



**Journal of Engineering  
Sciences Assiut University  
Faculty of Engineering**

**Vol. 49, No. 4**

**July 2021**

**PP. 476–508**



## **AN INTELLIGENT DETECTION SYSTEM FOR COVID-19 DIAGNOSIS USING CT-IMAGES**

**Amira M. Hasan<sup>1</sup>, Hala M. Abd El-Kader<sup>2</sup>, Aya Hossam<sup>3</sup>**

<sup>1</sup> *Engineer, Electrical Engineering Department  
Alexandria Higher Institute of Engineering Technology (AIET), Alex, Egypt  
E-mail: [amira.mohamed@aiet.edu.eg](mailto:amira.mohamed@aiet.edu.eg)*

<sup>2</sup> *Professor, Electrical Engineering, Department, Faculty of Engineering  
(Shoubra), Benha University, Cairo, Egypt  
E-mail: [hala.mansour@feng.bu.edu.eg](mailto:hala.mansour@feng.bu.edu.eg)*

<sup>3</sup> *Lecturer, Electrical Engineering, Department, Faculty of Engineering  
(Shoubra), Benha University, Cairo, Egypt  
E-mail: [Aya.ahmed@feng.bu.edu.eg](mailto:Aya.ahmed@feng.bu.edu.eg)*

*Received 1 February 2021; Revised 29 May 2021; Accepted 2 June 2021*

### **Abstract**

Early classification of the Coronavirus disease (COVID-19) is necessary to control its rapid spread and save patients' lives. The fast spread of COVID-19 has increased the diagnostic encumbrance of radiologists. Therefore, clinicians need to quickly assess if a patient has COVID-19 or not. Artificial Intelligence (AI) has shown promising results in healthcare. So, this paper proposed a computer-aided intelligence model that can identify positive COVID-19 cases. It presented the pipeline of medicinal imaging and examination methods involved in COVID-19 image acquirement, segmentation, and diagnosis, using Computed Tomography (CT) images. This paper introduced two effective models for single machine learning (SML) and ensemble machine learning (EML) with 10-fold cross validation, to detect cases of COVID-19. The first

classification model (SML) was applied with different algorithms, such as Decision Tree (DT), Artificial Neural Networks (ANN), and Support Vector Machines (SVM). Results showed that the performance of the SVM surpassed other classifiers with a 98.8% accuracy. The second classification model (EML) was applied with several algorithms, such as Random Forest (RF), Voting, and Bagging, to increase its accuracy up to 99.60%, especially using the Bagging classifier. Finally, the results of the two proposed models showed better performance compared with other recent studies. However, the EML showed an even better performance than SML and is recommended for use in real-time.

**Keywords:** Artificial intelligence (AI), COVID-19, Machine learning (ML), Segmentation method, Ensemble Machine Learning (EML).

---

## 1. Introduction

COVID-19 is produced by an infective virus, which causes severe lung disease. It was named Coronavirus because it is seen in the form of a crown under the electron microscope [1]. This Virus infects both humans and animals [2]. The strain of Coronavirus that first started in Wuhan, China, caused serious symptoms, including respiratory complications (pneumonia), renal disorder, and fluid accumulation in the lungs [3]. The World Health Organization (WHO) publicized this on-going outbreak as a worldwide health crisis on the 30<sup>th</sup> of January 2020 and raised the danger of infection to very high, on a global level in March 2020 [4]. On March 2, 88,948 COVID-19 cases, with 3,043 deaths, were reported, of which 80,174 were in China and 8,774 in 64 other countries, as shown in Figure (1) [5]. On June 24, 2020, while writing this paper, the total number of confirmed COVID-19 cases had reached 103,213,392, including about 2,231,154 deaths, as reported to the WHO. In Egypt, there were 164,871 confirmed cases and about 9,217 deaths at that time. Due to this large number of cases and deaths, medical imagery seems to be an essential tool for detecting cases, in CT and X-ray images [6]. Current studies addressed the necessity of medical image examination in COVID-19 patients. Compared to chest X-ray, CT has clearer information. It also offers better accuracy, because of its high sensitivity and low noise. In recent studies, dynamic radiological patterns in CT were identified and summarized in three phases for COVID-19 [7]. Phase one, zero to four days of the initial symptoms, presents the early

stage of the disease. Sub-pleural fluid causing ground-glass opacity (GGO) in the inferior portion of the lungs may be detected individually or together [8]. The progressive stage of the disease takes place between days five and eight of symptoms. In this phase, GGO, the crazy-paved pattern, and even consolidated glass are formed. During the peak period (nine to thirteen days), dense consolidations appear [9]. Consolidation and the crazy-paved patterns are gradually absorbed and only the GGO is left. Important CT features include GGO, consolidation, and reticulation/thickening of interlobular septa. In the intermediate stage (4–14 days) mad-paved patterns and probable Halo signs have also become essential features [10]. Isolating COVID-19-positive persons is of vital importance through observing Halo signs. Machine learning (ML) is necessary in advancing learning algorithms, which learn mechanically without humanoid assistance. ML techniques include rotation learning, significant learning, analog learning, and inductive learning.

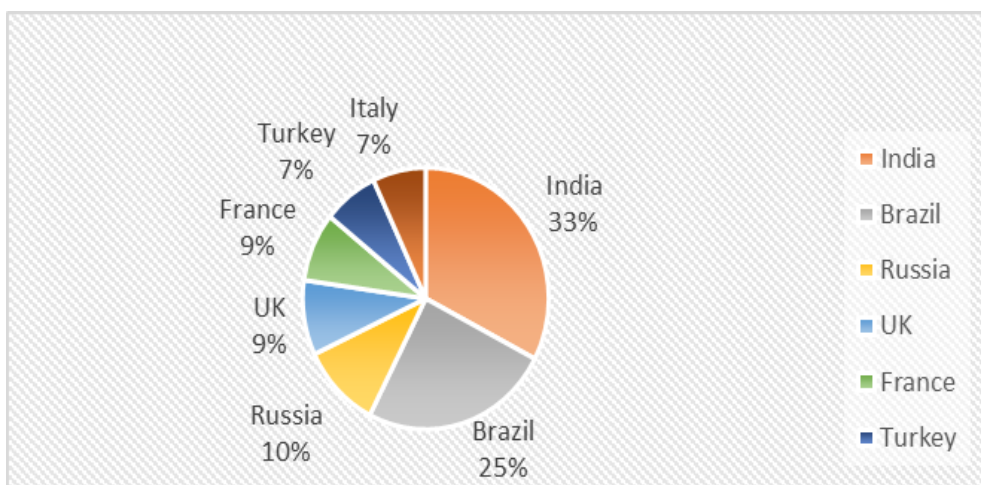


Figure (1) Distribution of Coronavirus-infected cases in several countries [5].

A radiologist would use an ML broadcast system, firstly to recognize cases of COVID-19 and secondly to compute and segment the manifestations of the infection in CT scans. So, there are specific key points discussed in this paper, which are summarized, as follows:

- This paper suggests a new learning procedure for segmentation and extraction of region of interest (ROI), which results in better positioning of CT images.

- This paper applies SML and EML algorithms to improve and increase the accuracy of the classification models, the application of the two proposed models has indicated that they are highly accurate. It is necessary to construct an automated model based on the ML approach to initiate a corrective action, following the decision of the treating physician. Supervised machine learning techniques were used for classifying cases for two categories, COVID-19, and non-COVID-19. An ensemble of learning techniques was also applied for classification. Section one of this paper presents the introduction; Section two provides the literature survey related to the proposed work. The framework for detecting Coronavirus from clinical text data is discussed in Sections three and four, along with the experimental results of the proposed framework. Sections five and six provide the conclusion of this paper.

## **2. Related Works**

Machine learning and natural language processing use big data-based models for pattern recognition, explanation, and prediction. According to [Bullock et al, \[11\]](#), machine learning and deep learning can replace humans by giving an accurate diagnosis. The perfect diagnosis can save radiologists' time and can be cost-effective than standard tests for COVID-19. X-rays and CT scans can be used for training machine learning models. Several initiatives are underway in this regard. [Wang and Wong et al, \[12\]](#) developed COVID-Net, a deep convolutional neural network, which can diagnose COVID-19 from chest radiography images. Once a case of COVID-19 is detected, the question becomes whether and how intensively that person will be affected. Not all COVID-19 positive patients will need rigorous attention. Being able to prognosticate who will be affected more severely can help in directing assistance and planning medical resource allocation and utilization. [Yan et al, \[13\]](#) used machine learning to develop a prognostic prediction algorithm to predict the mortality risk of a person, who has been infected, using data from 29 patients only, at Tongji Hospital in Wuhan, China. [EL-Bana et al, \[14\]](#) proposed a multi-task pipeline that takes advantage of the growing advances in neural network models for COVID-19 recognition, using X-ray and CT scans. Since less work is being done on diagnosis and prediction using medical images, machine learning and

ensemble learning models were used in this paper to classify clinical reports for the detection of COVID-19 cases.

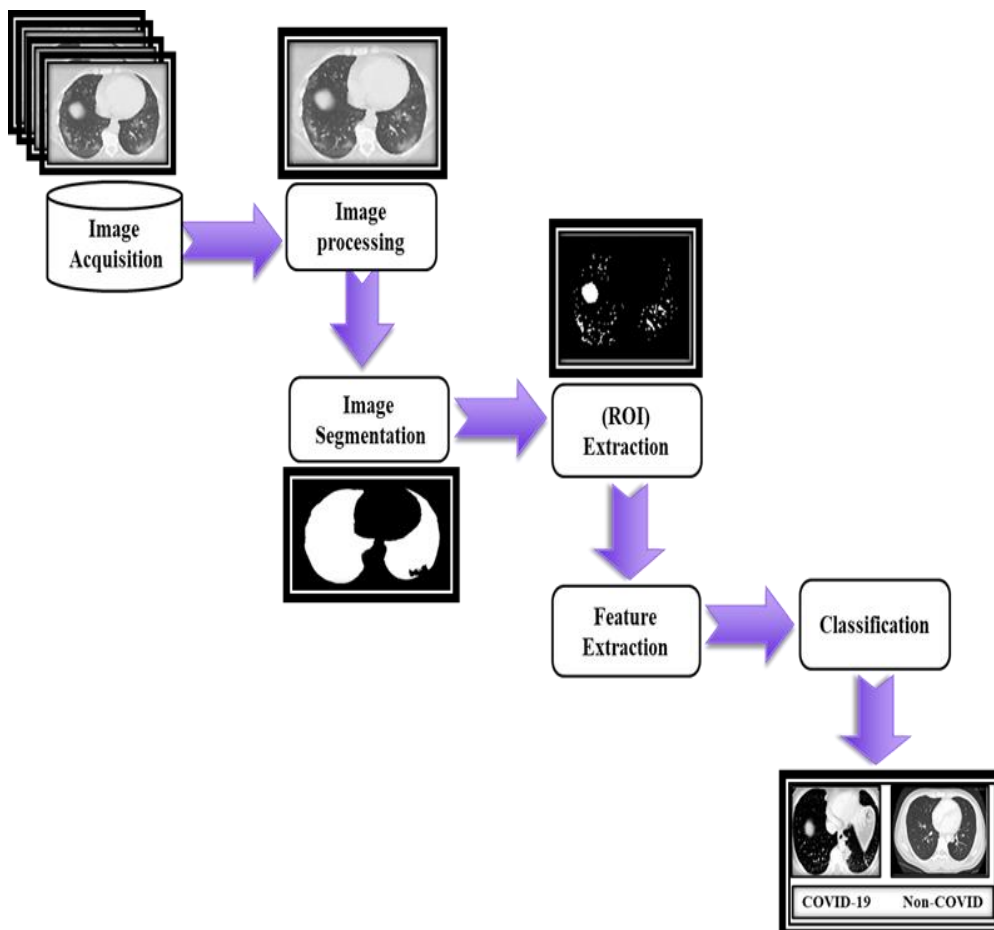


Figure (2) - Graphical diagram of the proposed model

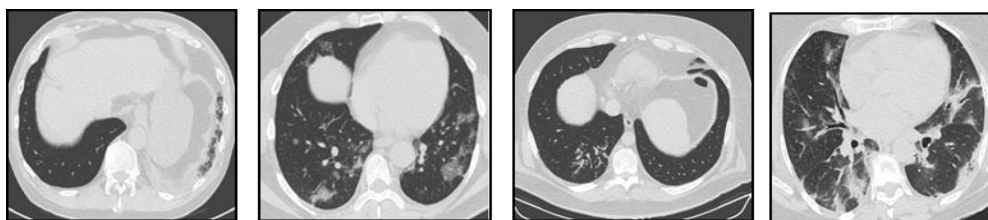
### 3. The Proposed Model

A machine learning-empowered system for COVID-19 diagnostics inherently involves multiple steps as shown in Figure (2). The system would first be required to recognize COVID-19 in CT scans. In this paper, authors have used a comparatively large dataset of COVID-19 and non-COVID-19 CT images. First, the Wiener filter and Histogram Equalization (HE) for image enhancement were employed. Second, the segmentation step was applied. It was divided into two categories, lung separation and ROI detection. Otsu's Thresholding and Morphological Processing were employed to isolate lung regions. Then, Edge Detection

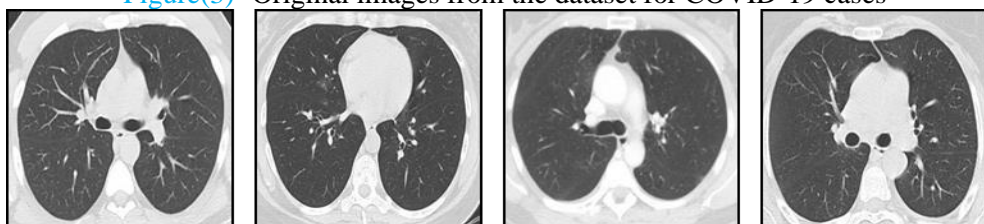
and Active Contour algorithms were applied. They produced lesion-segmentation masks, which highlighted the regions corresponding to infection manifestations (ROI). In the third step, all infection features were extracted from segmented images of ROI. Finally, to evaluate the proposed models, these features were fed into SML and EML classifiers to build the classification models.

### 3.1. Image Acquisition

The main stage of any visualization system is the acquisition of CTs. This is because if the image is not clear, then the expected tasks cannot be carried out. Samples of COVID-19 and non-COVID-19 CT images from a dataset as shown in Figure (3) and in Figure (4), respectively.



Figure(3)- Original images from the dataset for COVID-19 cases



Figure(4)- Original images from the dataset for Non-COVID-19 cases

### 3.2. Image Pre-processing

Image pre-processing techniques are used to enhance ROI and to eliminate erroneous data which are of no interest such as noise and body movements which affects medical images.

#### A. Wiener Filter

The Wiener filter is a commonly used optimization technique for cancel the Gaussian noise and improve the blurring edges of the image. For optimum segmentation, wiener filter of size 3\*3 is used to remove the noise present in the CT image. Gaussian noise is caused by the poor light or high temperature in the image where  $x$  and  $y$  are the position of every pixel on the image,  $f(x, y)$  which denotes the original input image is the



pixel for every position. After the Gaussian noise is added, the image is defined as follows:

$$S(x, y) = F(x, y) + G(x, y) \quad (1)$$

Meanwhile, Wiener filter is often used to cancel the Gaussian noise and better solve the blurring edge for image segmentation. Wiener Filter function here is defined by:

$$W(x, y) = F(S(x, y)) H(S(x, y)) \quad (2)$$

F is the Fourier transform for the input image and H is the blurring function. Wiener filter is used the linear estimation to make the mean square error (MSE) between the W (x, y) and F (x, y) minimal, i.e., the Gaussian noise G (x, y) is removed. This filter gives an estimate of the original uncorrupted image with minimum mean square error which is the nonlinear function of the corrupted image. The error metric is used to evaluate the filter with mean square error (MSE), which is the cumulative squared error between the compressed and the original image. A lower value for MSE means lesser error [15].

### *B. Histogram Equalization*

HE is an effective method used to improve image contrast. This is achieved by spreading out the most frequent values of intensity [16]. This is done by extending the range of image intensity. This method increases the overall image contrast, as shown in [Figure \(5\)](#).

### **3.3. Image Segmentation**

The segmentation step is necessary to isolate the lung structure from all other untargeted structures. This is done to extract the target of ROI. Segmentation methods can be mainly grouped into two categories: lung-region-oriented techniques and lung-lesion-oriented techniques.

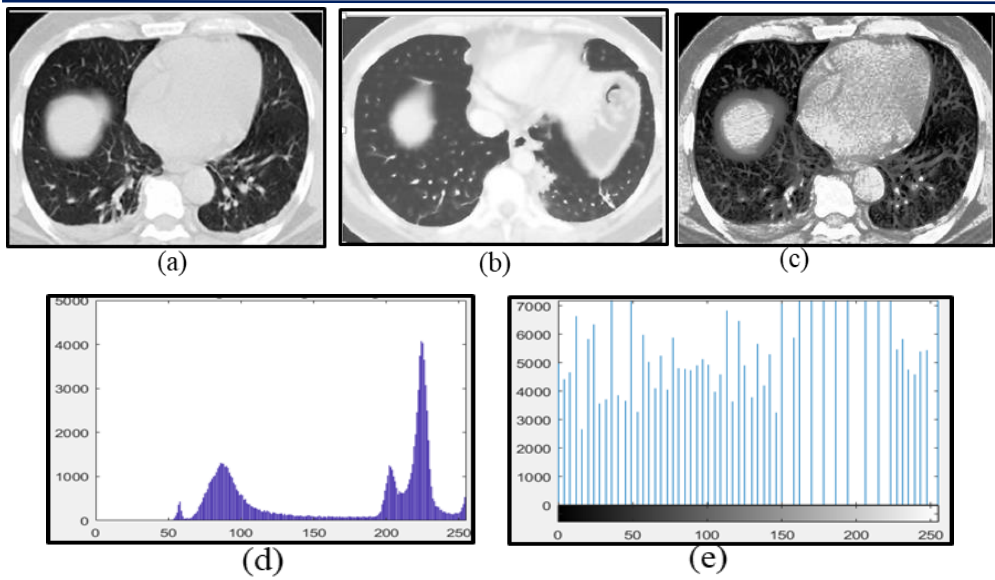


Figure (5)- (a) Original image, (b) Wiener filter, (c) Histogram equalized image, (d) Histogram of the original image, (e) Histogram of the equalized image

### 3.3.1. Lung-Region-Oriented Methods

These techniques aim to isolate lung regions, as complete lungs, and lobes, from other regions in the CT. These techniques help clinicians to distinguish the size of only lung because this size distinguishes which one stage of COVID-19 the patient had, as shown in Figure (6) [17].



Figure (6)-(a) Original image; (b) The image after Thresholding; (c) Lung mask.

#### A. Otsu's Thresholding

It is a powerful instrument for segmentation. This process converts a gray image into a binary one, where binary steps are given to pixels below or above a certain threshold value [18]. A threshold image  $g$  is defined as:



$$g(x, y) = \begin{cases} 1 & \text{IF } F(x, y) > T \\ 0 & \text{IF } F(x, y) \leq T \end{cases} \quad (3)$$

$g(x, y)$  = output image;  $F(x, y)$  = input image;  $T$  = threshold value

### B. Morphological Processing

Morphological processing is a collection of processes associated with the shape of the features of COVID-19. These operations rely solely on the relative ordering of pixel values, not on their numerical values, and are therefore particularly suited for smoothing the contour of the features and removing thin protrusions [19].

#### 3.3.2. Lung-Lesion-Oriented Methods

These methods are applied to separate lesions from other lung regions [20]. These lesions could be small with a different shape and texture from uninfected parts of the lung, as shown in Figure (7).

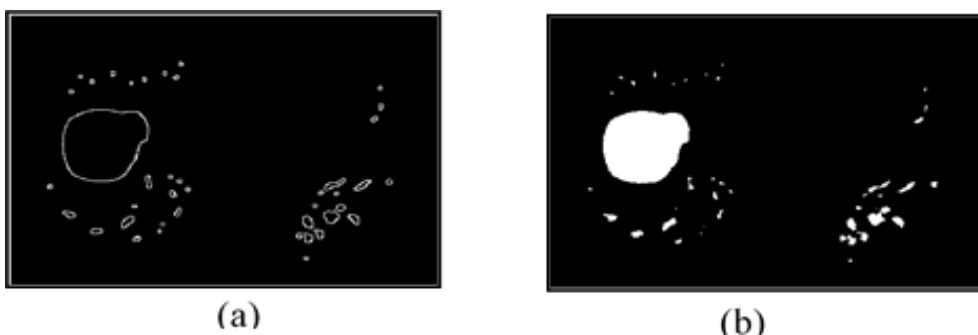


Figure (7) - (a) Output of edge detection; (b) Lesion mask (ROI).

### A. Edge Detection

Boundary processing is necessary to distinguish lung boundaries. So, the recognition of edges, using the canny method, is an important step [21].

### B. Active Contour

This is the procedure of generating mini regions with joint boundaries. It depends on moving deformed contours using forces, which help resist deformation, shape path boundaries, and nodes, accurately [22].

In segmentation steps, we applied Otsu's Thresholding, the morphological process and active contour algorithms to extract ROI, as shown in Figure (8).

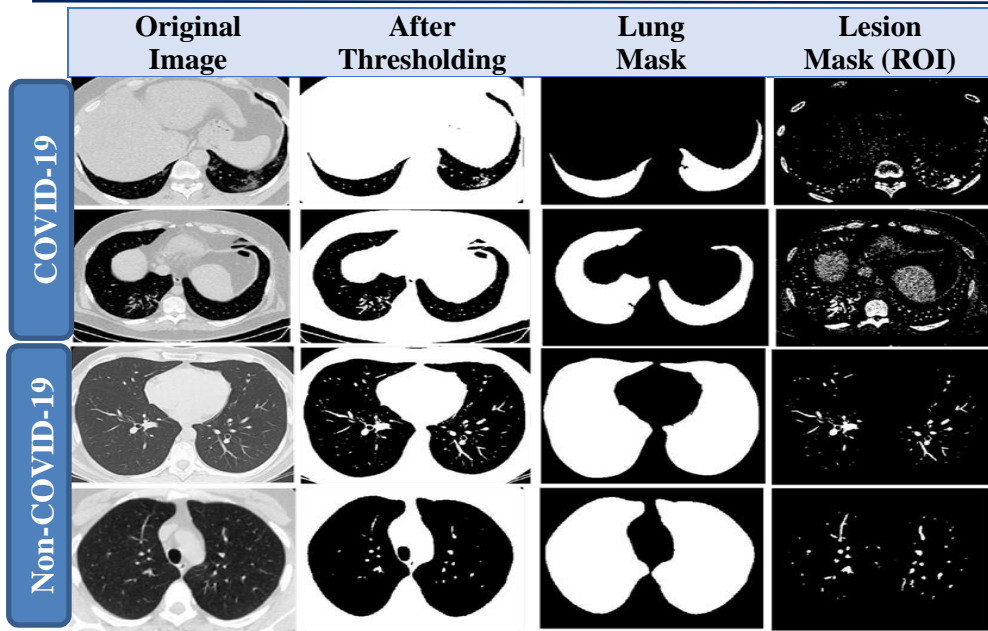


Figure (8) - (a) The Original image, (b) Thresholding (c) Lung mask, (d) Lesion mask (ROI)

### 3.4. Feature Extraction Techniques

The feature extraction stage is necessary for processing methods that use techniques to isolate various desired portions or forms (features) of an image. A feature is a vital part of information which extracted from the enhanced ROI of COVID-19 image. A set of three features namely, intensity histogram, gray-level co-occurrence matrix (GLCM) and gray-level run-length matrix (GLRLM), were extracted from each of the total images of the dataset.

#### 3.4.1. Intensity Histogram Features

The intensity-level histogram is a function showing (for each intensity level) the number of pixels in the whole image, which have this intensity. Here  $Z$  represents the pixel intensity and  $Z(S)$  represents the pixel intensity of  $S$  value. The required statistical features such as mean, variance, std division, kurtosis, Skewness and Entropy are listed in [Table \(1\)](#). Discrete values  $S = 0, 1 \dots L-1$ ,  $L$  is the total number of intensity levels in the image [23].

**Table (1):** Intensity Histogram features

| Features            | Formula  |
|---------------------|--|
| <b>Mean</b>         | $m = \sum_{S=0}^{L-1} S Z(S)$ (4)                                |
| <b>Variance</b>     | $\mu_v = \sum_{S=0}^{L-1} (S - m)^2 Z(S)$ (5)                    |
| <b>Std Division</b> | $\sigma = \sqrt{\mu_v}$ (6)                                      |
| <b>Kurtosis</b>     | $\mu_k = \frac{1}{\sigma^4} \sum_{S=0}^{L-1} (S - m)^4 Z(S)$ (7) |
| <b>Skewness</b>     | $\mu_s = \frac{1}{\sigma^3} \sum_{S=0}^{L-1} (S - m)^3 Z(S)$ (8) |
| <b>Entropy</b>      | $H = \sum_{S=0}^{L-1} Z(S) \log_2 Z(S)$ (9)                      |

### 3.4.2. Grey-Level Co-occurrence Matrix (GLCM)

The texture features have been deduced from the Gray Level Co-occurrence Matrix (GLCM). For texture features, the texture of an image region is calculated according to the gray level distribution over the pixels in the region. The GLCM is a statistical technique for investigating texture that takes the spatial relationship of pixels into account. It is a two-dimensional array that considers a pixel's specific position in relation to neighboring pixels. The co-occurrence matrix P (I, J) calculates the co-occurrence of the pixels with the values I, J. The concept of this method is based on determining the second order statistical probability values represented by the matrix element P (I, J | d,  $\theta$ ) based on variations in grey levels I and J at a given distance d and angle ( $\theta$ ). Each element (I, J) of GLCM represents the addition of the times that the pixel with the value I placed in the specific spatial relationship to a pixel with value J in the input COVID-19 image. Several texture features can be extracted from GLCM with parameters' values d=1 and ( $\theta$ ) are computed in four directions: horizontal, vertical, and two diagonal directions that are commonly used.  $P_x(I) - I^{th}$  Input in the marginal-probability matrix acquired by adding the rows of P (I, J).  $P_y(J) - J^{th}$  input in the marginal-probability matrix acquired by adding the columns of P (I, J).  $\mu_x, \mu_y, \sigma_x$  and  $\sigma_y$  are the means and standard deviations of  $P_x$  and  $P_y$  [24]. Finally, the extracted features are listed in Table (2). Firstly, the following notations are used in various formulas of GLCM features:

$$P_x(I) = \sum_{j=0}^{N-1} P(I, J) \tag{10}$$

$$P_y(J) = \sum_{l=0}^{N-1} P(I, J) \tag{11}$$

$$P_{x+y}(k) = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} P(I, J) \cdot (I + J) = k, \quad k = 2.3.4. \dots .2N \tag{12}$$

$$P_{x-y}(k) = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} P(I, J) \cdot |I - J| = k, \quad k = 0.1. \dots .N - 1 \tag{13}$$

$$\mu_x = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} lP(I, J) = \sum_{l=0}^{N-1} lP_x(I) \tag{14}$$

$$\mu_y = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} jP(I, J) = \sum_{j=0}^{N-1} jP_y(J) \tag{15}$$

$$\sigma_x = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} (I - \mu_x)^2 P(I, J) \tag{16}$$

$$\sigma_y = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} (J - \mu_y)^2 P(I, J) \tag{17}$$

Entropies of  $P_x$  and  $P_y$

$$HXY = - \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} P(I, J) \log (P(I, J)) \tag{18}$$

$$HXY1 = - \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} P(I, J) \log (P_x(I)P_y(I)) \tag{19}$$

$$HXY2 = - \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} P_x(I)P_y(I) \log (P_x(I)P_y(I)) \tag{20}$$

Table (2): Features of GLCM matrix

| Features           | Formula   |
|--------------------|---|
| Energy             | $\sum_{l=0}^{N-1} \sum_{j=0}^{N-1} P(I, J)^2$ (21)  |
| Contrast           | $CON = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} (I - J)^2 P(I, J)$ (22)  |
| Correlation        | $CORR = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} \frac{(I \times J) \times P(I, J) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$ (23) |
| Cluster Shade      | $CH = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} \{I + J - \mu_x - \mu_y\}^3 P(I, J)$ (24)   |
| Cluster Prominence | $CP = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} \{I + J - \mu_x - \mu_y\}^4 P(I, J)$ (25)   |
| Homogeneity        | $HOM = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(I, J)}{1 + (I - J)^2}$ (26)  |
| Dissimilarity      | $DIS = \sum_{l=0}^{N-1} \sum_{j=0}^{N-1}  I - J  P(I, J)$ (27)  |

|                                      |   |
|--------------------------------------|---|
| Maximum Probability                  | $MAX = \max_{I,J} P(I, J)$ (28)   |
| Inverse Difference Normalized        | $IDN = \sum_{I=0}^{N-1} \sum_{J=0}^{N-1} \frac{P(I, J)}{1+ I-J /N^2}$ (29)      |
| Inverse Difference Moment Normalized | $INN = \sum_{I=0}^{N-1} \sum_{J=0}^{N-1} \frac{P(I, J)}{1+(I-J)^2/N^2}$ (30)    |
| Information measure of correlation1  | $IMC1 = \frac{HXY - HXY1}{\max\{HX, HY\}}$ (31)                                 |
| Information measure of correlation2  | $IMC2 = (1 - \exp[-2(HXY2 - HXY)])^{\frac{1}{2}}$ (32)                          |
| Autocorrelation                      | $AUT = \sum_{I=0}^{N-1} \sum_{J=0}^{N-1} (IJ)P(I, J)$ (33)                      |
| Difference Entropy                   | $Diff\ entropy \sum_{I=0}^{N-1} P_{x-y}(I) \log (P_{x-y}(I))$ (34)              |
| Sum Entropy                          | $Sum\ entropy = \sum_{I=0}^{2N-2} P_{x+y}(I) \log (P_{x+y}(I))$ (35)            |
| Difference variance                  | $Diff\ variance = \sum_{I=0}^{N-1} I^2 P_{x-y}(I)$ (36)                         |
| Sum variance                         | $Sum\ variance \sum_{I=0}^{2N-2} (I - sum\ entropy)^2 P_{x+y}(I)$ (33)          |
| Sum of Squares                       | $Sum\ of\ squares = \sum_{I=0}^{N-1} \sum_{J=0}^{N-1} (I - \mu)^2 P(I, J)$ (38) |
| Sum average                          | $Sum\ average = \sum_{I=0}^{2N-2} I P_{x+y}(I)$ (38)                            |

### 3.4.3. Grey-Level Run-Length Matrix (GLRLM)

In GLRLM, statistic in concern is the number of pairs of gray level value and its length of runs in a certain ROI. A gray level run is a set of pixels having the same gray level value, which are consecutively and collinearly distributed in the ROI along some given directions. Number of pixels in that set is called the length of the gray level run. Thus, a gray level value and its length of a gray level run together characterize such a set. A GLRLM is a 2D histogram in form of a matrix that records the occurrence of all various combinations of gray level values and gray level runs in an ROI for a given direction. Conventionally, gray level values and gray level runs are denoted as keys of rows and columns, respectively of the matrix. A run length matrix R is defined as follows:

each element  $R(I, J)$  represents the number of runs with pixels of gray level intensity equal to  $I$  and run of length equal to  $J$  along a specific orientation. Hence, the  $(I, J)$ -th entry in the matrix specifies the number of combinations whose gray level value is  $I$  and whose run length is  $J$ . GLRLM is a 2D matrix in which element “ $R(I, J)$ ” gives the total number of consecutive runs of length “ $J$ ” at grey level “ $I$ ”. Here “ $M$ ” represents the number of gray levels and “ $K$ ” represents the maximum run length (here a run length is several neighboring pixels that possess the same grey intensity in a particular direction).  $n_r$  is the total number of runs and  $n_p$  is the number of pixels in the image [25]. GLRLM features are listed in Table (3).

**Gray level run length pixel number matrix:**

$$R_p(I, J) = R(I, J) \cdot J \quad (40)$$

Each element of the matrix represents the number of pixels of run length  $J$  and gray-level  $I$ . Compared to the original matrix, the new matrix gives equal emphasis to all length of runs in an image.

**Gray – level run – number vector:**

$$R_g(I) = \sum_{J=1}^K R(I, J) \quad (41)$$

This vector represents the sum distribution of the number of runs with gray level  $I$

**Run-length run-number vector:**

$$R_r(J) = \sum_{I=1}^M R(I, J) \quad (42)$$

This vector represents the sum distribution of the number of runs with run length  $J$ .



**Table (3):** Features of GLRLM matrix

| Features                            | Formula  |
|-------------------------------------|--|
| Short-Run Emphasis (SRE)            | $SRE = \frac{1}{n_r} \sum_{j=1}^K \frac{R_r(j)}{j^2} \quad (43)$ |
| Long-Run Emphasis (LRE)             | $LRE = \frac{1}{n_r} \sum_{j=1}^K R_r(j) \cdot j^2 \quad (44)$   |
| Gray-Level Non- uniformity (GLN)    | $GLN = \frac{1}{n_r} \sum_{l=1}^M R_g(l)^2 \quad (45)$           |
| Run Length Non- uniformity (RLN)    | $RLN = \frac{1}{n_r} \sum_{j=1}^K R_r(l)^2 \quad (46)$           |
| Percentage Run (PR)                 | $PR = n_r / n_p \quad (47)$                                      |
| Low Gray-Level Run Emphasis (LGRE)  | $LGRE = \frac{1}{n_r} \sum_{l=1}^M R_g(l) / l^2 \quad (48)$      |
| High gray-level Run Emphasis (HGRE) | $HGRE = \frac{1}{n_r} \sum_{l=1}^M R_g(j) \cdot l^2 \quad (49)$  |

### 3.5. Feature Selection (FS)

While feature extraction techniques are used to extract as many image parameters that identify COVID-19 characteristics as possible, a feature selection algorithm is required to select a subset of those extracted features that are most significant and best describe the COVID-19 characteristics. A total of 32 features were extracted from each image during the feature extraction process, but due to the large number of features, all these features cannot be supplied to classifiers. Although each feature is important in classification, only a few of them are extremely important in classifying and identifying disease conditions. In this technique, many feature selection algorithms were used such as Genetic Search Algorithm. Most feature selection algorithms employ a search method that spans the entire space. Many search methods calculate the importance of individual features and rank them accordingly and provide the best features from a given set [26]. This algorithm can simplify the data realization, decrease over fitting problem and the size of data storage. It also reduces over fitting, improves Accuracy, and reduces training time. This is accomplished by removing features that are irrelevant, redundant, or noisy. It selects the subset of features that will yield the best results in terms of precision and computation time. We reduce features to 26 features and use them as inputs for all classifiers. The selected features are listed in Table (4).

Table (4): Features selected by WEKA software

| Feature Category                    | NO of features Before FS | NO of features after FS | Selected Features   |
|-------------------------------------|--------------------------|-------------------------|---|
| <b>Intensity Histogram Features</b> | 6                        | 3                       | Variance, Skewness, Kurtosis  |
| <b>GLCM Features</b>                | 19                       | 18                      | Contrast, Correlation, Cluster Shade, Cluster Prominence, Homogeneity, Maximum probability, Sum of squares, Diff variance, Sum variance, Sum average, Inverse Difference Normalized, Inverse Difference Moment Normalized, Information measure of correlation-1, Information measure of correlation-2, Difference Entropy, Sum Entropy, Dissimilarity and Autocorrelation |
| <b>GLRLM Features</b>               | 7                        | 5                       | Short Run Emphasis, Long-Run Emphasis, Run Length Non-uniformity, Percentage Run and Low Gray-Level Run Emphasis  |
| <b>Total Features</b>               | <b>32</b>                | <b>26</b>               |   |

### 3.6. Classification

Classification is the last stage applied to distinguish cases of COVID-19. The relevant features were identified for this disease. They are used as inputs for building models. Classification is usually achieved through different ML models. In this paper, two classification categories have been used, namely the Single ML and the Ensemble ML. Three different algorithms were applied and analyzed in each category to detect infection.

#### 3.6.1. Single Machine Learning (SML)

This process is used to estimate unknown dependencies or structures in a system. It is suitable for data mining requests to recover hidden information and for decision-making in different applications, because of its powerful forecasting abilities and high accuracy [27].

##### A. Decision Tree (DT)

DT is an active instrument for classification and forecasting. It is composed of a diagram, like the construction of a tree. Every single one

of its internal nodes signifies a test attribute. Each branch signifies a test effect, and every leaf node controls a class of labels. The DT learns by separating a basis set into subsections, depending on the value of a test element [28].

#### *B. Support Vector Machine (SVM)*

SVM is a powerful classification algorithm that performs classification tasks, by constructing hyper-planes in a space that splits the values of several class labels. In this classifier, it is easy to have a linear hyper-plane among two modules [29].

#### *B. Artificial Neural Network (ANN)*

ANN is an effective technique, which has spurred from human intelligence. It consists of three layers; input, hidden, and output. The nodes in the network are interconnected. At first, the images associated with a nodule classification or detection are read by the input layer. Then, the hidden layer extracts some designs associated with the method being analyzed. The third layer (the output layer) delivers the results from the previous layer [30].

### **3.6.2. Ensemble Machine Learning (EML)**

It is robust method. It combines several individual models to produce one optimal prognostic model. It is recommended to use independent classifiers in such methods to achieve the best outcome [31]. Using this method will increase accuracy, as more than one classifier is applied.

#### *A. Random Forest (RF)*

This is a classification of the Ensemble learning technique [32]. It takes place by building decision-making trees during training periods and producing classifications. This technique represents a group of random and unpruned decision trees that have been trained, as an outcome of bootstrap training data.

#### *B. Majority Voting*

Voting is a technique that combines forecasts from several other models. For classification purposes, the forecasts for each label are summed and the label with the most popular vote is predicted [33].

### *C. Bagging*

A Bagging classifier is a set of meta-estimators. It fits the base classifiers on each random subset of the original dataset and then aggregates their individual predictions to form the final prediction [34].

## **4. Performance Evaluation**

The suggested model involves dividing all data into two groups, for experimental training and testing, to explore generalization and decrease over-fitting. The dataset is randomly divided into ten folds during this procedure, 9 of which are used to train the model, and the remaining fold is hidden and used to test the performance and confidence of the predictions after training. This process is repeated in a way that every fold is utilized as the test set. This increases the computational cost but enhances the significance of the result. The final accuracy is obtained by calculating the mean accuracy derived from each testing fold. This classification model has been implemented using two software tools. Firstly, MATLAB is used in image processing stage, image segmentation and feature extraction stage. Then WEKA is used in feature selection stage and classification stage to obtain the results of the classification model. Results were obtained using a laptop with an Intel Core i5 processor and 4GB of RAM.

### **4.1. Database Description**

During the time of the COVID-19 pandemic, researchers and hospitals have given open access to data on this Virus and everything associated to it. In this paper. An overall number of 2,110 CT scans were used, which included 1,110 CT images of people infected by COVID-19 and 1,000 images of uninfected persons. This dataset was collected from the open-source data repository GitHub and Kaggle [35], [36].

### **4.2. The Evaluation Model**

The presented models have three stages. In the first stage, following data collection, some pre-processing techniques were utilized. In the second stage, features of ROI were identified. The output of this stage was features, applied to classifiers in the third stage. In this later stage, single classifiers applied included the Decision Tree, the SVM, and the ANN. As for ensemble classifiers, they included Random Forest, Voting, and Bagging, as shown in [Figure \(9\)](#).

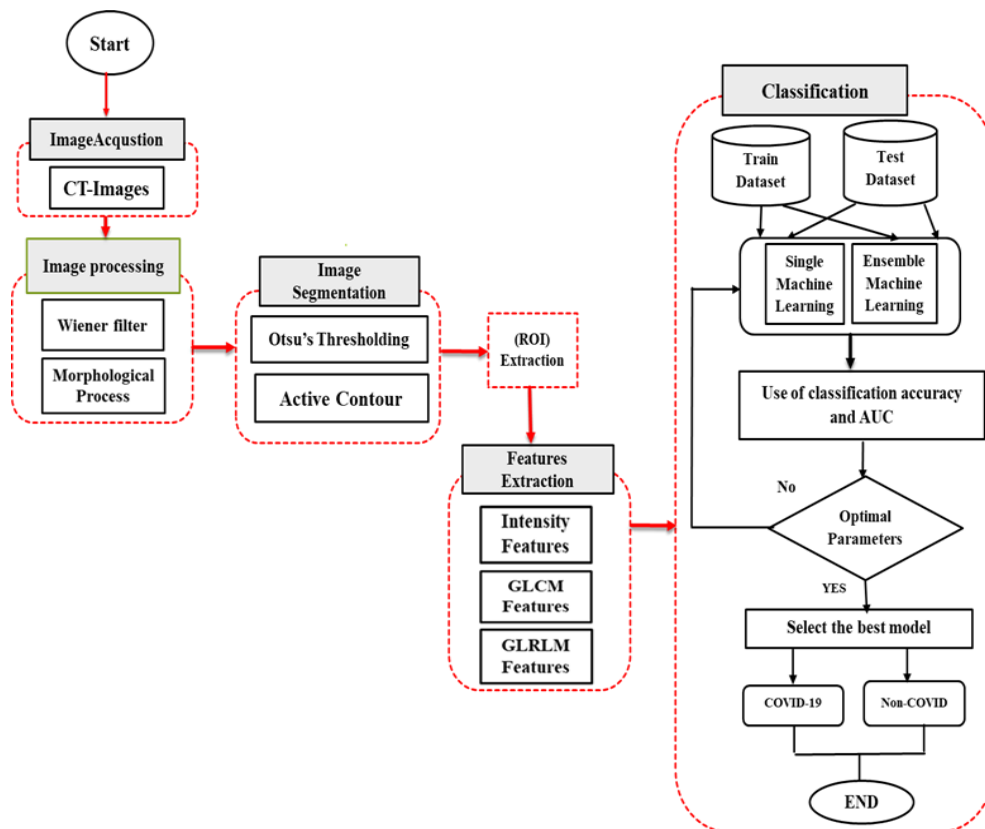


Figure (9) - Flow chart of the proposed classification model.

### 4.3. Performance Measures

It is necessary to measure the models' performance, by separating the dataset into one set for training and another for testing, in percentages of 70 and 30, respectively. The evaluation procedure was performed with 10-fold-cross-validation. Performance is described in terms of accuracy, specificity, sensitivity, the F-measure, root mean square error (RMSE), and the Kappa statistic, by applying a confusion matrix. This matrix consists of four indices: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

#### True Positive (TP)

- The predicted value matches the actual value.
- The actual value was positive, and the model predicted a positive value.

#### True Negative (TN)

- The predicted value matches the actual value.

- The actual value was negative, and the model predicted a negative value.

#### False Positive (FP) – Type 1 error

- The predicted value was falsely predicted.
- The actual value was negative, but the model predicted a positive value.

#### False Negative (FN) – Type 2 error

- The predicted value was falsely predicted.
- The actual value was positive, but the model predicted a negative value.

Additionally, the receiving operating characteristic (ROC) is a 2D graph, where the true positive rate (TPR) is drawn against the false-negative rate (FNR). Also, to compare between different classifiers, the area under the curve (AUC) is calculated [37, 38]. All metrics are listed in Table (5).

Table (5): Evaluation the metrics of (ML)

| Metrics                      | Formula   |
|------------------------------|---|
| <i>Accuracy</i>              | $Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (50)$                   |
| <i>Precision</i>             | $Precision = \frac{TP}{TP + FP} \quad (51)$                                 |
| <i>Specificity (FP-Rate)</i> | $FP - Rate = \frac{TN}{TN + FP} \quad (52)$                                 |
| <i>Sensitivity (Recall)</i>  | $TP - Rate (Recall) = \frac{TP}{TP + FN} \quad (53)$                        |
| <i>F-measure</i>             | $F - measur = 2 * \frac{precision * Recall}{Precision + Recall} \quad (54)$ |

#### 4.4 Objective Validation

This part presents the process of objective validations for the results of benchmarking classification techniques. To make sure that the benchmarking results of COVID-19 classification techniques are valid, this paper utilizes two statistical approaches: mean and standard deviation. The mean  $\pm$  standard deviation can be calculated for each class of data to ensure that the set of COVID-19 classification techniques is subjected to systematic ordering. The mean and standard deviation of the models for the experimental work with differently randomly selected training and testing data are listed in Table (6) and Table (7). Moreover, the performance of the additional validation set is like that of the training



and test sets. The results show that the model can accurately predict the outcome of patients [39].

The mean is the average result, and it is calculated by performing a deviation of the sum of the observed results over the result numbers with the use of the following equation:

$$Mean = \frac{1}{n} \sum_{i=1}^n x_i \quad (55)$$

**Table (6):** The Mean and Standard Deviation for the Results of SML Classifiers

| Evaluation Metrics (mean (%) $\pm$ std) | Single Classifier/Weighted Avg |                                  |   |
|---|--------------------------------|----------------------------------|---|
|   | DT                             | ANN                              | SVM   |
| TP- Rate                                | 0.823 $\pm$ 0.0179             | 0.986 $\pm$ 4.3 $\times 10^{-3}$ | 0.987 $\pm$ 2 $\times 10^{-3}$                |
| FP- Rate                                | 0.3232 $\pm$ 0.134             | 0.011 $\pm$ 4.3 $\times 10^{-3}$ | 8.2 $\times 10^{-3}$ $\pm$ 1 $\times 10^{-3}$ |
| Precision                               | 0.8262 $\pm$ 0.014             | 0.986 $\pm$ 4.3 $\times 10^{-3}$ | 0.988 $\pm$ 4.1 $\times 10^{-3}$              |
| Recall                                  | 0.823 $\pm$ 0.0179             | 0.986 $\pm$ 4.3 $\times 10^{-3}$ | 0.987 $\pm$ 2 $\times 10^{-3}$                |
| F-measure                               | 0.8246 $\pm$ 0.016             | 0.986 $\pm$ 4.3 $\times 10^{-3}$ | 0.987 $\pm$ 2 $\times 10^{-3}$                |
| MCC                                     | 0.6148 $\pm$ 0.461             | 0.969 $\pm$ 3.6 $\times 10^{-3}$ | 0.973 $\pm$ 3 $\times 10^{-3}$                |

**Table (7):** The Mean and Standard Deviation for the Results of EML Classifiers

| Evaluation Metrics (mean (%) $\pm$ std) | Ensemble Classifier/Weighted Avg  |   |   |
|---|-----------------------------------|---|---|
|   | RF                                | Voting  | Bagging                                     |
| TP- Rate                                | 0.964 $\pm$ 3 $\times 10^{-3}$    | 0.988 $\pm$ 2 $\times 10^{-3}$                  | 0.996 $\pm$ 1.4 $\times 10^{-3}$            |
| FP- Rate                                | 0.068 $\pm$ 6 $\times 10^{-3}$    | 7.8 $\times 10^{-3}$ $\pm$ 1.7 $\times 10^{-3}$ | 6 $\times 10^{-3}$ $\pm$ 4 $\times 10^{-3}$ |
| Precision                               | 0.966 $\pm$ 3 $\times 10^{-3}$    | 0.9886 $\pm$ 1.6 $\times 10^{-3}$               | 0.988 $\pm$ 1.4 $\times 10^{-3}$            |
| Recall                                  | 0.964 $\pm$ 3 $\times 10^{-3}$    | 0.9884 $\pm$ 1.6 $\times 10^{-3}$               | 0.996 $\pm$ 1.4 $\times 10^{-3}$            |
| F-measure                               | 0.9634 $\pm$ 3.5 $\times 10^{-3}$ | 0.9884 $\pm$ 1.6 $\times 10^{-3}$               | 0.996 $\pm$ 1.4 $\times 10^{-3}$            |
| MCC                                     | 0.921 $\pm$ 3.5 $\times 10^{-3}$  | 0.974 $\pm$ 3.9 $\times 10^{-3}$                | 0.966 $\pm$ 3.2 $\times 10^{-3}$            |

The standard deviation is used to determine the dispersion or variation amount in the set of values and is calculated as follows:

$$Std\ deviation = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (56)$$

n= the number of values in dataset;  $x_i$ = each value in the dataset;  $\bar{x}$  = the mean of all values in dataset

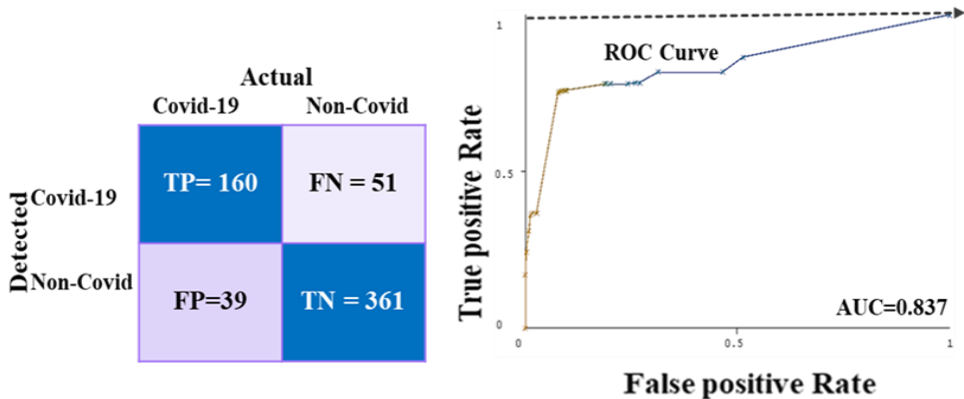
## 4.5 Subjective Validation

This section describes the subjective validation process as the performance of the classifier is verified using a ten-fold cross validation and the best value among the trials are chosen for the assessment. Here we present the best results for the repeated runs of the experimental work with differently randomly selected training and testing data. The experts can prove the effectiveness of the benchmarking results of classification techniques which obtained by our proposed decision-making approach by examining the values of all evaluation criteria used.

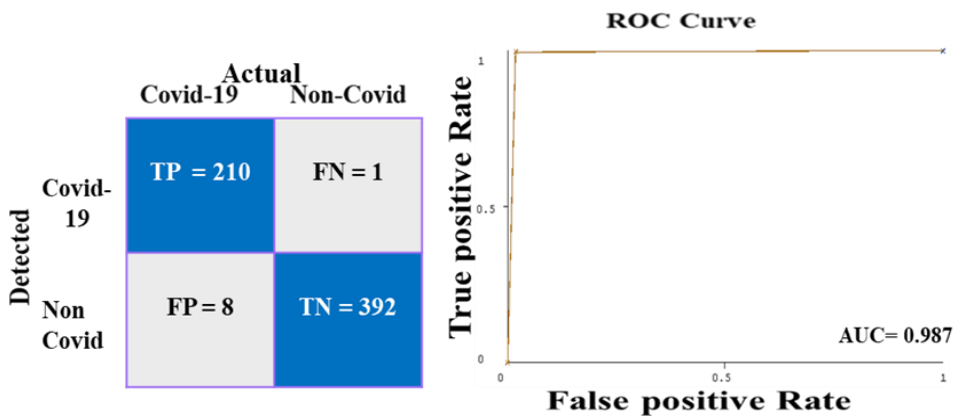
### A. *Single Classification Model (SML)*

This model applied a single classifier, such as DT, ANN and SVM. The outcomes of the classifiers are listed in [Table \(8\)](#) and [Table \(9\)](#). The confusion matrix and ROC are shown in [Figure \(10\)](#). Finally, a graph of the single classifiers is shown in [Figure \(11\)](#). We fine-tuned the parameters to follow for each single classifier.

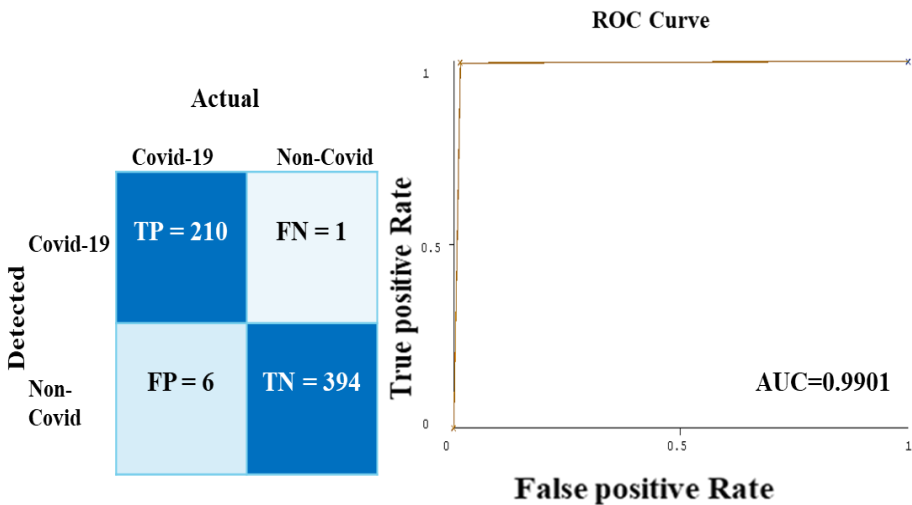
- Decision Tree: Batch size = 100, confidence factor = 0.25, number of decimal places = 2
- Artificial Neural Network: Hidden layers = 10, number of neurons = 8, maximum epochs = 1000, learning rate = 0.05
- Support Vector Machine: Batch size = 100, kernel = poly-kernel, calibrator = SMO, random state = 0, C = 1.0, random seed = 1, epsilon =  $1 \times 10^{-12}$ , number of decimal places = 2



(a) DT



(b) ANN



© SVM

Figure (10) ROC and confusion matrix of Single classifiers (a) DT, (b) ANN, (c)SVM.

Table (8): Evolution metrics of Single Classifiers for each class

| Single Classifier | DT    |           | ANN   |           | SVM   |           |
|-------------------|-------|-----------|-------|-----------|-------|-----------|
|                   | COVID | Non-COVID | COVID | Non-COVID | COVID | Non-COVID |
| TP Rate           | 0.758 | 0.903     | 0.995 | 0.980     | 0.995 | 0.985     |
| Weighted Avg      | 0.853 |           | 0.985 |           | 0.989 |           |
| FP Rate           | 0.098 | 0.242     | 0.020 | 0.005     | 0.015 | 0.005     |
| Weighted Avg      | 0.192 |           | 0.010 |           | 0.008 |           |
| Precision         | 0.804 | 0.876     | 0.963 | 0.997     | 0.972 | 0.997     |
| Weighted Avg      | 0.851 |           | 0.986 |           | 0.989 |           |
| Recall            | 0.758 | 0.903     | 0.995 | 0.980     | 0.995 | 0.985     |
| Weighted Avg      | 0.853 |           | 0.985 |           | 0.989 |           |
| F-measure         | 0.780 | 0.889     | 0.979 | 0.989     | 0.984 | 0.991     |
| Weighted Avg      | 0.852 |           | 0.985 |           | 0.989 |           |
| MCC               | 0.670 | 0.670     | 0.968 | 0.968     | 0.975 | 0.975     |
| Weighted Avg      | 0.670 |           | 0.968 |           | 0.975 |           |
| ROC Area          | 0.837 | 0.837     | 0.988 | 0.988     | 0.990 | 0.990     |
| Weighted Avg      | 0.837 |           | 0.988 |           | 0.990 |           |

Table (9): Performance metrics of the Single Classifiers

| Single Classifier                | DT     | ANN     | SVM    |
|----------------------------------|--------|---------|--------|
| Test Time                        | 2.6 s  | 3.62 s  | 2.17 s |
| Correctly Classified Instances   | 521    | 602     | 604    |
| Incorrectly Classified Instances | 90     | 9       | 7      |
| Kappa statistic                  | 0.6698 | 0.9677  | 0.9748 |
| Mean absolute error              | 0.2025 | 0.01047 | 0.0115 |
| RMSE                             | 0.3742 | 0.1214  | 0.107  |
| Sensitivity (%)                  | 75.8%  | 99.5%   | 99.5%  |
| Specificity (%)                  | 90.25% | 98%     | 98.5%  |
| AUC (%)                          | 83.7%  | 98.7%   | 99%    |
| Accuracy (%)                     | 85.27% | 98.52%  | 98.85% |

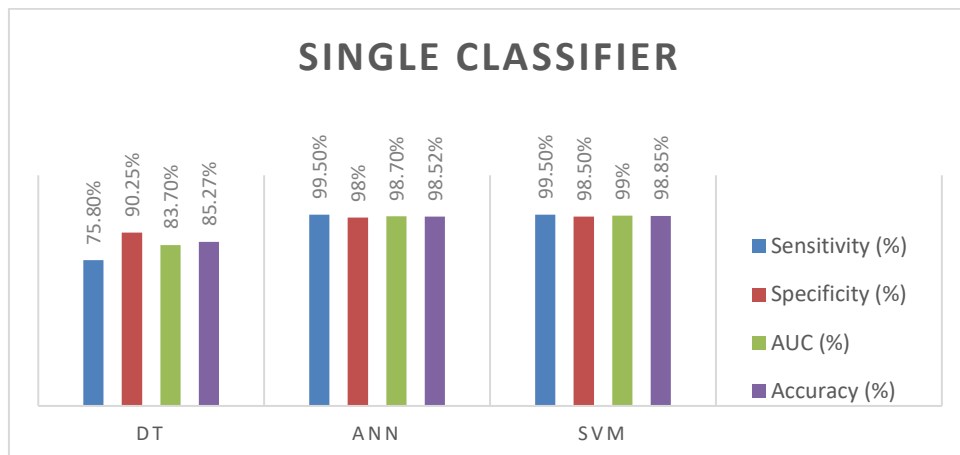
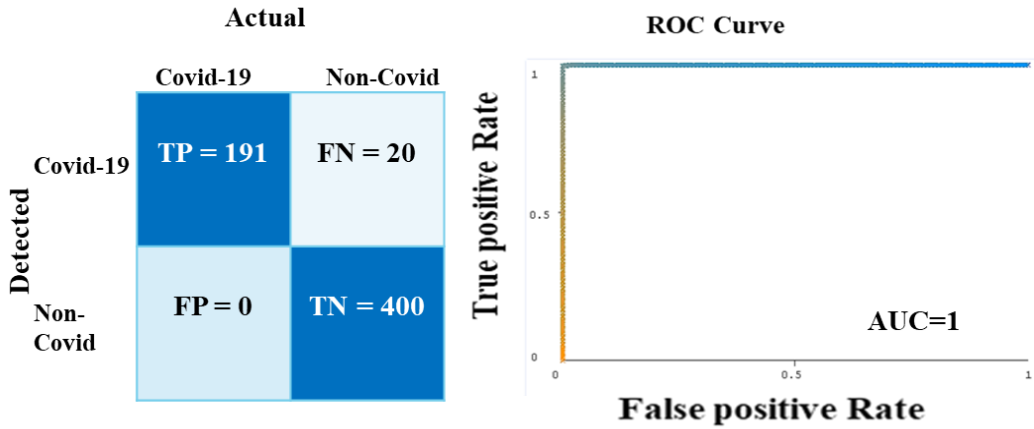


Figure (11) - Graph representation of the performance metrics of Single Classifiers

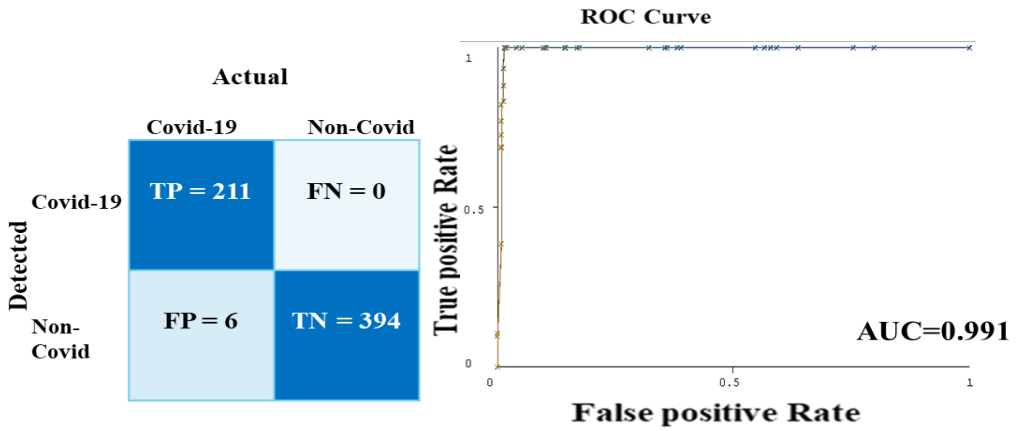
### ***B. Ensemble Classification Model (EML)***

Here, Ensemble classifiers were applied. This included Random Forest, Voting, and Bagging. The confusion matrix and ROC are shown in Figure (12). Finally, a graph of the ensemble classifiers is shown in Figure (13). The performance metrics for Ensemble classifiers are listed in Table (10) and Table (11). We fine-tuned the following parameters for each Ensemble classifier.

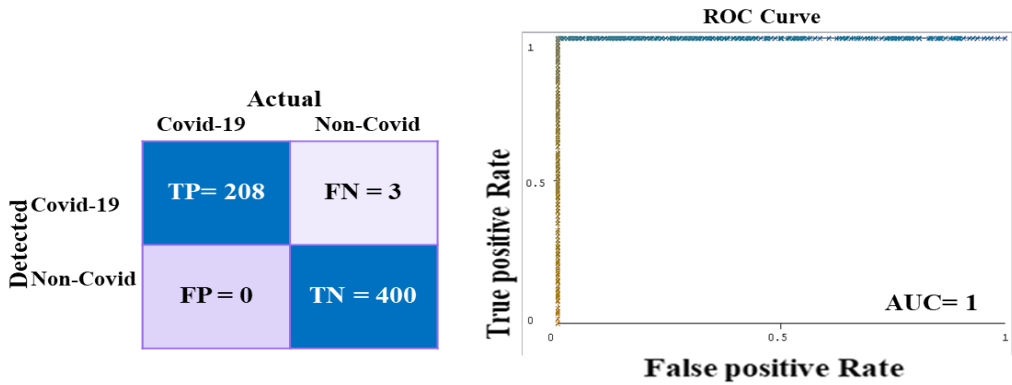
- Random Forests: Batch size = 100, bag size (percent) = 100, number of iterations = 100, number of decimal places = 2, seed = 2
- Voting: Batch size = 100, number of decimal places = 2, classifiers = SVM and DT, combination rule = average of probabilities, seed = 1
- Bagging: Batch size = 100, bag size (percent) = 100, number of iterations = 10, number of decimal places = 2, seed = 1



(a) Random Forest



(b) Voting



© Bagging

Figure (12) Confusion metrics for Ensemble classifiers: (a) RF, (b) voting, (c) Bagging.

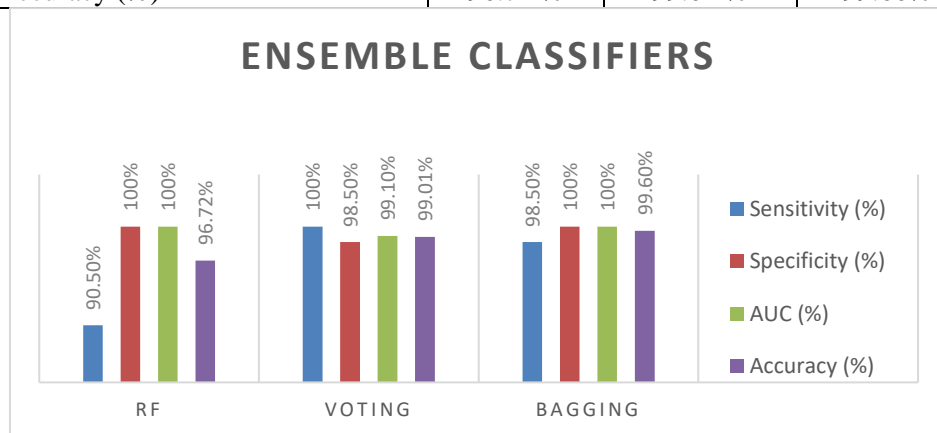


**Table (10):** The Results for Ensemble classifiers for each class

| Ensemble Classifier | RF        |           | Voting    |           | Bagging   |           |
|---------------------|-----------|-----------|-----------|-----------|-----------|-----------|
|                     | COVID -19 | Non-COVID | COVID -19 | Non-COVID | COVID -19 | Non-COVID |
| TP Rate             | 0.905     | 1.000     | 1.000     | 0.985     | 0.986     | 1.000     |
| Weighted Avg        | 0.967     |           | 0.990     |           | 0.995     |           |
| FP Rate             | 0.000     | 0.095     | 0.015     | 0.000     | 0.000     | 0.014     |
| Weighted Avg        | 0.062     |           | 0.005     |           | 0.009     |           |
| Precision           | 1.000     | 0.952     | 0.972     | 1.000     | 1.000     | 0.993     |
| Weighted Avg        | 0.969     |           | 0.990     |           | 0.995     |           |
| Recall              | 0.905     | 1.000     | 1.000     | 0.985     | 0.986     | 1.000     |
| Weighted Avg        | 0.967     |           | 0.990     |           | 0.995     |           |
| F-measure           | 0.950     | 0.976     | 0.986     | 0.992     | 0.993     | 0.996     |
| Weighted Avg        | 0.967     |           | 0.990     |           | 0.995     |           |
| MCC                 | 0.928     | 0.928     | 0.979     | 0.979     | 0.989     | 0.989     |
| Weighted Avg        | 0.928     |           | 0.979     |           | 0.989     |           |
| ROC Area            | 1.000     | 1.000     | 0.992     | 0.992     | 1.000     | 1.000     |
| Weighted Avg        | 1.000     |           | 0.992     |           | 1.000     |           |

**Table (11):** Performance metrics of Ensemble Classifiers

| Ensemble Classifier              | RF      | Voting  | Bagging |
|----------------------------------|---------|---------|---------|
| Test Time                        | 2.3 s   | 2.5 s   | 3.3 s   |
| Correctly Classified Instances   | 591     | 605     | 608     |
| Incorrectly Classified Instances | 20      | 6       | 3       |
| Kappa statistic                  | 0.9259  | 0.9784  | 0.9891  |
| Mean absolute error              | 0.2141  | 0.1408  | 0.168   |
| RMSE                             | 0.2428  | 0.2149  | 0.1931  |
| Sensitivity (%)                  | 90.5 %  | 100%    | 98.5 %  |
| Specificity (%)                  | 100%    | 98.5 %  | 100%    |
| AUC (%)                          | 100%    | 99.1 %  | 100%    |
| Accuracy (%)                     | 96.72 % | 99.01 % | 99.60%  |

**Figure (13) -** Graph representation of the performance of Ensemble Classifiers

## 5. Decision

This paper suggested two classification models using SML and EML classification algorithms and compared them with previous studies. The EML classification model proved to be far better, with more improved results than the SML model. Using EML improved classification accuracy and effectively reduced the rate of false negatives. This model demonstrated competitive performance than models. Finally, a comparison of all classifiers is shown in [Figure \(14\)](#).

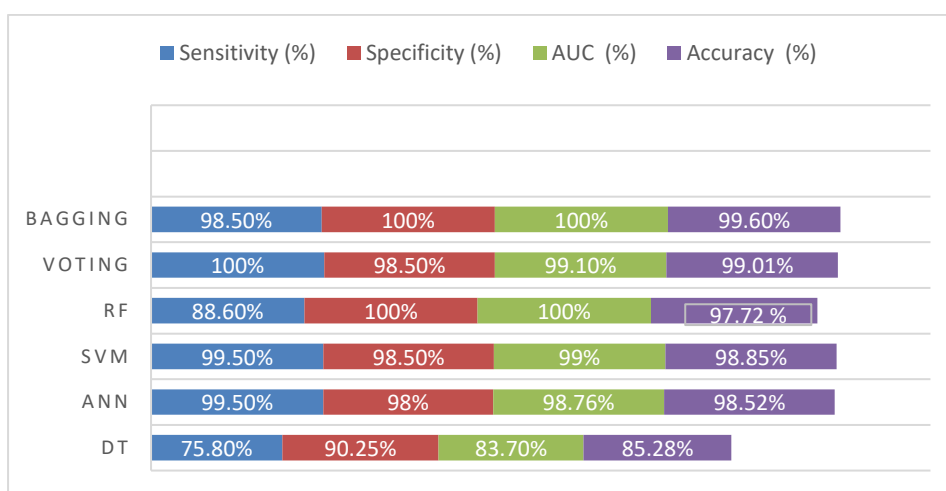


Fig.14. Comparison of classifiers using accuracy and AUC measures

## 6. Conclusion

COVID-19 is a pandemic, which has already caused thousands of casualties and has infected several million people worldwide. This paper proposed a rapid and intelligent COVID-19 disease classification model to detect infected patients from chest CT images. Initially, several pre-processing techniques were applied on the collected images to remove noise and improve contrast. These techniques included the HE technique. Then, the Active Contour algorithm was applied for the segmentation of ROI. The dataset of chest CTs of COVID-19-infected patients was divided into training and testing data groups. The training dataset was utilized for

building the COVID-19 classification model. In this paper, two classification models were used: Single ML and Ensemble ML with 10-fold cross validation. The performance of the proposed EML-based model, as a hybrid-classifier, was compared to single classifiers' performance, with a sequence of procedures from image pre-processing to classification. The proposed EML classification model achieved an accuracy of 96.72 %, 99.01 % and 99.60 %, using the Random Forest, Voting, and Bagging classifiers, respectively. On the other hand, the proposed SML-based classification model achieved accuracies that reached 85.27%, 98.52%, and 98.85% using DT, ANN, and SVM, respectively. The experimental results showed that the accuracy of the proposed EML classification model was better than that of the SML classification model. Future steps will involve using AI to recognize COVID-19, using several deep learning methods on X-ray images of patients suffering from different types of pneumonia.

## **References**

- [1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, & B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, vol.395, no.10223, pp. 497-506, 2020.
- [2] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han & L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The lancet*, vol.395, no.10223, pp.507-513, 2020.
- [3] D. S. Hui, E. I. Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, & E. Petersen, the continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *International journal of infectious diseases*, vol.91, no. 264-266, 2020.
- [4] J. F. W. Chan, K. K. W. To, H. Tse, D. Y. Jin, & K. Y. Yuen, Interspecies transmission, and emergence of novel viruses: lessons from bats and birds. *Trends in microbiology*, vol. 21, no.10, pp. 544-555, 2013.
- [5] World Meters: Covid-19 Coronavirus Pandemic; Accessed on: January 2021, <https://www.worldometers.info/coronavirus/>
- [6] World Health Organization; Coronavirus disease (COVID-19) advice for the public. Accessed on: April 11, 2020, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public>.

- 
- [7] A. Narin, C. Kaya & Z. Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. ArXiv preprint ArXiv: 2003.10849.
- [8] A. Hossam, A. Magdy, A. Fawzy & S. M. Abd El-Kader. An integrated IOT system to control the spread of COVID-19 in Egypt. In International Conference on Advanced Intelligent Systems and Informatics, Springer, Cham, pp. 336-346, 2020.
- [9] A. Hossam & A. Fawzy, A rapid diagnosis tool based on LASER for fighting COVID-19. International Journal of Microwave and Optical Technology, vol.15, no, 5, 2020.
- [10] K. Li, Y. Fang, W. Li, C. Pan, P. Qin, Y. Zhong, X. Liu, M. Huang, Y. Liao, and S. Li, 2020. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). European radiology, vol. 30, no. 8, pp.4407-4416, 2020.
- [11] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, & M. Luengo-Oroz, Mapping the landscape of artificial intelligence applications against COVID-19. Journal of Artificial Intelligence Research, vol. 69, no. 807-845, 2020.
- [12] L. Wang, Z. Q. Lin, & A. Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports, vol. 10, no.1, pp. 1-12, 2020.
- [13] L. Yan, H. T. Zhang, Y. Xiao, M. Wang, Y. Guo, C. Sun & Y. Yuan, Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan, 2020.
- [14] S. El-bana, A. Al-Kabbany & M. Sharkas. A multi-task pipeline with specialized streams for classification and segmentation of infection manifestations in covid-19 scans. PeerJ Computer Science, vol.6, e303, 2020.
- [15] J. Wang, J. Wu, J., Z. Wu, J. Jeong, & G. Jeon. Wiener filter-based wavelet domain denoising. Displays, 46, 37-41, 2017.
- [16] P. Garg & T. Jain. A comparative study on histogram equalization and cumulative histogram equalization. International Journal of New Technology and Research, vol.3, no.9, pp. 263242, 2017.
- [17] X. Chen, L. Yao, & Y. Zhang, Residual attention u-net for automated multi-class segmentation of covid-19 chest CT images. ArXiv preprint ArXiv: 2004.05645, 2020.
- [18] A. E. Hassanien, L. N. Mahdy, K. A. Ezzat, H. H. Elmousalami, & H. A. Ella, Automatic x-ray covid-19 lung image classification system based on multi-level Thresholding and support vector machine. MedRxiv, 2020.
- [19] J. H. Uhl, S. Leyk, Y. Y. Chiang, W. Duan & C. A. Knob lock. Spatializing uncertainty in image segmentation using weakly supervised

- convolutional neural networks: a case study from historical map processing. *IET Image Processing*, vol.12, no. 11, pp. 2084-2091, 2018.
- [20] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han & Y. Shi. Lung infection quantification of COVID-19 in CT images with deep learning. *ArXiv preprint ArXiv: 2003.04655*, 2020.
- [21] D. P. Fan, T. Zhou, G. P. Ji, Y. Zhou, G. Chen, H. Fu & 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.
- [22] J. Ma, Z. Nie, C. Wang, G. Dong, Q. Zhu, J. He & X. Yang. Active contour regularized semi-supervised learning for COVID-19 CT infection segmentation with limited annotations. *Physics in Medicine & Biology*, vol.65, no.22, pp. 225034, 2020.
- [23] T. Bhatia. An image-processing method to detect sub-optical features based on understanding noise in intensity measurements. *European Biophysics Journal*, vol. 47, no. 5, pp. 531-538, 2018.
- [24] S. Varela-Santos, & P. Melin. A new approach for classifying coronavirus COVID-19 based on its manifestation on chest X-rays using texture features and neural networks. *Information sciences*, vol.545, pp. 403-414, 2021.
- [25] H. Zhang, C. L. Hung, G. Min, J. P. Guo, M. Liu & X. Hu, GPU-accelerated GLRLM algorithm for feature extraction of MRI. *Scientific reports*, vol. 9, no.1, pp.1-13, 2019
- [26] P. K. Chandrakar, A. K. Shrivastava, & N. Sahu, Design of a Novel Ensemble Model of Classification Technique for Gene-Expression Data of Lung Cancer with Modified Genetic Algorithm. *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 7, no.25, pp. e2, 2021.
- [27] M. Zivkovic, N. Bacanin, K. Venkatachalam, A. Nayyar, A. Djordjevic, I. Strumberger & F. Al-Turjman, COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustainable Cities and Society*, vol.66, pp.102669, 2021.
- [28] M. Zaman & A. Hassan, Fuzzy Heuristics and Decision Tree for Classification of Statistical Feature-Based Control Chart Patterns. *Symmetry*, vol.13, no.1, pp.110, 2021.
- [29] A. M. Ali, K. Z. Ghafoor, H. S. Maghdid, & A. Mulahuwaish. Diagnosing COVID-19 Lung Inflammation Using Machine Learning Algorithms: A Comparative Study. In *Internet of Medical Things for Smart Healthcare*, pp. 91-105, Springer, Singapore, 2020.
- [30] M. Ilyas, H. Rehman & A. Naït-Ali. Detection of covid-19 from chest x-ray images using artificial intelligence: An early review. *ArXiv preprint ArXiv: 2004.05436*, 2020.

- 
- [31] A. R. M. T. Islam, S. Talukdar, S. Mahato, S. Kundu, K. U. Eibek, Q. B. Pham, & N. T. T. Linh. Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, vol.12, no.3, pp. 101075, 2021.
- [32] R. Mostafiz, M. S. Uddin, N. A. Alam, M.M. Reza, & M. M. Rahman, M. Covid-19 Detection in Chest X-ray through Random Forest Classifier using a Hybridization of Deep CNN and DWT Optimized Features. *Journal of King Saud University-Computer and Information Sciences*, 2020.
- [33] T. Bahadur Chandra, K. Verma, B. Kumar Singh, D. Jain, & S. Singh Netam, Coronavirus Disease (COVID-19) Detection in Chest X-Ray Images using Majority Voting Based Classifier Ensemble. *Expert Syst Appl*, 113909-113909, 2020.
- [34] A. Wadhawan, Phonemer at WNUT-2020 Task 2: Sequence Classification Using COVID Twitter BERT and Bagging Ensemble Technique based on Plurality Voting. *ArXiv preprint ArXiv: 2010.00294*, 2020.
- [35] [:https://www.kaggle.com/luisblanche/covidct?select=CT\\_NonCOVID](https://www.kaggle.com/luisblanche/covidct?select=CT_NonCOVID)
- [36] [GitHub - UCSD-AI4H/COVID-CT: COVID-CT-Dataset: A CT Scan Dataset about COVID-19](#)
- [37] S. Kadry, V. Rajinikanth, S. Rho, N. S. M. Raja, V.S. Rao & K. P. Thanaraj, Development of a machine-learning system to classify lung CT scan images into normal/covid-19 class. *ArXiv preprint ArXiv: 2004.13122*, 2020.
- [38] A. Hossam, H. M. Harb, & H. M. Abd El Kader. Automatic image segmentation method for breast cancer analysis using thermography. *JES. Journal of Engineering Sciences*, vol.46, no.1, pp.12-32, 2018.
- [39] Z. Fang, Z. Xu, T. Jang, F. Zhou & S. Huang, Standard deviation Quantitative Characterization and Process Optimization of the Pyramidal Texture of Mon crystalline Silicon Cells. *Materials*, vol.13, no.3, pp.564, 2020.

## نظام كشف ذكي لتشخيص كوفيد-١٩ باستخدام الصور المقطعية

ان التصنيف المبكر لمرض فيروس كورونا (كوفيد-١٩) ضروري للسيطرة على انتشاره السريع وإنقاذ حياة المرضى. أدى الانتشار السريع ل(كوفيد-١٩) إلى زيادة العبء التشخيصي لأخصائي الأشعة. لذلك، يحتاج الأطباء إلى تقييم سريع إذا كان المريض مصابًا ب (كوفيد-١٩) أم لا. أظهر الذكاء الاصطناعي نتائج واعدة في مجال الرعاية الصحية. لذلك، تقترح هذه الورقة نموذجًا استخباراتيًا بمساعدة الكمبيوتر يمكنه تحديد حالات كوفيد-١٩ الإيجابية. يقدم خط أنابيب التصوير الطبي وطرق الفحص المتضمنة في الحصول على صورة (كوفيد ١٩) وتجزئة وتشخيص المرض باستخدام صور التصوير المقطعي المحوسب. تقدم هذه الورقة نموذجين فعالين للتعلم الآلي الفردي والتعلم الآلي المجمع، لاكتشاف حالات (كوفيد-١٩). يتم تطبيق نموذج التصنيف الأول الفردي باستخدام خوارزميات مختلفة، مثل شجرة القرار، والشبكات العصبية الاصطناعية، وآلات المتجهات الداعمة. أظهرت النتائج ان أداء آلة المتجهات الداعمة يفوق المصنفات الأخرى بنسبة دقه %٩٨,٨٥. تم تطبيق نموذج التصنيف الثاني الجماعي بعدة خوارزميات: مثل شجره القرار والتصويت والتكليس لزيادة دقته حتى %٩٩,٦٠ باستخدام مصنف التكليس وأخيرًا أظهرت نتائج النموذجين المقترحين أداءً أفضل مقارنة بالدراسات الأخرى الحديثة. ومع ذلك، أظهر نموذج التصنيف الجماعي أداءً أفضل من نموذج التصنيف الفردي ويوصى باستخدامه في الوقت الفعلي.