Contents lists available at ScienceDirect

Measurement: Sensors

COVID-19 detection and classification for machine learning methods using human genomic data

Mohd Thousif Ahemad, Mohd Abdul Hameed [*], Ramdas Vankdothu

Department of Computer Science and Engineering, Osmania University Hyderabad, India

A B S T R A C T

Coronavirus is a disease connected to coronavirus. World Health Organization has declared COVID-19 a pandemic. It has an impact on 212 nations and territories worldwide. Examining and identifying patterns in X-Ray pictures of the lungs is still necessary. Early diagnosis may help to lessen a person's virus exposure and prevent it. Manual diagnosis is a time- and labor-intensive process. Since the COVID-19 virus has the potential to infect individuals all around the world, its finding is extremely concerning. The purpose of this study is to apply machine learning to identify and classify coronaviruses. The COVID-19 is anticipated to be discriminated and categorized in CT-Lung screening and computer-aided diagnosis (CAD). Several machine learning methods, including Decision Tree, Support Vector Machine, K-means clustering, and Radial Basis Function, were utilised in conjunction with clinical samples from patients who had contracted corona. While some medical professionals think an RT-PCR test is the most reliable and economical way to detect Covid-19 patients, others think a lung CT scan is more precise and less expensive. Serum samples, respiratory secretions, and whole blood samples are examples of clinical specimens. As a result of the earlier clinical evaluations, these tissues are used to assess 15 different parameters. As part of the proposed four-phase CAD system, the CT lungs screening collection is followed by a pre-processing step that enhances the appearance of ground-glass opacities (GGOs) nodules, which are initially extremely fuzzy and poorly contrasting due to the absence of contrast. These zones will be found and segmented using a modified K-means technique. Support vector machines (SVM) and radial basis functions (RBF) will be used as the input and target data for machine learning classifiers with a 50x50 pixel resolution to categorise the contaminated zones found during the detection phase (RBF). The 15 input items gathered from clinical specimens may be entered into a graphical user interface (GUI) tool that has been created to help doctors receive accurate findings.

1. Introduction

The novel coronavirus is a disease that is spread from person to person. With an estimated 3 lakh fatalities worldwide2 and a number of confirmed cases nearing 5 million, the new virus is putting the human race's survival in jeopardy [1,2]. The situation is concerning, as the number of new cases has climbed by almost five times in the previous month3 and there has been a dramatic spike of one million cases in the last two weeks. 3 High fever, cough, sore throat, and trouble breathing are all frequent coronavirus symptoms. It begins with a fever and may progress to pneumonia. Shortness of breath, chest tightness, and other serious infections are all possible consequences [5]. According to a study4, 80.9% of cases are moderate, 14% are severe, and 5% are serious. The risk of mortality is higher among the elderly and those with pre-existing conditions. 4 Cardiology, asthma, diabetes, and hypertension patients are at a greater risk. Only a few instances have been reported among youngsters [6,7].

Global specialists are working tirelessly to combat the epidemic and discover the most effective remedies. The COVID-19's unregulated spread might be attributable to a variety of factors. A patient may have little or no symptoms at first. A person may have infected multiple more people by the time the virus is discovered. Second, it is similar to the flu in that it causes no symptoms in the early stages. Symptoms might take anywhere from 2 to 14 days to show. 5 Third, there aren't enough testing kits to test a large enough population. Furthermore, there has yet to be created a vaccine to cure the 2019 coronavirus, so treatment takes time and control of its spread is challenging. Deep Learning technologies have made illness identification and categorization in medical imaging

more automated and efficient. Convolutional neural networks (CNNs) have a wide range of applications in medical image processing, including tumour classification, skin lesion segmentation, pneumonia detection, and many more. Pre-trained CNN models are utilised to categorise chest X-ray images and determine the likelihood of infection with COVID-19 in the proposed model. The results might aid in the early detection of people who need immediate treatment.

The healthcare industry is in a state of panic due to an unusual rise in the number of COVID-19 cases. With this unexpected surge of cases, a lack of resources is becoming an issue in many nations worldwide to fight this epidemic. To make the most use of the little healthcare resources available, early and precise diagnosis of infected individuals is a critical first step. Patients infected with Ebola may be identified using Polymerase Chain Reaction (PCR). Instead of relying on the body's immune response, or antibodies, PCR tests are used to directly detect the existence of the COVID-19 antigen. If someone has the virus, the tests can determine right away because they detect viral RNA present in the body before symptoms or antibodies develop.

When there is a lack of suitable training data and computing resources, deep learning is an effective way to train CNN. 16 For weight initialization, parameters learned from big datasets such as ImageNet are employed. Because training and classification take less time, transfer learning can be done on CPUs with no particular GPU needs. Low-level (edge, colour) information is frequently found in the characteristics of a pre-trained network's prior layers. High-level qualities, especially categorical information, are found in the subsequent levels. The suggested network architecture is shown in Fig. 1.

Deep Convolutional Neural Networks (CNN) are difficult to train from scratch since they need many training data. Furthermore, CNN models need a significant amount of time, perhaps days or even weeks. The model weights from pre-trained models may be reused to circumvent these constraints. Six models have been trained on over one million photos and can classify images into 1000 categories. High-performance models may be utilised directly, or they can be combined for a new challenge.

However, this is a time-consuming and labor-intensive manual procedure. This procedure is in limited supply due to the large number of instances that have been documented. Early studies [13,19] have shown that individuals infected with COVID-19 have radiographic scans that are aberrant. Using chest radiography images (e.g. X-rays and CT scans) to identify visual signs associated with COVID-19 infection as an alternative to PCR testing may now be possible. The capacity of radiologists to correctly detect and interpret the minute signs of COVID-19 is one of the most difficult hurdles in this route.

With this limitation in mind, we have proposed a unique deep learning model that can identify the signs of these illnesses in chest X-ray pictures with near-perfect accuracy. While this model may be used as a sorting tool to help doctors improve screening for COVID-19, we do not intend to use it as a replacement for traditional testing techniques.

A new coronavirus (CoV) known as "2019-nCoV" or "2019 novel coronavirus" or "COVID-19″ is responsible for the recent pneumonia epidemic that began in Wuhan Town in Hubei Province, China, in early December 2019. (WHO). According to the WHO, healthy persons who are in close contact with virus-infected individuals and who have their respiratory tracts touched by those who are sick are at risk of contracting COVID-19. There are a number of methods it may be conveyed, some of which specialists are currently debating. People who have been infected will begin to show symptoms within two to 14 days, including fever, cough, and pneumonia. The introduction of COVID-19, a groundbreaking coronavirus, ushered in a new era in global health care. The global economy, education, and transportation systems have all been affected, in addition to healthcare (Fong, Dey, and Chaki 2021). If treated correctly, this viral disease may cause severe respiratory illness. Human-to-human transmission and widespread infection are the virus's most dangerous negative effects.

An artificial intelligence (AI) forecast is feasible to find cluster situations using this strategy. It is also possible to utilize past clinical data to produce this forecast. AI is capable of operating in a manner akin to that of the human mind. Additionally, AI is capable of comprehending and displaying the development of the COVID-19 vaccination (Kondziolka, Couldwell, and Rutka 2020). It is imperative that current patient surveillance and screening be done in order to accurately forecast COVID-19 cases, which will aid in the prediction of infected individuals in the near future. Finding novel COVID-19 aid chemicals has become an increasingly common use of artificial intelligence (AI). The search for novel therapies for the illness and the use of medical image processing of CT scans and X-ray photographs to identify disease-affected people are both ongoing research projects.

To confirm the sickness, the reverse-transcription polymerase chain reaction is employed (RT-PCR). Because it lacks the sensitivity needed to treat suspected patients, RT-PCR lacks the ability to identify illness (Sun et al. 2020). Using computed tomography, a specific lung abnormality may be discovered. Because COVID-19 affects the lungs, a CT scan may be used to provide a better picture of the situation. Early detection of the illness or the COVID-19 virus is possible using CT-image screening. The similarity between pneumonia and COVID-19 may be seen in this CT scan (Zebin and Rezvy 2021). An infected individual's virus spreads to anyone who comes into close contact with them. Once an infectious agent has contaminated a person's nose and mouth, the infection can spread via the air. Within a few hours, the virus had spread across the affected area. In order to identify the virus, the standard diagnostic process was used. Loo-mediated isothermal amplification or
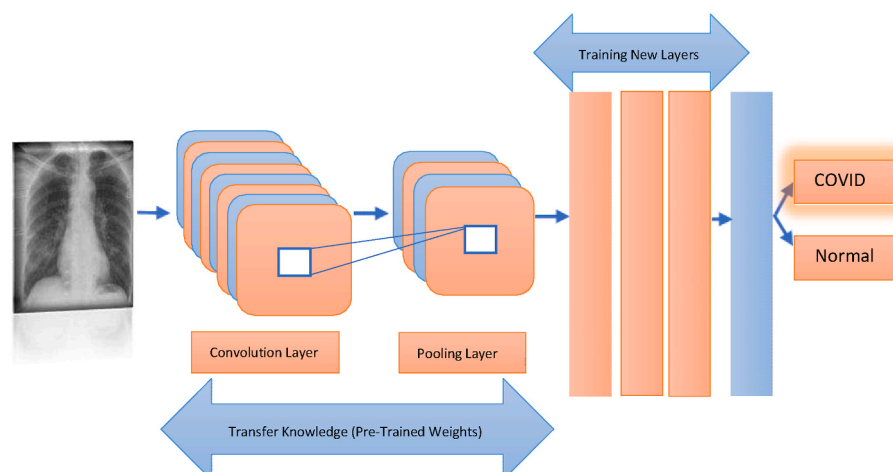

Fig. 1. The architecture of the proposed model.

Transcription-mediated amplification (TMA) or are two methods for amplifying the nasopharyngeal sample's nucleic acid (Khalifa et al. 2020). A number of measures were implemented in an effort to contain the corona virus' spread.

To identify a specific characteristic lung symptom, computed tomography is used. Since COVID-19 affects the lungs, this CT can be used to understand the situation. CT-image screening can be used to detect the COVID-19 virus or the illness at an early stage. The CT image of COVID-19 illustrates the similarity to the pneumonia condition (Zebin and Rezvy 2021). The virus spreads to another person when they come in contact with an infected individual. When an infected person breathes, coughs, or sneezes, it can spread since their mouth and nose are contaminated. The infected surface was immediately covered in the virus. The virus was found using the standard diagnostic process. Using RT-PCR, transcription-mediated amplification (TMA), or loo-mediated isothermal amplification, the nucleic acid from a nasopharyngeal swab is amplified (Khalifa et al. 2020). Numerous steps were made to protect the corona virus from spreading.

By examining the efficacy of big data and AI in combating laboratory findings generated from clinical samples of coronavirus suspects and examining current remedies, the study hopes to (Abbas, Abdelsamea, and Gaber 2021). A medical process known as computer-aided diagnosis (CAD) aids doctors in describing the clinical specimens of patients with coronavirus infections (Yuan et al. 2020). The problem is discovered when the doctor examines the patient's abdomen. A CT scan of the thorax. As the condition advances, the patient may have breathing problems, heart damage, and infections, among other things. Early detection of the corona is critical, since it may lead to mortality.

The suggested method's contribution:

➢ The COVID-19 virus is detected and classified using the suggested approach. A machine learning approach is the foundation of this system.
➢ This is done using the CT-lung screening method. The first organ to be affected by the covid-19 virus is the respiratory system. As a consequence, a lungs or respiratory system scan and diagnosis are necessary. Additionally, utilising clinical data, a machine learning model for identifying the Coronavirus in patients was developed.
➢ Creating a computer-aided design system that takes data from COVID-19 patients or suspects and decides whether or not they are contaminated. Modern machine learning algorithms are used to improve accuracy and access to data in less time.

2. Literature

New varieties of neural networks have outperformed earlier generations of algorithms for spotting irregularities in high-dimensional data. For example, tasks such as X-ray, CT scan and MRI segmentation, detection and classification [16] have fast become the "de-facto standard." In addition, neural network-based models dominate the clinical imaging landscape [27].

In 2021, Zou et al.proposed the goal of this research was to identify the variables that affect how often RT-PCR findings are positive. We examined the clinical data of patients with recurrent positive coronavirus disease 2019 (COVID-19) in many Wuhan medical institutions using a retrospective analysis. Patients are divided into two categories based on the findings of their RT-PCR tests: recurrent positives and nonrecurrent positives (non-RPos group). Two groups were created based on clinical features, updated content, and antibody titers. They measured the size of lung sections with various densities and examined pulmonary inflammatory exudation using AI-assisted chest increased computed tomography (HRCT) equipment.

Chest X-rays have been used in recent years to identify the presence of respiratory disorders such as pneumonia using several Deep Learning models [20,21,24,26]. Convolutional Neural Networks (CNNs) and DenseNet121 (DN121) are two of the most often used models. For detecting pneumonia in chest radiography pictures, CheXNet [20] is a benchmark model based on DenseNet121. A number of academics have been working on machine learning and deep learning models in order to keep up with the latest pandemic crisis. SVM and Random Forest have been utilised by Alqudah et al. [3] to identify COVID-19 instances with 90.5% and 81% accuracy, respectively, using machine learning methods. In order to identify COVID-19 from chest X-ray pictures, Ghoshal and Tucker [9] proposed a model based on Bayesian CNN and achieved an accuracy of 90%. For the detection of COVID-19 infection in chest radiographs, many authors [17,18,22,23] have developed CNN designs. According to Ref. [23], it is one of the first CNN-based models that can identify specific objects.

Pneumonia caused by both bacteria and viruses may be detected by COVID-19, which has a high sensitivity for COVID-19 detection This paradigm has the drawback of being exceedingly resource-intensive [17]. The COVIDAid model.

With a 90.5% success rate, the CheXNet [20]-based system works well when it comes to identifying COVID-19 in photos. For the categorization of COVID-19 chest X-ray pictures, Abbas and colleagues [4] suggested a DeTraC (Decompose, Transfer, and Compose) deep convolutional neural network and attained an accuracy of 93.1% utilising this architecture. Automatic identification of coronavirus infected individuals using chest X-ray pictures is proposed by the Covidx-net model [11]. Using VGG and DenseNet models, they were able to achieve an average accuracy of 91% for COVID instances.

Classification tasks are well-suited to deep learning models [8,14,15,25], which outperform machine learning models [3,9]. These DNN architectures, on the other hand, rely heavily on models like DenseNet and VGG for transfer learning. Personal digital assistants may not be able to use them because they are too complex, sluggish, and time-consuming. MobileNet [12] and ResNet50 [10] are two light frameworks that we use in this article to quickly identify the COVID-19 infection. For a fair comparison, we used two models - COVID-Net [23] and COVIDAid [17] - since we ran our tests on the same data split as the two models.

To put it another way, For the multiclass prediction problem, a suggested CAD system was tested using five-fold testing on COVID-19 and ChestX-ray8 datasets of chest X-ray images. To develop the recommended CAD system, researchers used a known collection of chest X-ray pictures as training data. Predictors developed for the identification and classification of COVID-19-related lesions were used to identify and classify locations on the full X-ray images with a detection accuracy of 96.31%. Nearly ninety percent of the test images from COVID-19 and other respiratory disease patients were correctly predicted using an IOU mean intersection over union (IOU). The total accuracy and F1-score of the COVID-19 diagnosis were increased by 6.64% and 12.17%, respectively, thanks to deep learning's regularisation of data balance and augmentation. Using the stated CAD approach, a chest X-ray diagnosis may be done in 0.0093 s. It's near to real-time because of the Design employed in this research, which can predict 108 frames per second (FPS). COVID-19 can be consistently distinguished from other respiratory disorders using the suggested deep learning CAD system. Health care systems, patients, and clinicians may all benefit from the recommended learning algorithm in the real world. MCKEE & ASSOCIATES, 2015) For the goal of this study, researchers wanted to find out whether the ACR Lung-RADS increases the incidence of incorrect results in a CT lung screening test. An investigation of how ACR Lung-RADS impacts the incidence of false-negative and false-positive results in a clinical CT lung screening was the purpose of this study. During the study period, a total of 2,180 high-risk patients had a baseline CT lung examination, with no clinical follow-up being provided to 577 of them. The total ACR Lung-RADS positive rate has dropped from 27.6% to 10.6%.

One year of monitoring the 152 patients categorized as benign has shown no false negatives. In 1,603 patients with follow-up, the predictive value of ACR Lung-RADS rose from 6.9% to 17.3%. Our CT lung screening group's positive predictive value improved by a factor of 2.5, to 17.3%, after implementing ACR Lung-RADS, while the frequency of

false-negative tests decreased.

(Mazzilli and colleagues, 2021) COVID-19 is an ongoing disease that affects millions of individuals worldwide. CT scans of the chest are the most often utilised imaging modality for correctly diagnosing and treating patients. An automated method based on individual HU thresholds was utilised to characterise the lungs of COVID-19 patients. The impact of the HU density calibration curve on inter-scanner variability was investigated. It was found that there was no inter-scanner variability. In comparison to the other two techniques, the highest gradient of the data has a substantially lower median value. A millimetre gradient on data approach was used to analyse our sample; the first peak averaged 853 56 Hu, and the second peak averaged 854 56 Hu; these three characteristics were quantified utilising the millimetre gradient on data method.

X-Ray images.

Large data analysis, picture categorization, face recognition, and sickness prediction have all benefited from deep learning and machine learning techniques. The application of artificial intelligence technologies to develop computer-based diagnostic tools will speed up early detection of COVID-19, easing the burden on health-care workers.

For COVID-19, Basu et al. proposed Domain Extension Transfer Learning as an alternative screening method. They extracted several distinguishing factors from the chest X-ray dataset. With 95.3% accuracy, the images were classified as normal, pneumonia, different diseases, and COVID-19.

3. Proposed system

Using machine learning techniques, this approach offers recognising and categorising COVID-19 CT-lungs screening and clinical specimens. Clinicians and patients must devote greater time and effort to infection identification in COVID-19 clinical specimen collections. This technique is used to diagnose infectious diseases in their early stages. Following that, image collecting will commence with a pre-processing phase to enhance the look of the ground glass opacities (GGOs) nodules, which had previously been fuzzy with fading contrast, and will be followed by CT image collection and the building of a classifier model (Aminisefat and Saravani 2020). A two-phase approach was developed to detect and categorise CT-Lungs screening, the first of which is the creation of a classifier model and the second of which is the evaluation of a fresh CT picture. The dataset is then used to train machine learning algorithms such Naive Bayes Classification, Decision Trees, Support Vector Machines, Radial Basis Function, and Kmeans Clustering to effectively distinguish between COVID-19-infected and non-infected patients in clinical specimens.

3.1. COVID-19 dataset

The data collection is used to identify infected and normal COVID-19 viral data as part of CT-lung screening. This data establishes a logical automated segmentation system and quantifies abnormal CT models. The dataset collecting is regarded a challenging endeavour since it necessitates a big number of ethical and privacy considerations on the part of the healthcare system (Zhao et al. 2020). The competent commissions approved this dataset based on ethics and legislation. The collection includes various kinds of the illness as well as scanning techniques. Based on the dataset, the patient will be classified as corona or non corona. The corona illness will infect at least half of a dataset of 100 samples. The full data set is obtained, classified by disease, and then processed to the next stage. They stored the information safely and securely for future use. The clinical specimens for SARS-CoV-2 testing are listed below.

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

3.2. Preprocessing

Creating a classifier model was part of this process. The sample is separated from the rest of the dataset and forwarded to the pre-processing stage for further analysis. The unstructured data from the clinical specimen must be enhanced before going on to the next detection stage. During the preprocessing stage, a number of things must be taken into account. Data samples that are incomplete or ambiguous should be discarded.

3.3. ROIs detection

The Regions of Interest (ROIs) are processed after the preprocessing test. In this approach, the RGB colour is used to move the window to the segment area of interest (ROIs) (Jiang, Wang, and Liu 2015). This is largely concerned with the position of the moving item as well as the repercussions of its rejection. The colour concentration is more important in this detection. The detection of the system is represented by the colours red, blue, and green. These colours might be used to show where the corona is located.

3.4. Feature extraction

As part of the capsules structure's feature-based component, each layer of the x-ray is eliminated, and a systematic output is produced. It's a technique for achieving a result by categorising and comparing one image to others.

3.5. Classification and detection

The increased attenuation of GGO nodules obscures the whole lung parenchyma while leaving the underlying arteries and bronchi visible. The many forms of GGO nodules are listed below.

- ✓ A nodule that contains GGO (part-solid nodule)
- ✓ A nodule containing GGO with pure localised GGO (non-solid nodule)
- ✓ The nodular nodules on the ground-glass Opacity, also known as the subtle nodule in clinical practise, is a technique that is utilised in combination with computed tomography (CT) scanning and image screening. The lungs scanning system employs this technology.

RBF networks have three layers: an internal layer, a hidden layer, and an outer layer. RBF networks are becoming increasingly frequent in neural networks with a wide range of applications, thus it's important to build proper beginning states for them. It is most likely the best multilayered perceptron available. The second classifier will be the vector supporting machine (SVM). SVM is an excellent classifier for digital image classification as well, especially when based on colour or other distinctive characteristics.

3.6. Classifier model

The last and most crucial stage in the classification process is to create a classifier model. This approach categorises the picture produced by feature extraction depending on the ailment that has been identified. When a fresh CT image is examined, the same procedure is followed, finally resulting to an unlabeled CT image. As a consequence, the individual who has been affected by the covid-19 has been recognized.

3.7. COVID-19 detection flowchart

The process for recognising COVID-19 is shown in the figure. The first stage of the test is the Covid-19 test, which is the first phase. Following that, a CT lung screening is performed (Zhang, Chu, and Zhao 2020). During CT-lung screening, a lung test is conducted, and the data

is then analysed utilising preprocessing and feature extraction. After the CT image has been created, the patient's diagnosis is made using a set of procedures referred to as feature extraction. The patients are then sorted into groups according to their illnesses and tested once again. The COVID-19 test is next performed, and the results are analysed. The two types of results are normal and consistently exceptional finds.

3.8. Training models

This stage entails evaluating the accuracy of the created dataset using a variety of machine learning models and comparing it to the accuracies of each model in order to choose the best model for real-time. Detection and categorization of COVID-19.

3.9. decision tree

A Decision Tree (DT) is a tree-like structure for representing choices and their results. The attribute test is represented by the internal nodes in a DT, whereas the result is represented by the branch. The class labels are represented by the leaf nodes (Kumar Dubey et al. 2021). DT might be useful in a number of situations since it does not need domain-specific knowledge to develop. DTs are also known as Classification Trees and Regression Trees (CART). The Decision Tree (DT) is a machine learning technique for categorization and prediction. The decision tree algorithm is simple to implement and change. The decision tree algorithm is un-rivalled in terms of efficiency and processing speed. As a consequence, in terms of accuracy, the decision tree machine learning technique beats the unit classification concept.

4. Result and discussion

4.1. Dataset description

The data comes from the Kaggle dataset repository17, which is an open-access collection of chest X-Ray pictures. It includes COVID-19 X-Ray pictures, bacterial and virulent pneumonia, and normal humans. COVID-19 and normal pictures are the two image classes evaluated in this study. There are 70 COVID-19 photos and 930 normal photographs in this collection. The data is weighted toward photographs with typical chests. As a result, 70 COVID-19 and 80 normal photos are considered for the suggested work. The size of each picture is varied. These photos are resized to 224x224 pixels before being input into the network. Both training and testing photos are included in the dataset. Validation is done using 20% of the data. COVID-19 and normal X-Ray pictures from the dataset are shown in Figs. 2–5 .

The performance metric was calculated using the classifier decision. The classifier assesses whether or not the individual is infected. This categorization may be broken down into four categories: infected, normal, and testing mistake. The first section is Test Positive and Negative, followed by False Positive and False Negative. The results of

the tests, both positive and negative, reveal whether or not a person is afflicted by covid. While a false negative suggests that someone is covid-19-negative but the test reveals otherwise, a false positive suggests that the person is normal but the test reveals otherwise. The results of the false-positive and false-negative tests demonstrate the accuracy of the sensitivity evaluation.

4.1.1. Confusion matrix
The proposed COVID-19 detection and classification algorithms may be assessed using a variety of performance metrics. A confusion matrix is built to govern how the recommended structure is presented. The Confusion Matrix is shown in Table 1. The performance measures that will be utilised to evaluate the classification model's performance might include accuracy, correctness, recall, and F1 score.

4.1.2. Accuracy
The accuracy of a proposed model is 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 given below Equation

$$\mathbf{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}} * \mathbf{100}$$

4.1.3. Precision
The performance measure accuracy is given in the ratio of True Positives (TP) to the total of False Positives (FP) and True Positives gathered from a specific dataset of clinical specimens is given below equation

$$\mathbf{Precision} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}} * \mathbf{100}$$

4.1.4. Recall
The ratio of True Positives (TP) to the total of False Negatives (FN) and True Positives produced from a dataset of clinical specimens is used to remember determine in this measure for gauging performance given below equation.

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP + FN}} * \mathbf{100}$$

4.1.5. F1-score
The F-measure, F˳, is calculated using precision and recall measures. The value of F˳ determines the metric F1-score. As demonstrated in the F1-score, the Precision and Recall on Harmonic Mean are represented in given equation.

$$\mathbf{F1 - Score} = \frac{\mathbf{2 * (Precision + Recall)}}{\mathbf{1 * (Precision * Recall)}} * \mathbf{100}$$

5. Conclusion

The study's purpose is to develop a method for identifying and categorising the covid-19 virus. In this case, the machine learning technology is employed to detect and categorise items. Because there were no particular therapies or vaccinations available at the time of the virus's introduction, the COVID-19's consequences surprised the world. Coronaviruses produce a deadly illness, and many research are being done to find a cure. These clinical samples were tested genetically, serologically, and biochemically. Clinical samples are analysed for a variety of characteristics, and CT-lung screening aids in speeding up the process and identifying those who have been infected with the coronavirus. The covid test is used to determine the accuracy, precision, memory, and F1-factor and distinguish between genuine positive, negative, and false positive, negative. Because the prediction rate is so high, it's a lot easier to find sick patients.To reduce the number of people who die. As a consequence, the optimal method for identifying and
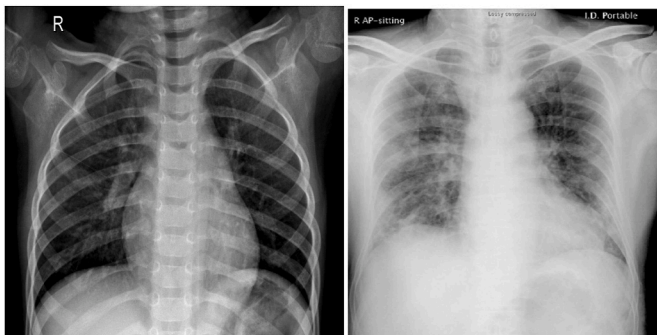

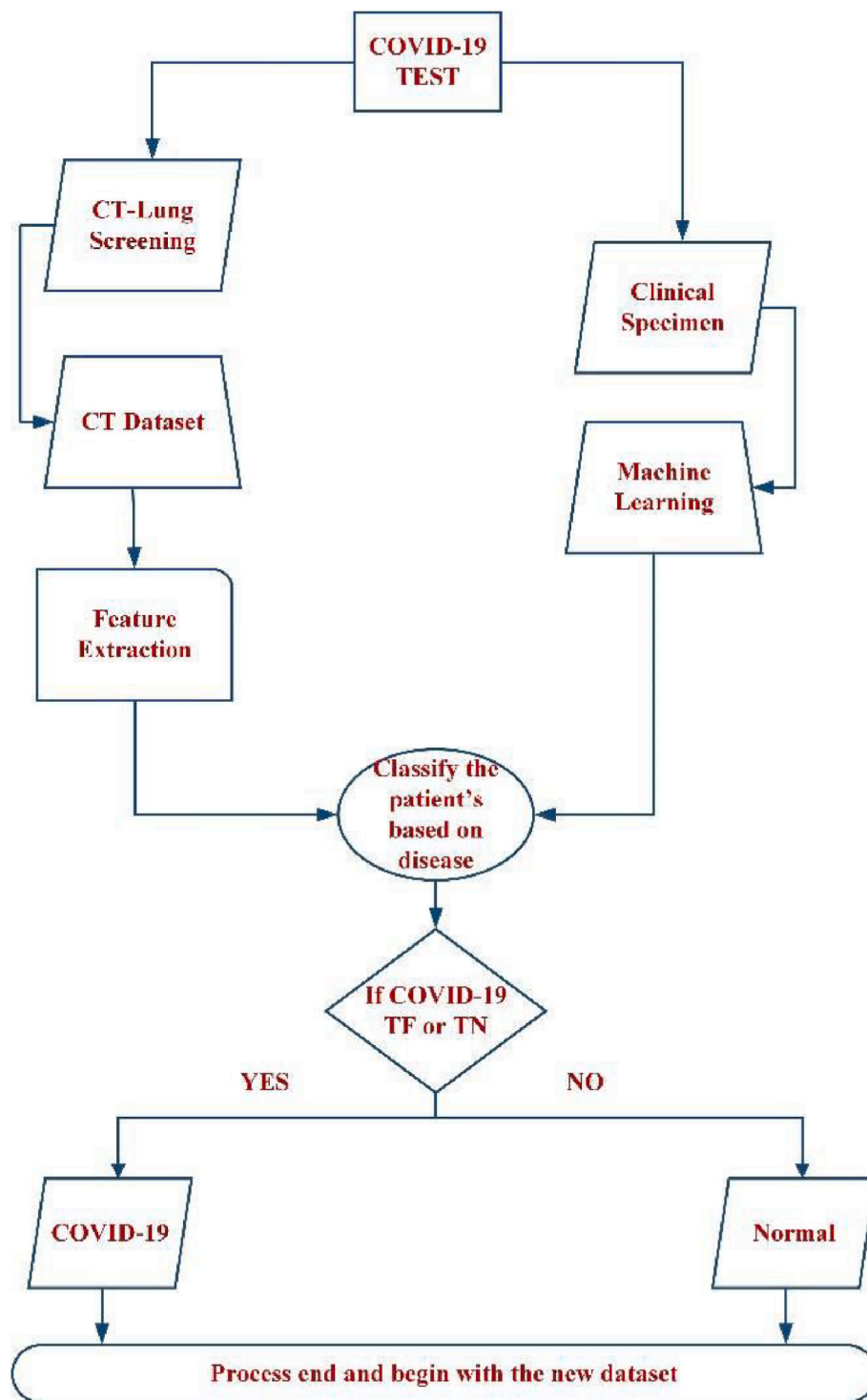**Fig. 2.** Examples of Normal, infected and COVID-19 infected chest.

Fig. 3. Flowchart for the COVID-19 detection.

categorising COVID-19 in CT lungs screening is to use machine learning to detect and categorise the virus. By increasing the number of samples utilised, the efficiency and accuracy of the recommended model may be improved. More feature engineering is needed for a better result, and deep learning may be used in the future.
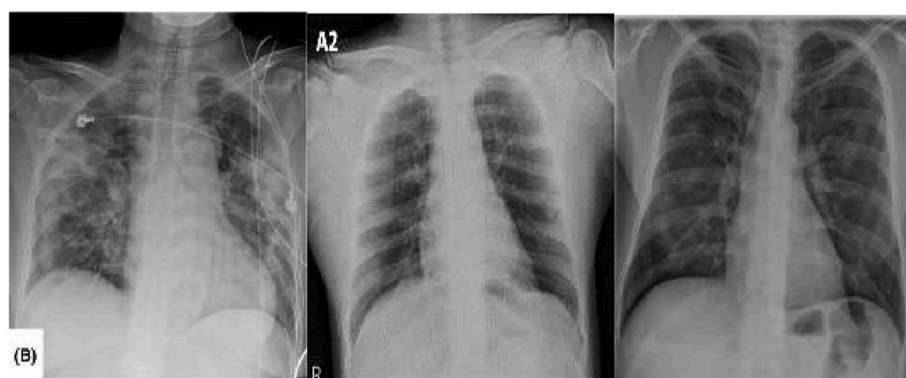
Fig. 4. COVID-19 chest X-Ray images.

Fig. 5. Normal chest X-Ray image.

Table 1
Confusion matrix.

Metrics	Test Result (Negative or Positive)	
Actual biopsy (True)	False Positive (FN)	True Positive (TP)
Actual biopsy (False)	True Negative (TN)	False Negative (FP))

CRediT authorship contribution statement

Mohd Abdul Hameed: Conceptualization, Supervision, Project administration, Formal analysis. **Ramdas Vankdothu:** Conceptualization, Methodology, Software, Validation, implementation, Writing – original draft, Writing – review & editing, Visualization.

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.

Data availability

No data was used for the research described in the article.

References

[1] Asmaa Abbas, Mohammed M. Abdelsamea, Mohamed Medhat Gaber, Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network, Appl. Intell. 51 (2) (2021) 854–864, https://doi.org/10.1007/s10489-020-01829-7.

[2] Nasiba M. Abdulkareem, Abdulazeez Adnan Mohsin, Zeebaree Diyar Qader, Dathar A. Hasan, COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms, vol. 6, 2020.

[3] Al-antari, A. Mugahed, Cam-Hao Hua, Jaehun Bang, Sungyoung Lee, Fast deep learning computer-aided diagnosis of COVID-19 based on digital chest x-ray images, Appl. Intell. 51 (5) (2021) 2890–2907, https://doi.org/10.1007/s10489-020-02076-6.

[4] Alireza Aminisefat, Khadijeh Saravani, A case study of hypertension and COVID-19, Gene, Cell and Tissue 7 (3) (2020).

[5] S. Dhamodharavadhani, R. Rathipriya, Jyotir Moy Chatterjee, COVID-19 mortality rate prediction for India using statistical neural network models, Front. Public Health 8 (August) (2020) 441, https://doi.org/10.3389/fpubh.2020.00441.

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

[7] Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee, A novel approach to predict COVID-19 using Support vector machine, in: Data Science for COVID-19, Elsevier, 2021, pp. 351–364, https://doi.org/10.1016/B978-0-12-824536-1.00014-9.

[8] Muhammad Imad, Naveed Khan, Farhat Ullah, Muhammad Abul Hassan, Adnan Hussain, COVID-19 Classification Based on Chest X-Ray Images Using Machine Learning Techniques, vol. 11, 2020.

[9] Emrah Irmak, COVID-19 disease severity assessment using CNN model, IET Image Process. 15 (8) (2021) 1814–1824, https://doi.org/10.1049/ipr2.12153.

[10] Nour Eldeen M. Khalifa, Mohamed Hamed N Taha, Aboul Ella Hassanien, Sally Elghamrawy, 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, 2020. *ArXiv Preprint ArXiv:2004.01184*.

[11] Doug Kondziolka, William T. Couldwell, James T. Rutka, Introduction. On pandemics: the impact of COVID-19 on the practice of neurosurgery, J. Neurosurg. 133 (1) (2020) 1–2, https://doi.org/10.3171/2020.3.JNS201007.

[12] Kumar Dubey, Ashutosh, Sushil Narang, Abhishek Kumar, Satya Murthy Sasubilli, Vicente Garc韄-D韄z, Performance estimation of machine learning algorithms in the factor Analysis of COVID-19 dataset, Comput. Mater. Continua (CMC) 66 (2) (2021) 1921–1936, https://doi.org/10.32604/cmc.2020.012151.

[13] Chenxi Li, Chengxue Zhao, Jingfeng Bao, Bo Tang, Yunfeng Wang, Bing Gu, Laboratory diagnosis of coronavirus disease-2019 (COVID-19), Clinica Chimica Acta; International . J. Clin. Chem. 510 (2020) 35.

[14] Aldo Mazzilli, Claudio Fiorino, Alessandro Loria, Martina Mori, Pier Giorgio Esposito, Diego Palumbo, Francesco de Cobelli, Antonella del Vecchio, An automatic approach for individual HU-based characterization of lungs in COVID-19 patients, Appl. Sci. 11 (3) (2021) 1238, https://doi.org/10.3390/app11031238.

[15] Majid Nour, Zafer Cömert, Kemal Polat, A novel medical diagnosis model for COVID-19 infection detection based on deep features and bayesian optimization,

Appl. Soft Comput. 97 (December) (2020), 106580, https://doi.org/10.1016/j.asoc.2020.106580.

[16] Nan-Nan Sun, Ya Yang, Ling-Ling Tang, Yi-Ning Dai, Hai-Nv Gao, Hong-Ying Pan, Bin Ju, A prediction model based on machine learning for diagnosing the early COVID-19 patients, Preprint. Infectious Diseases (except HIV/AIDS) (2020), https://doi.org/10.1101/2020.06.03.20120881.

[17] Wenling Wang, Yanli Xu, Ruqin Gao, Roujian Lu, Kai Han, Guizhen Wu, Wenjie Tan, Detection of SARS-CoV-2 in different types of clinical specimens, JAMA 323 (18) (2020) 1843–1844.

[18] Mingli Yuan, Yin Wen, Zhaowu Tao, Weijun Tan, Yi Hu, Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China, PLoS One 15 (3) (2020), e0230548.

[19] Tahmina Zebin, Shahadate Rezvy, COVID-19 detection and disease progression visualization: deep learning on chest X-rays for classification and coarse localization, Appl. Intell. 51 (2) (2021) 1010–1021, https://doi.org/10.1007/s10489-020- 01867-1.

[20] Junyong Zhang, Yingna Chu, Na Zhao, Supervised framework for COVID-19 classification and lesion localization from chest CT, Ethiop. J. Health Dev. 34 (4) (2020).

[21] Jinyu Zhao, Yichen Zhang, Xuehai He, Pengtao Xie, Covid-Ct-Dataset: A CT Scan Dataset about Covid-19, 2020. *ArXiv Preprint ArXiv:2003.13865* 490.

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

[23] Omar A. Alzubi, Jafar A. Alzubi, Ala' M. Al-Zoubi, Mohammad A. Hassonah, Utku Kose, An efficient malware detection approach with feature weighting based on Harris Hawks optimization, Cluster Computing J. (2021), https://doi.org/10.1007/s10586-021-03459-1.

[24] Jafar A. Alzubi, Rachna Jain, Anubhav Singh, Pritee Parwekar, Meenu Gupta, COBERT: COVID-19 question answering system using BERT, Arabian J. Sci. Eng. (2021), https://doi.org/10.1007/s13369-021-05810-5.

[25] Omar A. Alzubi, Jafar A. Alzubi, Mohammed Alweshah, Issa Qiqieh, Sara Al-Shami, Manikandan Ramachandran, An Optimal Pruning Algorithm of Classifier Ensembles: Dynamic Programming Approach, Neural Computing & Applications, 2020.

[26] Ramdas Vankdothu, Dr Mohd Abdul Hameed, Husnah Fatima, A brain tumor identification and classification using deep learning based on CNN-LSTM method, Comput. Electr. Eng. 101 (2022), 107960.

[27] Ramdas Vankdothu, Mohd Abdul Hameed, Adaptive features selection and EDNN based brain image recognition on the internet of medical things", Comput. Electr. Eng. 103 (2022), 108338. I.F:4.152.