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.

Computers in Biology and Medicine 136 (2021) 104729

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Fully automated unified prognosis of Covid-19 chest X-ray/CT scan images using Deep Covix-Net model

Dasari Naga Vinod [a], B. Rebecca Jeyavadhanam [b], Adamu Murtala Zungeru [c], S.R. S. Prabaharan [d,*]

[a] Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, SRM Nagar Kattankulathur, 603203, Kanchipuram, Chennai, TN, India
[b] Department of Computer Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, SRM Nagar Kattankulathur, 603203, Kanchipuram, Chennai, TN, India
[c] Department of Electrical, Computer and Telecommunications Engineering, Botswana International University of Science & Technology, Private Bag 16, Palapye, Botswana
[d] Directorate of Research, SRM Institute of Science and Technology, SRM Nagar Kattankulathur, 603203, Kanchipuram, Chennai, TN, India

ARTICLE INFO

Keywords:
Covid-19
Chest X-ray
Random forest
Fast fourier transform
Wavelet
Machine learning

ABSTRACT

SARS-COV2 (Covid-19) prevails in the form of multiple mutant variants causing pandemic situations around the world. Thus, medical diagnosis is not accurate. Although several clinical diagnostic methodologies have been introduced hitherto, chest X-ray and computed tomography (CT) imaging techniques complement the analytical methods (for instance, RT-PCR) to a certain extent. In this context, we demonstrate a novel framework by employing various image segmentation models to leverage the available image databases (9000 chest X-ray images and 6000 CT scan images). The proposed methodology is expected to assist in the prognosis of Covid-19-infected individuals through examination of chest X-rays and CT scans of images using the Deep Covix-Net model for identifying novel coronavirus-infected patients effectively and efficiently. The slice of the precision score is analysed in terms of performance metrics such as accuracy, the confusion matrix, and the receiver operating characteristic curve. The result leans on the database obtainable in the GitHub and Kaggle repository, conforming to their endorsed chest X-ray and CT images. The classification performances of various algorithms were examined for a test set with 1800 images. The proposed model achieved a 96.8% multiple-classification accuracy among Covid-19, normal, and pneumonia chest X-ray databases. Moreover, it attained a 97% accuracy among Covid-19 and normal CT scan images. Thus, the proposed mechanism achieves the rigorousness associated with the machine learning technique, providing rapid outcomes for both training and testing datasets.

1. Introduction

The novel coronavirus, or "Covid-19", could be an open-world emergency, according to the World Health Organisation (WHO). Currently, Covid-19 is rapidly spreading from its source in Wuhan City, China, throughout the globe [1–4]. The predicament of fighting the novel Covid-19 epidemic has brought an end to a few financial and civil problems in various social systems worldwide. Conversely, it has set off a torrent of research, both inside and outside the clinical space, to help networks conquer this test by limiting its unfavourable effects. The immensity of these logical endeavours and the speed at which the information regarding this matter has been produced presents issues for those attempting to stay current on these events [5]. Throughout the reaction to the Covid-19 emergency, numerous healthcare organisations have expanded their utilisation of broadcast communications. From clinics to residency chores, from understanding cooperation at home to those in isolation, virtual correspondence provides a safe manner to proceed with our obligations amid this epidemic [6–8]. Let us consider the way the coronavirus transmits. Contamination of Covid-19 is accomplished over droplets made when a tainted patient coughs, sneezes, or breathes out. These droplets are substantial, even when considering the observable suspension quickly falling to the floor [9]. Individuals are likely to be infected through close interaction with other individuals who are infected with Covid-19, either by touching an

* Corresponding author.
E-mail address: srsprabaharan1611@gmail.com (S.R.S. Prabaharan).

https://doi.org/10.1016/j.compbiomed.2021.104729
Received 30 July 2021; Accepted 31 July 2021
Available online 3 August 2021

infected surface or one's own eyes, nose, or mouth. Although Covid-19 impacts distinct people in several habits, highly infected people will make it difficult to coordinate syndromes. The primary side effects are fever, tiredness, and dry cough. A few individuals may encounter a "throbbing painfulness, nasal clog, runny nose, sore throat, Diarrhea" [10]. Amid an ordinary movement, it takes 6–8 days after someone is affected by the contamination for reactions to show up, and it can take up to 14–21 days as well. People with symptoms that are not severe who are otherwise normal should self-isolate in their homes. Clinical research is required in cases where an individual has fever, cough, and breathing problems.

In response to this unexpected pandemic, researchers and medical experts in hospitals have seriously considered a helpful screening technique and medicine for Covid-19. The domain of this study is not delimited to clinical or bioscience research, and other areas such as data science, deep learning, and artificial intelligence are needed to impede and manage this epidemic by assigning obtainable results to machine learning [11].

Data science has become increasingly common and has entirely replaced the outline of many research areas. In the medical domain, image databases such as retina images, chest X-rays, and CT scan images provide favourable outcomes amidst an increased accuracy in percentage enabled by deep learning mechanisms and artificial intelligence [12–14]. We know that X-ray and CT scan appliances accommodate expensive and expedited outcomes for screening several human organs in clinics. However, medical experts are frequently required to assess X-ray and CT scan images. As a data analyst, it is clear that processing those images by applying machine learning, deep learning, and artificial intelligence will assist clinical specialists in diagnosing COVID-19 individuals. This mechanism would enforce the developing regions where X-ray and CT scan facilities are accessible but specialists are unavailable. We intend to expand a machine learning model entitled a Deep Covix-Net model that can evaluate the X-ray and CT scan images of lungs and diagnose if the individual is infected with Covid-19 or pneumonia. Amid several machine learning classifiers, random forest acts as an extensive classification and clinical image diagnosis technique. The outcomes of machine learning models justify their use in scaling the picture database to an accurate and robust one. The lungs are a vital spot, and examining their variations can provide an exact outcome for Covid-19. This article's contributions include recommending various segmentation techniques that are adequate for detecting Covid-19-contaminated individuals and healthy lungs with the help of chest X-ray images enabled by machine learning algorithms. The dominant research contributions of this study are as follows:

- We have proposed a deep Covix-Net model to identify Covid-19, normal, and pneumonia individuals via chest X-ray and CT scan images.
- We rigorously examined the chest X-ray and CT scan images for individuals, aiming to screen for Covid-19, normal, and pneumonia. As a result, all the images enclosing the lesions were reaffirmed by qualified radiologists.
- The achieved outcomes were examined using three metrics: accuracy, receiver operating characteristic, and confusion matrix.
- The mechanism adopted in the present work facilitates the actual diagnosis of individuals affected by Covid-19, achieving a precision score of 99% for chest X-ray images. In comparison, it revealed a 98% recall score.
- The proposed technique gave a 96.8% accuracy for many image databases (9000 Chest X-ray), while revealing 97% accuracy in CT scan images for the qualitative study.

The remainder of this article is arranged as follows: Section 2 analyses current X-ray and CT scan image examinations utilising a deep learning-based convolutional neural network (CNN). Section 3 provides an illustration of the recommended mechanism and associated database. In Section 4, we describe the proposed model and the innovations behind the model. Section 5 describes the experimental results and analysis with the metrics of the classifier. Finally, the conclusions are drawn in Section 6.

2. Related work

A computer-aided diagnosis method primarily based on CT images with a large capacity is intended to differentiate contamination of Covid-19 from other types of unusual and viral pneumonia contamination. This confirms that ResNet-101 is a promising method to identify and diagnose Covid-19 contamination. This version no longer entails significant costs and can be used as a catalyst in CT imaging in radiology departments [15]. Hui et al. [16] developed a multi-view fusion version for preliminary screening of Covid-19 pneumonia. This version achieved a higher overall performance with a single-view version as well as a subgroup examination. This confirmed the great capacity to enhance the performance of the analysis as well as diminish the workload of radiologists. We anticipate that the multi-view deep learning positioning mechanism may be utilised by a subordinate radiologist to quickly and precisely discover Covid-19 pneumonia.

Elavarasan et al. [17] highlighted various innovations introduced to assist the public, government, and healthcare services in various facets of the war against Covid-19. They also addressed the technical rapidity that has emerged during the pandemic and its effect on the atmosphere and culture. In addition to the technology introduced, their study also discussed untapped mechanisms that have potential uses in managing pandemic conditions. Covid-19 identification with a deep learning approach has been performed. Because it is essential to classify the rapidly and widely spreading Covid-19 disease, artificial intelligence (AI) techniques are being applied to accomplish this task accurately and efficiently. Applying the pre-processing steps to an image is one of the innovative features of the suggested solution. More powerful functions can be deduced from the image data using pre-processing. With the piling method, identical images are superimposed on each pixel, and the low-efficiency pixels are improved. Effective properties can be obtained using the proposed technique and the SMO algorithm [18].

Regarding infectious ailments, data collection is ambitious at the very best of times. The rise of data science has brought impetus to researchers and systems scientists, and the capacity to store and work with them by critical surveillance of vast amounts of data has increased. Coded data are complex and difficult to propagate. However, even after the pandemic is over, data will still exist and will be helpful for the research community. Data collection may cease because there are no new instances, but there will still be time to study the data. The impact of big data has provided clinicians and researchers with the ability to store and work with large amounts of data [19]. Pereira et al. [20] attempted to use only CXR images to distinguish pneumonia caused by Covid-19 from other categories, as well as stable lungs. The following viewpoints suggest a classification scheme: i) multiclass classification and ii) hierarchical classification, as pneumonia can be considered a pyramid. Texture is one of the key visual aspects of CXR pictures, so we can use a well-known interpretative method and utilise a pre-trained CNN technique to derive our classification schema functionality [20].

Given the potential for future Covid-19 epidemics, Santosh et al. [21] presented and explored the relevance of AI-driven instruments and their relevant training and testing models. The fundamental result of their study show that AI scientists should not always delay training, verification, and analysis of the models for the entire dataset. Instead, from the outset of data collection, AI-driven tools must be applied in parallel with specialists in fields where constructive learning is utilised.

Oh et al. [28] showed that a patch-based CNN model improved classification execution using a relatively small number of trainable parameters. The results of the experiments show that the proposed model achieved an accuracy of 88.9%. The modified inception transfer learning model to establish the mechanism, recommended by Shuai

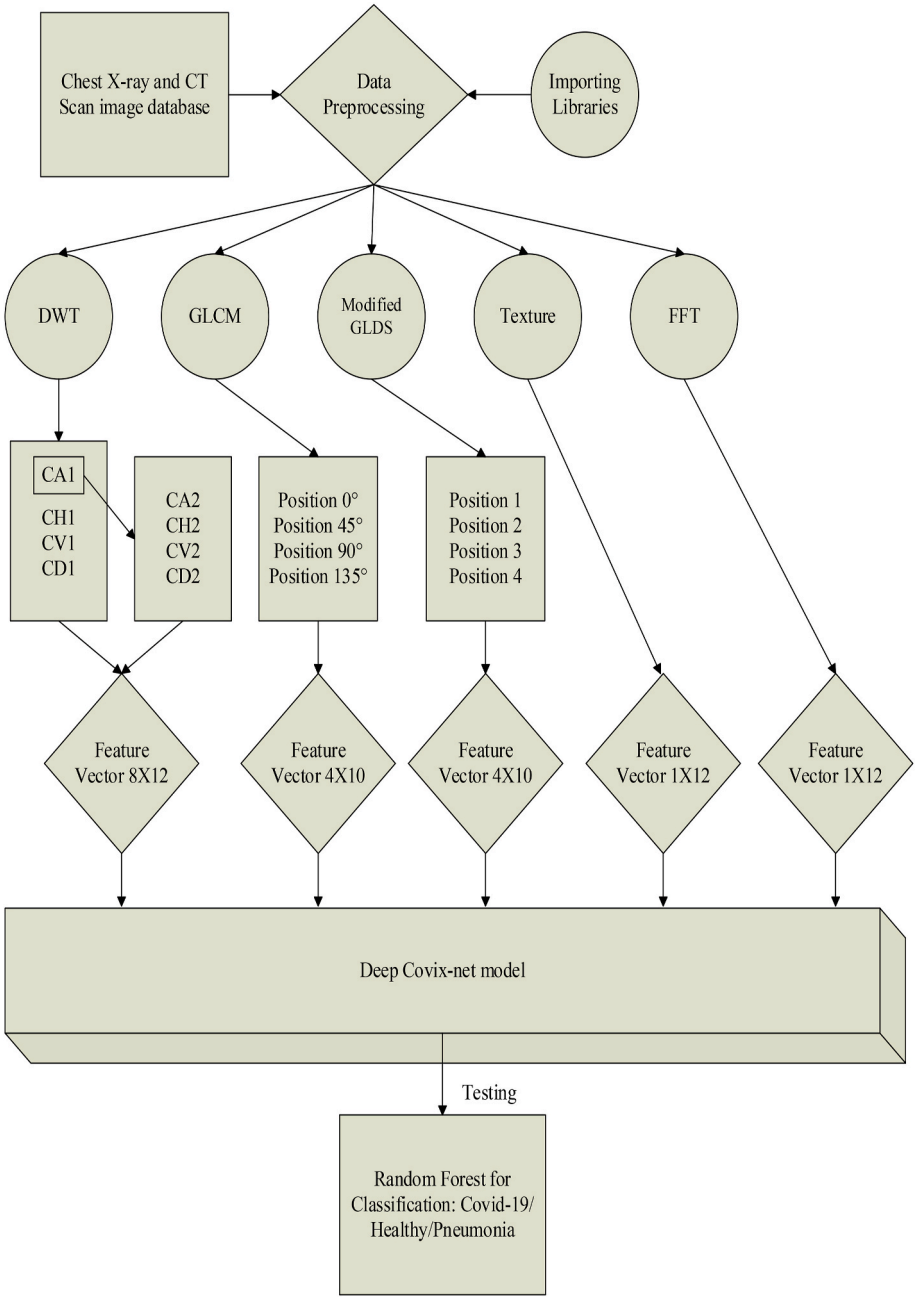


Fig. 1. Schematic workflow of the proposed model to diagnose Covid-19/pneumonia/normal individuals.

Wang et al. [29], attained an accuracy of 89.5% for internal validation and 79.3% for external validation. A generative adversarial network (GAN) with deep transfer learning for Covid-19 diagnosis in chest X-ray images was proposed by Loey et al. [30] to enhance the overall performance using AlexNet, GoogLeNet, and ResNet 18 models. The recommended mechanism attained an accuracy of 80.6%. Hemdan et al. [31] suggested Covid-19 classification with an f1-score of 89%. A deep convolution network based on the concatenation of the Xception and ReNet50V2 models proposed by Rahimzadeh et al. [32] achieved an overall average accuracy of 91.4% for all classes. Ten well-known CNNs (Alexnet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and Xception) were utilised for the prognosis of Covid-19 in [33], in which the researchers concluded that the ResNet-101 and Xception models attained the best performance, with an AUC of 0.994. COVID-Net, based on a tailored deep CNN recommended by Wang et al. [34], attained an accuracy of 93.3%. Zhang et al. [35] developed a new deep anomaly Covid-19 detection model for fast and reliable screening. The recommended model attained 96% sensitivity for Covid-19 cases and 70.65% sensitivity for non-Covid-19 cases.

Horry et al. [36] recommended a deep-based semi-automated image pre-processing method to detect Covid-19. The proposed mechanism attained 83% precision. Tsiknakis et al. [37] recommended a transfer learning model that attained an accuracy of 92% with 5-fold cross-validation. It was recently reported that a prudent methodology can help identify Covid-19-infected people among normal individuals by utilising CT scans and chest X-ray images using AI [38]. The image diagnosis tool utilises a decision tree classifier to identify a novel coronavirus-infected person. The recommended mechanism achieves an accuracy of 87% in chest X-rays, while it reveals 82% accuracy in the CT scan image database. Mei et al. [39] proposed a CNN model for diagnosing Covid-19 using CT scan images, a multilayer perceptron (MLP)

Input: Input Image I(e, f, g) $_{(e, f, g)\ \varepsilon\ (1,2,\ldots n)^3,\ e = f = g}$
Output: Output Image O(e, f, g) $_{(e, f, g)\ \varepsilon\ (1,2,\ldots n)^3,\ e = f = g}$
Begin
For each Input image I do
For {e, f, g} = 1 to n do
Convert input image I into grey and resize the image to 224*224
Apply min-max normalisation and return the output image (O)
End For
End For

model for diagnosing Covid-19 with clinical information, and a joint model for diagnosing Covid-19 using both CT scan images and clinical information. These models achieved accuracies of 79.6%, 74.2%, and 83.5%, respectively. Mizuho et al. [40] proposed a computer-aided diagnosis model for the classification of Covid-19, pneumonia, and healthy chest X-ray images. This model utilised VGG16 as a pre-trained model and a combination of conventional and data augmentation methods. The recommended model attained an accuracy of 83.6%.

3. Materials and methods

3.1. Data acquisition

The database utilised in this task is an open-source Kaggle and GitHub repository [41], which presently consists of over 9000 chest X-ray images, with 3000 for Covid-19-positive individuals, 3000 normal chest X-ray images, and 3000 pneumonia chest X-ray images. The CT scan database contains over 6000 images (3000 for normal and 3000 for Covid-19). The repository of the image database is open to image segmentation approaches, and recent images are regularly incorporated. The entire image database has been verified and elucidated, containing verdicts regarding the X-ray and CT scan images.

3.2. Data resizing

Because the images in the database are not flexible, the X-ray and CT images are distinct extents; accordingly, we have reformed all the pictures to a similar extent of 224 × 224 pixels. RGB reverting was enforced, and the eventual input to the recommended mechanism was provided as a 224 × 224 × 3 image. Subsequently, we applied image-segmentation techniques to this model.

3.3. Feature extraction and feature selection

We utilised advanced image segmentation techniques, such as texture, grey-level co-occurrence matrix (GLCM), grey-level difference method (GLDM), fast Fourier transform (FFT), and discrete wavelet transform (DWT), to process an aggregate of 200 lineaments in both the structural and recurrence areas. We actualised GLCM for an individual class or every division; we registered 12 lineaments by employing similar factual capacities. We estimated the following 12 statistical features: average, skewness, kurtosis, energy, average deviation, dimension, RMS, consistency, average gradient, std gradient [22], min, and median. The feature elicitation technique brought about 200 lineaments for every X-ray and CT scan picture together (12 lineaments from texture, 12 lineaments from FFT, 40 lineaments from GLCM, 40 lineaments from GLDM, and 96 lineaments from DWT).

4. Implementation

The training model was implemented using an Anaconda-Jupyter notebook with a tensor flow platform to train the extensive image database. Importation of the necessary libraries was performed at the initial stage to access the code from other modules. Then, the image database is uploaded to the path, and the features are extracted using spatial and frequency domains such as texture, GLCM, GLDM, DWT, and

Input: Input Image O(e, f, g) $_{(e, f, g)\ \varepsilon\ (1,2,\ldots n)^3,\ e = f = g}$
Output: Deep Covix_net model
Begin
For each Input image O do
For {e, f, g} = 1 to n do

IF file ← e, Then
Label ← 0
ELIF file ← f,
Label ← 1
ELSE
Label ← 2
ENDIF

End For
End For
F ← {Calculating features on input image O (Area, Mean, Std, Skewness, Kurtosis, Energy, Mean Absolute Deviation, Median, Range, Uniformity, RMS and Minimum)}
Textual, T ← Compute F on input image, O FFT in eq. (1), F ← Compute F on input image, O WT,d1 ← Compute F on input image, O DWT, d2 ← Compute F on Approximation of d1 image D ← d1 + d2 GLDM, G ← Compute F on input image O in four directions GLCM, g ← Compute F on input image O in four directions Deep Covix_net model ← $\sum_{i=1}^{n}\{\mathbf{T}, \mathbf{F}, \mathbf{D}, \mathbf{G}, \mathbf{g}\}$ ε O (Total F=200) Save the model

Input: Deep Covix_net model, number of instances = 100, testing ratio = 20%, Training ratio = 80%, number of trees =100, Classifier = Random Forest,
Output: Confusion Matrix, Visualisation, Performance metrics.
Begin
For Input Deep Covix_net model do
x = concatenate {(e(F), f(F), g(F))} ε $(1,2,\ldots n)^3$ (inputs) y = concatenate {(e(F), f(F), g(F))} ε $(1,2,\ldots n)^3$ (target labels)
End For

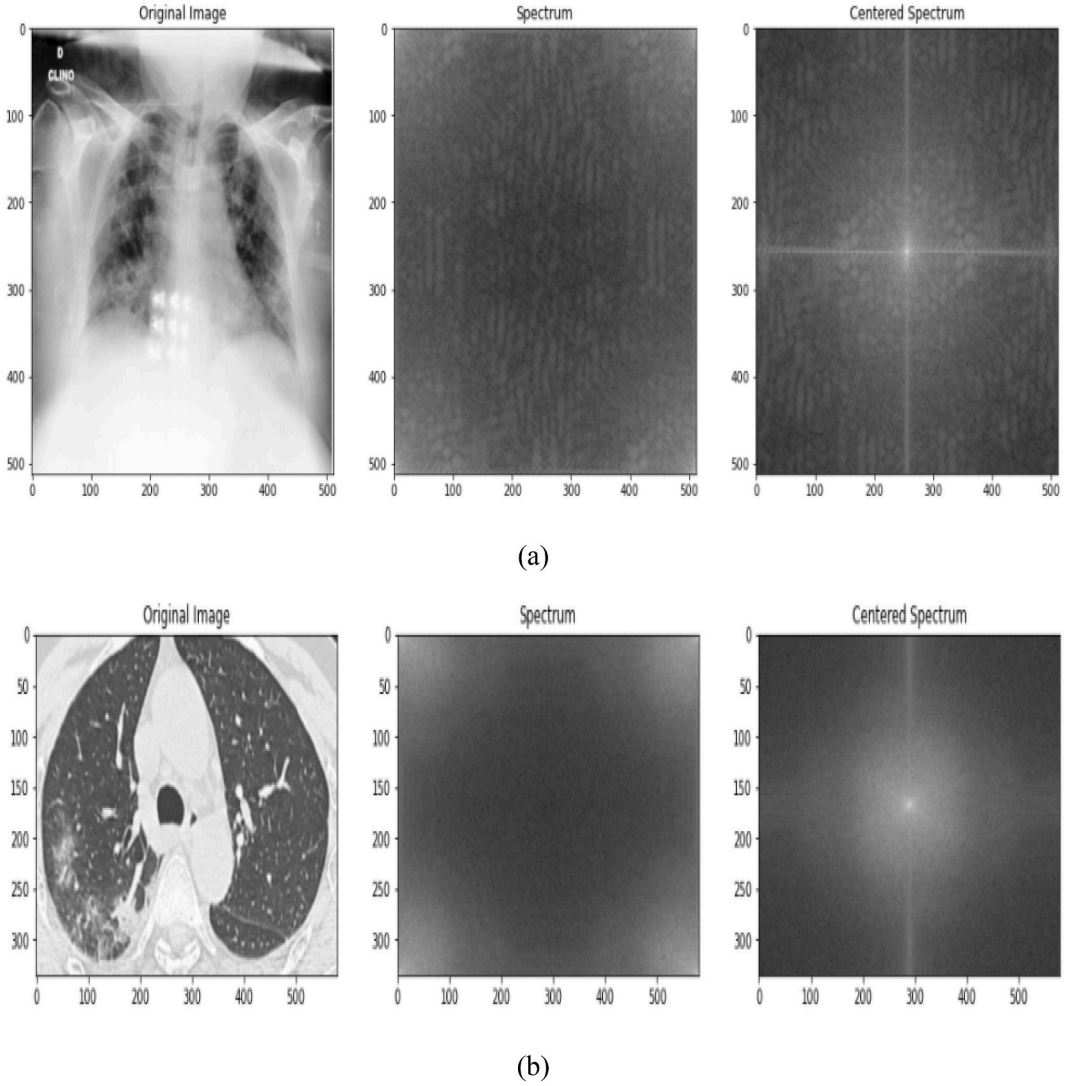


Fig. 2. (a) FFT segmented images of a sample Covid-19 chest X-ray; (b) FFT segmented images of a sample Covid-19 CT scan.

FFT for segmentation. Finally, we combined all 200 statistical features in both domains and utilised a Deep Covix-Net model to diagnose COVID-19 patients enabled by the random forest classifier, as illustrated in Fig. 1. The image pre-processing mechanism is shown in algorithm-1, and feature extraction of the images is shown in algorithm-2. Furthermore, the classification and performance measures of the model are shown in algorithm-3 for a chest X-ray image database with multiple classifications. The same process was followed in the CT scan image database with binary classification.

Algorithm 1. Chest X-ray Image for Pre-processing
where e, f, and g are the classes in the chest X-ray image database, and n is the total number of images.

Algorithm 2. Feature Extraction Technique for Chest X-ray Images

Algorithm 3. Train the Chest X-ray image database enabled by Deep Covix-Net model

4.1. FFT positioned disjuncture

Fast Fourier transform (FFT) estimates the discrete Fourier transform (DFT) and its inverse. The FFT is utilised to transform a digital signal (s) among range (L) from the time region into a frequency region (D), considering the amplitude of vibration based on its progress against the frequency as the signal emerges [23].

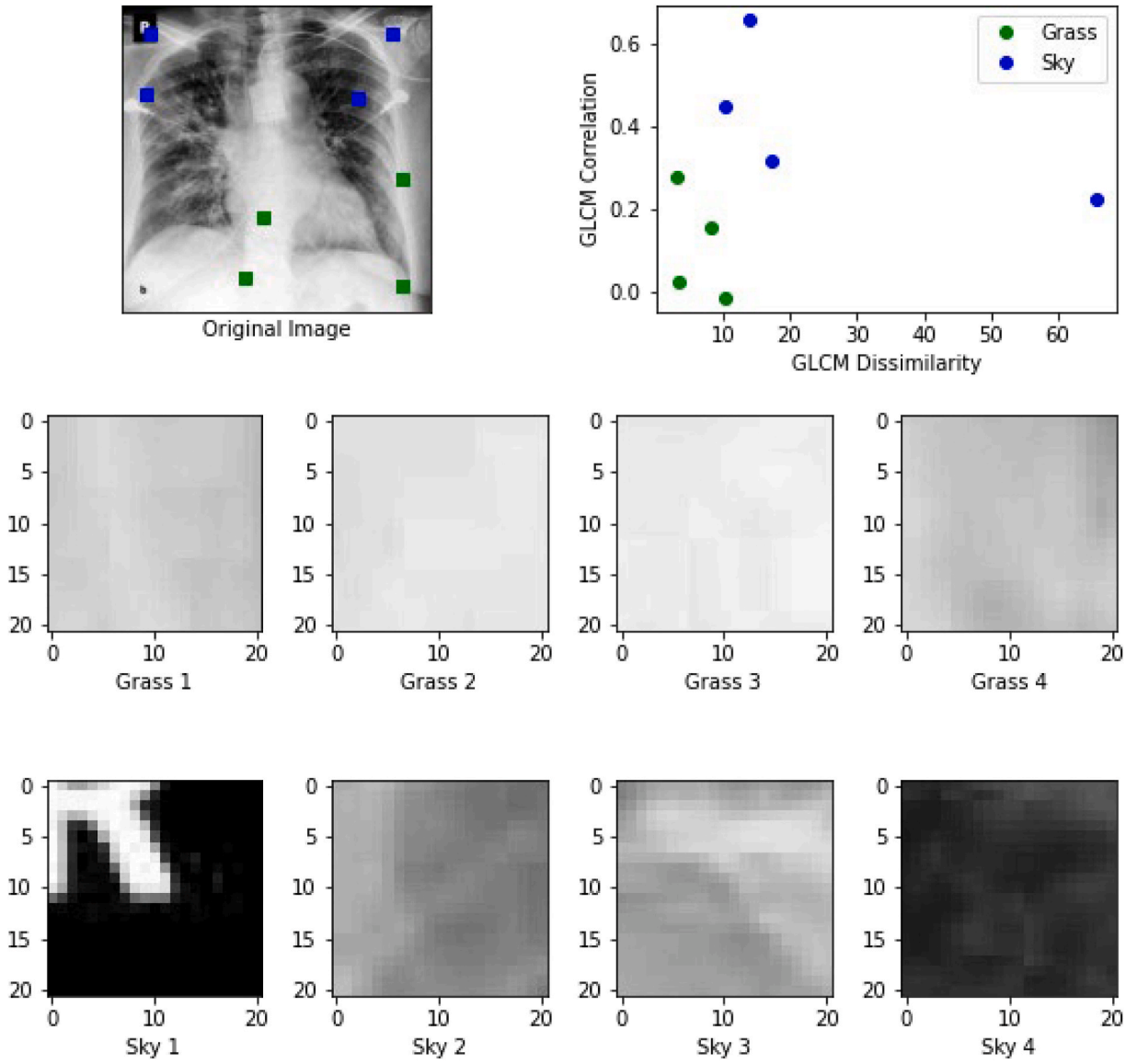


Fig. 3. Grey-level co-occurrence matrix segmentation on sample Covid-19 chest X-ray images.

$$D[k] = \sum_{j=0}^{L-1} \left(s(j) W_L^{jk}\right),\ W_L = e^{-2\pi J/L},\ for\ k = 0, 1, 2, \ldots. \quad (1)$$

The frequency spectrum vector is split into various frequencies to robotise the selection method of the sensitive frequencies to the fault under analysis. The average of each extent is then taken as the sensory aspect of the entity. The FFT technique was used on original images of both chest X-ray and CT scan images to observe the spectrum and centred spectrum of the images, as shown in Fig. 2 (a) and (b). Using FFT, we calculated 12 statistical features in all images.

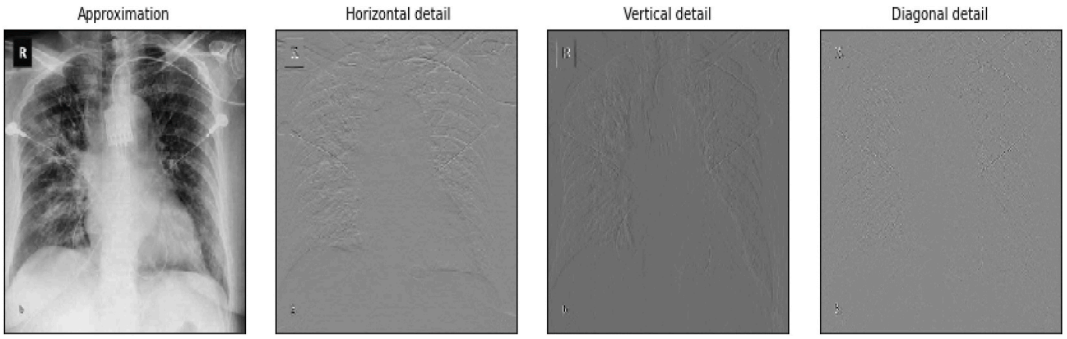


Fig. 4. Wavelet transform technique on a sample Covid-19 chest X-ray image.

4.2. GLCM positioned disjuncture

The grey-level co-occurrence matrix (GLCM) is an analytical process broadly utilised to describe images and has been shown to be significant for second harmonic generation (SHG) collagen picture classification. This system considers the spatial connection between the image pixels at a particular point. Typically, it is determined in four directions at explicit distances. Textural highlight work is determined over this grid. Generally, various directions are contrasted or found in the middle value to obtain a novel measurement boundary [24]. GLCM segmentation of a sample Covid-19 chest X-ray image with GLCM dissimilarity, grass, and sky in four positions is shown in Fig. 3.

The co-occurrence matrix is formally defined as the probability of grey level m occurring in the neighbourhood of another grey level n at a distance e in direction ö,R (m, n |e, ö), where e is a displacement vector, e =(Δa, Δb), and the direction ö is one of eight orientations. The difference between opposite directions is often ignored, and symmetric probability matrices can be employed only for four orientations, i.e., 0°, 45°, 90°, and 135°. Statistical measures extract image features from this matrix. In the present study, the correlation (C) textural feature function was used. This is expressed in Eq. (2):

$$C = \sum_{m,n} \frac{(m-\gamma_m)(n-\gamma_n)R(m,n)}{\sigma_m \sigma_n} \tag{2}$$

where

$$\gamma_m = \sum_{m,n} m.\, r(m,n) \tag{3}$$

$$\gamma_n = \sum_{m,n} n.\, r(m,n) \tag{4}$$

$$\sigma_m = \sqrt{\sum_{m,n} (m-\gamma_m)^2 .r(m,n)} \tag{5}$$

$$\sigma_n = \sqrt{\sum_{m,n} (n-\gamma_n)^2 .r(m,n)} \tag{6}$$

Using GLCM, we calculated the first ten statistical features in four spatial locations for all images.

4.3. Modified GLDM and texture positioned disjuncture

Texture as an image highlight is extremely valuable in many image handling and computer vision applications. A broad survey on texture examination in image refining revealed the essential parameters of order, division, and union. Texture highlights have been utilised in various applications, such as satellite and aeronautical image examination, clinical image investigation for identifying variations from abnormalities, and recently in image recovery utilising surface as a descriptor. This section presents a method to deal with portraying texture utilising a multi-band disintegration of an image with application to characterisation, division, object recognition, and picture recovery [25]. In the texture analysis, we calculated the 12 statistical features from all the images.

The modified grey level difference method (GLDM) interaction computes the grey-level difference statistical likelihood thickness capacities for the pre-processed grey picture. This technique is utilised to remove the entire surface highlights of an advanced image. Contrast is characterised as an adjustment in density among the most noteworthy and least dense stages in an image. Therefore, the neighbourhood varieties are on the grey level. The angular second moment is the proportion of the homogeneity. If the contrast between grey levels over an area is low, these areas are expressed as having better angular second moment (ASM) values. The mean offers normal force estimation [26]. We calculated the first ten statistical features in four spatial locations with distance d = 8 from the reference and neighbour pixels (p, q) in all images in the database.

$$y\,(p,q) = |x(p,q) - x(p,q+d)| \tag{7}$$

where x is the input image, y is the result of image x, and d is the distance for the modified GLDS calculation.

4.4. Wavelet positioned disjuncture

A discrete wavelet transform is characterised as a non-redundant tested CWT. The wavelet transform is intended to address a discrete-time arrangement y(n) as a bunch of wavelet coefficients. These coefficients are inspected from a CWT to yield a symmetrical (or biorthogonal) set of premise capacities. Wavelet plans are abundant, with varying qualities. In this section, consideration is limited to symmetrical wavelets with little assistance.

The utilisation of symmetrical bases guarantees that the portrayal is non-redundant. Typically, the utilisation of symmetrical portrayals prompts straightforward calculations for both remaking and disintegration. However, productive wavelet calculations require computational burdens that are frequently not expected to actualise a quick Fourier transform [27].

There are a few comparative perspectives from which wavelets can be studied. Here, we discuss the wavelet through the idea of a channel bank. A couple of finite impulse response (FIR) channels with M coefficients are characterised. One of these channels is a high-pass channel, while the second is a low-pass channel, which cuts on/off at a large portion of the inspecting recurrence. The wavelet transform can be characterised by utilising these channels and applying them recursively. The channels are first applied to the input time arrangement to yield low-pass and high-pass segments, $Y_1(n)$ and $Y_2(n)$, individually:

$$Y_1(n) = \sum_{l=0}^{M-1} e_l y(n-l) \tag{8}$$

$$Y_2(n) = \sum_{l=0}^{M-1} f_l y(n-l) \tag{9}$$

where e_l and f_l are the coefficients of the low-pass and high-pass filters, respectively. It is expected to build a high-pass channel dependent on the low-pass channel, which is generally accomplished by utilising the rotating flip plan. Hence, the two arrangements of channel coefficients are connected as follows:

$$f_l = (-1)^l e_{M-l} \tag{10}$$

The yield of the two channels is a large portion of the input sequence transmission capacity, such that $Y_1(n)$ involves a lower recurrence range and $Y_2(n)$ is the upper band. The yields of each channel constitute a large portion of the first data transmission of Y(n) to ensure that these double cross arrangements contain excess data. Therefore, one can subsample the two-channel yields to a large portion of the first inspection rate. Subsampling is accomplished by disposing of substitute examples in each arrangement. Such resampling implies that the two signals indeed involve full data transmission. Because of the defective plan of the two FIR channels, the activity of resampling is associated with the two segments.

Here, we performed two-way sequential coefficient operations such as CA1, CH1, CV1, and CD1 from CA1 and again calculated other wavelet coefficients such as CA2, CH2, CV2, and CD2 for every sequential coefficient, finally calculating the 12 statistical features (see Fig. 4).

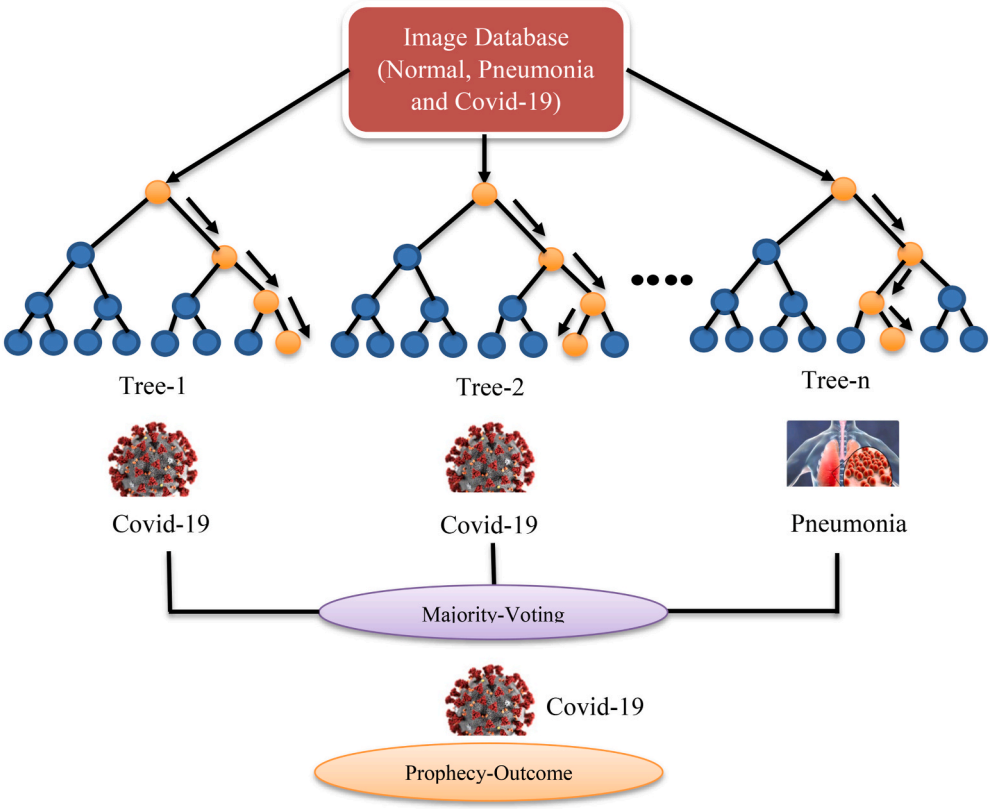


Fig. 5. Illustration of random forest and its classification.

4.5. *Applying random forest classification*

Random forest is a supervised learning mechanism used for classification and regression problems. However, it is primarily utilised for classification issues. A forest comprises trees, and more trees imply a more vigorous forest. Essentially, the random forest mechanism makes decision trees from information tests and subsequently obtains the forecast from every one of them, eventually choosing the best outcome through voting. A group strategy is superior to a solitary decision tree because it diminishes overfitting by averaging the outcome. An illustration of the operation of the random forest classifier is shown in Fig. 5.

Functioning of Random Forest Mechanism:

Stage 1: Initially, we start with the determination of random data samples from a given dataset.
Stage 2 – Next, this mechanism develops a decision tree for each data sample. Then, at that point, we obtain the expected result from each decision tree.
Stage 3: In this progression, voting is performed for each prophecy outcome.

The advantages of the Random Forest mechanism are as follows:

- Random forest prevents overfitting by averaging or joining the outcomes of various decision trees.
- It performs better for an enormous scope of data samples than does a solitary decision tree.
- It is truly adaptable and offers exceptionally high accuracy.
- Scaling of data is not required for a random forest mechanism. This results in great precision, even after providing information without scaling.
- The random forest mechanism maintains a high precision, even when an enormous amount of data disappears.

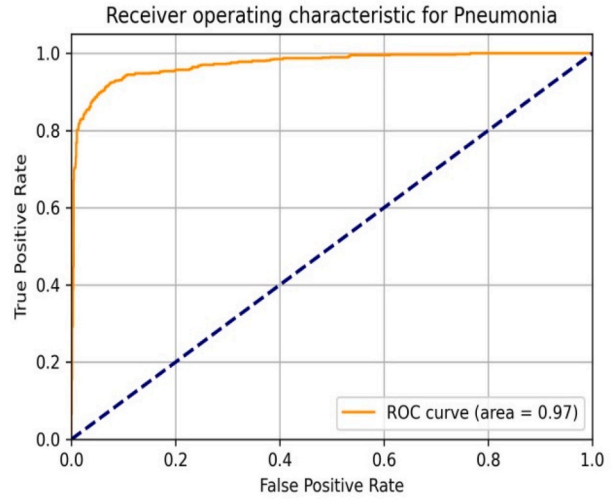
The random forest mechanism—an ensemble machine learning method broadly known for its superior performance over other machine learning techniques—was preferred for our model. The number of trees and the preferred number of instances, i.e., 100 are used in the random forest classifier to process if batch prediction is being performed. More or fewer instances may be provided, but this allows implementations a chance to specify a preferred batch size. We applied random forest classification to subgroups of Covid-19 individuals, delineating them by normal and pneumonia individuals.

5. Results and analysis

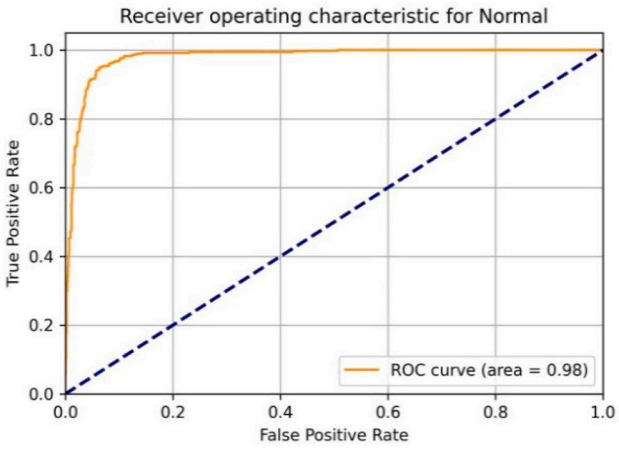
While training the system, the testing set attained an approximately 0.32 loss score and 96.8% accuracy in chest X-ray images, while the CT scan images and testing set attained an approximately 0.3 loss score and 97% accuracy. The loss score demonstrated a solid match between training and testing, confirming that our system does not experience overfitting or underfitting in both X-ray and CT scan databases. Then, we determined a receiver operating characteristic (ROC) curve and area under the ROC (AUC) to additionally evaluate the efficiency of our system, as shown in Fig. 6. (a, b, c). A comparison of Covid-19 chest X-ray images with non-Covid-19 and normal demonstrated that our mechanism achieved 98% precision and 99% recall when assessed on a test set of 1800 chest X-ray images, as shown in Table 1.

For CT scan images, the proposed mechanism achieved 97% accuracy among normal and Covid-19 individuals. The output visualisation of the CT scan image database is shown in Fig. 7. The depiction estimates utilised for recognising the ideal division approach for Covid-19, normal, and pneumonia pictures ensue.

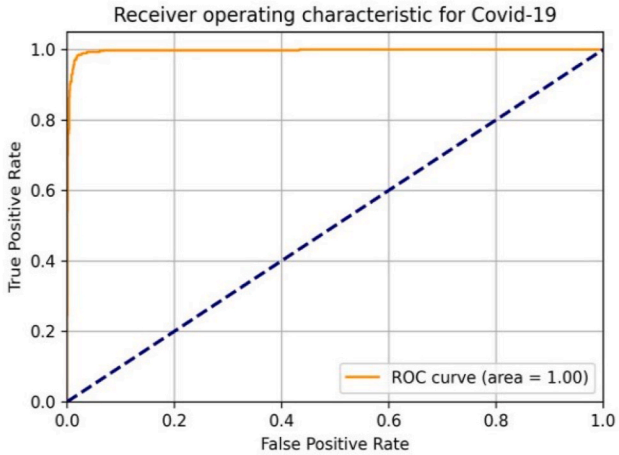
$$A = \frac{\sum_{u=1}^{u=k} c\,[u][u]}{N} \tag{11}$$



(a)



(b)

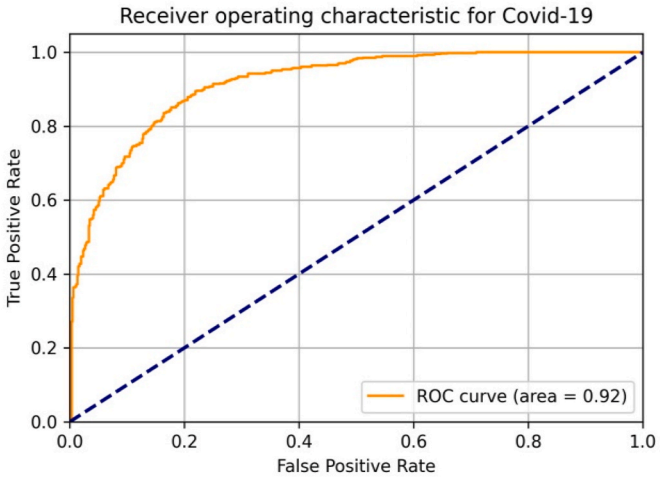


(c)

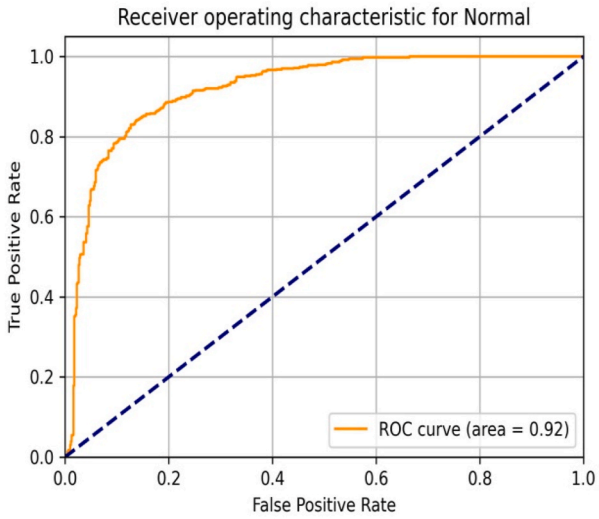
Fig. 6. Chest X-ray image database performance: (a) ROC graph for 1800 test cases, with pneumonia as the target class; (b) ROC graph for 1800 test cases, with normal as the target class; (c) ROC graph for 1800 test cases, with Covid-19 as the target class.

Table 1
Confusion matrix for 1800 test cases with and without Covid-19.

Class	Covid-19	Pneumonia	Normal	Support
Covid-19	603	4	0	607
Pneumonia	10	562	30	602
Normal	0	13	578	591
Total				1800



(a)



(b)

Fig. 7. CT scan image database performance: (a) ROC graph for 1800 test cases, with Covid-19 as the target class; (b) ROC graph for 1800 test cases, with normal are the target class.

$$\rho(\mathrm{v}) = \frac{c\,[v][v]}{\sum_{u=1}^{u=k} c\,[u][v]} \tag{12}$$

$$\prod(v) = \frac{c\,[v][v]}{\sum_{u=1}^{u=k} c\,[v][u]} \tag{13}$$

$$\partial(v) = \frac{c\,[v][v]}{\sum_{u=1}^{u=k} c\,[v][v]} \tag{14}$$

Table 2
Performance metrics for the proposed model: chest X-ray images.

Class	Precision	Recall	F-Measure	MCC
Covid-19	0.98	0.99	0.98	0.98
Pneumonia	0.97	0.93	0.95	0.92
Normal	0.95	0.97	0.96	0.94

Table 3
Confusion matrix for 1800 test cases with and without Covid-19.

Class	Covid-19	Normal	Support
Covid-19	609	22	631
Normal	13	556	569
Total			1200

Table 4
Standard depiction of the features in five categories.

Methods	Chest X-ray Image Database		CT Scan Image Database	
	Accuracy (%)	Value of Kappa	Accuracy (%)	Value of Kappa
Texture (T)	88.8	0.83	96.4	0.92
FFT (F)	90.5	0.85	95.8	0.91
DWT (D)	95.3	0.93	96.5	0.92
GLCM (G)	94.3	0.91	96.2	0.92
GLDM (g)	93	0.89	96.9	0.93
Proposed Deep Covix-net model (T + F + D + G + g)	**96.8**	**0.94**	**97**	**0.94**

Table 5
Comparison of different machine learning classifiers on Chest X-ray and CT Scan image database.

Machine Learning Algorithms	Chest X-ray Accuracy (%)	CT Scan Accuracy (%)
SVM	92.5	86.2
Naive Bayes	77.6	59.3
Decision Tree	91.3	90.4
Random Forest	**96.8**	**97**
Logistic	96.3	89.7

$$F1\ Score = \frac{2 * \frac{c[v][v]}{\sum_{u=1}^{u=k} c[u][v]} * \frac{c[v][v]}{\sum_{u=1}^{u=k} c[v][u]}}{\frac{c[v][v]}{\sum_{u=1}^{u=k} c[u][v]} + \frac{c[v][v]}{\sum_{u=1}^{u=k} c[v][u]}} \quad (15)$$

where k indicates the number of classes, c is the confusion matrix, N is the total number of elements, and [u][u] denotes elements on the diagonal corresponding to the correct prediction. In addition, α represents accuracy, ρ(v) represents precision for the vth class, Π(v) represents recall for the vth class, and ∂(v) denotes the specificity for the vth class. The performance metrics of the proposed model in terms of the chest X-ray image database are listed in Table 2. The confusion matrix for the CT scan image database is shown in Table 3.

Table 4 compares the performance of five segmentation techniques in terms of the accuracy and kappa value of the features. The proposed deep Covix-Net model approach has better predictions than other segmentation approaches correlated with Covid-19 and non-Covid-19. It exhibits the significance of obtaining a recurrence space and infers that those lineaments are pertinent to identifying Covid-19 disease on chest X-ray images. The proposed model improves the accuracy by 8%, 6.3%, 1.5%, 2.5%, and 3.8% compared with the texture, FFT, DWT, GLCM, and GLDM feature extraction techniques, respectively, in the chest X-ray image database. For the CT scan image database, the proposed model

Table 6
Comparison of various percentages of train-test splitting on chest X-ray and CT scan image database.

Train-Test Splitting with Random Forest Classifier	Image Count in Chest X-ray (Train-Test)	Chest X-ray Accuracy (%)	Image Count in CT Scan (Train-Test)	CT Scan Accuracy (%)
90%–10%	8100–900	97.3	5400–600	98
80%–20%	7200–1800	**96.8**	4800–1200	**97**
70%–30%	6300–2700	95.8	4200–1800	96.9
60%–40%	5400–3600	95	3600–2400	96.4
50%–50%	4500–4500	94.7	3000–3000	96.1

Table 7
Distinct mechanisms applied to identify Covid-19 individuals via chest X-ray and CT scan images.

Reference	Category of Chest Images	Mechanism	Accuracy (%)
Oh et al. [28]	X-ray	Patch-based CNN	88.9
S. Wang et al. [29]	CT Scan	Modified Inception Transfer Learning	92.4
Loey et al. [30]	X-ray	Googlenet	80.56
Hemdan et al. [31]	X-ray	CovidX- Net	90
Rahimzadeh et al. [32]	X-ray	Xception+ ResNet-50V2	91.4
Ali et al. [33]	CT scan	Alexnet	82.6
Wang et al. [34]	X-ray	COVID-Net	'93.3
Zhang et al. [35]	X-ray	CNN	96
Horry et al. [36]	X-ray	VGG 19	81
Tsiknakiset al. [37]	X-ray	Inception V3	92
Vinod et al. [38]	X-ray	Decision Tree	87
	CT Scan		82
Mei et al. [39]	CT Scan	CNN	79.6
		MLP	74.2
		Joint	83.5
Mizuho et al. [40]	X-ray	Computer-aided diagnosis	83.6
Deep Covix-net model (Proposed Method)	**X-ray CT Scan**	**FFT + DWT + GLCM+ GLDM + Texture**	**96.8 97**

improves to the tune of 0.6%, 1.2%, 0.5%, and 0.8%, with a 0.1% accuracy, with respect to the techniques mentioned earlier. The proposed mechanism provides a recall of 97.9%, specificity of 96.1%, precision of 96.5%, and F1 score of 97.1% in the CT scan image database.

Furthermore, we applied different machine learning classifiers and a CNN to predict Covid-19 individuals. The random forest classifier provides more accuracy when compared with other classifiers, as shown in Table 5, while CNN provides 95% accuracy in chest X-ray and 86.1% accuracy in CT scan image database. We applied different training and testing splitting mechanisms, such as 90–10, 80–20, 70–30, 60–40, and 50-50, as shown in Table 6. Finally, we also applied the validation mechanism with 5-fold, 10-fold, and splitting techniques (training: 80% and testing: 20%) and achieved more accuracy with more images in the testing set than the cross-validation method and various splitting percentages.

Table 7 shows various mechanisms applied in the image database to diagnose Covid-19 individuals with performance metrics such as accuracy (%) for qualitative analysis. We have applied our model in the same image database from the GitHub and Kaggle repository [41], used in other recent research. We have fully used the machine learning model and achieved better performance than the other researchers' deep learning mechanism.

6. Conclusion

The recommended Deep Covix-net model can identify a Covid-19-positive individual with a precision score of 98% and recall score of

99% for the X-ray database, while 97% accuracy for the CT scan is achieved. We used a large database size and obtained an output of approximately 96.8% accuracy for multiclassification of chest X-ray images. To the best of our knowledge, the novel methodology of the Deep Covix-Net model developed and proposed in the present work is the first of its kind to diagnose Covid-19 and non-Covid-19 individuals. We obtained good execution with balanced and large image databases compared with other existing models, as shown in Table 7. Our mechanism was completely robotised using an end-to-end system without the requirement for physical component extraction. It is feasible to develop a deep Covix-net model for multiple classifications of Covid-19, normal, and pneumonia with various segmentation approaches augmented by chest X-ray and CT scan image databases. The proposed mechanism could aid emergency clinic organisations and clinical specialists in making significant strides to deal with Covid-19, normal, and pneumonia individuals after obtaining this quick prognosis.

Acknowledgment

One of the authors (DNV) is indebted to SRMIST for the research fellowship. The authors are grateful to SRMIST management for their support. DNV is indebted to Dr. V. Dinesh Ram and Dr. S. Vinayagam at SRM Medical College Hospital and Research Centre for their expert advice and assistance in examining the chest X-ray and CT scan image database.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.104729.

References

[1] T. Singhal, A review of coronavirus disease-2019 (COVID-19), Indian J. Pediatr. 87 (4) (2020) 281–286, https://doi.org/10.1007/s12098-020-03263-6.
[2] J.F. Chan, S. Yuan, K.H. Kok, K.K. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C. Yip, R. W. Poon, H.W. Tsoi, S.K. Lo, K.H. Chan, V.K. Poon, W.M. Chan, J.D. Ip, J.P. Cai, V. C. Cheng, H. Chen, C.K. Hui, K.Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to person transmission: a study of a family cluster, Lancet 395 (10223) (2020) 514–523, https://doi.org/10.1016/S0140-6736(20)30154-9.
[3] Who, Coronavirus disease (Covid, Situation reports-54. https://www.who.int/docs/defaultsource/coronaviruse/situation-reports/20200314-sitrep-54-covid-19.pdf?sfvrsn=dcd46351_2, 2019. (Accessed 15 March 2020). Accessed.
[4] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E. C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature 44 (59) (2020) 265–269, https://doi.org/10.1038/s41586-020-2008-3.
[5] M. Haghani, M.C.J. Bliemer, F. Goerlandt, J. Li, Journal pre-proofs. https://doi.org/10.1016/j.ssci.2020.104806, 2020.
[6] A. Elkbuli, H. Ehrlich, M. McKenney, The effective use of telemedicine to save lives and maintain structure in a healthcare system: current response to COVID-19, The American journal of emergency medicine 44 (2020) 468–469, https://doi.org/10.1016/j.ajem.2020.04.003.
[7] M. Dewey, P. Schlattmann, Deep learning and medical diagnosis, Lancet 394 (2019) 1710–1711, https://doi.org/10.1016/S0140-6736(19)32498-5.
[8] S. Salehi, A. Abedi, S. Balakrishnan, A. Gholamrezanezhad, Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 Patients, AJR Am. J. Roentgenol. 215 (2020) 87–93, https://doi.org/10.2214/AJR.20.23034.
[9] C. Long, H. Xu, Q. Shen, X. Zhang, B. Fan, C. Wang, B. Zeng, Z. Li, X. Li, H. Li, Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? Eur. J. Radiol. 126 (2020) 108961, https://doi.org/10.1016/j.ejrad.2020.108961.
[10] L. Li, Z. Yang, Z. Dang, C. Meng, J. Huang, H. Meng, D. Wang, G. Chen, J. Zhang, H. Peng, Y. Shao, Propagation analysis and prediction of the COVID-19, Infectious Disease Modelling 5 (2020) 282–292, https://doi.org/10.1016/j.idm.2020.03.002.
[11] J.B. Long, J.M. Ehrenfeld, The role of augmented intelligence (ai) in detecting and preventing the spread of novel coronavirus, J. Med. Syst. 44 (59) (2020), https://doi.org/10.1007/s10916-020-1536-6.
[12] M.P. Wilson, A.S. Jack, Coronavirus disease (COVID-19) in neurology and neurosurgery: a scoping review of the early literature, Clin. Neurol. Neurosurg. 193 (April) (2020) 105866, https://doi.org/10.1016/j.clineuro.2020.105866.
[13] C. Hani, N.H. Trieu, I. Saab, S. Dangeard, S. Bennani, G. Chassagnon, M.P. Revel, COVID-19 Pneumonia: a review of typical CT findings and differential diagnosis, Diagnostic and Interventional Imaging 101 (2020) 263–268, https://doi.org/10.1016/j.diii.2020.03.014.
[14] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, et al., Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, Radiology 296 (2020) 32–40, https://doi.org/10.1148/radiol.2020200642.
[15] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795.
[16] X. Wu, H. Hui, M. Niu, L. Li, L. Wang, B. He, Y. Zha, Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study, Eur. J. Radiol. 128 (2020) 109041.
[17] R.M. Elavarasan, R. Pugazhendhi, Restructured society and environment: a review on potential technological strategies to control the COVID-19 pandemic, Sci. Total Environ. 725 (2020) 138858.
[18] M. Togacar, B. Ergen, Z. Comert, COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches, Comput. Biol. Med. 121 (2020) 103805.
[19] S. Callaghan, COVID-19 is a data science issue, Patterns 1 (2) (2020).
[20] R.M. Pereira, D. Bertolini, L.O. Teixeira, C.N. Silla Jr., Y.M. Costa, COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios, Comput. Methods Progr. Biomed. 194 (2020) 105532.
[21] K.C. Santosh, AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data, J. Med. Syst. 44 (5) (2020) 1–5.
[22] A.Z. Khuzani, M. Heidari, S.A. Shariati, COVID-classifier: an Automated Machine Learning Model to Assist in the Diagnosis of COVID-19 Infection in Chest X-Ray Images, medRxiv, 2020.
[23] F.P.G. Márquez, M. Papaelias, An Overview of Wind Turbine Maintenance Management. *Non-destructive Testing and Condition Monitoring Techniques for Renewable Energy Industrial Assets*, 2020, pp. 31–47.
[24] A.V. Alvarenga, W.C. Pereira, A.F.C. Infantosi, C.M. Azevedo, Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images, Medical physics 34 (2) (2007) 379–387.
[25] B.S. Manjunath, G.M. Haley, W.Y. Ma, S.D. Newsam, Multiband techniques for texture classification and segmentation, in: Handbook of Image and Video Processing, Academic press, 2005, pp. 455–470.
[26] D.C.R. Novitasari, A. Lubab, A. Sawiji, A.H. Asyhar, Application of feature extraction for breast cancer using one order statistic, GLCM, GLRLM, and GLDM, Adv. Sci. Technol. Eng. Syst. J. 4 (4) (2019) 115–120.
[27] P.M. Bentley, J.T.E. McDonnell, Wavelet transforms: an introduction, Electron. Commun. Eng. J. 6 (4) (1994) 175–186.
[28] Y. Oh, S. Park, J.C. Ye, Deep learning COVID-19 features on CXR using limited training data sets, IEEE Trans. Med. Imag. 39 (8) (2020) 2688–2700, https://doi.org/10.1109/TMI.2020.2993291.
[29] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, B. Xu, A Deep Learning Algorithm Using CT Images to Screen for Corona Virus Disease (COVID-19), medRxiv, 2020.
[30] M. Loey, F. Smarandache, N.E.M. Khalifa, Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning, Symmetry 12 (4) (2020), https://doi.org/10.3390/SYM12040651.
[31] E.E.D. Hemdan, M.A. Shouman, M.E. Karar, Covidx-Net, A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images, 2020 arXiv preprint arXiv:2003.11055.
[32] M. Rahimzadeh, A. Attar, A modified deep convolutional neural network for detecting COVID-19 and Pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2, Informatics Med. Unlocked 19 (2020), https://doi.org/10.1016/j.imu.2020.100360.
[33] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795.
[34] L. Wang, A. Wong, COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images, 2020 arXiv preprint arXiv:2003.09871.
[35] J. Zhang, Y. Xie, Y. Li, C. Shen, Y. Xia, COVID-19 screening on chest X-ray images using deep learning based anomaly detection. 2020, aeXiv 1 (2002) 12338.
[36] M. Horry, et al., X-ray image based COVID-19 detection using pre-trained deep learning models. https://doi.org/10.31224/osf.io/wx89s, 2020.
[37] N. Tsiknakis, et al., Interpretable artificial intelligence framework for COVID-19 screening on chest X-rays, Exp. Ther. Med. 20 (2020) 727–735, https://doi.org/10.3892/etm.2020.8797.
[38] D.N. Vinod, S.R.S. Prabaharan, Data science and the role of Artificial Intelligence in achieving the fast diagnosis of Covid-19. *Chaos*, Solitons & Fractals 140 (2020) 110182.
[39] X. Mei, H.C. Lee, K.Y. Diao, M. Huang, B. Lin, C. Liu, Y. Yang, Artificial intelligence–enabled rapid diagnosis of patients with COVID-19, Nat. Med. 26 (8) (2020) 1224–1228.
[40] M. Nishio, S. Noguchi, H. Matsuo, T. Murakami, Automatic Classification between COVID-19 Pneumonia, Non-COVID-19 Pneumonia, and the Healthy on Chest X-Ray Image: Combination of Data Augmentation Methods in a Small Dataset, 2020 arXiv preprint arXiv:2006.00730.
[41] Processed dataset is available for public use at. https://github.com/mhorry/SegmentedLungCXRs. (Accessed 24 March 2021). https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. Accessed.