Vol.12 -1 (2025) 24–42
International Journal of Computers and Information (IJCI)
Available online at *https://ijci.journals.ekb.eg/*

COVID-19 Detection Based on CT scan Using Meta-Heuristic Feature Selection Method

Abdelghany Fathy, Hatem Abdelkader , Amira Abdelatey

Dept. of Information Systems, College of Computers and Information - Menoufia University, Shibin Al Kawm, Menoufia 32511, Egypt
abdalghany.fathy@ci.menofia.edu.eg, hatem.abdelkader@ci.menofia.edu.eg, amira.abdelatey@ci.menofia.edu.eg

Abstract

The rapid and widespread transmission of COVID-19 has necessitated the development of efficient diagnostic tools. While RT-PCR remains the standard method for diagnosis, its limitations in terms of time and resource intensity highlight the need for alternative solutions. This study addresses this gap by proposing a three-stage hybrid methodology for the rapid identification of COVID-19 using CT scans. In the first stage, pre-trained convolutional neural networks (CNNs), including Vgg-16, ResNet50, and MobileNet-v2, are utilized to extract relevant features from COVID-19-affected lungs. The second stage enhances feature selection through the application of meta-heuristic techniques such as genetic algorithms (GA) and particle swarm optimization (PSO), optimizing the feature set for improved accuracy. Finally, the selected features are classified using four distinct classifiers, achieving remarkable classification accuracies of 99.57% and 98.42% on the COVIDx-2A CT and SARS-CoV-2 CT-Scan datasets, respectively. The novelty of this approach lies in the integration of multiple CNNs and meta-heuristic methods to enhance feature selection and classification performance. Our contributions include the development of a robust diagnostic tool that significantly improves the speed and accuracy of COVID-19 detection, offering a viable alternative to traditional RT-PCR methods.
Keywords: COVID-19, CT scan, Deep features, CNN, Meta-heuristic feature selection, Classification;

1. Introduction

The COVID-19 outbreak began in China in late 2019 and spread globally. The World Health Organization (WHO) gave the virus the formal name SARS-CoV-2 (Severe Acute Respiratory Syndrome-Coronavirus-2)[1]. The World Health Organization has called it a "pandemic" due to the virus's fast spread and the number of deaths. It is well understood that air and physical touch are the most important factors in viral transmission. According to reports, this virus targets explicitly the lungs and causes a severe form of pneumonia [2]. Around 700 million individuals worldwide have been infected with COVID-19, resulting in approximately 7 million deaths as of July 20th, 2024. COVID-19 therapy and management rely heavily on early diagnosis. Real-time polymerase chain reaction (RT-PCR) and imaging tests, such as chest X-rays and computerized chest tomography (CT), are reliable techniques for the diagnosis of COVID-19[3]. According to research, some people may see changes in their X-ray and CT scans before they have any symptoms of COVID-19. Specifically, chest CT scans have shown typical radiographic features for COVID-19 patients and have provided quick and efficient findings[4, 5].

However, there may not be enough physicians and radiologists to examine CT scans if the number of COVID-19 patients continues to rise at its current rate, straining public health systems to overload. Another potential solution to aid professionals in medical diagnosis within this specific context is the utilization of computer-aided diagnostic (CAD) systems. When a clinician is faced with a difficult diagnosis that the human eye cannot resolve, these technologies can step in and offer a second opinion by processing and analyzing images using computational approaches[6-8]. Currently, many individuals employ deep learning (DL) to

autonomously generate feature representations from the provided data. This is mostly because computers have gotten better over the past few years [6, 9].

Convolutional neural networks (CNNs) are highly proficient in analyzing CT scans and generating accurate predictions regarding a patient's COVID-19 status. While CNN architectures specialize in image classification, building a CAD system utilizing a CNN necessitates extensive datasets and substantial processing capacity to get satisfactory outcomes. CNN offers a variety of models, such as ResNet [10], AlexNet [11], VGG-Net [12], and GoogLeNet [13]. Classification accuracy in CNN models is proportional to the number of convolution layers that are expanded [14].

Due to the time necessary to search in a wide search field, selecting the finest or essential item becomes quite tough whenever there is an abundance of it. In the same way, there could be duplication of effort in a feature set used for categorization, therefore not all of the characteristics might be necessary. Given the sheer volume of potential feature combinations, selecting the optimal one from the initial set can be a time-consuming and resource-intensive process. The goal of feature selection (FS) is to improve the learning model's performance by extracting the most relevant characteristics from a pool of available information [15]. Among the many effective approaches to radiology application challenges, meta-heuristic algorithms stand out. The logical behavior of physical algorithms seen in nature is the inspiration for most of these algorithms. These optimization strategies usually find acceptable solutions with minimal computing work and in a fair amount of time [16]. Patients have a better chance of making a full recovery if coronavirus is caught early and contained. Consequently, many AI methods for early COVID-19 identification have been suggested in scientific literature.

Using four cascaded steps, this study proposes a framework for COVID-19 categorization. Data loading and preprocessing, including scaling and cropping, was the initial stage. The second phase involved extracting the characteristics using three distinct convolutional neural network (CNN) models ResNet18, Vgg-16, and MobilNet-v2. Step three involves selecting features. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) techniques were employed for feature selection. Finally, four distinct classifiers—SVM, k-NN, MLP, and Ensemble learning—were used to determine the COVID-19 and other class detection performances. Two kinds of CT datasets are used in the experiments to test the proposed framework.

In summary, the main contributions of the proposed framework are as follows:

- Research that relied on deep features rather than conventional feature extraction techniques was conducted utilizing deep learning models.
- In order to further improve the identification of COVID-19, the deep features that were collected from three separate CNN models have been integrated.
- Three convolutional neural network (CNN) models, two metaheuristic feature selection algorithms, and four classifiers were the components of a hybrid study that was presented.
- Among the metaheuristic approaches, PSO and GA were favored for detecting irrelevant and uninformative features due to their low computing burden and small parameter requirements.
- Radiologists now have access to a very accurate decision-making system for the identification and follow-up of COVID-19.

This paper is organized in the following way: After this introduction, Section 2 shows the work that was previously mentioned. In Section 3, the suggested framework is shown. In Section 4, The results of the experiment are discussed. Section 5 is the conclusion of this paper.

2. **Related Work**

Several current approaches to COVID-19 identification that utilize ML and DL models are outlined here. Models based on deep learning and machine learning make the identification of many chronic diseases quite straightforward. As an example, the COVID-Net model [17, 18] has been effective in training deep neural network models with open-source freely accessible CT image datasets (COVIDx-CT) and in optimizing micro and macro architecture through machine-driven design exploration to construct a problem-specific architecture that is both efficient and effective. Zhao et al. [19] proposed a modified version of ResNet CNN A deep learning model for COVID-19 detection using CT images was developed Which focuses on the use of transfer learning,

and 4 models were produced that gave good results, specifically the BIT-M model which gave an accuracy of 92.2%.

Loddo et al. [20] proposed a novel approach to diagnosing COVID-19 that utilizes various CNN architectures to ensure precise identification of the virus. For the purpose of evaluating the proposed model, the COVIDx CT-2A and COVID-CT CT scan image datasets were incorporated during the formulation of the proposed method. The proposed method was assessed utilizing various evaluation metrics. Compared to other CNN architectures, the VGG19 achieved an accuracy of 98.87% on the COVIDx CT-2A dataset.

Mishra et al.[21] proposed a various transfer learning (TL) algorithms for the detection of COVID-19 patients. Additionally, they integrated outcomes from multiple deep convolutional neural network (CNN) algorithms through a decision fusion-based approach to generate a conclusive result. Their experimentation involved 744 CT scan images, comprising 347 COVID-19 cases and 397 non-COVID-19 cases, resulting in an average accuracy of 88.34%, an area under the curve (AUC) of 88.32%, and an F1 score of 86.7%.

Carvalho et al. [22] proposed an approach to identifying COVID-19 from CT images scan. After suggesting a convolutional neural network design for feature extraction from CT scans, they use a tree Parzen estimator to fine-tune the network's hyperparameters. After that, they use a genetic algorithm to apply certain features. Lastly, four classifiers with distinct techniques are used to achieve the classification. On the SARS-CoV-2 CT-Scan dataset, the suggested approach obtained an accuracy of 0.997, while on the COVID-CT dataset, it was 0.987.

By combining deep feature extraction with feature selection, Basu et al. [15] created an end-to-end architecture. In order to extract the deep features from CT scan images, they utilized three distinct pre-trained Deep learning models (DenseNet, ResNet, and XceptionNet).To improve performance, they used a combination of Adaptive β-Hill Climbing (AβHC), a local search method, and Harmony search, a meta-heuristic optimization strategy. The SARS-COV-2 CT scan dataset, which includes CT scan images, was used to evaluate the approach. This method outperformed all others on the given dataset, with an accuracy level of 98.87%.

In order to determine which COVID-19 dataset features were most important, Dey et al. [23] used ResNet and GoogleNet as pre-trained models to deep features then develop a hybrid meta-heuristic feature selection technique they called the manta ray foraging based golden ratio optimizer (MRFGRO). In this model, three datasets were utilized: MOSMED, SARS-COV-2 CT scan, and COVID-CT. This approach was able to get a 99.42% accuracy rate on the SARS-COV-2 CT scan dataset.

El-Kenawy et al. [24] introduce a three-phase framework that utilizes two optimization methods to perform feature selection and classification of COVID-19. The first step is extracting features from CT scans using the AlexNet model. This is followed by the utilization of the Guided Whale Optimization Algorithm (Guided WOA), which is based on the Stochastic Fractal Search (SFS) method. Lastly, a voting classifier is employed, which is based on the Particle Swarm Optimization (PSO) technique. The model attains an Area Under the Curve (AUC) value of 0.995.

Hassan E. et al. [25] assessed the effectiveness of various pre-trained transfer learning models, including ResNet-50, VGG-19, VGG-16, and Inception V3, for the classification of CT images. The study utilized the binary cross-entropy metric to distinguish COVID-19 cases from normal cases. To mitigate overfitting, they employed the Stochastic Gradient Descent and Adam optimization techniques. The pre-trained models demonstrated high accuracy, achieving 99.07%, 98.70%, 98.55%, and 96.23%, respectively.

Singh et al [26] developed a feature selection approach for classifying COVID-19 and healthy individuals' chest CT images. From a dataset of 2,482 images, 213 features were extracted and reduced through a two-step process: a Chi-square test to select 75% of the features, followed by optimization using three nature-inspired algorithms. The reduced feature set was then classified using five machine learning models, with XGBoost yielding the best results. Their approach achieved 95.99% accuracy, a mean intersection over union of 0.9655, and an area under the curve of 0.9966, highlighting the importance of effective feature selection in medical image analysis.

Hossain M. M. et al. [27] introduced a COVID-19 detection model using chest CT scans, where features from VGG-19 and ResNet-50 CNNs were fused and optimized through techniques like Recursive Feature Elimination, PCA, and LDA. The optimized features were classified using a Max Voting Ensemble, achieving

high performance metrics, including 98.51% accuracy and 99.49% sensitivity, following 5-fold cross-validation. For more clarification Table 1 presents the pros and cons of the related work.

In previous studies, the focus was on binary classification for detecting COVID-19 disease. However, in this study, our approach is expanded to include three classes: Normal, pneumonia, and COVID-19. With the proposed framework, the selection of datasets includes the 194922 CT image. In addition, the proposed framework examines the utilization of feature selection and optimization techniques like PSO and GA.

Table 1: Comparison of related work

Ref	Year	Pros	Cons
[18]	2020	High accuracy, clinical relevance, and explainability	Generalization to other data sources is unknown; limited explainability analysis.
[17]	2021	High accuracy of 98.1% in detecting COVID-19	Complex architecture requiring significant computational resources.
[19]	2021	Effective use of transfer learning, improving model performance.	Subtle visual differences in CT images reduce accuracy; limited generalizability.
[20]	2021	Comparative study of CNN architectures and explores transfer learning for COVID-19 diagnosis.	Limited modifications to architectures; VGG19 misclassifies some cases.
[21]	2020	Combines predictions from multiple models, enhancing efficiency.	Variability in CT image sizes may affect performance; risk of overfitting due to small dataset.
[22]	2021	Efficient hyperparameter optimization using the Tree Parzen Estimator.	High accuracy on a limited dataset suggests potential overfitting; computational complexity.
[15]	2022	Utilizes state-of-the-art CNN models (DenseNet, ResNet, Xception) for feature extraction.	Risk of overfitting due to high-dimensional features without proper regularization.
[23]	2021	Hybrid approach combining multiple meta-heuristic algorithms, demonstrating high accuracy in COVID-19 screening.	Requires further validation across different datasets; significant computational resources needed for scalability.
[24]	2020	Adaptable method that can be modified for different datasets.	Lack of interpretability, especially for non-experts; approach may not generalize well to different datasets without adjustments.
[25]	2024	Evaluation of multiple pre-trained models with high accuracy, sensitivity, and specificity.	Dataset may not cover all COVID-19 variations; some models require significant computational resources.
[26]	2024	Innovative use of nature-inspired algorithms for feature selection.	Performance heavily depends on parameter tuning, which is time-consuming and requires expertise; computational complexity.
[27]	2024	Ensemble classification approach improves robustness and accuracy; integrates optimized deep features for better COVID-19 detection.	Computationally intensive processes generalization of results may be limited to the specific dataset used.

3. Proposed Work

The proposed framework for the COVID-19 detection procedure is illustrated in Fig. 1. The CT scan images are preprocessed before being sent into a convolutional neural network (CNN) model for feature extraction. Essentially, the features refer to the output of the preceding layer that contains the prediction probabilities. Following feature extraction, PSO and genetic algorithms are used to choose the most relevant features. After that, the classifiers are trained using the selected features to get the final predictions. The following sections provide an in-depth analysis of the aforementioned pipeline's main stages.

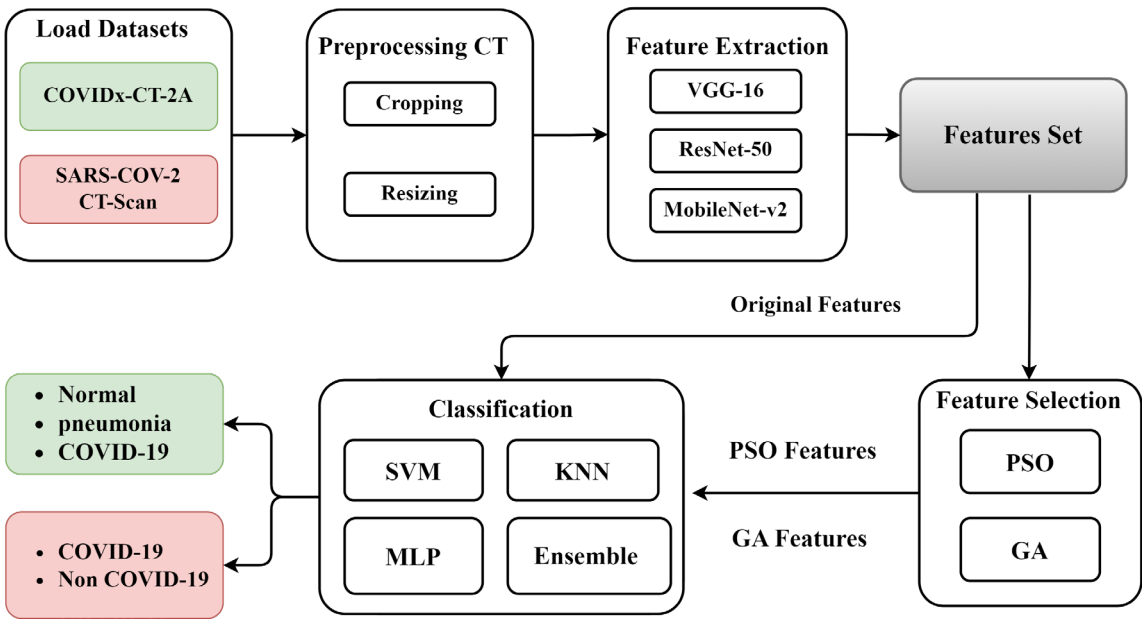


Fig. 1: Proposed Framework

3.1. Datasets

The datasets included in this work are COVIDx CT-2A and SARS-COV-2 CT-Scan, both of which are publicly available. Below is our comprehensive portrayal of them.

3.1.1 COVIDx-2A CT scan

COVIDx-2A CT [28] images dataset is a valuable open-access resource for researchers and healthcare professionals studying COVID-19. It extends the original COVIDx dataset, focusing specifically on CT (computed tomography) imaging. The dataset comprises a substantial collection of CT images meticulously selected and verified to ensure high-quality and clinically significant data.

At the time of writing, the COVIDx-2A CT images dataset contains a vast array of 194,922 CT images obtained from 3,745 patients across 15 different countries. The patients' ages range from 0 to 93 years, with a median age of 51. Expert pathologists have strongly clinically validated the findings associated with each image, thus providing reliable and accurate diagnoses.

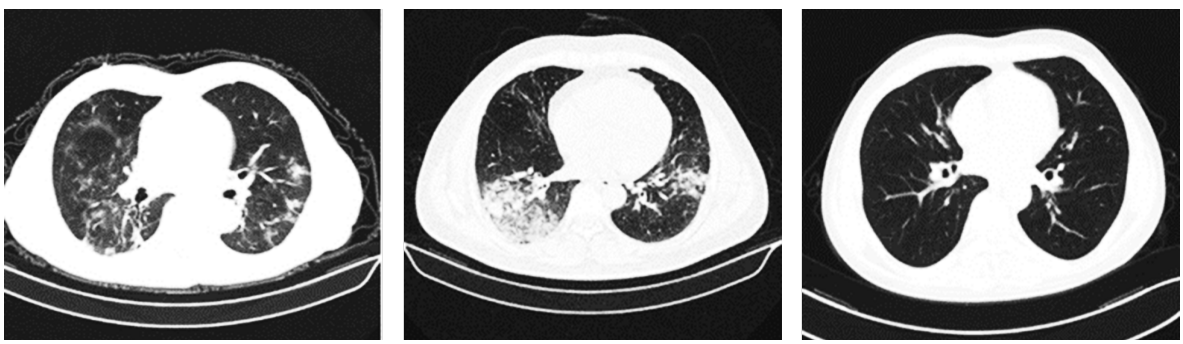
The dataset is organized into distinct classes, enabling researchers to categorize and analyze the images effectively. These classes include COVID-19, representing CT images of patients confirmed

positive for COVID-19; pneumonia, showing CT scans of people with pneumonia who have nothing to do with COVID-19; and normal, signifying CT images of patients with no abnormalities or underlying conditions. One sample per class is shown in Fig. 2.

3.1.2 SARS-CoV-2 CT-Scan Dataset

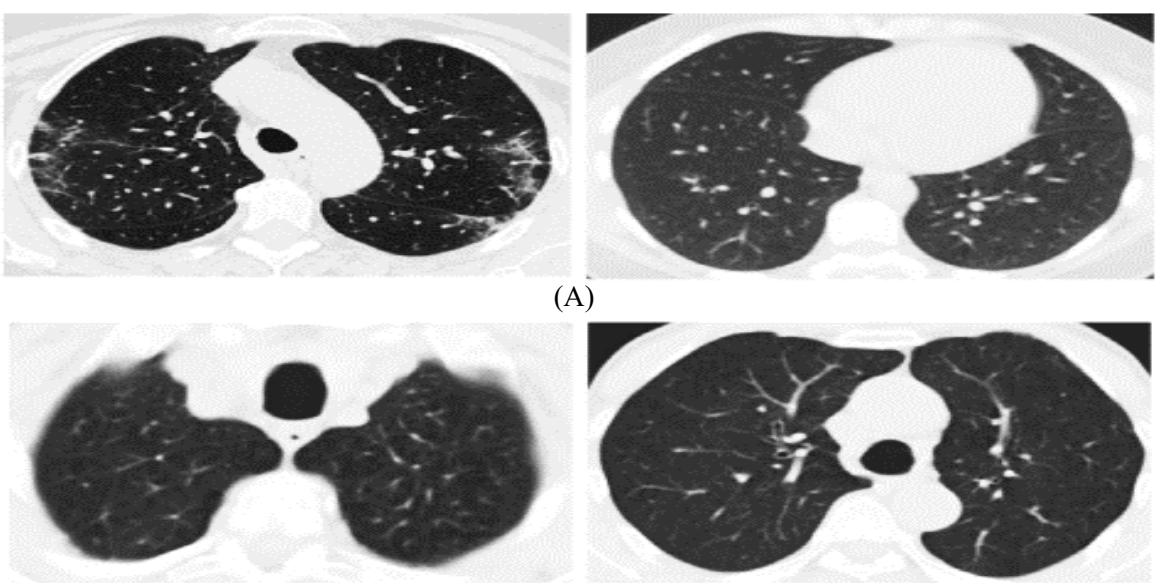
The SARS-CoV-2 CT-Scan Dataset [29] is a valuable resource that contains 2,482 images of CT scans. This dataset was collected from hospitals in Sao Paulo, Brazil, and included scans of 120 patients of both genders. The dataset comprises 1,252 CT scans of patients infected with SARS-CoV-2 and 1,230 CT scans of patients with other lung diseases.

The availability of this dataset provides researchers, medical professionals, and data scientists with a comprehensive collection of CT scan images related to SARS-CoV-2 and other lung diseases. These images can be utilized for various purposes including diagnostic research, disease progression analysis, and the development of machine learning algorithms for accurate detection and classification. The image's sizes range from 119 × 104 to 416 × 512. Fig. 3 shows some sample images from the dataset.



(A) (B) (C)

Fig. 2 : Example chest CT images from the COVIDx-CT dataset, illustrating (A) COVID-19 pneumonia cases, (B) non-COVID-19 pneumonia cases, and (C) normal control cases.



(A)

(B)

Fig. 3 : Some sample CT images from the SARS-COV-2 CT-Scan Dataset, illustrating (A) Scans from COVID-19 positive patients and (B) Scans from COVID-19 negative patients

3.2. Preprocessing

The preprocessing stage in detecting COVID-19 from CT scan images is crucial to the success and accuracy of the subsequent analysis. Reducing or removing data variability's effect on model performance, enhancing picture contrast, and making the disease zone more noticeable relative to the original image are all part of the process. Each image of the two datasets underwent a cropping technique to highlight the illness region in the lung. Subsequently, the CT images were resized to 224 × 224 pixels to be used as input for the CNN models.

3.3. Feature Extraction

A CNN-based model is employed for feature extraction. Hybrid techniques are produced by combining the retrieved characteristics with classical classifiers. Typically, CNN networks are composed of three layers: a convolution layer, a pooling layer, and a fully connected layer. The convolution layer is a crucial component of the model, serving as its foundational layer. The incoming patterns are processed through filters, resulting in the formation of feature maps. By moving these filters together with the pattern, a wide range of characteristics can be identified. The more convolution layers you have, the more in-depth attributes can be acquired. Typically, a flattening or global pooling layer is used after the last convolutional layer to create a one-dimensional tensor. Subsequently, there are one or more substantial layers. The last dense layer generates a one-dimensional tensor with a size equivalent to the number of output classes and often uses a softmax activation function. The function encodes a probability distribution that represents the likelihood of the picture belonging to each of the classes. Furthermore, it is utilized in various layers such as normalization and dropout layers[30].

Our primary emphasis has been on extracting profound characteristics utilizing pre-trained Convolutional Neural Network (CNN) models. For the purpose of extracting deep features, Three commonly used pre-trained Convolutional Neural Networks (CNNs) have been selected: VGG16 [12], ResNet50[31], and MobileNet-v2 [32]. In this study, both the training and testing images are run through the model, and features are extracted from the final layer. Table 2 illustrates the number of extracted deep features using various CNNs.

Furthermore, to assess the amalgamated deep features derived from various CNNs collectively, Combinations of different CNNs were examined by merging their feature sets. This fusion process entails concatenating the features extracted from the aforementioned CNNs to construct the ultimate feature vector, which is then evaluated using our proposed algorithm for feature selection (FS).

Table 2 : Number of deep features extracted using various CNN models when applied in datasets.

Pre-trained CNN model	Number of features Extracted
VGG16	512
ResNet50	2048
MobileNet-v2	1280

3.4. Feature selection

Features from multiple CNNs have been combined, resulting in a significantly expanded feature set. This large feature vector raises concerns about potential overfitting of classifiers and the presence of redundant features. Feature selection aims to identify the optimal subset of features by removing irrelevant data from these extensive datasets. While it doesn't guarantee improved estimation rates, achieving comparable performance with fewer features is considered a favorable outcome. In this study, the use of some meta-heuristic FS algorithms, such as PSO and GA, is proposed.

3.4.1. Particle swarm optimization (PSO)

The Particle Swarm Optimization (PSO) algorithm draws inspiration from the coordinated movements observed in animal herds striving to fulfill their basic needs, enabling them to efficiently achieve their objectives. PSO [33] operates by managing a group of individuals, referred to as "particles," each possessing velocity and position information, collectively forming a "swarm". Through the mathematical evaluation of fitness functions, the optimization of particles is guided. The state of a particle that is closest to the solution is referred to as "Pbest" (personal best), whereas the state of the particle that is closest to the solution within the whole swarm is called "Gbest" (global best). Modifying these variables determines the rate of variation and displacements of each particle [34].

Each particle in the population has two properties: position vector $X_i = (X_{i1} , X_{i2} , \ldots\ldots , X_{id})$ and velocity vector $V_i = (V_{i1} , V_{i2} , \ldots\ldots , V_{id})$, where d denotes the dimension. and to update the position and velocity of each particle is calculated by the following equations.

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{1}$$

$$V_i(t+1) = w \times V_i(t) + c_1 \times r_1 \times (pbest_i - x_i(t)) + c_2 \times r_2 \times (gbest - x_i(t)) \tag{2}$$

The variables V_i and X_i denote the velocity and position vectors of the i^{th} (i = 1, 2, ..., N) particle. The top limit for each dimension is set to 1, while the lower limit is set to 0. The inertia parameter, denoted as w, is a non-negative value. In terms of acceleration, we have c_1 and c_2. c_1 is the user-specified personal learning parameter, while c_2 is the global learning parameter that controls the particle search scope. r_1 and r_2 are two random integers that fall into the interval [0, 1]. $pbest$ is best position, $gbest$ is global best.

3.4.2. Genetic Algorithm (GA)

Genetic algorithms (GAs) [35] have emerged as a powerful tool for feature selection in various machine learning applications. In the realm of feature selection, GAs leverage principles inspired by natural evolution to efficiently explore and optimize feature subsets. This involves encoding potential feature subsets into binary strings, where each bit represents the presence or absence of a feature. The evaluation of these subsets is facilitated by a fitness function, typically based on performance metrics like classification accuracy or model complexity. Mathematically, the fitness function *f(x)* quantifies the suitability of a feature subset x, often defined as *f(x)=evaluation metric(x),* and then applying genetic operations like as crossover and mutation to generate offsprings that make up the next generation of solutions. This iterative technique tries to steadily enhance the quality of the feature subsets over time, finally converging to an ideal solution. The implementation of Genetic Algorithm based feature selection is shown in Fig. 4.

3.5. Classification Algorithm

Support Vector Machine (SVM) [36], K-Nearest Neighbor (KNN), and MultiLayer Perceptron (MLP) [37] were the three classification algorithms utilized in the proposed framework. The algorithms were then subjected to ensemble voting. The algorithms were selected for their track records of success in previous studies, and the ensemble voting learning method was selected for its ability to harness the power of each classifier individually. In a situation where there are n classifiers (classifier 1, classifier 2, ..., and classifier n), the outcome is not dependent on any one classifier. Instead, the results from each classifier are fed into a fuser, which oversees the ultimate decision-making process.

With each pre-trained model and concatenated between them, these classifier algorithms were run three times: once without feature selection, once with PSO selecting features, and once with GA selecting features. Different algorithms include different sets of hyperparameters. Every algorithm is tested on all three sets of features—the original features, the features selected by PSO, and the features selected by GA—using identical

parameters and constraints. To achieve the best possible results, the model was fine-tuned using hyperparameters. Table 3 shows the updated hyperparameters for each classifier.

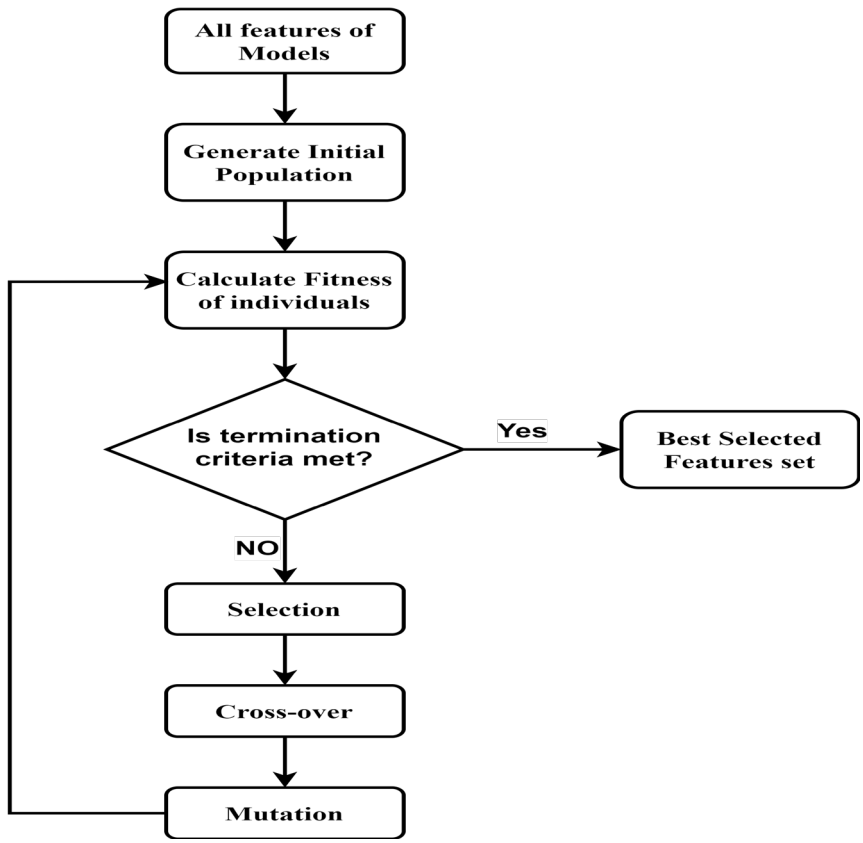


Fig. 4 : The flowchart of genetic algorithms (GA) based feature selection.

Table 3 : The updated hyperparameters for each classifier.

Classifier	Parameters
SVM	*kernel= poly, C=1.0*
KNN	*n_neighbors = 5,weights =distance*
MLP	*activation= relu , solver = adam, learning_rate=constant, alpha=0.001*

3.6. Evaluation Metrics

All four of these statistical measures—accuracy (A), recall (R), precision (P), and F-score—are widely utilized in the literature to assess the outcomes (F1). All these criteria have been considered to assess the suggested model in a broader sense. The metrics are defined using the following equations, which are based on the values of TP (true positive), TN (true negative), FP (false positive), and FN (false negative) [38].

- Accuracy (A) is a metric that evaluates a model's probability to identify correctly categorized data. It computes several accurately categorized negative and positive samples as in Equation (3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

- Recall (R) is calculated by dividing the number of cases accurately classified into the positive class by the total count of positive class items. In other terms, it indicates the number of positive cases that were accurately classified as in Equation (4).

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

- Precision (P) is a quantitative measure utilized to evaluate the accuracy with which a model predicts positive data samples. The proportion of positive samples that were predicted precisely is calculated using Equation (5).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

- F1-score (F1) is measuring the harmonic mean of Precision and Recall and is calculated using Equation (6).

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

4. Experimental Results and Discussion

The experimental results offer useful information about the performance and efficacy of the suggested framework. This part outlines the technique used to evaluate our proposed strategy, as described in section 4.1. The findings of our suggested approach are presented in section 4.2. Additionally, a comparison study with current research on the same dataset is reviewed in section 4.3.

4.1. Experimental Setup

Four different pre-trained deep convolutional neural network models - VGG-16, ResNet-50, and MobileNet-v2 - are used as feature extractors in the experimental setup. These models were pre-trained on the ImageNet dataset. Four distinct machine learning classifiers - SVM, KNN, MLP, and Ensemble - are utilized. The incoming photos are preprocessed, followed by the extraction of deep features. The features are combined, followed by a genetic and PSO selection procedure. The studies were carried out on a personal computer with a GPU 16 GB NVIDIA GeForce GTX 1070, utilizing Jupyter IDE python 3 and TensorFlow 2.10 deep learning framework.

4.2. Results

The experiments were carried out using two separate datasets. The first dataset, COVIDx-2A CT, was employed to classify CT images into three categories: COVID-19, pneumonia, and normal. The second dataset, SARS-CoV-2 CT-Scan, was used to differentiate between normal and COVID-19 CT scans.

Once the CT images are pre-processed, the next step involves extracting features using pre-trained CNN models, including VGG-16, ResNet-50, and MobileNet-v2. The extracted features are then concatenated to enhance the feature vector representation, all possible concatenations between the extracted features from each two pre-trained CNN models are performed and finally, the concatenation is performed using the three CNN model's features, after that, the concatenated features subjected to a rigorous selection process, where two robust feature selection algorithms, PSO and GA, are employed to identify the most relevant and critical features from the original feature set. These selected concatenated features are subsequently fed into different machine-learning classifiers to execute the classification process.

Several experiments were conducted, with the first focusing on using the original concatenated feature set as input to various machine learning classifiers without any prior feature selection. The purpose of this experiment was to obtain an initial assessment of the performance of the features extracted by the CNN models. The findings from the first experiment across both datasets are presented in Table 4. As detailed in Table 4, the MLP classifier demonstrates superior performance on the COVIDx-2A CT dataset, attaining an accuracy of 0.9839 through the concatenation of features from ResNet-50 and MobileNet-v2 models. Also, on the SARS-CoV-2 CT-Scan dataset, the KNN classifier achieves the highest accuracy of 0.9732 by employing concatenated features from VGG-16 and MobileNet-v2.

Building on the results of the first experiment, it is recommended to implement feature selection on the concatenated features. This approach aims to identify the most significant features within the extracted set, which could substantially enhance the performance outcomes.

In the second experiment, two powerful selection algorithms are utilized, PSO and GA, to identify the most critical feature sets from the individual CNN models as well as from the combined features (VGG-16 + ResNet-50, VGG-16 + MobileNet-v2, ResNet-50 + MobileNet-v2, and VGG-16 + ResNet-50 + MobileNet-v2). Table 5 details the number of extracted and concatenated features before and after the selection process. The results indicate that both PSO and GA effectively reduced the original feature sets by a significant percentage, thereby decreasing the required computation time and overall complexity of the proposed framework.

Subsequently, the selected features were utilized as inputs for various machine learning classifiers, including SVM, KNN, MLP, and Ensemble, to assess their performance. Table 6 and Table 7 present the results of this experiment using PSO and GA, respectively. According to Table 6, the Ensemble classifier outperformed the others, achieving an accuracy of 0.9957 with PSO-selected concatenated features from the ResNet-50 + MobileNet-v2 models on the COVIDx-2A CT dataset. Additionally, the Ensemble classifier achieved the highest accuracy of 0.9758 using PSO-selected concatenated features from all three CNN models on the SARS-CoV-2 CT-Scan dataset.

As shown in Table 7, when GA is employed as the feature selection algorithm, the ensemble classifier demonstrates the highest performance, achieving an accuracy of 0.9951 using GA-selected concatenated features from the ResNet-50 + MobileNet-v2 models. Furthermore, it outperforms the other classifiers with an accuracy of 0.9842 when utilizing GA-selected concatenated features from all three CNN models.

Our results indicate a significant performance improvement, particularly after applying the PSO and GA selection algorithms. Notably, the highest accuracy of 0.9957% was achieved on the COVIDx-2A CT dataset using the ensemble classifier with PSO-selected features from the ResNet-50 + MobileNet-v2 models. Moreover, the highest accuracy was obtained by using the concatenated features from only two models, rather than all three. This framework not only reduces the required time and resources but also simplifies the process by working with a smaller set of features.

For a more detailed analysis, Table 8 provides a comprehensive overview of the average time required for each process in the proposed approach. It can be observed from Table 8 that the SARS-CoV-2 CT-Scan Dataset requires less processing time compared to the COVIDx-2A CT dataset, particularly in feature extraction and

selection, due to its smaller size. Moreover, variations in processing time may also be influenced by the complexity of the model structure. Models with more complex architectures tend to demand more time for each process. This detailed time analysis underscores the importance of considering both the dataset size and the complexity of the model architecture when evaluating the proposed approach's efficiency.

Table 4: Results of the proposed framework without feature selection.

Features Set	Classifiers	Classifiers with original features set							
		COVIDx-2A CT dataset				SARS-CoV-2 CT-Scan Dataset			
		A	P	R	F1	A	P	R	F1
VGG-16	SVM	0.9251	0.9258	0.9166	0.9209	0.9356	0.9223	0.9519	0.9368
	KNN	0.9533	0.9577	0.9427	0.9484	0.9624	0.9553	0.9706	0.9629
	MLP	0.9465	0.9401	0.9464	0.9431	0.8926	0.8808	0.9091	0.8947
	Ensemble	0.9559	0.9541	0.9520	0.9528	0.9517	0.9447	0.9599	0.9523
ResNet-50	SVM	0.9326	0.9331	0.9251	0.9289	0.9463	0.9489	0.9439	0.9464
	KNN	0.9545	0.9569	0.9457	0.9496	0.9477	0.9491	0.9465	0.9478
	MLP	0.9844	0.9839	0.9828	0.9833	0.9423	0.9584	0.9251	0.9415
	Ensemble	0.9689	0.9697	0.9635	0.9662	0.9503	0.9566	0.9439	0.9502
MobileNet-v2	SVM	0.9161	0.9156	0.9087	0.9118	0.9181	0.9173	0.9198	0.9186
	KNN	0.9489	0.9512	0.9400	0.9441	0.9490	0.9492	0.9492	0.9492
	MLP	0.9718	0.9744	0.9663	0.9702	0.8872	0.8737	0.9064	0.8898
	Ensemble	0.9577	0.9578	0.9523	0.9544	0.9315	0.9307	0.9332	0.9319
VGG-16 + ResNet-50	SVM	0.9283	0.9295	0.9202	0.9246	0.9450	0.9393	0.9519	0.9456
	KNN	0.9563	0.9599	0.9471	0.9519	0.9651	0.9579	0.9733	0.9655
	MLP	0.9695	0.9720	0.9620	0.9663	0.9329	0.9286	0.9385	0.9335
	Ensemble	0.9632	0.9675	0.9563	0.9615	0.9597	0.9550	0.9652	0.9601
VGG-16 + MobileNet-v2	SVM	0.9215	0.9234	0.9116	0.9171	0.9356	0.9382	0.9332	0.9357
	KNN	0.9544	0.9587	0.9442	0.9497	**0.9732**	0.9634	0.9840	0.9735
	MLP	0.9728	0.9713	0.9705	0.9708	0.9329	0.9263	0.9412	0.9337
	Ensemble	0.9534	0.9605	0.9381	0.9474	0.9597	0.9550	0.9652	0.9601
ResNet-50 + MobileNet-v2	SVM	0.9368	0.9368	0.9307	0.9336	0.9544	0.9620	0.9465	0.9542
	KNN	0.9576	0.9596	0.9502	0.9534	0.9530	0.9520	0.9545	0.9533
	MLP	**0.9839**	0.9814	0.9840	0.9827	0.9423	0.9534	0.9305	0.9418
	Ensemble	0.9735	0.9737	0.9692	0.9710	0.9570	0.9647	0.9492	0.9569
Combined	SVM	0.9290	0.9303	0.9212	0.9255	0.9477	0.9443	0.9519	0.9481
	KNN	0.9566	0.9601	0.9477	0.9524	0.9664	0.9604	0.9733	0.9668
	MLP	0.9551	0.9483	0.9604	0.9539	0.9490	0.9468	0.9519	0.9493
	Ensemble	0.9720	0.9734	0.9665	0.9695	0.9624	0.9727	0.9519	0.9622

Table 5 : Number of features selected by PSO and GA

Features Set	Original Features	COVIDx-2A CT dataset		SARS-CoV-2 CT-Scan Dataset	
		PSO (%)	GA (%)	PSO (%)	GA (%)
VGG-16	512	249 (49%)	214 (42%)	265 (52%)	188 (37%)
ResNet-50	2048	1032 (50%)	907 (44%)	1019 (50%)	869 (42%)
MobileNet-v2	1280	640 (50%)	560 (44%)	624 (49%)	561 (44%)
VGG-16 + ResNet-50	2560	1263 (49%)	1128 (44%)	1216 (48%)	1142 (45%)
VGG-16 + MobileNet-v2	1792	903 (50%)	751 (42%)	876 (49%)	754 (42%)
ResNet-50 + MobileNet-v2	3328	1665 (50%)	1512 (45%)	1643 (49%)	1520 (46%)
Combined	3840	1927 (50%)	1774 (46%)	1886 (49%)	1694 (44%)

4.1. Comparative Analysis

This section evaluates the accuracy of our suggested framework in comparison to recent and existing studies in literature. Our proposed framework has compared with prior research conducted by Gunraj et al [17] [18], Zhao et al [19], and Loddo et al [20] on the COVIDx CT-2A dataset, and by Carvalho et al [22], Basu et al [15], and Dey et al [23] on the SARS-COV-2 CT-Scan dataset. Table 8 summarizes the results, indicating that the suggested framework achieves higher accuracy in classifying COVID-19 detection compared to prior studies. The suggested framework attained the greatest accuracy of 99.57% in the COVIDx CT-2A dataset, exceeding the values reported in prior studies on the same dataset. The suggested framework failed to achieve the highest accuracy in the SARS-COV-2 CT-Scan dataset, reaching 98.12%.

5. **Conclusions**

The COVID-19 pandemic has caused unprecedented global losses and challenges, further exacerbated by the emergence of new variants, raising significant concerns. In response, this paper introduces a robust framework for detecting COVID-19 in CT scans, leveraging two publicly available image datasets, one of which distinguishes between pneumonia cases and COVID-19 patients. The proposed framework utilizes convolutional neural network (CNN) architectures—specifically VGG-16, ResNet-50, and MobileNet-v2—to extract relevant features. Following this, an effective feature selection (FS) stage, employing Particle Swarm Optimization (PSO) and Genetic Algorithm (GA), refines the feature vectors by filtering out irrelevant data. Subsequently, four different classification algorithms are applied.

The results demonstrate that our approach achieves impressive accuracy rates of 99.57% and 98.42% on the COVIDx-2A CT dataset and the SARS-CoV-2 CT-Scan dataset, respectively. The use of advanced feature selection techniques not only enhanced the accuracy of COVID-19 detection but also enabled clear differentiation from pneumonia. Moreover, our framework effectively mitigates the overfitting issues commonly associated with small datasets, performing competitively against pre-trained architectures and other state-of-the-art methods. Consequently, this framework has the potential to provide valuable second opinions in COVID-19 diagnosis and could be seamlessly integrated into computer-aided diagnostic systems.

The limitation of the proposed framework lies in its potential difficulty in accurately identifying COVID-19-positive CT scans during the early stages of infection. Consequently, the convolutional neural networks (CNNs) employed may struggle to extract the most relevant features at these critical junctures.

Future research will focus on addressing this challenge by enhancing the feature extraction process. This can be achieved by integrating optimization techniques to boost performance and applying meta-heuristic methods to fine-tune CNN parameters. Additionally, employing a combined feature selection approach will help in identifying the most effective features for classification. Exploring advanced techniques such as ensembling, pruning, and attention mechanisms will further strengthen the robustness of the feature extractors.

Table 6 : Results of the proposed framework with PSO feature selection.

Features Set	Classifiers	Classifiers with PSO feature selection							
		COVIDx-2A CT dataset				SARS-CoV-2 CT-Scan Dataset			
		A	P	R	F1	A	P	R	F1
VGG-16	SVM	0.9848	0.984	0.9841	0.984	0.9436	0.9826	0.9037	0.9415
	KNN	0.9919	0.9918	0.9908	0.9912	0.9436	0.9278	0.9626	0.9449
	MLP	0.9806	0.9785	0.9801	0.9793	0.9275	0.9124	0.9465	0.9291
	Ensemble	0.9903	0.9899	0.9895	0.9897	0.957	0.9647	0.9492	0.9569
ResNet-50	SVM	0.9932	0.993	0.9932	0.9931	0.8899	0.9899	0.7888	0.878
	KNN	0.9928	0.9924	0.9923	0.9923	0.9409	0.9388	0.9439	0.9413
	MLP	0.9881	0.9862	0.9887	0.9874	0.9436	0.9462	0.9412	0.9437
	Ensemble	0.9947	0.9943	0.9949	0.9946	0.9584	0.9831	0.9332	0.9575
MobileNet-v2	SVM	0.9879	0.9876	0.9871	0.9874	0.8832	0.9767	0.7861	0.8711
	KNN	0.9891	0.9888	0.9884	0.9885	0.9342	0.938	0.9305	0.9342
	MLP	0.9792	0.9777	0.9781	0.9779	0.9047	0.9151	0.893	0.9039
	Ensemble	0.9896	0.9892	0.9891	0.9891	0.9329	0.9765	0.8877	0.93
VGG-16 + ResNet-50	SVM	0.9887	0.9884	0.9884	0.9884	0.9235	0.9938	0.8529	0.918
	KNN	0.9933	0.993	0.9928	0.9929	0.9544	0.9474	0.9626	0.9549
	MLP	0.9906	0.9906	0.99	0.9903	0.9477	0.9539	0.9412	0.9475
	Ensemble	0.994	0.9933	0.9945	0.9939	0.9651	0.9807	0.9492	0.9647
VGG-16 + MobileNet-v2	SVM	0.9827	0.9826	0.9814	0.982	0.9248	0.9789	0.869	0.9207
	KNN	0.9932	0.9931	0.9926	0.9928	0.9584	0.9598	0.9572	0.9585
	MLP	0.9826	0.9811	0.9817	0.9814	0.949	0.9541	0.9439	0.9489
	Ensemble	0.9917	0.992	0.9902	0.9911	0.9624	0.9753	0.9492	0.9621
ResNet-50 + MobileNet-v2	SVM	0.9947	0.9945	0.9948	0.9946	0.906	0.9872	0.8235	0.898
	KNN	0.9939	0.9936	0.9935	0.9935	0.9503	0.9469	0.9545	0.9507
	MLP	0.9922	0.9914	0.9922	0.9918	0.945	0.9537	0.9358	0.9447
	Ensemble	**0.9957**	0.9955	0.9957	0.9956	0.9624	0.9832	0.9412	0.9617
Combined	SVM	0.9898	0.9895	0.9895	0.9895	0.9745	0.9863	0.9626	0.9743
	KNN	0.9936	0.9933	0.9931	0.9931	0.9705	0.9731	0.9679	0.9705
	MLP	0.989	0.9892	0.9876	0.9884	0.9732	0.9758	0.9706	0.9732
	Ensemble	0.9945	0.994	0.9945	0.9942	**0.9758**	0.9837	0.9679	0.9757

Table 7 : Results of the proposed framework with GA feature selection

Features Set	Classifiers	Classifiers with GA feature selection							
		COVIDx-2A CT dataset				SARS-CoV-2 CT-Scan Dataset			
		A	P	R	F1	A	P	R	F1
VGG-16	SVM	0.9847	0.9842	0.9838	0.984	0.9423	0.9715	0.9118	0.9407
	KNN	0.9922	0.9921	0.9912	0.9916	0.9369	0.9269	0.9492	0.9379
	MLP	0.9777	0.9751	0.9771	0.9761	0.9101	0.8828	0.9465	0.9135
	Ensemble	0.9898	0.9893	0.9892	0.9892	0.953	0.9593	0.9465	0.9529
ResNet-50	SVM	0.9938	0.9937	0.9937	0.9937	0.9248	0.9877	0.861	0.92
	KNN	0.9932	0.9928	0.9926	0.9927	0.9383	0.9409	0.9358	0.9383
	MLP	0.9894	0.9887	0.9888	0.9887	0.9342	0.9357	0.9332	0.9344
	Ensemble	0.9941	0.995	0.995	0.995	0.9503	0.9746	0.9251	0.9492
MobileNet-v2	SVM	0.9876	0.9873	0.9866	0.9869	0.9007	0.963	0.8342	0.894
	KNN	0.9887	0.9884	0.9883	0.9883	0.9503	0.9469	0.9545	0.9507
	MLP	0.9769	0.9748	0.9757	0.9753	0.906	0.9064	0.9064	0.9064
	Ensemble	0.9898	0.9892	0.9895	0.9893	0.9342	0.9577	0.9091	0.9328
VGG-16 + ResNet-50	SVM	0.9897	0.9896	0.9894	0.9895	0.9248	0.9877	0.861	0.92
	KNN	0.9941	0.9938	0.9935	0.9936	0.9624	0.9651	0.9599	0.9625
	MLP	0.9903	0.9889	0.9905	0.9897	0.953	0.952	0.9545	0.9533
	Ensemble	0.9941	0.9938	0.994	0.9939	0.9597	0.9804	0.9385	0.959
VGG-16 + MobileNet-v2	SVM	0.9856	0.9853	0.9848	0.985	0.9436	0.9882	0.8984	0.9412
	KNN	0.9935	0.9933	0.9928	0.993	0.9705	0.9681	0.9733	0.9707
	MLP	0.9875	0.9862	0.9875	0.9869	0.9477	0.9373	0.9599	0.9485
	Ensemble	0.9929	0.9928	0.9925	0.9926	0.9678	0.9861	0.9492	0.9673
ResNet-50 + MobileNet-v2	SVM	0.994	0.9939	0.994	0.9939	0.9195	0.9846	0.8529	0.914
	KNN	0.9937	0.9933	0.9934	0.9933	0.9436	0.9462	0.9412	0.9437
	MLP	0.9893	0.9882	0.9892	0.9887	0.9503	0.9566	0.9439	0.9502
	Ensemble	**0.9951**	0.9948	0.9951	0.995	0.9638	0.9806	0.9465	0.9633
Combined	SVM	0.9902	0.99	0.9901	0.99	0.9772	0.9786	0.9759	0.9772
	KNN	0.9945	0.9942	0.9941	0.9941	0.9745	0.9708	0.9786	0.9747
	MLP	0.9912	0.9899	0.9914	0.9906	0.9718	0.9783	0.9652	0.9717
	Ensemble	0.9947	0.994	0.9952	0.9946	**0.9842**	0.9839	0.9786	0.9812

Table 8 : Average Execution Time in Minutes

Process	COVIDx-2A CT dataset	SARS-CoV-2 CT-Scan Dataset
Preprocessing	27.18	0.41
Feature Extraction	205.32	2.87
Feature concatenation	9.37	0.08
PSO selection	399.37	115.43
GA selection	176.94	2.11
Classification Without selection	655.72	0.73
Classification with PSO selection	126.05	0.36
Classification with GA selection	167.85	0.27

Table 9 : Comparative analysis of accuracy in prior research

Ref	Dataset	Accuracy
[18]	COVIDx CT-2A	0.945
[17]	COVIDx CT-2A	0.981
[19]	COVIDx CT-2A	0.992
[20]	COVIDx CT-2A	0.9887
[22]	SARS-COV-2 CT-Scan	0.975
[15]	SARS-COV-2 CT-Scan	0.973
[23]	SARS-COV-2 CT-Scan	0.9942
Proposed	COVIDx CT-2A	0.9957
Proposed	SARS-COV-2 CT-Scan	0.9812

References

[1] G. Lippi, M. Plebani, and B. M. Henry, “Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis,” *Clinica chimica acta,* vol. 506, pp. 145-148, 2020.

[2] Y. Wu, X. Xu, Z. Chen, J. Duan, K. Hashimoto, L. Yang, C. Liu, and C. Yang, “Nervous system involvement after infection with COVID-19 and other coronaviruses,” *Brain, behavior, and immunity,* vol. 87, pp. 18-22, 2020.

[3] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z. A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, and K. Li, “Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection,” *Radiology,* vol. 295, no. 3, pp. 685-691, 2020.

[4] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, “Inf-net: Automatic covid-19 lung infection segmentation from ct images,” *IEEE transactions on medical imaging,* vol. 39, no. 8, pp. 2626-2637, 2020.

[5] E. Hussain, M. Hasan, M. A. Rahman, I. Lee, T. Tamanna, and M. Z. Parvez, "CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images," *Chaos, Solitons & Fractals,* vol. 142, pp. 110495, 2021.

[6] E. D. Carvalho, O. Antonio Filho, R. R. Silva, F. H. Araujo, J. O. Diniz, A. C. Silva, A. C. Paiva, and M. Gattass, "Breast cancer diagnosis from histopathological images using textural features and CBIR," *Artificial intelligence in medicine,* vol. 105, pp. 101845, 2020.

[7] E. D. Carvalho, A. O. de Carvalho Filho, A. D. de Sousa, A. C. Silva, and M. Gattass, "Method of differentiation of benign and malignant masses in digital mammograms using texture analysis based on phylogenetic diversity," *Computers & Electrical Engineering,* vol. 67, pp. 210-222, 2018.

[8] A. S. V. de Carvalho Junior, E. D. Carvalho, A. O. de Carvalho Filho, A. D. de Sousa, A. C. Silva, and M. Gattass, "Automatic methods for diagnosis of glaucoma using texture descriptors based on phylogenetic diversity," *Computers & Electrical Engineering,* vol. 71, pp. 102-114, 2018.

[9] E. D. Carvalho, E. D. Carvalho, A. O. de Carvalho Filho, A. D. de Sousa, and R. de A. L. Rabúlo, 'COVID-19 diagnosis in CT images using CNN to extract features and multiple classifiers', in 2020 IEEE 20th international conference on bioinformatics and bioengineering (BIBE), 2020, pp. 425–431.

[10] S. Yu, L. Xie, L. Liu, and D. Xia, "Learning long-term temporal features with deep neural networks for human action recognition," *IEEE Access,* vol. 8, pp. 1840-1850, 2019.

[11] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: a survey," *IEEE Signal Processing Magazine,* vol. 35, no. 1, pp. 84-100, 2018.

[12] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] A. Al-Dhamari, R. Sudirman, and N. H. Mahmood, "Transfer deep learning along with binary support vector machine for abnormal behavior detection," *IEEE Access,* vol. 8, pp. 61085-61095, 2020.

[14] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging,* vol. 9, pp. 611-629, 2018.

[15] A. Basu, K. H. Sheikh, E. Cuevas, and R. Sarkar, "COVID-19 detection from CT scans using a two-stage framework," *Expert Systems with Applications,* vol. 193, pp. 116377, 2022.

[16] M. A. Al-Qaness, A. A. Ewees, H. Fan, and M. Abd El Aziz, "Optimization method for forecasting confirmed cases of COVID-19 in China," *Journal of clinical medicine,* vol. 9, no. 3, pp. 674, 2020.

[17] H. Gunraj, A. Sabri, D. Koff, and A. Wong, "COVID-Net CT-2: Enhanced deep neural networks for detection of COVID-19 from chest CT images through bigger, more diverse learning," *Frontiers in Medicine,* vol. 8, pp. 729287, 2022.

[18] H. Gunraj, L. Wang, and A. Wong, "Covidnet-ct: A tailored deep convolutional neural network design for detection of covid-19 cases from chest ct images," *Frontiers in medicine,* vol. 7, pp. 608525, 2020.

[19] W. Zhao, W. Jiang, and X. Qiu, "Deep learning for COVID-19 detection based on CT images," *Scientific Reports,* vol. 11, no. 1, pp. 14353, 2021.

[20] A. Loddo, F. Pili, and C. Di Ruberto, "Deep learning for COVID-19 diagnosis from CT images," *Applied Sciences,* vol. 11, no. 17, pp. 8227, 2021.

[21] A. K. Mishra, S. K. Das, P. Roy, and S. Bandyopadhyay, "Identifying COVID19 from chest CT images: a deep convolutional neural networks based approach," *Journal of Healthcare Engineering,* vol. 2020, 2020.

[22] E. D. Carvalho, R. R. Silva, F. H. Araújo, R. de AL Rabelo, and A. O. de Carvalho Filho, "An approach to the classification of COVID-19 based on CT scans using convolutional features and genetic algorithms," *Computers in biology and medicine,* vol. 136, pp. 104744, 2021.

[23] A. Dey, S. Chattopadhyay, P. K. Singh, A. Ahmadian, M. Ferrara, N. Senu, and R. Sarkar, "MRFGRO: a hybrid meta-heuristic feature selection method for screening COVID-19 using deep features," *Scientific reports,* vol. 11, no. 1, pp. 24065, 2021.

[24] E.-S. M. El-Kenawy, A. Ibrahim, S. Mirjalili, M. M. Eid, and S. E. Hussein, "Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images," *IEEE access,* vol. 8, pp. 179317-179335, 2020.

[25] E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, “Detecting COVID-19 in chest CT images based on several pre-trained models,” *Multimedia Tools and Applications*, pp. 1-21, 2024.
[26] L. K. Singh, M. Khanna, H. Monga, and G. Pandey, “Nature-inspired algorithms-based optimal features selection strategy for COVID-19 detection using medical images,” *New Generation Computing*, pp. 1-64, 2024.
[27] M. M. Hossain, M. A. A. Walid, S. S. Galib, M. M. Azad, W. Rahman, A. Shafi, and M. M. Rahman, “Covid-19 detection from chest ct images using optimized deep features and ensemble classification,” *Systems and Soft Computing,* vol. 6, pp. 200077, 2024.
[28] H. Gunraj, A. Sabri, D. Koff, and A. Wong, "COVID-Net open source initiative-COVIDx CT-2 dataset," 2020. Available online: *https://www.kaggle.com/hgunraj/covidxct* (accessed on 30 October 2021).
[29] E. Soares, P. Angelov, S. Biaso, M. H. Froes, and D. K. Abe, “SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification,” *MedRxiv*, pp. 2020.04. 24.20078584, 2020.
[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*: MIT press, 2016.
[31] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep residual learning for image recognition’, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: inverted residuals and linear bottlenecks. In proceedings of the 2018 IEEE conference on computer vision and pattern recognition," IEEE Salt Lake City, America, 2018.
[33] J. Kennedy and R. Eberhart, ‘Particle swarm optimization’, in Proceedings of ICNN’95-international conference on neural networks, 1995, vol. 4, pp. 1942–1948.
[34] B. Ji, X. Lu, G. Sun, W. Zhang, J. Li, and Y. Xiao, “Bio-Inspired Feature Selection: An Improved Binary Particle Swarm Optimization Approach,” *IEEE Access,* vol. PP, pp. 1-1, 05/06, 2020.
[35] J. Yang, and V. Honavar, “Feature subset selection using a genetic algorithm,” *IEEE Intelligent Systems and their Applications,* vol. 13, no. 2, pp. 44-49, 1998.
[36] V. Vapnik, "The support vector method of function estimation," *Nonlinear modeling: Advanced black-box techniques*, pp. 55-85: Springer, 1998.
[37] P. Naraei, A. Abhari, and A. Sadeghian, ‘Application of multilayer perceptron neural networks and support vector machines in classification of healthcare data’, in 2016 Future technologies conference (FTC), 2016, pp. 848–852.
[38] M. Wageh, K. M. Amin, A. Zytoon, and M. Ibrahim, “Brain Tumor Detection Based on A Combination of GLCM and LBP Features with PCA and IG,” *IJCI. International Journal of Computers and Information,* vol. 10, no. 2, pp. 43-53, 2023.

المجلد (١٢) العدد (١) (لسنة ٢٠٢٥)
المجلة الدولية للحاسبات والمعلومات
متاح الإنترنت على هذا الرابط : https://ijci.journals.ekb.eg/

الكشف عن كوفيد -١٩ بناء على الأشعة المقطعية باستخدام طريقة اختيار الميزات الوصفية

عبدالغنى فتحي ، حاتم عبدالقادر ، أميرة عبدالعاطي

قسم نظم المعلومات – كلية الحاسبات والمعلومات – جامعة المنوفية

abdalghany.fathy@ci.menofia.edu.eg, hatem.abdelkader@ci.menofia.edu.eg, amira.abdelatey@ci.menofia.edu.eg

ملخص البحث

قد استلزم الانتقال السريع والواسع النطاق لكوفيد-١٩ تطوير أدوات تشخيصية فعالة. في حين يظل تفاعل البوليميراز المتسلسل هو الطريقة القياسية للتشخيص ، فإن حدوده من حيث الوقت وكثافة الموارد تسلط الضوء على الحاجة إلى حلول بديلة. تعالج هذه الدراسة هذه الفجوة من خلال اقتراح منهجية هجينة من ثلاث مراحل للتعرف السريع على كوفيد -١٩ باستخدام الأشعة المقطعية. في المرحلة الأولى ، يتم استخدام الشبكات العصبية التلافيفية المدربة مسبقا ، بما في ذلك VGG-16 ، *ResNet-50*، و *MobileNet v2* ، لاستخراج الميزات ذات الصلة من الرئتين المتضررة من كوفيد-١٩. المرحلة الثانية تعزز اختيار ميزة من خلال تطبيق تقنيات الكشف عن مجريات الأمور الفوقية مثل الخوارزميات الجينية (*GA*) و سرب الجسيمات الأمثل (*PSO*) ، وتحسين مجموعة ميزة لتحسين الدقة. وأخيرا ، يتم تصنيف الميزات المختارة باستخدام أربعة مصنفات متميزة ، وتحقيق دقة تصنيف ملحوظة تصل إلي ٩٩,٥٧ ٪ و ٩٨,٤٢ ٪ على مجموعات البيانات *COVIDx-2A CT* و *SARS-CoV-2 CT-Scan* ، على التوالي. تكمن حداثة هذا النهج في دمج عدة شبكات عصبية تلافيفية *CNNs* والأساليب الفوقية لتحسين أداء اختيار الميزات والتصنيف. تتضمن مساهماتنا تطوير أداة تشخيصية قوية تُحسن بشكل كبير من سرعة ودقة الكشف عن كوفيد -١٩، مما يوفر بديلاً فعالًا لطرق *RT-PCR* التقليدية.