobiny, A., Cicalese, P. A., Zare, S., Yuan, P., Abavisani, M., Wu, C. C., Ahuja, J., de Groot, P. M., & Nguyen, H. V. (2020). Radiologist-level covid-19 detection using CT scans with detail-oriented capsule networks.
 17. Singh, D., Kumar, V., Vaishali, & Kaur, M. (2020). Classification of covid-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. *European Journal of Clinical Microbiology Infectious Diseases*.
 18. Xue, B., Fu, W., & Zhang, M. (2014). Multi-objective feature selection in classification: A differential evolution approach. In: Dick, G., Browne, W. N., Whigham, P., Zhang, M., Bui, L. T., Ishibuchi, H., Jin, Y., Li, X., Shi, Y., Singh, P., Tan., K. C., Tang, K. (Eds.) *Simulated Evolution and Learning*, pp. 516–528. Springer International Publishing.
 19. Babatunde, O. H., Armstrong, L., Leng, J., & Diepeveen, D. (2014). A genetic algorithm-based feature selection. *British Journal of Mathematics Computer Science*.
 20. Khan, A., & Baig, A. R. (2015). Multi-objective feature subset selection using non-dominated sorting genetic algorithm. *Journal of Applied Research and Technology*, 13(1), 145–159.
 21. Ferriyan, A., Thamrin, A. H., Takeda, K., & Murai, J. (2017). Feature selection using genetic algorithm to improve classification in network intrusion detection system. In *2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, pp 46–49.
 22. Sindhiya, S., & Gunasundari, S. (2014) A survey on genetic algorithm based feature selection for disease diagnosis system. In *Proceedings of IEEE International Conference on Computer Communication and Systems ICCCS14*, pp 164–169.
 23. Ballard, D. H. (1987). Modular learning in neural networks. In *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1, AAAI'87*, pp. 279–284. AAAI Press.
 24. Autoencoders, P. B. (2011). Unsupervised learning and deep architectures. In *Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning Workshop - Volume 27, UTLW'11*, pp. 37–50. JMLR.org.
 25. Hou, B., Yan, R. (2018). Convolutional auto-encoder based deep feature learning for finger-vein verification. In *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–5.
 26. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197.
 27. Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11, 169–198.
 28. Breiman, L. (1996). Bagging predictors. *Machine Learning*, pp. 123–140.
 29. Soares, E., Angelov, P., Biaso, S., Higa Froes, M., & Kanda Abe, D. (2020). SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *medRxiv*.
 30. Coello Coello, C. A., & Lechuga, M. S. (2002). Mopso: A proposal for multiple objective particle swarm optimization. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, Vol. 2, pp. 1051–1056.