DOI: https://dx.doi.org/10.21123/bsj.2023.7403

# A New Model Design for Combating COVID -19 Pandemic Based on SVM and CNN Approaches

*Sura Monther Alnedawe* [1*] *Hadeel K. Aljobouri* [2]

[1]Computer Engineering Department, College of Engineering, Al-Nahrain University, Baghdad, Iraq.
[2]Biomedical Engineering Department, College of Engineering, Al-Nahrain University, Baghdad, Iraq.
*Corresponding author: bslaub2011@gmail.com
E-mail addresses: hadeel_bme77@yahoo.com

Received 9/5/2022, Revised 5/9/2022, Accepted 7/9/2022, Published Online First 20/1/2023, Published 1/8/2023



## Abstract:

In the current worldwide health crisis produced by coronavirus disease (COVID-19), researchers and medical specialists began looking for new ways to tackle the epidemic. According to recent studies, Machine Learning (ML) has been effectively deployed in the health sector. Medical imaging sources (radiography and computed tomography) have aided in the development of artificial intelligence(AI) strategies to tackle the coronavirus outbreak. As a result, a classical machine learning approach for coronavirus detection from Computerized Tomography (CT) images was developed. In this study, the convolutional neural network (CNN) model for feature extraction and support-vector machine (SVM) for the classification of axial lung CT-scans into two groups (COVID-19 and NonCOVID-19) had been proposed. A dataset used is 960 slices of CT scan collected from Iraqi patients /Ibn Al-Nafis teaching hospital. The performance metrics are used in this study (accuracy, recall, precision, and F1 scores). The results indicate that the proposed approach generated a high-quality model for the collected dataset, with an overall accuracy of 98.95% and an overall recall of 97 %.

**Keywords**: CNN, COVID-19, Machine Learning, Medical Imaging, RNN, SVM.

## Introduction:

ML is a subset of AI technology in which algorithms search through large data sets for patterns, learn from them, and perform jobs without being told how to fix the problem. ML algorithms have a lot of potential in the healthcare industry because of the volume of data collected for each patient. As a result, there are currently a number of effective machine learning implications in healthcare [1, 2]. Coronavirus is one of the contagious diseases generated by the coronavirus (SARS-CoV-2). Because until the beginning of 2021, there is no vaccine or licensed medicine for coronavirus, social isolation has been identified as the most effective control and prevention strategy and it is an important aspect of managing and preventing the virus's spread [3]. Diagnostic tests and antibody tests are the two types of standardized tests used to identify coronavirus. These procedures are time-consuming, necessitate the use of specific materials and tools, and are poor at generating true positive rates. As a result, current diagnostic and tracking procedures for coronavirus illness are useless [4]. The secondary diagnosis of coronavirus relies mostly on chest CT imaging. It may take a few seconds for the coronavirus chest CT-assisted diagnosis approach based on ML to provide reliable test findings. ML models are one of the most promising tools for assisting radiologists in early coronavirus detection. Moreover, it reduces radiologists' workload, improves detection accuracy and efficiency, and provides coronavirus patients with timely and accurate treatment [5].

## Applications of Machine Learning in COVID-19 Fighting:

ML methods can be separated into supervised and unsupervised learning. Under these methods, there are many common machine learning techniques used in various applications such as K-nearest neighbor (KNN), SVM, decision trees (DT ), K-means, and principle component analysis (PCA), Fig. 1 shows a classification of ML methods[6]. In this paper, the use of the supervised and unsupervised algorithms in combating COVID-19 and limiting its spread will be introducing.

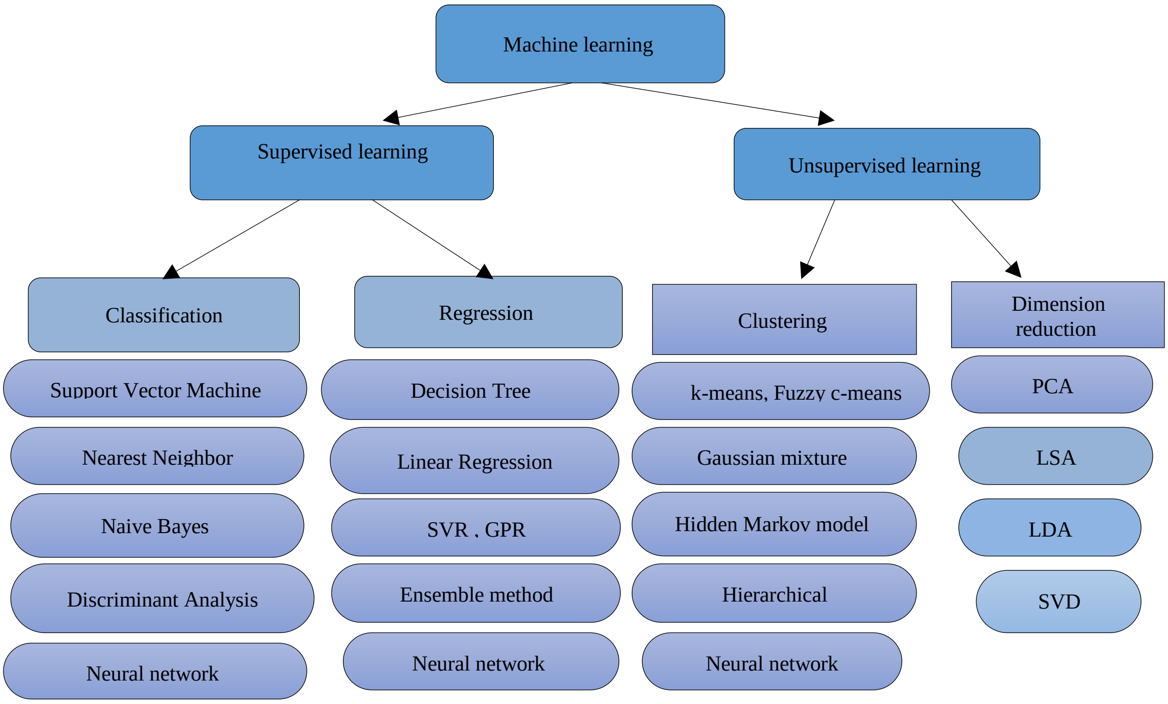


**Figure 1. Classification of ML methods**

## Supervised learning techniques

Supervised learning techniques are machine learning approaches that are based on past data and current data to predict future events. The learning process for machines begins with a process of training the data set to predict output values where the examples are correctly labeled according to the category to which they belong [5]. Implementation of some of the most significant supervised ML approaches, such as SVM, logistic regression (LR), and KNN, in tackling the coronavirus pandemic will be discussed.

## Support vector machine (SVM)

SVM is a controlled ML algorithm and powerful tool that can be deployed for both classification and regression challenges. It is widely used in the health sector because of its excellent precision and performance. As a result of its excellent performance, SVM has recently been utilized to battle the coronavirus epidemic [7,8]. Initially, M.Turkoglu built a model named COVIDetectioNet that relied on SVM to detect coronavirus using X-ray pictures. The dataset consisted of 6092 images categorized as normal, COVID-19, and Pneumonia, and the suggested model attained an accuracy of 99.1%.[9]. For accurate detection of coronavirus patients using X-ray images the authors use SVM as a classifier, the maximum accuracy, sensitivity, and specificity were achieved when fusion SVM with CNN and the sobel filter were used, with 99.02 %, 100 %, and 95.23 %, respectively[10]. Furthermore, C. Zhou et al proposed a model for coronavirus detection, they used a mix of image regrouping and ResNet-SVM to achieve their goal. SVM was used in the final step for the recognition of the feature extracted, this model accomplished an accuracy of 93% and sensitivity of 88% [11]. In another study SVM frame was utilized to ride the issues of early diagnosis of coronavirus, It is the first research to design an SVM frame to perform simultaneous multi-job and multi-modal learning for medical image analysis using CT images the accuracy accomplished by this model reached 92.57% and the sensitivity of 91.44% [12]. In the presented work, SVM is proposed as a classifier of CT images into normal and abnormal COVID-19 groups.

## Logistic Regression

LR is a ML technique for modeling the prospect of a specific category or occurrence. This strategy is used when the data are linearly distributed and the outcome is binary or dichotomous. That is to say, for binary classification problems, logistic regression is commonly utilized [13]. LR was recently used to mitigate the coronavirus pandemic. The classification of patients with COVID-19 and now whether the infected has risked or not using dataset collected from Kaggle and achieving an accuracy of 92% [14]. On the other hand, the LR and Multinomial Nave Bayes were employed to classify a clinical report, and they used classical and ensemble ML methods to divide textual clinical reports into four categories. Term frequency/inverse document frequency (TF/IDF), Bag of words (BOW), and report length were utilized to implement feature engineering. The results revealed that LR and Multinomial Nave Bayes had

96.2 % performance in identifying covid-19, outperforming other ML techniques[15]. Another study used LR for forecasting the spread of coronavirus based on the confirmed data [16].

**K Nearest-Neighbored (KNN)**

The KNN algorithm is a supervised ML technique that could be utilized to handle both classification and regression issues. It saves all patients' cases diagnosed and classifies new data or cases using a similarity score. It's most typically used to classify a data point depending on its neighbors' classifications [17]. Several researchers used the KNN algorithm for fighting COVID-19, Mukherjee et al. proposed an enhanced KNN for the detection of coronavirus, they used a mathematical function to determine the value of k not randomly choose it. This algorithm was applied to 7 benchmarks COVID-19 datasets collected from different countries. The outcome showed that the enhanced KNN classifier is well than the KNN classifier and the enhanced KNN along with features selection is well than KNN without feature selection [18]. Another study uses nine machine learning algorithms, KNN with SVM, bagging, boosting, SVM, bidirectional long short-term memory, DT, naïve Bayes, RF, and multinomial LR as the classifier. These algorithms used with Noise filter algorithms are used to eliminate noise from the datasets applied in this research to show its impact of it on the machine learning algorithms. The result showed that ML models have produced good results, KNN achieved an accuracy of 100% after performing noise filtering operations for the prediction of coronavirus cases [19]. Hasoon et al. used KNN as a classifier with features extractions operators (LBP, HOG, and Haralik) the models were tested depending on test samples of 5000 images for classification and detection of coronavirus. The result showed that the (KNN-LBP) model exceeds other models with just an average accuracy of 98.66 %, a sensitivity of 97.76 %, a specificity of 100 %, a precision of 100 %, an error rate of 1.34 %, and zero false positives[20].

**Decision Tree (DT)**

The DT algorithm is a type of supervised learning algorithm. The aim of using a DT is to establish a training model that can be used to forecast the category or value of the input variables by learning simple decision rules learned from prior data (training data). Because of its ease of use and resilience, DT is widely employed in a variety of fields [21]. Recently, the DT algorithm was known in the medical field. For example, a decision tree model was established to detect the severity of the condition of children with COVID-19. The clinical laboratories and epidemiological information of 105 unwell children were obtained from a Chinese hospital from February 1 to March 3, 2020. COVID-19 positivity was found in 105 children, including 41 females and 64 males. The female hitting rate (39.05%) is lower than the male infection rate (60.95 %). The proposed approach performed admirably, receiving a perfect F1 score of 100 %[22]. Yoo et al. suggested a deep learning-based decision tree classifier to detect coronavirus from CXR images. The classifier's first binary decision tree differentiates between normal and anomalous CXR images, while the second tree detects aberrant photos with symptoms of tuberculosis, and the third tree does the same for coronavirus images. The first and second decisions were both 98 % and 80 % accurate, respectively, while the third decision was 95 % accurate on average [23]. In order to forecast the death rate in patients infected with coronavirus, a comparative analysis of different ML approaches including DT, RF, KNN, SVM, LR, and ANN was used. This research employed a dataset of 117,000 coronavirus ill cases from both males and females. For the prediction of the death rate, the model had a 93% accuracy rate. While employing 10-fold cross-validation, DT was able to reach an accuracy of 90.63% [24]. A summary of some applications of supervised ML methods for coronavirus fighting is shown in Table. 1.

**Unsupervised Learning Technique**

Unsupervised learning analyzes and clusters unlabeled datasets utilizing ML methods. These algorithms detect underlying patterns or data groups even without human participation. Because of its ability to find similarities and divergences in information, it is the greatest choice for exploratory data analysis, cross-selling tactics, customer segmentation, and image recognition. Implementation of some of the important unsupervised ML methods which are PCA and K-means in fighting the coronavirus pandemic will be discussed.

**Principle Component Analysis(PCA)**

PCA is a widely used unsupervised ML approach in a wider range of usage including exploratory data analysis, dimension reduction, information fusion, and data de-noising. The primary objective of PCA is to produce the most significant data features. It guarantees that numerous feature variables are decreased without sacrificing information [25]. The authors used PCA for reducing the features of images extracted from CNN, this method achieved good accuracy of 97.84% [26]. In another research aligned big genome sequences were subjected to PCA, and numerical values were transformed from symbols using an established approach for protein sequence cluster analysis. This research developed a

Baghdad Science Journal
2023, 20(4): 1402-1413
P-ISSN: 2078-8665
E-ISSN: 2411-7986

quick tool for analyzing high-volume genome sequences like coronavirus, which was effectively used for over 20,000 sequences and might give mutation direction data for coronavirus studies [27]. The RNA expression patterns of 16 coronavirus patients and 18 healthy control subjects were analyzed using PCA-based unsupervised feature extraction (PCAUFE). From 60,683 potential probes, 123 genes were identified as crucial for COVID-19 development, including immune-related genes [28]. Ozturk et al classified coronavirus from X-ray and CT images. They used four feature extraction methods to extract the features vector and the feature vector is decreased in size by PCA to refuse interconnected features in the feature vector. Since the PCA architecture can work independently from the number of samples, it has provided very efficient classification [29].

**K-means Learning**

One of the most basic and often used unsupervised ML algorithms is k-means learning. In other terms, the K-means algorithm finds k centroids and then assigns each data point to the cluster with the fewest centroids. In simple terms, k-means clustering allows us to automatically cluster data into numerous groups by identifying distinct types of groups in unlabeled datasets without the need for data training. The purpose of this centroid-based strategy is to minimize the summation of distances among data points and their associated clusters by connecting each cluster to a centroid [30]. The authors utilized unsupervised ML algorithms (k-means) to identify groups of countries that depend on the data.

**Table 1. Summary of some supervised ML for fighting COVID-19**

| Author | Methodology | Sample type | Outcome results |
|---|---|---|---|
| p.Sethy et al [8] | Deep features of different 13 models of CNN plus SVM | X-ray images(COVID-19,pneumonia,and normal) | Highest accuracy achieved by ResNet50 plus SVM=95.33% |
| M.Turkoglu et al [9] | SVM for diagnoses | X-ray images, (Normal,coronavirus, and Pneumonia) | Accuracy=99.18 |
| D. Sharifrazi et al [10] | CNN-SVM with Sobel filter | X-ray images, distributed as Normal (healthy), coronavirus | Accuracy=99.02%,Sensitivity=100 % Specificity =95.23% |
| C. ZHOU et al [11] | SVM for recognition. | COVID-19 X-ray images | Accuracy=93.1% |
| R.Hu et al [12] | classification and regression using the Multi-Modality SVM (MM-SVM) | CT images of COVID-19 | Accuracy=92.57, Sensitivity =91.44% |
| A. B. Majumde et al [14] | LR to find whether the infection of COVID-19 risk or not | two types of tomography-EIT electrical impedance tomography)and UST(ultrasound transmission tomography) | Accuracy=92% |
| A.Khanday et al [15] | LR for detecting of coronavirus | clinical text data | Accuracy =96.2% |
| W. M. Shaban et al [17] | Hybrid feature selection and KNN classifier | CT scan images | Accuracy= 96% |
| R.Mukherjee et al [18] | KNN for detection of coronavirus | clinical data | Accuracy=95.75% |
| D. Oyewola et al [19] | KNN for predication coronavirus | epidemiological data of coronavirus infection cases | Accuracy=100% |
| J. N. Hasoon et al [20] | KNN for classification coronavirus | X-ray images (healthy and pneumonia) | Accuracy =98.66%. Sensitivity=97.76%, Specificity = 100% Precision = 100% |
| H.Yu et al [22] | DT for detection coronavirus | Clinical data | F1 score=100 |
| S. Yoo et al [23] | deep learning-based DT tree classifier | CXR images | Accuracy=95% |
| M.Pourhomayoun et al [24] | DT to furcating the death rate in patients infected with coronavirus | Clinical data | Accuracy=90.63 |
| A.M. Ali & et al [31] | KNN for detection COVID-19 | CT scan images | Accuracy= 91.304 % |
| H. Arslan et al [32] | CpG island features and KNN classifier | genomic sequence of human coronaviruses | Precision =98.4%,Recall =99.2% F-measure =98.8%,Accuracy =98.4% |
| G.Feng et al [33] | LR for diagnose COVID-19 | Time series | Validation cohort,AUC = 0.8409 Precision = 40%,Recall = 100% |
| L.YANG et al [34] | LR for Prediction COVID-19 | CT images(52 patients) | Sensitivity = 100% ,Specificity =89% |
| **Proposed model** | **CNN for feature extraction+ SVM as a classifier** | **CT images** | **Accuracy= 98.95%** |

Disease incidence rates, air pollution measures, socioeconomic status, and health system coverage were all submitted to the algorithm as inputs. The model was shown to be able to stratify nations based on the number of verified coronavirus cases (p0.001) [35]. In another study, the K-means method was used in clustering coronavirus cases in Indonesia's provinces using November 2020 data. This study looked more into the statistical side. There are three variables originally applied in this paper, i.e. numbers of coronavirus certain cases, recovered, and death [36]. J. Wu and S. Sha used k-mean to learn about the pattern of the coronavirus pandemic in the United States and the implications for mitigating the disease. They applied the k-means algorithm to discover the pattern of disease curves in different cases [30]. In this study, the 3M k-means algorithm was utilized to identify those who had been in close contact with COVID-19 infected patients, as well as to determine the risk of infection. This algorithm decreases the amount of data used by ignoring the excluded risk categories and significantly reducing computational time, achieves significant cost savings, and provides psychological security for working people to resume work. Therefore, the use of this algorithm can help in controlling the epidemic coronavirus [37]. On other hand L. M. Abdulrahman et al was used to study the adverse reactions of the global COVID-19 vaccine based on the cluster ML algorithms. The K-means, Density-Based Algorithm, Hierarchical Clustering Algorithm, and Expectation and Maximization (EM) algorithms were compared in this study. K-means have outperformed the rest of the algorithms by using them extensively in different types of the coronavirus vaccine reverse vaccine dataset [38]. A summary of some applications of unsupervised ML methods for coronavirus fighting is shown in Table 2.

**Table 2. Summary of some unsupervised ML for fighting COVID-19**

| Author | Methodology | Sample type | Outcome results |
|---|---|---|---|
| Aswathy A.L et al [26] | CNN+PCA | CT images (COVID-19 –ve&COVID-19 +ve images) | Accuracy=97.84% |
| B.Wag et al [27] | PCA | genome sequences of coronavirus | The PCA approach is supposed to offer a quick analytical method with minimal data cleaning needs, however classifications may be constrained. |
| K.Fujisawa et al [28] | PCA-based unsupervised feature extraction (PCAUFE) | RNA expression profiles of 16 coronavirus patients and 18 healthy | successfully and robust methodology |
| S. Öztürk et al[29] | PCA for feature reduction + four feature extraction technique +SVM as classifier | chest X-ray and CT images of patients with coronavirus | with 260 samples and 20 features (Accuracy= 88.46%, Precision= 89.75%. Specificity= 97.32%) with 495 samples and 20 features (Accuracy= 94.23%, Precision= 96.73%. Specificity= 98.54%) |
| J.Wu et al [30] | K-means | COVID-19 case data with time series | (Global Moran's I = 0.2288, z-score = 185.4885, p-value < 0.0001) |
| R. Carrillo-Larco et al [35] | K-means for classification | Text data | p < 0.001 |
| F. Virgantari et al [36] | 4 clustering algorithms | Vaccine Adverse Reactions data is linked to consumer information includes 7 characteristics and 5471 cases. | The K-means algorithm requires the shortest time (0.01) seconds and has the best accuracy (67 %). |
| X. Liang et al [37] | 3M K-means | traffic data | simple and effective algorithm speeds up the collection process nearly 80 times for a city of over 5 million people |

### The Proposed Model Design

The overall algorithmic framework is given in Fig. 2.

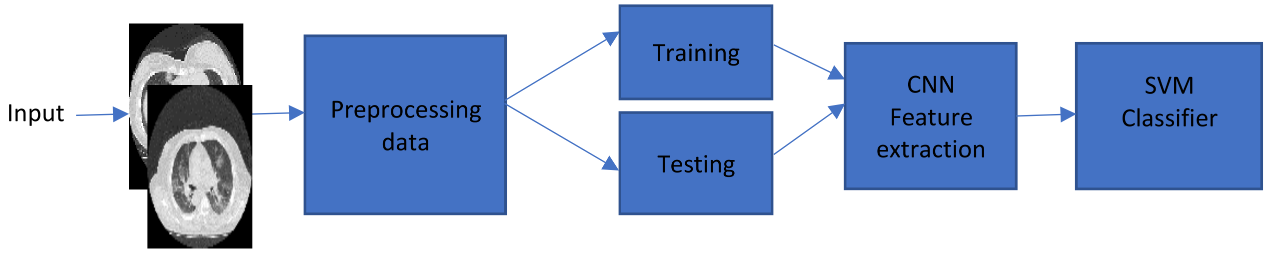


**Figure 2. The block diagram of the proposed model**

## Materials and Methods:

### COVID-19 Dataset

### Data collection and Acquisition

The data were collected from Iraqi patients at Ibn Al-Nafis teaching hospital in Baghdad, Iraq. The scanning device was Toshiba (Aquilion 64-slice multi-slice helical CT). The following scanning parameters were used to capture the CT images:

- Tube voltage 120 kV
- Current 500 mA
- Slice thickness of 5 mm
- Scanning with no contrast introduced.

The reconstructed image is in DICOM format and has a spatial resolution with a size of 512× 512 (rows and columns). The overall chest CT scans composed were of 217 cases, these included (52, 89, 76) for coronavirus, Non coronavirus, and excluded cases respectively. The ground truth concerning this data was established based on the diagnosis decisions of three experienced radiologists. Cases are eliminated based on a variety of criteria, including circumstances in which the contrast was employed for purposes other than diagnosis, the total cases for each class are listed in Table 3.

**Table 3. Details of the dataset**

| Classes | Total cases | Male | Age range | Female | Age range |
|---|---|---|---|---|---|
| COVID-19 | 52 | 28 | (26-71) | 24 | (32-75) |
| Non-COVID-19 | 89 | 40 | (19-74) | 49 | (19-79) |

### Data Preparation and Interpretation

The total axial CT images acquired from collected cases included 480 images for each class, 480 images for the COVID-19 case, and 480 images for the Non-COVID19. Three radiological professionals with more than 10 years of experience chose and annotated the number of slices for each patient depending on the major symptoms of COVID-19 and their diffusion in the lung area. These symptoms vary depending on the severity of the illness in each patient:

- Early-stage: inflammatory infiltration with patchy or segmental pure Ground Glass Opacities in the subpleural or peri-broncho-vascular sections of one side or both lungs and Vascular dilation.
- The introductory stage: more widespread pure GGOs in numerous lobes of both lungs, consolidation of some lesions, thickening of the interlobular septum, and the appearance of a crazy-paving pattern; air bronchograms are normal.
- The advanced stage: the lesions were disseminated in both lungs, consisting mostly of GGOs and united. Parenchymal bands and a small degree of pleural effusion characterize the advanced stage.

CT images are obtained in DICOM format, which has a depth of 16 bits. Because high-resolution photographs need significant computational power, considerable processing time, and huge memory. When dealing with deep learning, the images were transformed to bitmap format (.bmp). This reduced the size of the image. Another crucial step is data anonymization, which safeguards patient-related private data. During image conversion, the metadata is stripped from the photos. All the images used in the classification model were pre-processed by resizing the images to (256 x 256) to obtain impersonation by a different set of features. Samples of the collected CT scans are shown in Fig. 3.

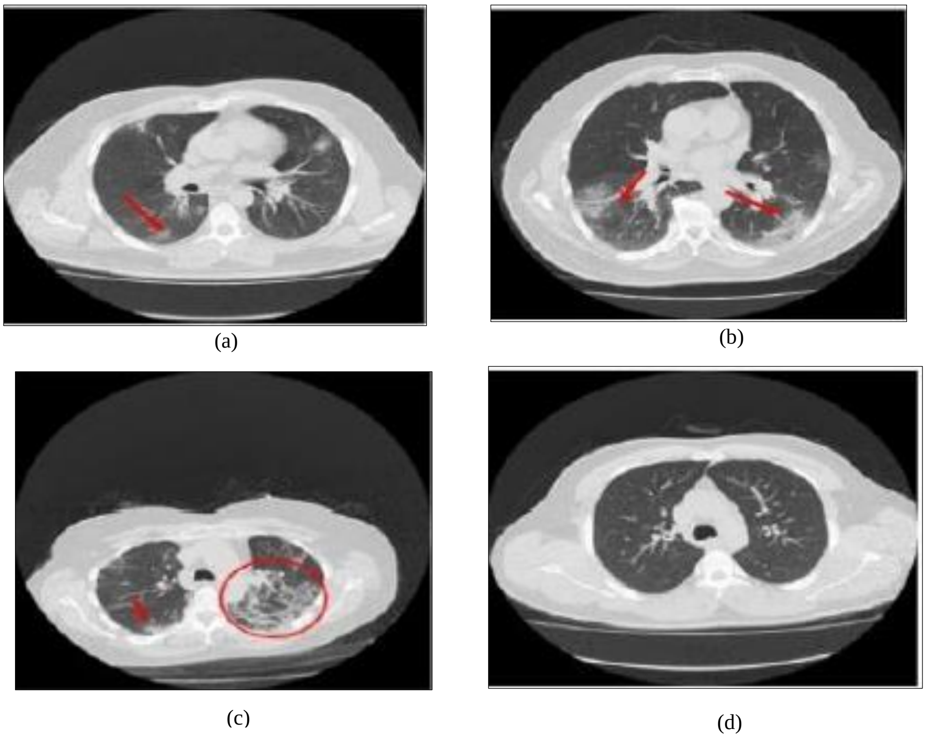

**Figure 3. Sample images of collected data: (a) early-stage COVID-19 infection with sub-pleural GGO, (b) progressive stage of COVID-19 infection with more diffused GGO and consolidation of some lesions (c) advanced stage of COVID-19 infection with signs of pleural effusion, and (d) Healthy CT scan**

## Methods:

### Image preprocessing

Pre-processing is a very important step before the images are fed into the models, in order to increase efficiency and reduce complexity [39]. The size of input images was chosen experimentally taking into account the computational power of the system to train the model, therefore images were resized to a smaller size equal to (256×256), and rescale the values of the pixels for the 8 bits image (between 0 and 255) to the scale of [0, 1]. The data set is separated into 80% for training and 20% for testing.

### Feature Extraction

Convolutional neural networks (CNNs) are artificial neural networks that are utilized to extract features from data and classify it. The structure of CNN consists of two parts, the initial part work for feature extraction included (the input layer, convolutional layer, activation layer, pooling layer, and the flatten layer), and another part work for classification including the dense fully connected layer and the output layer [40 41]. The benefit of CNN is that it can extract features on its own using its convolution kernels; further, there are a set of filters in the convolutional layers known as kernels or convolutional filters. These filters convolve with the input image in order to give the features map that uses after that to train the model.

In our model, the number of filters was used 10,10, 20, and 30 respectively for every layer to yield a tensor of feature maps. The most common activation function is Rectifier Linear Unit (ReLU) was used in this work for calculating the feature maps that are produced from a convolutional layer, The Relu function is defined as[42]:-

$$F(x) = \begin{cases} x & as\ x \geq 0 \\ 0 & as\ x < 0 \end{cases} \qquad \dots 1$$

The pooling layer comes to minimize the dimensions of the features map while trying to preserve the contained information via the pooled maximum or average value for the defined kernel size (usually 2x2). In this work we used (max pooling) with a kernel size of 2x2 and strides of 2, 2, stride refer to the number of steps that the filter moves along both dimensions(x,y)vertically and horizontally.

Finally, a flattened layer was used to obtain a feature vector of 27000 lengths, that represents the number of features extracted from CNN layers. Fig.4 shows the deep CNN as feature extraction with four convolutional layers and three max pooling.

Baghdad Science Journal
2023, 20(4): 1402-1413
P-ISSN: 2078-8665
E-ISSN: 2411-7986

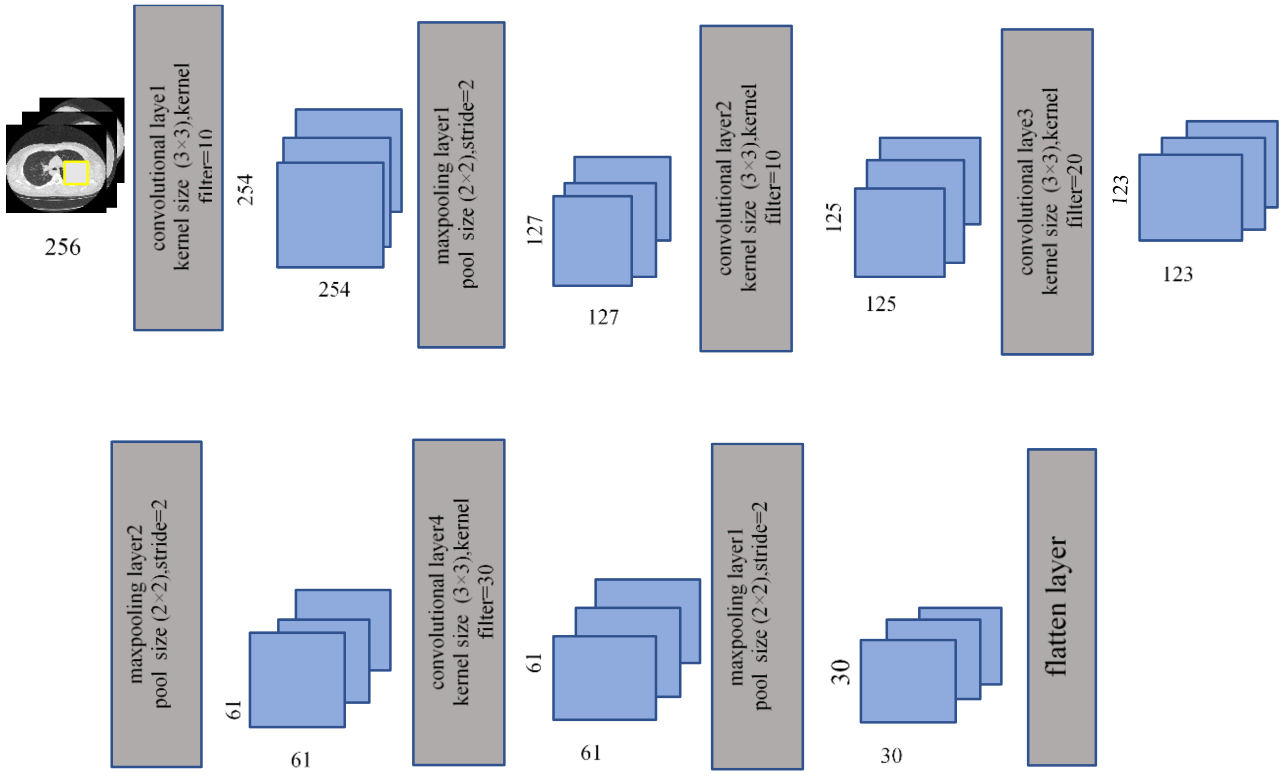


**Figure 4. Architecture of CNN deep model**

## SVM Classification

SVM is a statistical learning algorithm established by Vapnik to fix classification problems by finding the best hyperplane value and the results obtained from the optimal classification [43].

As illustrated in Fig. 5, the SVM classifier depends on calculating the best hyperplane to differentiate between two classes. The hyperplane is considered to be optimal if the distance between the hyperplane and the neighboring data from every class is the greatest distance possible. [9] . In a classification issue that containing two classes and n elements, learning data is X = {$x_i$, $y_i$} and i = {1,2,……, n}, where x represents the class labels corresponding to -1 or + 1 as a result of the calculations shown in Eq.2 and y represents the input vector of the features in size N. [11]

$$Linear\ SVM = \begin{cases} wx_i + \mathrm{b} \ \geq \ 1 \text{ for } y_i \geq \mathbf{1} \\ wx_i \ + \mathrm{b} \ \leq \ -1 \text{ for } \ y_i \ \leq -1 \end{cases} \quad \dots 2$$

where w and b denote the weight vector and hyperplane respectively. The basic method of SVM work is linear classification, the hyperplane is found in the classification problems that can be linearly separated which ensures the greatest possible separation between the training data of two classes that are near to one another, as an outcome of the optimization solution the problem appears in Eq.3 [12].

$$\frac{1}{2}||w||^2 \text{ for } \ y_i(wx_i\text{+b}) \geq +1 \qquad \dots 3$$

Thus, the purpose of SVM is to compute the highest value of *d* when the value of $\frac{1}{2}||w||^2$ is minimum, where

d denotes the distance between the supports vector and the hyperplane.

In classification problems a positive slack variable (ξ ) and a smoothing parameter (C) are added for handling the non-linearity as displayed in Eq.4[12].

$$y_i\ (wx_i\text{+ b}\ ) \geq 0\ -\xi \quad \text{for i=1,2,3…….m} \qquad \dots 4$$

In this study SVM model was configured with a linear kernel, and the regularization parameter C=1.0 was suggested by[44] and give the best classification result.

Baghdad Science Journal
2023, 20(4): 1402-1413
P-ISSN: 2078-8665
E-ISSN: 2411-7986

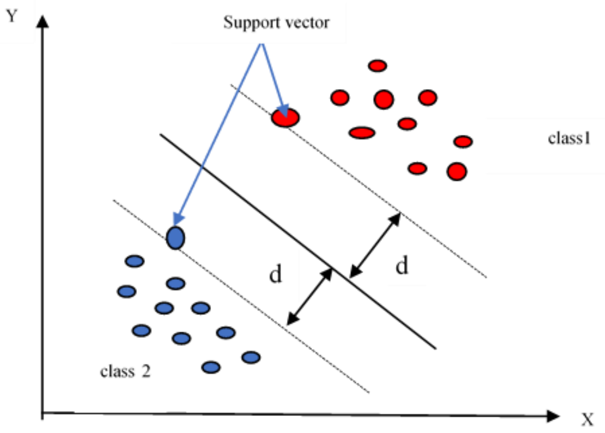


**Figure 5. Support Vector Machine**

## Results:

The dataset was divided in this study into two separate datasets at random, 80% and 20 % used for training and testing, respectively. These models were performed using HP laptop Intel® Core™ i7-10510U, RAM (8 GB), and Microsoft Windows 10 Pro. The code was done by Jupyter Notebook which is a web application available online using python programming language and Keras API with TensorFlow platform for ML.

The SVM classifier uses the deep feature had been extracted from CNN network. The classification is then carried out, and the performance of all classification models is evaluated.

For our investigation, the confusion matrices were employed to assess the model's performance. The following formulas were utilized to determine the Accuracy, Precision, Recall (sensitivity), and F1 score for the implemented models based on the predicted values:

Accuracy = ((TP+TN) / (TP+FP+FN+TN)) *100% … 5

Precision = (TP / (TP+FP)) *100% … 6

Recall = (TP / (TP+FN)) *100% … 7

F1 Score = (2*(Recall * Precision) / (Recall + Precision)) *100% … 8

TP, FP, TN, and FN are the acronyms for True Positive, False Positive, True N negative, and False Negative, respectively. TP is the ratio of positive (Covid-19) that the model properly identifies as Covid-19 for a particular test dataset and model. FP is the ratio of negative (normal) to (positive) that the model erroneously classifies (Covid-19). The ratio of negative (normal) that is rightly categorized as normal is TN, whereas the ratio of positive (coronavirus) that is wrongly labeled as negative is FN (normal).

Fig. 6 specifies the accuracy of the training and testing phases. The accuracy obtained by the proposed model is 98.95%. The precision, recall and f1 scores are 100 %, 97 % and 98.94%, respectively

The confusion matrix gives us an idea about the performance of the model by calculating the accuracy of the model, the confusion matrix in Fig.7 shows the good performance obtained in this study, Among 192 images used to test the model, two images were misclassified by our model these images belong to the NonCOVID-19 cases. The threshold value used in this study is 0.5, any value greater than threshold refers to the COVID-19 class and under it belongs to the NonCOVID-19

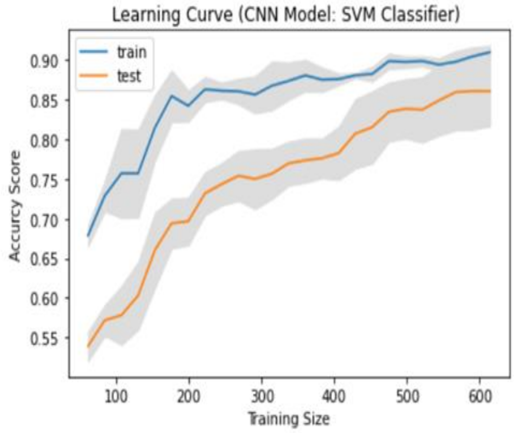


**Figure 6. Plot of the SVM Learning curves CNN features**

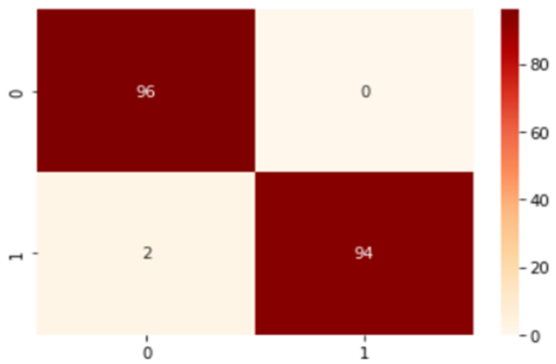


**Figure 7. Confusion matrix**

## Discussion:

In this article first, the applications of ML in coronavirus fighting have been offered. The applications of ML methods include supervised ML methods which are SVM, LR, KNN and unsupervised ML methods such as PCA and K-means have been introduced. All the common and practical information including writer's name, methodology, sample kind and outcome result of the proposed model, related to the employment of ML for fighting COVID-19 have been abstracted and offered in Tables 1 and 2. The research based on traditional supervised and unsupervised ML techniques (SVM, DT, K-means, PCA, etc.) showed low performance when applied alone. While it is performing well when combined with other techniques. Therefore, it was found that the grouping of CNN-based features and supervised classifier algorithm gives good results for classifying axial lung CT-images due to the intrinsic features provided by CNN feature extraction layers followed by the optimal parameter values of the SVM classifier. Combining the features obtained by pretrained network Resnet with SVM classifier usually had a good performance and reasonable accuracy levels on classification 95.33% when applied to X-ray images of lung[8]. Features that were chosen using the Relief feature selection algorithm from the layers of the Alexnet architecture, and combined with SVM classifier have further enhanced classification accuracy 99.18% on X-ray images, reflecting the efficiency of CNN features[9].

Using SVM classifier to classify a set of CNN features, were provided good results (98.24%) on the classification of X-ray images, and when fusion CNN with SVM and sobel filter on the same data gave excellent results (99.02)[10]. In this study, a CNN deep learning algorithm that was built from scratch was Used to extract features and feed it's to SVM as a classifier for the discovery of coronavirus from lung CT images. The dataset utilized for model development was locally taken from the Iraqi health center. When comparing the proposed model in this study and other related studies listed in Table. 1, it has been stated that the proposed model has achieved relatively good results for the data set used in this study. Table 4 below shows the performance metrics of some previous methods introduced for solving similar classification problems in comparison with the model proposed in this study.

**Table 4. Performance analysis for the purpose of comparing to previous methods.**

| Authors | Methodology | Accuracy% | Precision% | Recall% | F1score% |
|---|---|---|---|---|---|
| P.Sethy et al | Resnet+svm | 95.33 | ——— | 95.33 | 95.43 |
| M.Turkoglu et al | Alexnet+svm | 99.18 | 99.48 | 99.13 | 99.3 |
| D.Sharifrazi et al | CNN+SVM | 98.24 | —— | 98.8 | 98.39 |
| | CNN+sobel filter+SVM | 99.02 | —— | 100 | 99.35 |
| Proposed model | CNN+SVM | 98.95 | 100 | 97 | 98.94 |

## Conclusion:

The application of AI-based technologies, such as ML models, is one of the most promising strategies for assisting radiologists in the early identification of coronavirus. Besides, it decreased the workload of the technicians, increased the accuracy and efficiency of the detection, given appropriate time response and exact medicament for the patients of coronavirus. This study highlights the latest usages of ML to combat coronavirus. It was discovered that the grouping of CNN-based features and ML classifier algorithms were mostly used for coronavirus pandemic prediction and give good results for classifying axial lung CT-images. For that, this paper offered an overview of several research and studies to mitigate the coronavirus pandemic and limit its spread. In our methodology, a model for processing computed tomography images of the lung and detecting coronavirus disease was presented based on extracted features using CNN model and fed to SVM for classification CT-images. In comparison to the other related works, this work used a real dataset from Iraqi health centers. So, it is worth noting that many obstacles faced the study when compared to other works that used free available datasets; including the process of data collection, archiving, and avoiding poor quality data that leads to inaccurate predictions. Also, the most important thing is the risk of infection virus during the data collection. This work is vitally important for Iraqi healthcare centers by finding quick and accurate methods to mitigate the overload lies on the radiologist for detecting COVID-19 where it achieved a 98.95% accuracy rate in diagnosing the disease. For future works more data can be collected from multi-healthy centers further generalizing the model in its performance for COVID-19 diagnosis.

## Acknowledgment

The authors of this paper would like to thank all of Dr. Jawad AlKaaby and Dr. Zahraa AlKhafajy at Ibn Al-Nafis Teaching Hospital; Dr. Zainab Kassim Aljobouri at Kadhimiya Teaching Hospital. All of them are specialist radiologist who act to conduct data preparation, annotation providing the ground truth labels for the dataset.

## Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Authors sign on ethical consideration’s approval.
- Ethical Clearance:–This study was conducted according to the approval of the Research Ethics Committee of the Biomedical Engineering at Al-Nahrain University (02/2020)

## Authors' contributions statement:

S.M.A. study conception and design, acquisition of data, analysis and interpretation of data, drafting of manuscript. H.K.A. study conception and design, analysis and interpretation of data, and critical revision.

## References:


1. Perumal V, Narayanan V, Rajasekar SJS. Prediction of COVID-19 with Computed Tomography Images using Hybrid Learning Techniques. Dis Markers 2021; 2021: 1–15.
2. Mosa ZM, Ghae NH, Ali AH. Detecting keratoconus by using SVM and decision tree classifiers with the aid of image processing. Baghdad Sci J 2019; 16(4): 1022–1029. https://doi.org/10.21123/bsj.2019.16.4(Suppl.).1022
3. Dogan O, Tiwari S, Jabbar MA, Guggari S. A systematic review on AI/ML approaches against COVID-19 outbreak. Complex Intell Syst 2021; 7: 2655–2678.
4. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert Rev Mol Diagn 2020; 20: 453–454.
5. Khan M, Mehran MT, Haq ZU, Ullah Z, Naqv SR, Ihsan M, et al. Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review. Expert Syst Appl 2021; 185: 115695.
6. Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Computer Science; 2. Epub ahead of print 2021. https://dx.doi.org/10.1007/s42979-021-00815-1 .
7. Enireddy V, Kumar MJK, Donepudi B, Karthikeyan C. Detection of COVID-19 using Hybrid ResNet and SVM. IOP Conf Ser Mater Sci Eng. 2020; 993: 012046. https://dx.doi.org/10.1088/1757-899X/993/1/012046 .
8. SethyPK, Behera SK, Ratha PK, Biswas P. Detection of coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine. Int J Math Eng Manag Sci. 2020;5(4):643-651.
9. Turkoglu M. COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble. Appl Intell 2021; 51: 1213–1226.
10. Sharifrazi D, Alizadehsani R, Roshanzamir M, Joloudari JH, Shoeibi A, Jafari M, et al. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. Biomed Signal Process Control 2021; 68: 102622.
11. Zhou C, Song J, Zhou S, Zhang Z, Xing J. COVID-19 Detection based on Image Regrouping and ResNet-SVM using Chest X-ray Images. IEEE Access. 2021; 9: 81902-81912. https://dx.doi.org/10.1109/ACCESS.2021.3086229 .
12. Hu R, Gan J, Zhu X, Liu T, Shi X. Multi-task multi-modality SVM for early COVID-19 Diagnosis using chest CT data. Inf Process Manag. 2022; 59(1):102782. https://dx.doi.org/10.1016/j.ipm.2021.102782 .
13. Rymarczyk T, Kozłowski E, Kłosowski G, Niderla K. Logistic regression for machine learning in process tomography. Sensors (Switzerland) 2019; 19: 1–19.
14. Majumder AB, Gupta S, Singh D, Majumder S. An intelligent system for prediction of COVID-19 case using machine learning framework-logistic regression. J Phys Conf Ser. 2021; 1797. https://dx.doi.org/10.1088/1742-6596/1797/1/012011.
15. Khanday AMUD, Rabani ST, Khan QR, Rouf N, Din MMU. Machine learning based approaches for detecting COVID-19 using clinical text data. Int J Inf Technol 2020; 12: 731–739.
16. Almeshal AM, Almazrouee AI, Alenizi MR, Alhajeri SN. Forecasting the Spread of COVID-19 in Kuwait Using Compartmental and Logistic Regression Models. Appl Sci 2020; 10: 3402.
17. Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud MA. A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowl.-Based Syst. 2020; 205: 106270.
18. Mukherjee R, Kundu A, Mukherjee I, Gupta D, Tiwari P, Khanna A, et al. IoT-cloud based healthcare model for COVID-19 detection: an enhanced k-Nearest Neighbour classifier based approach. Computing. 2021:.1-21 https://dx.doi.org/10.1007/s00607-021-00951-9 .
19. Oyewola DO, Dada EG, Misra S, Damaševicius R. Predicting covid-19 cases in south korea with all k-edited nearest neighbors noise filter and machine learning techniques. Info.2021; 12(12):528. DOI: https://dx.doi.org/10.3390/info12120528 .
20. Hasoon JN, Fadel AH, Hameed RS, Mostafa SA, Khalaf BA, Mohammed MA. COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images. Results Phys 2021; 31: 0–7.
21. Milanović M, Stamenković M. CHAID Decision Tree: Methodological Frame and Application. Econ Themes 2016; 54: 563–586.
22. Yu H, Shao J ,Guo Y, Xiang Y, Sun C, Yuan Y. Data-driven discovery of a clinical route for severity detection of COVID-19 paediatric cases. IET Cyber-systems Robot 2020; 2: 205–206.
23. Yoo SH, Geng H, Chiu TL,Yu SK, Cho DC,Choi MS, et al. Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. Front Med 2020; 7: 1–8.
24. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. Smart Health. 2021; 20: 1-8. https://dx.doi.org/10.1016/j.smhl.2020.100178 .
25. Ma J, Yuan Y. Dimension reduction of image deep feature using PCA. J Vis Commun Image Represent 2019; 63: 102578.
26. Aswathy A.L., S. A.H., S.S. V.C. COVID-19 diagnosis and severity detection from CT-images using transfer learning and back propagation neural network. J Infect Public Health 2021; 14(10): 1435–1445.
27. Wang B, Jiang L. Principal Component Analysis

Applications in COVID-19 Genome Sequence Studies. Cogn Comput. 2021; 13: 1-12. https://doi.org/10.1007/s12559-020-09790-w .
28. Fujisawa K, Shimo M, Taguchi YH, Ikematsu S, Miyata R. PCA-based unsupervised feature extraction for gene expression analysis of COVID-19 patients. Sci Rep 2021; 11: 1–11.
29. Öztürk Ş, Özkaya U, Barstuğan M. Classification of Coronavirus (COVID-19) from X-ray and CT images using shrunken features. Int J Imaging Syst Technol 2021; 31: 5–15.
30. Wu J, Sha S. Pattern recognition of the covid-19 pandemic in the united states: Implications for disease mitigation. Int J Environ Res Public Health 2021; 18: 1–13.
31. Ali AM, Ghafoor KZ, Mulahuwaish A, Maghdid HS, Hussein SM, Mohammed MA. COVID-19 Pneumonia Level Detection using Deep Learning Algorithm. Techrxiv 2020; 0–16.
32. Arslan H, Arslan H. A new COVID-19 detection method from human genome sequences using CpG island features and KNN classifier. Eng Sci Technol an Int J 2021; 24: 839–847.
33. Feng C, Wang L, Chen X, Zhai Y , Zhu F, Chen H,et al. A Novel Triage Tool of Artificial Intelligence-Assisted Diagnosis Aid System for Suspected COVID-19 Pneumonia in Fever Clinics.medRxiv.2021. https://doi.org/10.1101/2020.03.19.20039099 .
34. Yan L, Zhang HT, Goncalves J, Xiao Y, Wang M, GuoY,et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. medRxiv. 2020. DOI: https://dx.doi.org/10.1101/2020.02.27.20028027 .
35. Carrillo-Larco RM, Castillo-Cara M. Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach. Wellcome Open Res 2020; 5: 1–22.
36. Virgantari F, Faridhan YE. K-Means Clustering of COVID-19 Cases in Indonesia’s Provinces. Int J Eng Nat Sci. 2020; 5(2): 1–7.
37. Liang X, Feng Z, Xue Y, et al. A 3M K-means algorithm for fast and practicably identifying COVID-19 close contacts. MATEC Web Conf 2021; 336: 06009.
38. Rahimi HM, Nadimi M, Langeroudi AG, Taheri M, Fard. Application of Machine Learning in Diagnosis of COVID-19 Through X-Ray and CT Images: A Scop SGing Review. Front Cardiovasc Med.2021;8: 638011 . https://dx.doi.org/10.3389/fcvm.2021.638011 .
39. Sadoon TAU-M, Ali MH. Coronavirus 2019 (COVID-19) Detection Based on Deep Learning. Al-Nahrain J Eng Sci 2020; 23: 408–415.
40. Abdulkareem JF, Aljobouri HK, Hasan AM. Chest CT Images Analysis with Deep- Learning and Handcrafted Based Algorithms for COVID-19 Diagnosis. Des Eng 2021; 6246–6262.
41. Hasan AM, Ibrahim RW, Jalab HA. Breast Cancer MRI Classification Based on Fractional Entropy Image Enhancement and Deep Feature Extraction. Baghdad Sci J. 2022; 20(1):221–234. https://doi.org/10.21123/bsj.2022.6782
42. Lin G, Shen w. Research on convolutional neural network based on improved Relu piecewise activation function. Procedia Comput. Sci.2018; 131: 977–984
43. Sitaula C, Aryal S. New bag of deep visual words based features to classify chest x-ray images for COVID-19 diagnosis. Heal Inf Sci Syst 2021; 9: 1–12.
44. Maia M, Pimentel JS, Pereira IS, Gondim J, Barreto ME, Ara A. Convolutional support vector models: Prediction of coronavirus disease using chest x-rays. Inf 2020; 11: 1–19.

## تصميم موديل حديث لمواجهة جائحة كوفيد-19 بناءا على نموذج SVM و CNN

**سرى منذر النداوي**[1] **هديل قاسم الجبوري**[2]

[1]قسم هندسة الحاسوب، كلية الهندسة، جامعة النهرين، بغداد، العراق.
[2]قسم هندسة الطب الحياتي، كلية الهندسة، جامعة النهرين، بغداد، العراق.

**الخلاصة:**
اثناء الازمة الصحية العالمية الناجمة عن فيروس كورونا (كوفيد-19), بدأ الباحثون والخبراء الطبيون بالبحث عن طرائق جديدة لمحاربة الوباء. وفقا للدراسات الحديثة فقد استخدم التعلم الالي بشكل فعال في القطاع الطبي. لقد ساعدت مصادر التصوير الطبي منها (التصوير الشعاعي والتصوير المقطعي المحوسب) في تطوير استراتيجيات الذكاء الاصطناعي من اجل التصدي لانتشار فيروس كورونا. ونتيجة لذلك ، تم تطوير نموذج تعلم آلي كلاسيكي لاكتشاف فيروس كورونا من صور التصوير المقطعي المحوسب (CT). تم اقتراح نموذج الشبكة العصبية التلافيفية (CNN) والة ناقلات الدعم(SVM) لتصنيف التصوير المقطعي المحوري للرئة الى مجموعتين في هذه الدراسة(COVID-19,NonCOVID-19) كانت مجموعة البيانات المستخدمة في هذه الدراسة هي 960 شريحة من الأشعة المقطعية تم جمعها من المرضى العراقيين / مستشفى ابن النفيس التعليمي. مقاييس الاداء المستخدمة هي (الدقة والانضباط والاسترجاع ودرجة ال F1) أظهرت النتائج أن النهج المقترح أنتج نموذجًا عالي الجودة ، بدقة إجمالية تبلغ 98.95٪ ، واستدعاء كلي بنسبة 97٪ لمجموعة البيانات المجمعة.

**الكلمات المفتاحية:**CNN، COVID-19،تعلم الالة، التصوير الطبي، RNN.