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

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

Osama R. Shahin a,⁎, Hamoud H. Alshammari b, Ahmed I. Taloba a, Rasha M. Abd El-Aziz a

a Department of Computer Science, College of Science and Arts in Gurayat, Jouf University, Saudi Arabia
b Information Systems Department, College of Computer and Information sciences, Sakaka, Jouf University, Saudi Arabia

ARTICLE INFO

Keywords:
Covid-19 analysis
Radial basis function
CAD system
Clinical specimens, SVM

ABSTRACT

As people all over the world are vulnerable to be affected by the COVID-19 virus, the automatic detection of such a virus is an important concern. The paper aims to detect and classify corona virus using machine learning. To spot and identify corona virus in CT-Lung screening and Computer-Aided diagnosis (CAD) system is projected to distinguish and classifies the COVID-19. By utilizing the clinical specimens obtained from the corona-infected patients with the help of some machine learning techniques like Decision Tree, Support Vector Machine, K-means clustering, and Radial Basis Function. While some specialists believe that the RT-PCR test is the best option for diagnosing Covid-19 patients, others believe that CT scans of the lungs can be more accurate in diagnosing corona virus infection, as well as being less expensive than the PCR test. The clinical specimens include serum specimens, respiratory secretions, and whole blood specimens. Overall, 15 factors are measured from these specimens as the result of the previous clinical examinations. The proposed CAD system consists of four phases starting with the CT lungs screening collection, followed by a pre-processing stage to enhance the appearance of the ground glass opacities (GGOs) nodules as they originally lock hazy with fainting contrast. A modified K-means algorithm will be used to detect and segment these regions. Finally, the use of detected, infected areas that obtained in the detection phase with a scale of 50×50 and perform segmentation of the solid false positives that seem to be GGOs as inputs and targets for the machine learning classifiers, here a support vector machine (SVM) and Radial basis function (RBF) has been utilized. Moreover, a GUI application is developed which avoids the confusion of the doctors for getting the exact results by giving the 15 input factors obtained from the clinical specimens.

1. Introduction

The recent pneumonia pandemic, which started in early December 2019 around Wuhan Town in Hubei Province, China, is produced by a novel coronavirus (CoV) designated either by World Health Organization as "2019-nCoV" or "2019 novel coronavirus" or "COVID-19" (WHO). COVID-19 is a virus that can cause disease [1]. According to phylogenetic study using known whole genome sequences, bats are related with the COVID-19 viral reservoir, but the intermediate host(s) has yet to be discovered [2]. It posed a

⁎ Corresponding author.
E-mail addresses: orshahin@ju.edu.sa (O.R. Shahin), hhalshammari@ju.edu.sa (H.H. Alshammari), aitaloba@ju.edu.sa (A.I. Taloba), rmhassanien@ju.edu.sa (R.M.A. El-Aziz).

https://doi.org/10.1016/j.compeleceng.2022.108055
Received 30 September 2021; Received in revised form 22 April 2022; Accepted 27 April 2022
Available online 29 April 2022
0045-7906/© 2022 Elsevier Ltd. All rights reserved.
global threat, similar to the Acute Respiratory Distress Syndrome (ARDS) and Severe Acute Respiratory Syndrome (SARS), both of which are Coronavirus-related diseases. The WHO acknowledged the COVID-19 epidemic as a "public health emergency," stating that the virus spreads to healthy people who come into contact with virus-infected people through the respiratory tract. It can also be transferred in a variety of different ways that experts are still trying to figure out. Infected people's symptoms, such as fever, cough, and pneumonia, will appear in 2 to 14 days.

COVID-19, a revolutionary coronavirus, revolutionized the world's healthcare system. In addition to healthcare, global economics, education, and transportation have all been altered. This virus disease can cause serious respiratory sickness, but with correct treatment, it can be healed. However, the virus's most hazardous side effects include human-to-human transmission and community proliferation. In cluster cases, a prediction based on the artificial intelligence (AI) is possible to detect them using this method. Additionally, previous clinical data can be used to make this prediction. AI can work in a fashion that is similar to the human brain. Furthermore, AI can comprehend and portray the progress of the COVID-19 vaccine development. The present patient tracking, screening, analyzing, and predicting should be done for an accurate prediction of COVID-19 instances, which can help in the future prediction of infected patients [3]. AI is now routinely utilized to find new compounds for the development of a COVID-19 assistance.

Many studies are being conducted to find new treatments for curing the disease, as well as computing to detect disease-affected persons using medical image processing of CT scans and X-ray pictures.

### Algorithm 1

**The sickness is confirmed using the reverse-transcription polymerase chain reaction (RT-PCR).** The RT-PCR sensitivity is insufficient to detect disease because it lacks the sensitivity required to proceed with the treatment of suspected patients. Computed tomography is utilized to discover a certain distinctive manifestation in the lungs. This CT can be utilized to perceive the case since COVID-19 affects the lungs. CT-image screening can be utilized to identify the disease or COVID-19 virus at an early stage. COVID-19's CT picture demonstrates the pneumonia disease's resemblance [4]. Once one person gets into close touch with someone who is afflicted, the virus transmits to the other. It can spread through when an infected individual breathes, coughs, or sneezes, their nose and mouth is infected. The virus quickly spread throughout the infected surface. The conventional diagnostic procedure was employed to detect the virus. The nucleic acid is amplified from a nasopharyngeal swab using RT-PCR, transcription-mediated amplification (TMA), or llo-mediated isothermal amplification. To stop the corona virus spreading, many precautions were taken.

The research work goals to establish effectiveness of big data and AI to use these technologies to combat laboratory results derived from clinical specimens of coronavirus suspects, as well as to analyze recent solutions. Computer-aided diagnosis (CAD) [5] is a medical procedure that assists clinicians in explaining the clinical specimens of coronavirus patients. The abdominal problem is discovered while looking at the patient’s CT chest imaging. When the condition becomes more advanced, the patient may have breathing issues, heart damage, and other infections. This can result in death, so it's critical to catch the corona early on.

The contribution of the proposed method:

- The proposed method is held in detecting and classifying the COVID-19 virus. The basic method used in this system is a machine learning method.
- The CT-lung screening approach is used to do this. The respiratory system is the first organ to be impacted by the covid-19 virus. As a result, a scan and diagnosis of the lungs or respiratory system are required. Also, using clinical samples to develop a machine-learning model for detecting the Coronavirus in the patients.
- Building a CAD system that collects data from COVID-19 patients or suspects and determines if the patient is infected or not. To enhance accuracy and access to data in less time, modern machine learning methods are applied.

The rest structure of the article starts with related works in section 2. The proposed work is showcased in Section 3. The classification and detection process is elaborated in Section 4. The results and discussion part is showcased in section 5. Finally, the conclusion part is depicted in section 6.

### 2. Related Work

The authors in [6] proposed CAD system was assessed using five-fold tests for the multi-class prediction issue utilizing two independent databases of chest X-ray images, COVID-19 and ChestX-ray8. The suggested CAD system was trained using an identified training set of chest X-ray images. The proposed CAD predictor was employed to determine and categorize regions on entire X-ray images with COVID-19-related lesions, with overall detection and diagnosis accuracy of 96.31 percent and 97.40 percent, accordingly. With such a mean intersection over union (IOU) of better than 90%, the most of test photographs from COVID-19 and other respiratory ailments patients were accurately predicted. Deep learning regularizes data balancing and enhancement improved COVID-19 diagnosis efficiency by 6.64 percent and 12.17 percent, respectively, in terms of overall accuracy and F1-score. A diagnosis based on individual chest X-ray photographs can be made in 0.0093 seconds using the specified CAD method. As a result, the Design used in this study can forecast at 108 frames a second (FPS), which is close to real-time. Using proposed deep learning CAD system, COVID-19 may be reliably identified from other respiratory illnesses. In the real world, the suggested learning algorithm looks to be a reliable tool for supporting health care systems, consumers, and physicians.

The purpose of the research in [7] was to determine whether ACR Lung-RADS affects the rate of false-negative and positive findings in a clinical CT lung screening test. The goal of this review was to determine how ACR Lung-RADS affects the rate of false-negative and positive findings in a clinical CT lung screening. A number of 2,180 high-risk individuals had a baseline CT lung examination during
the research period, with 577 patients receiving no clinical follow-up. The overall positive rate for ACR Lung-RADS was reduced from 27.6% to 10.6%. There are still no false negatives among the 152 patients classified as benign after a 12-month follow-up. Using ACR Lung-RADS increased the predictive value for detecting cancer from 6.9% to 17.3 percent in 1,603 patients with follow-up. The introduction of ACR Lung-RADS increased our CT lung screening group’s positive predictive value by a factor of 2.5, to 17.3 percent, while lowering the number of false-negative tests.

The purpose of the study in [8] was to determine what factors impact the frequency of positive RT-PCR results. Using a retrospective analysis, we looked at the clinical information of recurring positive coronavirus disease 2019 (COVID-19) patients in multiple medical facilities in Wuhan. Based on their RT-PCR results, patients are separated into two groups: recurrent positives and non-recurrent positives (non-RPos group). Clinical characteristics, updated content, and antibody titers was placed in two groups. The authors used AI-assisted chest increased computed tomography (HRCT) equipment to examine pulmonary inflammatory exudation and assess the size of lung sections with varied densities. This study included 122 COVID-19 participants. In terms of age, gender, past diseases, clinical manifestations, clinical classification, clinical history, medication regimens, or serum-specific antibodies, there are still no significant differences between the two groups. COVID-19 recurrence is associated with subpleural exudation towards the lung periphery and severe respiratory failure at discharge.

The researchers in [9] provides a unique COVID-19 illness severity classification approach based on a convolutional neural network (CNN). Using chest X-ray images as input, an automated CNN model is constructed and proposed to split COVID-19 patients into four severity classes: mild, moderate, severe, and critical with an average accuracy of 95.52 percent. Experiments on a sufficiently large number of chest X-ray images illustrate the efficacy of the CNN model constructed with the suggested framework. It’s the first COVID-19 injury severity evaluation study comprising 4 stages, using a sufficient huge number on X-ray data sets and a CNN with virtually all hyper-parameters dynamically adjusted mostly by variable selection optimization, as far as the researcher is aware.

Millions of individuals [10] around the world are affected by the ongoing COVID-19 pandemic. Chest computed tomography (CT) is the most widely used imaging modality, and it is critical for patient diagnosis and treatment. The lungs of COVID-19 patients were described using an automated methodology based on individual adapted Hounsfield unit (HU) thresholds. The HU-density calibration curve’s impact on inter-scanner variability was explored. Inter-scanner variability was found to be insignificant. Individual thresholds th1 had median values of 768, 780, and 798 HU for the three techniques, respectively. In comparison to the other two methods, the maximum gradient of the data had a substantially lower median value. Three factors of our electorate were quantified using a
millimeter gradient on the data method; aerated, intermediate and consolidation components had median values of 793 499 cm3, 914 291 cm3, and 126 111 cm3, respectively; the first peak had an average value of 853 56 HU, and the second peak had an average value of 854 56 HU.

3. Proposed system

This system proposed detection and classification of CT-lungs screening and clinical specimens of COVID-19 using machine learning methods. The clinical procedure for infection diagnosis in the COVID-19 clinical specimen samples takes more time and effort by the doctors and also for the patients. This method is used for early infectious disease detection. Then the process of image collection, starting with a pre-processing stage to improve the appearance of the ground glass opacities (GGOs) nodules, which were previously fuzzy with fading contrast, the process of gathering CT images and creating a classifier model will take place. To detect and classify the CT-Lungs screening has been proposed in two-phase the one is the building of classifier model and the other is testing a new CT image [11,12]. Then detect the COVID-19 in clinical specimens is supervised by machine learning techniques such as Naive Bayes Classification, Decision Trees, Support vector machines, Radial basis function, and K-means clustering are used and trained on the dataset to effectively distinguish between infected and non-infected COVID-19 cases in clinical test samples. The following sections explain the detection and classification of COVID-19 are shown in Fig. 1.

Several machine-learning models based on several supervised learning approaches such as Naive Bayes Classification, Decision Trees, Support vector machines, Radial basis function, and K-means clustering are used and trained on the collected dataset in the third phase. Fig. 1 also depicts the proposed CAD design.

3.1. COVID-19 dataset:

As part of CT-lung screening, the data set is used to detect affected and normal COVID-19 virus data. This information provides a clear automatic segmentation scheme and quantifies anomalous CT models [13]. The collection of the dataset is considered a difficult task because they require a large number of ethical and privacy of the hospital system. Based on ethics and law, the appropriate commissions accepted this dataset. Different forms of the disease and scanning methods are included in the collection. The patient will be categorized as corona or not corona based on the given dataset. If a dataset has 100 samples, the corona disease will infect at least half of them. The entire data set is gathered, categorized according to the ailment, and then processed to the next phase. They preserved and secured the data for subsequent use. The clinical specimens for SARS-CoV-2 testing are given as follows:

- Whole blood specimens.
- Respiratory secretion.
- Serum Specimens.

3.2. Preprocessing

This step involved classifier model building. The sample is separated from the dataset is further processed to the preprocessing section. This section contains the CT scan slices from which the data is gathered. 2D and 3D sample slices are included in the slides [14]. The facility and medical images are dimensioned and normalized as part of the preprocessing step. In this step, the medical image is diagnosed and presented. This leads to the following stage, which is the identification of ROIs. On clinical specimen preprocessing, the unstructured data must be refined before proceeding to the next detection stage. Several elements must be taken into account during the preprocessing stage. Samples with insufficient or ambiguous data should be discarded. To improve accuracy, the unstructured texts in the dataset will be preprocessed for punctuation, lemmatizations, symbols, stop words, and URLs [15].

3.3. ROIs detection

Following the preprocessing test, the Regions of Interest (ROIs) [16] are processed. The RGB color is employed in this technique to accomplish the moving window to the segment region of interest (ROIs). This focuses primarily on the location of the moving thing, with the consequences of the moving object being rejected. In this detection, the color concentration is more crucial. The colors red, blue, and green are utilized to indicate the system’s detection. These colors could be used to distinguish the areas that are influenced by the corona.

3.3. Feature extraction

Each layer in the x-ray is extracted and then a systematic result is created as part of the feature-based part of the capsules structure [17]. It is a procedure in which one image is categorized and compared to other images to obtain a result.

4. Classification and Detection

In CT imaging of a corona patient’s lungs, infected areas appear as Solid nodules and Ground Glass Opacity (GGO). The enhanced attenuation without obscuration of the underlying arteries and bronchi in Ground Glass Opacity (GGO) [18] nodules obscures the
whole lung parenchyma inside it. GGO nodules are categorized as follows:

- A GGO-containing nodule (part-solid nodule)
- A GGO-containing nodule with pure localized GGO (non-solid nodule)
- The nodular ground-glass Opacity, also known as the subtle nodule, is a technique that is utilized in clinical practice, computed tomography (CT) [19] scanning and image screening were widely employed. The lungs scanning system employs this technology.

In medicine, computer-aided detection (CAD) or computer-aided diagnostics (CAD) are tools that aid clinicians in deciphering medical images. The radiologist must analyze and estimate a huge amount of data in minimum amount of period using the imaging modalities including X-ray, MRI, Ultrasound, and CT diagnoses. CAD systems aid in the scanning of digital images to detect various lung locations, such as probable infections. Computer-aided diagnosis is described as the process of a person making a decision about his or her medical health and diagnosing that process (CAD). This computer output is used to process a version of an original image.

The suggested system’s major phase is the categorization model. This system must be able to distinguish between positive and negative images of Corona virus-infected lungs and other normal images. Numerous machine-learning algorithms may be used to create such a model. We’ll use two classifiers in this case. Radial basis function (RBF) networks [20,21], for example, have 3 layers: an indoor layer, hidden layer, and outside layer. It is critical to set suitable initial states for RBF networks since it is becoming increasingly popular in neural networks with various applications. It is probably the most competent to the multi-layered perceptron. The vector supporting machine will be utilized as the second classifier (SVM). SVM is also a good classifier for digital picture classification, especially when the categorization is based on colors or specific features.

4.1. Classifier Model

The classifier model is the last and most important step in the classification process [22,23]. The image extracted through feature extraction is categorized in this method based on the disease that has been diagnosed. The same process is repeated when a new CT

![Flowchart for the COVID-19 detection.](image)
picture is tested, eventually leading to an unlabeled CT image. The person who has been afflicted by the covid-19 has been identified as a result.

4.2. Flowchart for COVID-19 detection

The workflow for detection of COVID-19 is shown in Fig. 2. The first step is to perform the Covid-19 test, which is the first stage of the test. Continuing, a CT-lung screening is performed. In CT-lung screening, the person undergoes a lung test, which is then processed by preprocessing and feature extraction. The CT image is retrieved and then diagnosed through a series of techniques known as feature extraction. The patients are then categorized according to their ailment and tested again. The COVID-19 test is then performed based on this information, and the results are analyzed. There are two types of results: normal and covid positive.

4.4. Training Models

The generated dataset is analyzed using different machine learning models [24,25] in this step to check its accuracy and compare it to the accuracies of each model to determine the optimal model for real-time COVID-19 detection and classification.

4.4.1. Decision Tree

A Decision Tree (DT) is a tree-like structure used to model decisions and their outcomes. The internal nodes in a DT denote the attribute test, whereas the branch denotes the outcome. The leaf nodes represent the class labels. Because it does not require domain-specific expertise to develop, DT could be beneficial in a variety of contexts. Classification and Regression Trees are other names for DTs (CART) [26]. The Decision Tree (DT) is a classification and prediction machine learning algorithm. The decision tree algorithm is straightforward to create and modify. In terms of efficiency and processing speed, the decision tree algorithm performs admirably. As a result, as compared to the principle of unit classification, the accuracy of the decision tree machine learning algorithm is higher.

DT and LR are two common classification methods. Algorithm 3 depicts the key steps taken by the DT and LR classifiers for predicting the COVID-19 artificial individuals.

The decision tree is a predictive analysis approach that is supposed to work by creating a category based on the dataset it was fed with, which then creates its sub-categories, which again creates its sub-categories, and so on until the last node is created, which gives us the desired result or the programmed terminates. The model’s training dataset will be used to generate the prediction result. With the increasing length and depth of the decision tree, the complexity of its execution rises. The decision tree approach algorithm in
4.4.2. Naïve Bayes

Random forest is a classification and regression analysis algorithm that uses supervised learning. Tin Kam Ho used the random subspace method to build the algorithm. An ensemble of multiple decision trees, the random forest is composed. Based on the taking of majority votes in the tree class, each tree spits out a class prediction. The random forest’s basic steps are to take a random sample from dataset and build a decision tree from each tree to produce an estimate. Vote on the prediction tree’s final prediction and choose the one with the most votes.

The Bayes theorem is used to classify data, and the Naïve Bayes algorithm is based on it. This methodology was first utilized for text categorization in the 1960s under the way of text retrieval problems. The Bayes theorem allows you to compute the posterior probability of $P(x|y)$ by using $p(x)$, $p(y)$, and $P(y|x)$, as shown in equation 1.

$$P(x|y) = \frac{P(y|x)p(x)}{p(y)}$$

In Eq. (1) presents a posterior probability of $p(z/a)$ in the case of a specific predictor ($z$ is a target and $a$ is the attribute). The prior probability of the class and predictor, respectively, is $p(z)$ and $p(a)$. The probability of a particular class is given by $p(a/z)$. Fig. 4 shows the working of the Naïve Bayes Algorithm.

To determine the probability of a positive result in sample A with tokens $\{O_1, ..., O_n\}$, it is necessary to integrate tokens with varied positive rates to compute the overall rate of COVID-19 positive samples. To accomplish the categorization, we calculated the product of the infected rate for a single token and compared it to the product of the non-infected rate for that token directly.

If the total positive rate of the product, $P[D]$, is greater than the negative rate of the product, $N[D]$, the test sample is categorized as positive. As shown in Algorithm 3, the above equation is used in the classification of Naïve Bayes strategy for the COVID-19 organization.

4.4.3. Support Vector Machine

Detecting Unpredictable data makes determining COVID-19 from symptoms difficult. No suitable data set may therefore be used as a standard. According to the findings, the majority of COVID-19-infected persons were hospitalized with viral fever, respiratory infection, and trouble breathing. Once infected with COVID-19, Those with high blood pressure, heart problems, and a rapid pulse rate swiftly proceed to the next level. If the virus progresses to acute respiratory illness syndrome, it can cause respiratory failure, septic shock, and acute respiratory disease syndrome (ARDS). Using our proposed method, we can determine whether or not someone is infected based on their symptoms. Mildly infected, severely infected, and not infected are three classifications for the outcome, i.e.,
infected status. The numerical values have been mapped to the classes as follows in Fig. 5:

**Not Infected:**
Cases labeled as not infected only show signs of a single sickness, which is completely normal for any human being. People can get a moderate fever and a dry cough from certain illnesses, such as the common cold, but such symptoms alone aren’t enough to rule out COVID-19 infection.

**Mildly Infected:**
This classification denotes the symptoms do not conclusively indicate COVID-19, then they lead to the serious, if need the precautions are didn’t taken. In circumstances when the patient has a mild fever and a mild respiratory difficulty, this can speculate that patient is diseased with the COVID-19.

**Severely Infected:**
COVID-19 has shown favorable outcomes in the majority of patients with more than two or three symptoms, each of which is over the normal range. A high fever, rapid breathing, and acute respiratory syndrome are all signs of a serious illness for the patient. Next, the dataset is given to the SVM Classification model.

SVM was chosen for this problem because it employs the kernel method to transfer low-dimensional input space to high-dimensional space, effectively translating a non-separable issue into the separable issues. A train set has been created from the dataset. The SVM classifier uses a hyper-plane to linearly separate the data using the linear kernel. Parallel hyper-planes separate each data class, ensuring that the distance between them is as vast as possible. Detecting COVID-19 is a high-priority situation, hence we are looking for a hyper-plane with a narrower margin of error to better reliably classify the infected classes. The SVM working principle is in Fig. 6.

For the categorization of clinical samples, the C-SVM classification algorithm is used in this research. The C in C-SVM stands for the cost parameter, which controls the modeling error that happens when the function is an overfitted to sets of data points, resulting in a mistake. To begin the process of training, consider that a collection of data is being trained, with (C, γ) as a hypothetical merger parameter capable of developing a SVM superior classifier. The Grid exploration in yawnd C parameters is a practical strategy that is commonly used in the SVM classifier to find this fusion limitation. In the grid-search, the k-fold rotation estimation is used to find the SVM classifier with best rotation approximation accuracy prediction. Algorithm 3 is the SVM Classifier’s algorithm.

### 4.4.4. Radial Basis Function
It is an ANN that utilizes activation functions of radial basis, as illustrated in the Eq. 2. The RBFNN is a feed-forward neural network having a 3 levels. While the first layer transmits only the input signal, while the second layer uses non-linear Gaussian functions. Lastly, the third layer includes the Gaussian linear outputs. During training, only its tap weights between the hidden layer and the output layer were changed.

\[
f(a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2\sigma^2}}
\]  

The Gaussian function \( f(a) \) is used to approximate a function in Eq. 2. The real values are represented by a. The input a is used to determine the function’s dimensional parameters. In generalization issues, the radial basis function (RBF) network is an extensively used as Artificial Neural Network (ANN). The RBF network differs from other neural networks in that it learns more quickly and has universal approximate. It is made up of 3 layers: the input, the output, and the hidden, of all which are associated by a feed-forward network. Each layer has a distinct purpose.
When the error reaches the target value of 0.01 or the total iteration of the training reaches 500 times, the RBF model training can be ended. The RBF is chosen in such a way that the hidden layer of the RBF must have a total of 10 nodes. The Gaussian function is employed as the transfer function in the computational units. The RBF network is quick and efficient, taking less time to complete the training. Fig. 7 shows the working of the Radial Basis Function.

The RBF has two inputs, which are the coordinates of a point, and a single output in the approximation situation of a 2-variable function. The phrase is used to calculate the RBF output with the input $A = [a_1, a_2]^T$ in Eq. (3).

$$u(a) = \sum_{t=1}^{n_{RBF}} w_t \phi_t(a)$$

where, $n_{RBF}$ = Number of RBFs,
$w_t$ = Weight of the $t^{th}$ neuron,
$A$ = Input vector,
$\phi_t(X)$ = Value of $t^{th}$ RBF at this point.

A Gaussian function is used as an RBF, which, in its two-dimensional case, is written as:

$$\varphi(A) = \exp \left( -\frac{\|A - C\|^2}{2x^2} \right)$$

were, $C = [c_1, c_2]^T$ → vector of coordinates of the RBF center,
$x$ = width,
The Euclidean norm is given as (5)

$$\|A - C\| = \sqrt{(a_1 - c_1)^2 - (a_2 - c_2)^2}$$

Errors can be minimized by using Eq. (6).

$$H = \frac{1}{2} \sum_{v=1}^{n} e_v^2 = \frac{1}{2} \sum_{v=1}^{n} (u(P_v) - T_v)^2$$

where $n$ = number of test points,
$e_v$ = solution error at $j^{th}$ test point,
$P_v$ = coordinates of $j^{th}$ test point, in case of approximation of a 2-variable function $P_v = [p_{v1}, p_{v2}]^T$,
$u(P_v)$ = $u(a)$ at the $j^{th}$ test point,
$T_v$ = target value at $j^{th}$ test point, multiplier $\frac{1}{2}$ is used for simplification.

4.4.5. K-means Clustering

The k-means technique is used to classify or cluster the data collected from various countries. The dataset was clustered using a change of methods, it including a Density-based Spatial Clustering of Application with the Noise (DBSCAN), k-means, and the Hierarchical Clustering Algorithm. The k means technique is widely utilized because of its distance-based, quick processing, and linear
calculation. The k-mean can be determined using the steps below.

There are two stages to the K-means clustering technique. In the first phase, an average linkage technique is used to perform a preliminary categorization. The clustering centres obtained in the first phase as the initial clustering centre are employed in the typical K-means approach in subsequent classifications in the second phase. The expected number of clusters is given as K and is determined by the dataset’s characteristics.

To detect and classify the covid-19 data set, the k-means are calculated. When compared to the other algorithms, this is the most effective strategy.

First, initialize the clustering centers from the first phase in the second phase of the K-means clustering approach. The samples are then classified into clusters based on how similar their initial centers. Then, for successive iteration, compute the average value of the clusters as centers. These steps must be repeated until the clusters have been altered. The working of K-Means clustering is demonstrated in Fig. 8.

"The kNN" algorithm is one of the most extensively used and basic classification algorithms. The kNN algorithm is slow, nonparametric method. Let’s try to comprehend the idea of passive learning, we will find that it does not have a training period, unlike eager learning. When wish to make an estimate, it searches the entire dataset for the closest neighbors. A k value is determined during the algorithm study. The number of elements to be examined is represented by the k value. As soon as a value is obtained, the range between the k closest element is calculated. It is usual to utilize the Euclidean function in distance calculations. The Euclidean function can be replaced by the City Block, Minkowski, and Chebyshev functions. This is followed by sorting and assigning the incoming value to the relevant class.

5. Result and Discussion

The classifier decision was used to determine the performance measure. The classifier determines whether or not the person is affected by the disease. This classification can be separated into four categories, the first two of which are infected and normal, and the other two of which are due to a testing error. The first section is called Test Positive and Negative, while the second section is called False Positive and False Negative. The test positive and negative results indicate whether or not a person is impacted by covid. The false-positive indicates that the person is normal, but the test indicates that the person is not, while the false negative indicates that the

| Metrics                  | Test Result (Positive or Negative) |
|--------------------------|-----------------------------------|
| Actual biopsy (True)     | True Positive (TP)               |
| Actual biopsy (False)    | False Negative (FP)              |
Algorithm 1

DT and LR for detection of positive COVID-19 patients

**Input:** The following is a feature ranking.

**Output:** confusion matrix, classification report, accuracy

**Process:**

i. Using the StandardScaler() function, standardize the selected characteristics.

ii. Applying DT to the specified features using the DTClassifier (criterion='entropy', max_depth=5, random_state=0) function with some parameters.

iii. Select features are used to train the model.

iv. K-fold cv specifications: thresh=0.5, k_fold_seed=13, n_folds=10.

v. Using the test dataset, forecast the outcome.

vi. Predict the outcome using the test dataset.

vii. To evaluate FN, FP, TN, and TP by using confusion matrix()

viii. Calculaterecall, precision, and F1 score with classification report() function.
A person is impacted by the COVID-19, but the test indicates that the person is not. The sensitivity measurement can be seen in the false-positive and false-negative results.

The performance of the proposed COVID-19 detection and classification algorithms can be assessed using a variety of performance indicators. To control the presentation of the suggested structure, a Confusion Matrix is produced. Table 1 shows the Confusion Matrix.

Four metrics can be used to specify the performance metrics that will be used to evaluate the classification model: accuracy, correctness, recall, and F1 score.

The accuracy of a proposed model can be assessed by expressing the average of the True values in the obtained results from a given dataset of clinical specimens as the proportion of the sum of True Positives (TP) and Negatives (TN) to the total of True Positives (TP) and Negatives (TN) and False Positives (FP) and Negatives (FN), as shown in (7)

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100
\]  

(7)

The ratio of True Positives (TP) to the sum of False Positives (FP) and True Positives acquired from a particular dataset of clinical specimens is used to evaluate the performance metric precision is expressed in (8)

\[
\text{Precision} = \frac{TP}{TP + FP}
\]  

(8)
The metric for measuring performance the ratio of True Positives (TP) to the sum of False Negatives (FN) and True Positives derived from a dataset of clinical specimens is used to recall determine in (9).

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{9}
\]

Precision and Recall metrics are used to calculate the F-measure, \( F_\beta \). The measure F1-score is determined by the value of \( F_\beta \). The Precision and Recall on Harmonic Mean as indicated in the F1-score is expressed in (11).

\[
F_\beta = \left( 1 + \beta^2 \right) \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \left( \text{Precision} + \text{Recall} \right)} \tag{10}
\]

Substitute \( \beta = 1 \), after simplification,

\[
F_1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{1 \times (\text{Precision} + \text{Recall})} \tag{11}
\]

6. Conclusion

The goal of the research is to develop the technique for detecting and classifying the covid-19 virus. The machine learning method is used to recognize and classify objects in this manner. The COVID-19’s effects astounded the globe because there were no specific treatments or vaccines available at the time of the virus’s emergence. Several studies are being conducted to overcome this lethal disease caused by Coronavirus. These clinical specimens were subjected to genetic, serological, and biochemical testing. For further processing, several characteristics are collected from clinical samples, and also the CT-lung screening speeds up the process and aids in identifying people who have been infected with the coronavirus. Based on this, the covid test is performed, the genuine positive, negative, and false positive, negative are separated, and the accuracy, precision, recall, and F1-factor are calculated. Because the prediction rate is so high, it’s much easier to locate those who are afflicted. To lower the death rate. As a result, utilizing machine learning to detect and categorize COVID-19 in CT lungs screening is deemed the best way for detecting and classifying the virus. The efficiency and accuracy of the suggested model can be enhanced by increasing the number of samples used. For a better outcome, more feature engineering is required, and deep learning can be applied in the future.
Declaration of competing interest

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

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at Jouf University for funding this work through research grant No (DSR2020-04-2632).

References

[1] Abbas Asmaa, Abdelsamea Mohammed M, Gaber Mohamed Medhat. Classification of COVID-19 in Chest X-Ray Images Using DeTrAC Deep Convolutional Neural Network. Appl Intell 2021;51(2):854-64. https://doi.org/10.1007/s10489-020-02189-7.

[2] Abdullahreem, Nasiba M, Adnan Mohsin Abdulazeez, Diyar Qader Zeebaree, and Dathar A Hasan. “COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms.”

[3] Al-antairi Mugahed A, Hua Cam-Hao, Bang Jaehun, Lee Sungyoung. Fast Deep Learning Computer-Aided Diagnosis of COVID-19 Based on Digital Chest x-ray Images. Appl Intell 2021;51(5):2890–907. https://doi.org/10.1007/s10489-020-02076-6.

[4] Aminisefat Alineza, Saravani Khadijeh. A Case Study of Hypertension and COVID-19. Gene, Cell and Tissue 2020;7(3).

[5] Dhamodharavandhani S, Rathipriya R, Chatterjee Jyotir Moy. COVID-19 Mortality Rate Prediction for India Using Statistical Neural Network Models.”. Front in Public Health 2020;8(August):441. https://doi.org/10.3389/fpubh.2020.00441.

[6] by Feng Simon James, Dey Nilanjan, Chaki Jyotismita. An Introduction to COVID-19. In: Feng Simon James, Dey Nilanjan, Chaki Jyotismita, editors. Artificial Intelligence for Coronavirus Outbreak. Singapore: SpringerBriefs in Applied Sciences and Technology. Singapore: Springer; 2021. p. 1–22. https://doi.org/10.1007/978-981-15-5936-5_1.

[7] Guhakhumar K Sohaim, Kundu Souvik, Chakraborty Arpita, Banerjee Jyoti Sekhar. A Novel Approach to Predict COVID-19 Using Support Vector Machine. Data Science for COVID-19. Elsevier; 2021. p. 351–64. https://doi.org/10.1016/S2666-1393(20)60149-3.

[8] Imad, Muhammad, Naveed Khan, Farhat Ul Islam, Abul Hassan, and Adnan Hussain. 2020. “COVID-19 Classification Based on Chest X-Ray Images Using Machine Learning Techniques.”

[9] Irmak Emrah. COVID-19 Disease Severity Assessment Using CNN Model. IET Image Processing 2021;15(8):1814–24. https://doi.org/10.1049/iet-ipr.2021.0119.

[10] Jiang Guoquan, Wang Zhiheng, Liu Hongmin. Automatic Detection of Crop Rows Based on Multi-ROIs. Expert Sys with App 2015;42(3):2429–41.

[11] Islam Abdalla Mohamed, Ben Aissa Anis, Hussein Loay F, Taloba Ahmed I. A New Model for Epidemic Prediction: COVID-19 in Kingdom Saudi Arabia Case Study. In: Materials Today: Proceedings; 2021.

[12] Jiang Guoquan, Wang Zhiheng, Liu Hongmin. Automatic Detection of Crop Rows Based on Multi-ROIs. Expert Sys with App 2015;42(3):2429–41.

[13] Li Chenxi, Zhao Chengxue, Bao Jingfeng, Tang Bo, Wang Yunfeng, Gu Bing. Laboratory Diagnosis of Coronavirus Disease-2019 (COVID-19). Clinica Chimica Acta; Int J of Clin Chem 2020;510:35.

[14] Kumar Dubey Ashutosh, Narang Sushil, Ahmad Hassan, and Sally Elghamrawy. 2020. “Detection of Coronavirus (COVID-19) Associated Pneumonia Based on Generative Adversarial Networks and a Fine-Tuned Deep Transfer Learning Model Using Chest X-Ray Dataset.” ArXiv Preprint ArXiv:2004.01194.

[15] Kim Tae Jung, Goo Jin Mo, Lee Kyung Won, Park Chang Min, Lee Hyun Ju. Clinical, Pathological and Thin-Section CT Features of Persistent Multiple-Glassy Opaque Nodules: Comparison with Solitary Glassy-Opaque Nodule. Lung Cancer 2009;64(2):171–8.

[16] Kondziolka Doug, Coulbwell William T, Rutka James T. Introduction. On Pandemics: The Impact of COVID-19 on the Practice of Neurosurgery. J of Neurosurg 2020;133(1):1–2. https://doi.org/10.3171/2020.3.JNS201007.

[17] Kumar Dubey Ashutosh, Narang Sushil, Kumar Abhishek, Saubhili Satya Murthy, Garg Gaurav, Vicente. Performance Estimation of Machine Learning Algorithms in the Factor Analysis of COVID-19 Dataset. Comp, Mater & Continua 2021;66(2):1921–36. https://doi.org/10.32604/cmcc.2021.02151.

[18] Li Chenxi, Zhao Chengxue, Bao Jingfeng, Tang Bo, Wang Yunfeng, Gu Bing. Laboratory Diagnosis of Coronavirus Disease-2019 (COVID-19). Clinica Chimica Acta; Int J of Clin Chem 2020;510:35.

[19] Mazzilli Aldo, Florinio Claudio, Lorrick Alessandro, Mori Martina, Espoier Pietro Giorgino, Palumbo Dino, et al. An Automatic Approach for Individual H-Based Characterization of Lungs in COVID-19 Patients. Appl Sci 2021;11(3):1238. https://doi.org/10.3390/app11031238.

[20] McKee Brady J, Regis Shawn M, McKee Andrea B, Flacke Sebastian, Wald Christoph. Performance of ACR Lung-RADS in a Clinical CT Lung Screening Program. J of the Am Coll of Radiology 2015;12(3):273–6. https://doi.org/10.1016/j.jacr.2014.08.004.

[21] Islam Abdalla Mohamed, Ben Aissa Anis, Hussein Loay F, Taloba Ahmed I. A New Model for Epidemic Prediction: COVID-19 in Kingdom Saudi Arabia Case Study. In: Materials Today: Proceedings; 2021.

[22] Nour Majid, Comert Zafeer, Polat Kemal. A Novel Medical Diagnosis Model for COVID-19 Infection Detection Based on Deep Features and Bayesian Optimization. Appl Soft Comput 2020;97(December):106580. https://doi.org/10.1016/j.asoc.2020.106580.

[23] Sun, Nan-Nan, Ya Yang, Ling-Ling Tang, Yi-Ning Dai, Hai-Nv Gao, Hong-Ying Pan, and Bin Ju. 2020. “A Prediction Model Based on Machine Learning for Diagnosing the Early COVID-19 Patients.” Preprint. Infectious Diseases (except HIV/AIDS). https://doi.org/10.1101/2020.06.03.2020881.

[24] Wang Weiling, Xu Yanli, Gao Rujing, Lu Roujian, Han Kai, Wu Guizhen, Tan Wenjie. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. Jama 2020;323(18):1843–4.

[25] Yuan Mingli, Yin Wen, Tao Zhaoou, Tan Weijun, Hu Yi. Association of Radiologic Findings with Mortality of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. PloS One 2020;15(3):e0230548.

[26] Zebin Tahmina, Rezvy Shahadate. COVID-19 Detection and Disease Progression Visualization: Deep Learning on Chest X-Rays for Classification and Coarse Localization. Appl Intell 2021;51(2):1010–21. https://doi.org/10.1007/s10489-020-01867-1.

[27] Zhang Junyong, Chu Yingna, Zhao Na. Supervised Framework for COVID-19 Classification and Lesion Localization from Chest CT. Ethio J of Health Dev 2020; 34(4).

[28] Zhao Jinyu, Yichen Zhang, Xuehai He, and Pengtao Xie. 2020. “Covid-Ct-Dataset: A Ct Scan Dataset about Covid-19.” ArXiv Preprint ArXiv:2003.13865490.

[29] Zou You, You Jia-Ni, Zou Ya-Se, Wang Bin-Ru, Sun Liu, Xu Shan, Li Sheng-Lan, et al. Factors Affecting Recurrent Positive RT-PCR Results in Clinically Cured COVID-19 Patients. A Multicent Study 2021;9.

Osama R. Shahn is an Assistant professor in the Department of Computer Science at Jouf University, where he has been since 2016. He received his Ph.D. in Computer Science from the Menofia University in 2015. His research interests span Image Processing, Artificial Intelligence mainly through the application of data mining, statistics, and performance evaluation.

Hamoud H. Alshammari, is an assistant professor at the Computer Science and Information department at Jouf University in Saudi Arabia. He holds a Ph.D. in a Computer Science and Engineering from Bridgeport University in USA. His interest areas included but not limited to Artificial Intelligence, Data Analysis, Big Data and Hadoop, Internet of Things, and Text Mining.

Ahmed I. Taloba is an Assistant Professor of Information Systems Faculty at Computers and Information, Assiut University, Assiut, Egypt. He works now as an Assistant Professor of Computer Science Department, College of Science and Arts in Gyarayt, Jouf University Saudi Arabia. He received the PhD degree in 2015 from Faculty of Computers and Information, Assiut University, Egypt.
Rasha M. Abd El-Aziz is an Assistant Professor of Computer Science at Faculty of Computers and Information, Assiut University, Assiut, Egypt. She works now as an Assistant Professor of Computer Science Department, College of Science and Arts in Garayat, Jouf University Saudi Arabia. She received the PhD degree in 2017 from Faculty of Science, Assiut University, Egypt.