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.

Computer Methods and Programs in Biomedicine 226 (2022) 107109

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

A novel multimodal fusion framework for early diagnosis and accurate classification of COVID-19 patients using X-ray images and speech signal processing techniques

Santosh Kumar[a], Mithilesh Kumar Chaube[b], Saeed Hamood Alsamhi[c,d], Sachin Kumar Gupta[e], Mohsen Guizani[f], Raffaele Gravina[g,*], Giancarlo Fortino[g]

[a] Department of Computer Science and Engineering, International Institute of Information Technology, Naya Raipur, Chhattishgarh, India
[b] Department of Mathematical Sciences, International Institute of Information Technology, Naya Raipur, Chhattishgarh, India
[c] Insight Centre for Data Analytics, National University of Ireland, Galway, Ireland
[d] Faculty of Engineering, IBB University, Ibb, Yemen
[e] School of Electronics and Communication Engineering, Shri Mata Vaishno Devi University, Katra, India
[f] Machine Learning Department, Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates
[g] Department of Informatics, Modeling, Electronic, and System Engineering, University of Calabria, Rende 87036, Italy

ARTICLE INFO

Article history:
Received 13 May 2022
Revised 11 July 2022
Accepted 2 September 2022

Keywords:
Deep learning
COVID-19
Early detection
Speech processing
X-ray image classification
Multimodel fusion

ABSTRACT

Background and objective: COVID-19 outbreak has become one of the most challenging problems for human being. It is a communicable disease caused by a new coronavirus strain, which infected over 375 million people already and caused almost 6 million deaths. This paper aims to develop and design a framework for early diagnosis and fast classification of COVID-19 symptoms using multimodal Deep Learning techniques.

Methods: we collected chest X-ray and cough sample data from open source datasets, Cohen and datasets and local hospitals. The features are extracted from the chest X-ray images are extracted from chest X-ray datasets. We also used cough audio datasets from Coswara project and local hospitals. The publicly available Coughvid DetectNow and Virufy datasets are used to evaluate COVID-19 detection based on speech sounds, respiratory, and cough. The collected audio data comprises slow and fast breathing, shallow and deep coughing, spoken digits, and phonation of sustained vowels. Gender, geographical location, age, pre-existing medical conditions, and current health status (COVID-19 and Non-COVID-19) are recorded.

Results: The proposed framework uses the selection algorithm of the pre-trained network to determine the best fusion model characterized by the pre-trained chest X-ray and cough models. Third, deep chest X-ray fusion by discriminant correlation analysis is used to fuse discriminatory features from the two models. The proposed framework achieved recognition accuracy, specificity, and sensitivity of 98.91%, 96.25%, and 97.69%, respectively. With the fusion method we obtained 94.99% accuracy.

Conclusion: This paper examines the effectiveness of well-known ML architectures on a joint collection of chest-X-rays and cough samples for early classification of COVID-19. It shows that existing methods can effectively used for diagnosis and suggesting that the fusion learning paradigm could be a crucial asset in diagnosing future unknown illnesses. The proposed framework supports health informatics basis on early diagnosis, clinical decision support, and accurate prediction.

1. Introduction

COVID-19 pandemic has had an unprecedented economic and social impact worldwide. By early February 2022, with more than almost six million deaths and over 375 million infections, the pandemic is still a global concern, without showing any signs of nearing to an end. The number of infected people across the world is still increasing [25]. In dealing with the COVID-19 pandemic,

* Corresponding author.
E-mail addresses: santosh@iiitnr.edu.in (S. Kumar), mithilesh@iiitnr.edu.in (M.K. Chaube), saeed.alsamhi@insight-centre.org (S.H. Alsamhi), sachin.gupta@smvdu.ac.in (S.K. Gupta), mguizani@ieee.org (M. Guizani), r.gravina@dimes.unical.it (R. Gravina), giancarlo.fortino@unical.it (G. Fortino).

https://doi.org/10.1016/j.cmpb.2022.107109
0169-2607/

there is the need to design reliable early diagnosis and intervention methods and implement effective mitigation efforts. Therefore, the design of these early diagnosis strategies hinges on the effective prediction of surveillance of the disease's spatio-temporal evolution [1].

A reliable learning method of forecasting the spread of the virus and early diagnosis could significantly enhance the predictive surveillance capability and help designing disease containment policies. Prior research has suggested and evaluated different statistical, epidemiological, and Machine Learning (ML)-based forecasting models for COVID-19 [1,4,9] based on factors such as the number of current infections, fatalities, and recoveries. Other epidemic forecasting models, such as those suggested in Windmon et al. [10], Alsamhi et al. [11], Alsamhi and Lee [12], rely on human movement and within- and between-season data.

Although these models can forecast the initial outbreak and growth trajectories, they are restricted in their ability to capture many temporally dynamic and geographically variable aspects driving disease spread [1]. Therefore, governments are looking for disruptive technologies to promptly diagnose individuals and control the widespread COVID-19 pandemic [4,26,27,29,30]. Several clinical diagnosis-based frameworks and strategies have been deployed for testing, tracing, and treatment, helping to crush the pandemic curve across the world (e.g., in Singapore, South Korea, and China) in its early stages. The test works are similar to Polymerase Chain Reaction (PCR) to identify a portion of the COVID-19 ribonucleic acid (RNA) in the nasopharyngeal or oropharyngeal swab [5].

Based on World Health Organization (WHO) supervision, Nucleic Acid Amplification Tests (NAAT) like real-time Reverse Transcription (rRT-PCR) must be applied for routine confirmation of COVID-19 infection by detecting the unique sequences of virus RNA. Thanks to the enormous advancement of medical science and disruptive technologies, most existing diseases have been appropriately diagnosed based on available tools and methods. In addition, these methods allow for curing or preventing the individual from further consequences [5].

The traditional controlling mechanism is reported as manual procedures. However, traditional medical clinical procedures diagnosis methods take more time to process the samples for early diagnosis and predictions, favoring the spread of a highly communicable disease such as COVID-19. Social isolation of positive suspected subjects, strict quarantine policies, social distance between people, personal hygiene, and use of personal protection equipment such as face masks are globally implemented guidelines [6,7].

The collected databases/samples of infected people from the different communities are analyzed to find significant diagnosis measures using diversified methods, including mathematical modelling, statistical techniques, simulation, statistical modelling, various ML-based representation techniques, and data mining. These methods are also used to predict disease behaviour both at individual and community level.

Unfortunately, we are still struggling to stop the fast-spreading of infection at the community level because proper medical testing kits and procedural methods are not used at a large scale to provide vaccination to citizens of different countries [8]. Based on literature work, several statistical analysis-based models and mathematical modelling systems are used for early diagnosis. In addition, artificial intelligence methods are highly used to analyze collected data of infected people from different countries. The workflow of a statistical model is illustrated in Fig. 1. The statistical model consists of different components: (1) input data collection, (2) data pre-processing of the collected sample, (3) data modelling, (4) data analysis, (5) data visualization and representation.

The input data is collected from different individual patient and stored at a cloud server for further processing. The input data consists of patients information such as age, sex, infected or not infected status, medical procedures taken for complete diagnosis, and records of taken medicines. In pre-processing step, the collected input data is treated using pre-processing techniques to achieve essential data after removing noises and artifacts stored using statistical methods.

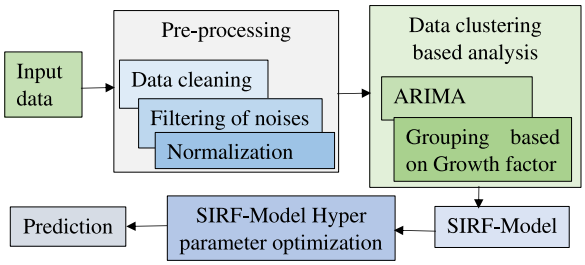

Fig. 1. SIR model using statistical learning.

Data modelling techniques and data representation and visualization steps identify data that can be grouped into different clusters to analyze the growth and prediction of infection rates using Autoregressive Integrated Moving Average (ARIMA) models (shown in Fig. 1). Based on the literature, several statistical analysis-based models and mathematical modelling systems are used for early diagnosis. However, collected datasets consist of different between-class and within-class labeled information on COVID-19. The shared labeled information of infected people cannot analyze by applying statistical machine learning techniques due to overlapped information, including Autoregressive Integrated Moving Average (ARIMA) models. These techniques and models cannot find the scatter matrix to separate these classes for better analysis of the infected patient.

1.1. Motivation

With the ever-growing requirement for expediting early diagnosis and screening millions of patients, huge efforts and technological innovation are deployed to face COVID-19 spread [1,4–6]. It is worth noting that COVID-19 outbreak is partially due to false-negative results in Reverse Transcription Polymerase chain Reaction and Reverse Transcription (RT-PCR) tests.

To tackle this issue, several interdisciplinary studies proposed frameworks for fast and accurate COVID-19 diagnosis and prediction. These frameworks have been used to estimate growth rates and infection rates to prevent acute infection due to the COVID-19 pandemic. However, the proposed methods and frameworks are unable to process massive amount of data [19]. As a result, several clinical testing labs and medical doctors have faced significant difficulties in disease prediction. Several research groups [5], and medical expert committees have recommended to follow the imaging of the chest of humans for early COVID-19 diagnosis. Based on routine use of scanning methods, it has been widely demonstrated that Chest X-Ray-based imaging (CXR), and Computed Tomography (CT) scans are highly reliable for COVID-19 early diagnosis. However, several researchers have recommended that CT-based analysis methods are not appropriate in many circumstances due to the use of contrast material (dye), making it inapplicable for patients with significant medical conditions such kidneys failure. In addition, CT scans are non-specific and overlap with other acute viral infections including influenza, H1N1, SARS, and MERS [10].

Based on the available literature, the analysis of chest X-ray images using ML techniques represents a valuable methodology for early and accurate prediction.

Moreover, the chest X-ray images are repeatedly obtained over time to monitor the evolution of lung disease. Wong et al. [8] showed that the severity of CXR results peaked at 10–

12 days from the date of symptom onset and proposed a model based on X-ray analysis obtaining 69% sensitivity. Artificial Intelligence (AI) techniques are attracting interest due to the high need for early COVID-19 diagnosis. In particular, Deep Learning (DL) has received prevalent attention for the analysis of chest X-ray images [22] and cough (audio) signals. However, there is a need of very effective computational approaches to realize fast, automated, effective computing systems and algorithms to detect abnormalities in chest X-rays images that are due to early COVID-19 infection.

To address this problem, we proposed a novel multimodal framework for COVID-19 patients early diagnosis and accurate prediction using DL techniques. The proposed framework uses a Convolutional Neural Network (CNN) to train the model based on the chest X-ray images database and cough (voice) samples. We employ the weighted sum rule method to fuse both the chest X-ray and cough (audio) model to predict COVID-19 accurately.

The benefits of the proposed multimodal framework are to use, for early diagnosis, non-invasive, fast prediction method and novel architectures. The proposed framework also uses voice samples to process the spectrograms of cough episodes for model training.

1.2. Contributions

The major contributions of this work are highlighted as follows:

1. A novel multimodal framework is proposed for COVID-19 patients early diagnosis and accurate prediction using DL techniques.
2. The proposed multimodal framework consists of chest X-ray images and cough (voice) based models on processing the chest X-ray images and cough sample database for extracting discriminatory features and performing early prediction of COVID-19 and non-COVID-19 patients.
3. The proposed framework applied the U Net DL technique, CNN speech processing techniques to perform segmentation and to extract features from the preprocessed chest X rays images by isolating the image portion containing the lung parts.
4. The experimental results show accurate classification based on different settings and protocols. This may help in analyzing data, predicting, and planning prospective applications for future pandemics.

The rest of the paper is organized as follows. Section 2 illustrates the method for early diagnosis of COPD. Section 2.6 depicts the feature extraction and classification techniques applied in this work. Section 3 shows the experimental results and performance evaluation based on different existing benchmark protocols and methods. Section 3.4 reports the results of statistical analysis carried out on the X-ray and cough audio datasets. Finally, conclusions and prospected future directions are drawn in Section 5.

2. Methods

This section presents the proposed multimodal learning framework for early diagnosis and fast prediction of COVID-19 patients. The framework consists of two learning models: (1) chest-X-rays image classification model, (2) cough (voice) based analysis model (see Fig. 2). The steps involved in the workflow are explained in the block diagram and detailed in the following.

2.1. Chest X-ray model

We collected the chest X-ray image database for extracting rich and distinct information. The proposed model extracts the discriminatory features from the captured chest X-ray images for prediction. In the proposed multimodal system, we used CNN learning model to classify COVID-19 and non-COVID-19 (non-infected subjects) based on the adopted chest X-ray image datasets.

The chest X-ray-based working model consists of the following steps: (1) collection of database description and preprocessing/cleaning, (2) segmentation and clustering of images, (3) feature extraction, and (4) classification. Since we focus on one disease, the chest X-ray model is binary, so the result is either positive to COVID-19 or negative. In many literature works, we found that with the increase of the number of classes, the classification becomes less accurate on average. We have divided the whole work of the chest X-ray model into several parts.

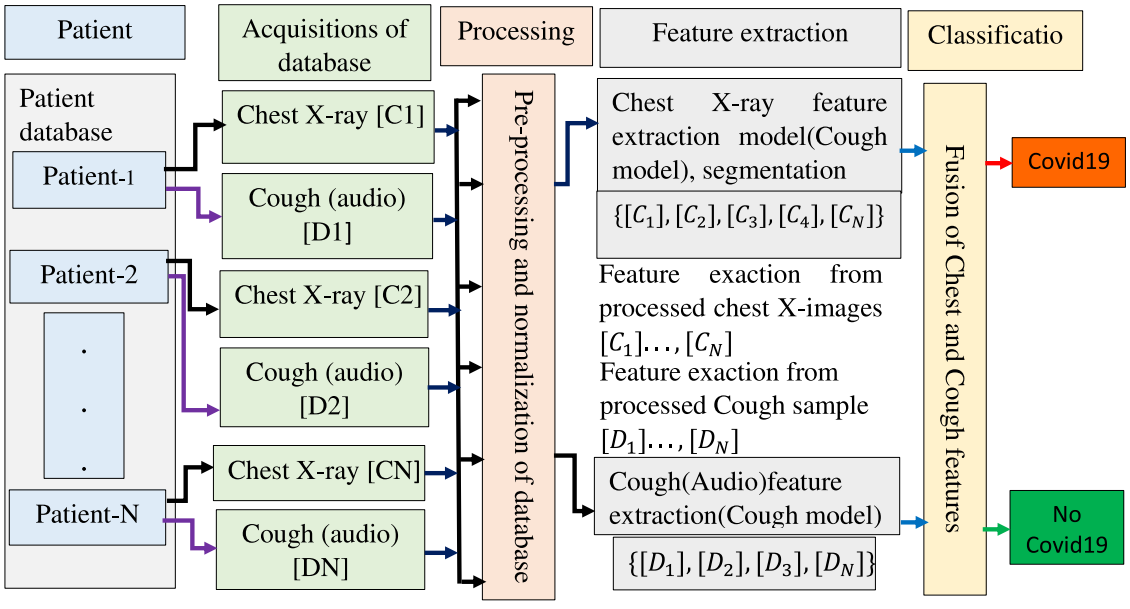

Fig. 2. Proposed model for early accurate prediction of COVID-19.

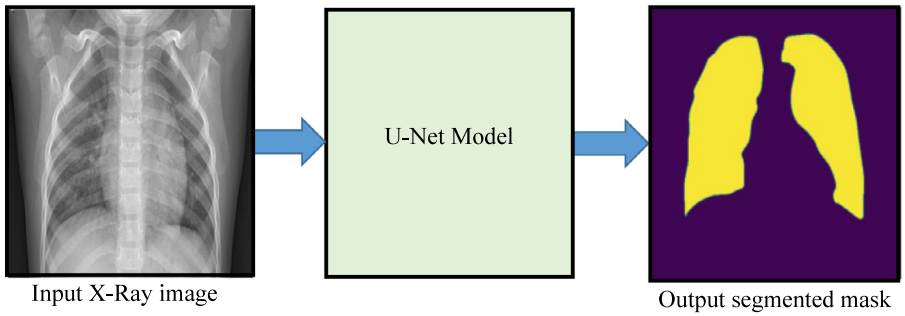

Fig. 3. Segmentation of chest X-ray image using U-net model.

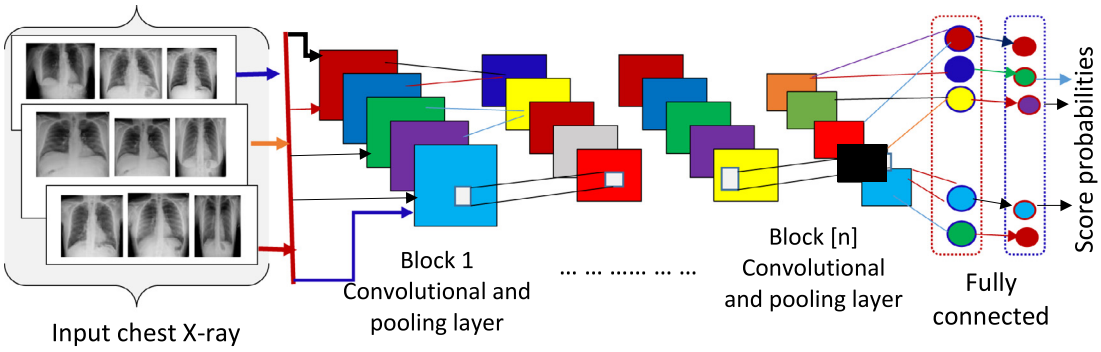

Fig. 4. Workflow of chest X-ray model using CNN architecture.

2.2. Data collection

We have collected the COVID-19 chest X-ray images and cough samples from open datasets [2,22]. Specifically, details on the X-ray images database are summarized in Table 1.

2.3. Segmentation of chest-X-ray image

Segmentation is the process of dividing the input image into the distinct region of interest. In the chest X-ray image processing model, the lung images of infected patients are divided into regions of interest using the deep U-Net-based DL architecture for classification (see Fig. 3).

We trained the U-Net model on the Shenzhen hospital X-rays dataset and validated the proposed framework based on the Montgomery County dataset since segmentation of the medical image is a significant challenge for obtaining an accurate region of interest. For example, the chest X-ray image separates the lungs region in a chest X-ray. Therefore, we have used the Montgomery and Shenzhen datasets with the lungs masks and trained the proposed model with the masks to get the lungs region as output (see Fig. 3).

Table 1
Database of chest X-ray images.

Dataset	Size	Used model
A	468	Used to make COVID-19 class data
B	5860	Used to make non-COVID-19 class data
C	566	Segmentation model
D	138	Used for validation of segmentation model
E	852	426N(190W + 236M) + D

A: IEEE-8023 CXR - Cohen dataset [23], B: Pneumonia and normal chest X-ray, C: Shenzhen CXR with Masks, D: Montgomery county CXR images, E: COVIDGR 1.0, W = Women, M = Men, N = Negative cases, P = positive D = 426P(239W + 187M) used training model.

2.4. Classification model

The proposed system employs the CNN technique and the U-Net learning model to classify COVID-19 and non COVID-19 patients based on provided four labeled classes (i.e. Normal, Bacteria infection, Tuberculosis (TB), Viral infection(VI), COVID-19, non finding (non COVID-19). The basic architecture of the U-Net learning-based CNN framework is shown in Fig. 4. The classification model contains convolution, pooling, and fully connected layers. We used the Darknet-19 model [13] for the classification of COVID-19. It consists of 19 convolutional and five max-pool layers. Each convolutional layer has a different number of filters with size 5×5. Thus, the number of convolution filters increased in the architecture. In the proposed framework, we used fewer layers for classification.

The proposed system consists of 17 convolutional layers, a pooling layer, and a fully connected layers. We used the batch normalization method and LeakyReLU non-linear activation function to train the proposed system's over-collected chest X-ray image database (see Fig. 4). The batch normalization technique is used to solve the over-fitting problem in the trained model for classifying COVID-19 and non-COVID-19 based on the chest X-ray image database. In LeakyRelu non-linear activation function, instead of function becoming zero for (x, 0), it has a slight negative slope (nearly 0.01), which prevents the dying ReLU. An activation function takes a value and performs a mathematical operation. Based on the output of the activation function, it decides which neuron has to activate.

The pooling layer reduces feature maps' dimension using down-sampling by summarizing the features using the down-sampling technique. The fully connected layer evaluates measured scores probability of the output COVID-19 and non COVID-19 classes. We have used several combinations of convolution layer and max layer according to the requirement of the proposed method for better training and testing of COVID-19 and non COVID-19 classes (see Fig. 4).

2.5. Cough based COVID-19 diagnosis model

The noticeable symptoms of COVID-19 infected patients include severe cough and breathing difficulties. Therefore, when these breathing and cough samples are analyzed using speech signal processing and ML techniques, we claim that the respiratory sounds of patients provide valuable insights, enabling the development of an early diagnostic DL tool.

2.5.1. Data collection and description

We used several open datasets from different sources.

1. **Collection of Coswara sound database:** The Coswara project, which is affiliated with IISc Banglore in India, provided us with a cough (audio) recording database. The main goal of this project is to create a diagnostic tool for detecting COVID-19 using respiratory, cough, and speech sounds. Such dataset includes breathing noises (rapid and slow), cough sounds (deep and shallow), phonation of sustained vowels (/a/ as in made, /i/,/o/), and counting numbers at a slow and fast rhythm.
2. Subjects were from all continents except Africa, and audio recordings were sampled at 44.1 KHz. The dataset comprised of various categories, namely cough (two kinds: heavy and shallow), breathing (two kinds: heavy and shallow), sustained vowel phonation (three kinds: *a*, *e*, *o*), and digit counting (two kinds: fast and regular) along with metadata information. The Coughvid [14], DetectNow [15], Sarcos, and Virufy [16] datasets are publicly available.
3. **Sarcos cough sample Dataset:** In the Sarcos cough database, the infected people (subjects) were motivated to record their cough (voice) using their smartphone's microphone. The audio samples are recorded by infected people participants online. The sampling rate of cough (voice) audio recordings is 44.1 KHz. All the cough sample audio recordings are considered from different people. They presented with a voluntary and anonymous questionnaire and provided informed consent.
4. The questionnaire includes information about instructions and set of phonetical phonemes and alphabet and constant collection, including age, disease history, and gender. Suspected cases are tested by an authorized COVID-19 testing center and diagnosed with the COVID-19 positive or negative. This information is stored in the databases along with the country of residence. The spectrogram representation of one recorded cough sample is shown in Fig. 5.

2.5.2. Pre-processing of cough (audio) database

The samples of recorded audio cough are manually segmented into binary classes, namely, positive and negative case. The cough samples are labeled as positive, mild, or positive, segmented as positive, and samples in which COVID-19 status was labeled as healthy or no respiratory illness exposed were segmented as negative. We down-sampled the audio recordings at 16 KHz. We created an amplitude of 100 Hz to get rid of dead spaces and tiny background noises from the audio signal. We divided cough audio samples into chunks of 4 seconds each and padded as needed. The significant challenges are that the Coswara audio sample database includes signals with significant amount of irrelevant data; furthermore, the recordings were variable in length [5].

Fig. 2 shows the cough model to classify audio recordings. It includes preprocessing of a wide-band spectrogram of heavy cough samples for feature extraction and classification. The cough model consists of the following steps: (1) collection of data, (2) preprocessing of data, (3) data cleaning, (4) feature extraction, and (5) classification. We used 80% of the Coswara sample dataset for training the proposed cough model, while the remaining 20% was used for testing and validation.

Algorithm 1: Feature extraction.

1. Initialization: Let y(n) be an input cough (voice signal): y[n] = x[n]-ax[n].
2. Processing and filtering noise of the input signal y[n] (shown in Eq.(1):

$$y[n] = \frac{1}{n}\sum_{i=0}^{n-1} y[n-i] \qquad (1)$$

3. **Segmentation using Hamming Window:** The hamming window method is used to reduce noises and ripple from cough sample.
4. **Mel Frequency Cepstral Coefficients (MFCCs) features:**
5. **Segment the signal into short frames:** Each frame's period gram was determined, and the power spectrum's period gram estimate was calculated.
6. Apply the Mel filter bank method and use algorithm of all filter bank energies.
7. **Computation of Discrete Cosine Transformation (DCT:** We used DCT [18] to calculate the log filter bank energies and kept DCT coefficients 2–13 while discarding the rest.
8. The MeI scale is given as follows:

$$\mathrm{M(f)} = 1125\ln(1+\mathrm{f}/700) \qquad (2)$$

1. **Data preprocessing:** Preprocessing techniques are used to remove noises from input images. The input cough voice data consists of several artifacts and noises. We employed the filter approach to reduce noise in the cough sample database after isolating these artefacts. To reduce noise in the collected nasal images, we used a low-pass filter. The filter method performs computation on $X[n]$ using MA filtering technique, which takes previous $(X[n-i]$, for $i = (1, 2, 3 \ldots, n)$ low illumination sample images from the unconstrained background. The output filtered image is denoted as $Y[n]$. Let I be a vector of pixel values of given images, then:

$$Y[n] = \frac{1}{n}\sum_{i=1}^{n} X[n-i] \qquad (3)$$

2.6. Feature extraction and classification

We computed the MFCCs feature of the Intrinsic Mode Functions (IMFs), and windowed cough signal samples from spectrograms used as MFCC features to classify cough using a deep CNN model. Because MFCCs features consists of complete human respiratory, characteristics varies with time. Therefore, the discriminatory MFCCs features are extracted from cough sample together with (△) first order derivatives, and second-order (△△) derivatives on measured short term power spectrum of cough/sound sample for early diagnosis

Moreover, in this work, we computed MFCCs features are used along with first-order derivatives (△) and second-order derivatives (△△) features for differentiating dry coughs from wet coughs and tuberculosis coughs based on computed delta coefficients(shown in Eq. (4). The computation of delta coefficient ($\triangle_t$) from cough samples and MFCC features are illustrated in Algorithms 1 and 2, respectively.

The proposed model extracted MFCC features from the preprocessed audio sample database using ML techniques. The primary objective is to select discriminatory features from extracted features using the subspace feature selection method and Principal Component Analysis (PCA) method [17]. Then, the extracted features are used to train the cough model to perform prediction for cough samples.

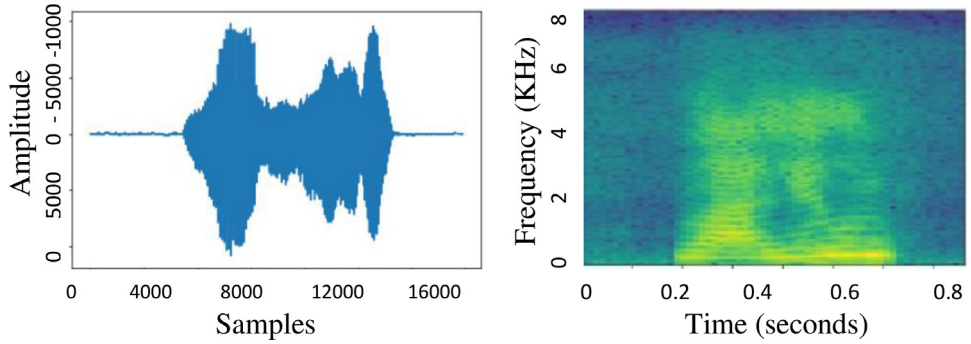

Fig. 5. Illustration of cough for volume and its power spectrum.

Algorithm 2: Calculation of delta-delta ($\triangle_t\triangle_t$) coefficient for cough.

1. The Discrete Fourier Transform (DFT) technique [18] is applied to transform the signal in the time domain from the frequency domain after dividing the speech signal into speech frames.
2. The power spectrum was obtained and triangular filters were used to map it onto the Mel scale. Figure 5 illustrates an example of a speech input volume power spectrum.
3. Log outputs are found using Discrete Cosine Transform technique.
4. Delta ($\triangle$) and delta-delta ($\triangle$) ($\triangle$) coefficients are calculated as follows:
5. Finally, let us denote the MFCC of a window frame (t) by C_t. The delta coefficient ($\triangle_t$) is computed as follows:

$$\triangle_t = \frac{\sum_{i=1}^{l} i\times(C_{t+i}-C_{t-i})}{2\times\sum_{i=1}^{l} i^2} \tag{4}$$

Algorithm 3: Segmentation of chest X ray images.

1. **Initialization:** Chest X-ray image I [M× N]=[$I_1(a_q, b_r), \ldots, I_N(a_q, b_r)$ for input image.
2. **Normalization:** the input images are resized into 500 × 500 pixels.
3. **Enhancement process:** The gray scale images are enhanced by histogram equalization technique.
4. Applying filter at center pixel (a_0, b_0) of each block:

$$C(a,b) = (a+qb)\times Gb(a,b) \tag{5}$$

where Gb(a,b) is defined as follows:

$$Gb(a,b) = e^{-}(a^2+b^2)/2\sigma^2 \tag{6}$$

where $Gb(x,y)$ is a Gaussian kernel window with $[m\times n]$. $\sigma = 1.2$ is Gaussian variance. Responses on implementing Gaussian kernel window filters on every patch/block is acquired as follows:

$$C(a,b) * D(a,b) = (e^{-}(a^2+b^2)/2\sigma^2(a+qb) * D(a,b)) \tag{7}$$

5. Filtered images are reconstructed. Highest responses from thee complex filter within already filtered image could be presumed as important features, where $D(a,b)$ is the oriented image of pixels.
6. **Output:** Segmented regions of chest X ray images.

The MFCC feature extraction includes windowing the signal, taking the Fourier Transform, wrapping the spectrum's powers into the Mel scale, and taking the powers' logs. The Mel log powers list is a signal applying discrete cosine transform to signals and results in amplitudes known as MFCCs. The steps involved in this process are further explained in Fig. 4.

After windowing and frame blocking the cough audio signal, DFT technique [18] is applied to each windowed frame to convert the audio signal to power spectrum moving from the time domain to frequency domain using the following formula.

$$X(k) = \sum_{n=0}^{N-1} x(n)e\frac{-j2\pi nk}{N} \tag{8}$$

where N is the number of points used to compute the DFT [18] and k ranges between 0 and $N-1$.

After obtaining the spectrum, we compute the log power spectrum, which gives the magnitude in decibels. This spectrum is a continuous signal with some periodic structures because the log power spectrum has some harmonic components. Hence, treating the signal as a time domain signal, we applied discrete inverse Fourier transform [18] to get a spectrum; namely, cepstrum is in a pseudo frequency domain known as Quefrency. The cepstrum represents how these quefrencies are present in the log power spectrum. The mathematical equation to obtain a cepstrum is the following.

$$C[x(t)] = F^{-1}[log(F[x(t)])] \tag{9}$$

where C is the obtained cepstrum and F^{-1} is the Inverse discrete Fourier transform, and F is the DFT technique [18]. The next step involves the computation of Mel spectrum. Mel is a unit of measure based on how the human ear perceives a frequency. Human auditory systems do not perceive pitch linearly in a physical frequency scale. The Mel approximation from physical frequency is expressed as follows:

$$f_{Mel} = 2595\log_{10}(1+\frac{f}{700}) \tag{10}$$

where f_{Mel} and f denote the perceived frequency and the physical frequency which is partitioned the physical frequency scale into bins and, using overlapping triangular filters, transform each bin into the corresponding bin in the Mel scale. A Mel spectrogram can be computed by multiplying each triangular Mel weighing filter with the magnitude spectrum.

We have considered the first 13 MFCC coefficients truncating the high order DCT coefficients to make the system robust. The zero*th* coefficient was removed since it represents the signal' aver-

Fig. 6. Segmentation of chest X-ray images using proposed model.

age log energy and contains little information. The following equation is used to calculate MFCC characteristics:

$$c(n) = \sum_{m=0}^{M-1} log_{10}(s(m))cos(\frac{\pi n(m-0.5)}{M}), [n = 0, 1 \ldots, C-1] \quad (11)$$

where (c) is the number of MFCCs of cough sample, M and $c(n)$ are respectively the number of cough samples and cepstral coefficients.

3. Results

The performance of the proposed framework is evaluated based on chest X-ray and cough sample data for accurate prediction of COVID-19 patients for early diagnosis.

1. **Performance evaluation based on chest X-ray and cough sample datasets:** The proposed system is trained with 566 images with lung masks using U-Net based DL architecture. The segmentation model provided the segmented input images to the classification model of chest X-ray with the validation accuracy of 98.46%. The input chest X-ray image, the actual masks, and the output predicted mask are highlighted in Fig. 6.
2. **Classification performance metrics:** To evaluate the performance of the classification methods, we calculated the following indicators based on confusion matrix based measures:

$$Accuracy(A) = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$Precision(P) = \frac{TP}{TP + F} \quad (13)$$

$$Recall(R) = \frac{TP}{TP + FN} \quad (14)$$

$$F1 - measure = \frac{2(P + R)}{P + R} \quad (15)$$

$$Specificity = \frac{TN}{TN + FP} \quad (16)$$

where TP are True Positive, FP are False Positive, TN are True Negative, and FN are False Negative cases.Among them, the F1 score was employed as the evaluation criteria for early halting. Finally, the overall metric scores of the algorithm were obtained by averaging each metric over numerous classes, as shown in Table 2.

The proposed multimodal framework for identifying COVID-19 and non COVID-19 discoveries uses the segmented chest X-ray images as input. For training the model, we have 468 images of COVID-19 positive patients and 720 images of non-COVID-19 patients. We employed 20% and 80% of the chest X-ray image database for training and validation of the proposed model, respectively, and for measuring system performance. A 5-fold cross-validation procedure was used to validate the suggested model. We used 20% of the total images for validation and the remaining 80% for the suggested model training scheme (see Fig. 9 shows fold 3 and fold 4 accuracy).

Table 2 shows the performance results. Fig. 7 depicts segmentation of chest X-ray images using the proposed model. The confusion matrix of the proposed method is shown in Fig. 8(a) for chest X-ray images and in Fig. 8(b) for cough audio signals.

3.1. Evaluation on chest X ray image dataset

The performance of the proposed framework is shown in Figs. 3 and 10, respectively. Our cough detection algorithm classifies cough sample for predicting COVID-19 with an overall accuracy of 82.30%. Table 3 shows the Resnet-50 classifier mode and other classifiers. These classification models provide better performance based on cough (audio) databases. The Restnet-50 model provided 97.60% accuracy for classification cough samples. The evaluation is done based on extracted 256-dimensional feature vectors. The proposed approach showed 95.30% accuracy, 93% sensitivity, and 98% specificity. This outperforms the WHO's baseline standards for a community-based triage test. The cough-based model, which used CNN and LSTM-based classifiers, also performed better in terms

Table 2
Performance of proposed chest X-ray model based on fold cross validation.

Folds	Sensitivity	Specificity	Precision	Accuracy	F1
1	0.9459	1.0000	1.0000	0.9890	0.9722
2	1.00009	1.0000	1.0000	1.0000	1.0000
3	0.9453	0.91235	0.9367	0.9578	0.9646
4	0.9354	0.94285	0.9669	0.9677	0.9879
5	0.9891	0.97685	0.9556	0.9576	0.9789

Table 3
Performance metrics of proposed multimodal framework on Sarcos sample.

Method	MFCC	Frames	HW	SP	ST	A	F1
LR	120	1024	120	67	91	75.5	73.62
LSTM	130	1024	75	63	74	68.30	72.88
CNN + SVM	60	148	1050	78	74	77.28	81.54
MLP + K-NN	175	2048	100	90	88	93.5	89.69
MLP+BM	200	1024	110	84	94	76.02	83.29
CNN+ Y	250	1024	140	92	90	96.57	95.30
CNN+G	78	300	100	93	90	97.35	94.99
Proposed	450	4024	250	94	95	97.57	98.90

MFCC = MFCC Features, SP = Specificity, ST = Sensitivity, A = Accuracy (%), F1 = F1 score, BM = Bayesian Model, Y = ReLu + SVM, G = SVM + LDA.

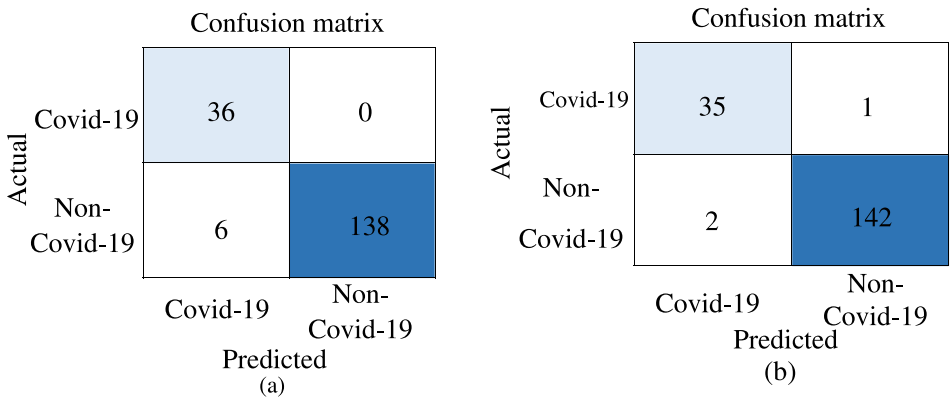

Fig. 7. Confusion Matrix features (a) chest X-ray images and (b) cough based diagnosis.

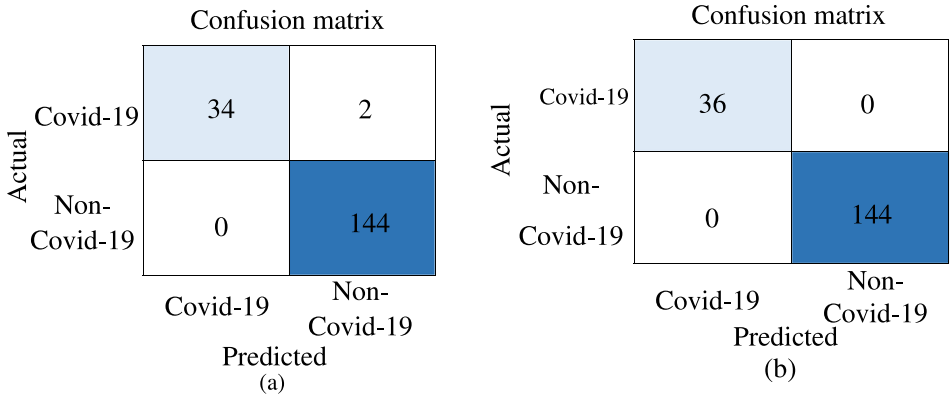

Fig. 8. Confusion Matrix for (a) chest X-ray images and (b) cough based diagnosis.

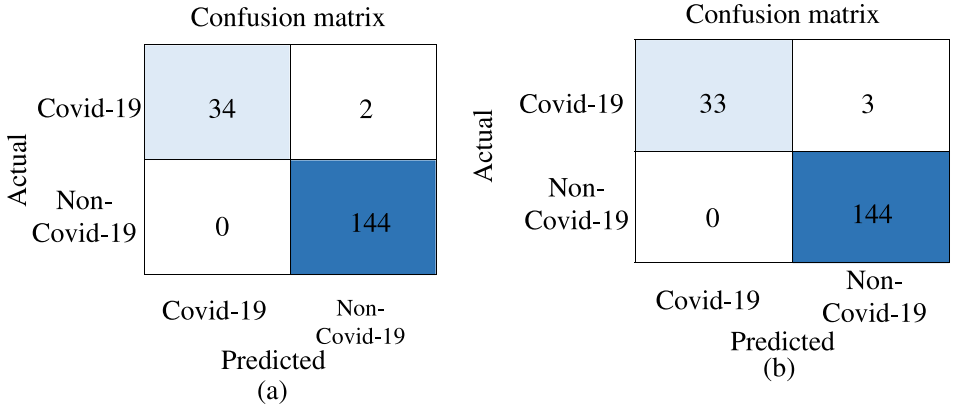

Fig. 9. (a) shows accuracy of fold 3 and (b) fold 4 for CXR classification.

of classification accuracy, obtaining 95.3% and 94.2% respectively. Therefore, the MultiLayer Perceptron (MLP) learning model performed better than other classification methods. The classification accuracy provided by MLP method is an AUC of 0.897%. To classify the cough sample database, we optimized the Logistic Regression (LR) and Support Vector Machine (SVM) classifiers. Both classifiers show substantially weaker performance, with AUCs of 73.60% and 81.50%, respectively. Based on overall observation, Table 4 shows the classification accuracy of different ML techniques for classifying cough (audio) sample databases. A more significant number of MFCCs features extracted from the cough sample consistently leads to improve the performance of the proposed method (see Fig. 7).

The spectral resolution is used to evaluate the 39-dimensional MFCCs features that surpass the human auditory system from the cough database. We conclude that the proposed multimodal framework uses cough features not generally perceivable by the human ear. The performance of the proposed framework is evaluated on chest X-ray images and its features for classifying COVID-19 and non COVID-19 using deep Darknet techniques [13]. The confusion

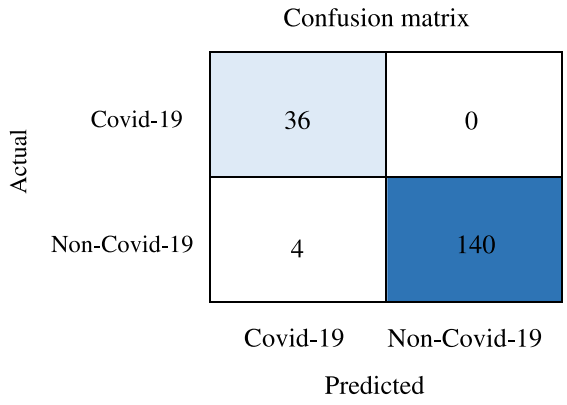

Fig. 10. Confusion matrix (fold5) for COVID-19 classification based on CXR images.

matrix of classifying COVID-19 and non-COVID-19 is depicted in Fig. 8.

3.2. Performance analysis on sarcos cough sample dataset

Table 3 illustrates performance matrices of the different ML techniques based on extracted MFCC features from the Sarcos cough (audio) sample database.

DL techniques are used to measure the performance of the proposed framework based on extracted MFCC features based on a selected number of frames and segmented cough samples from the Sarcos cough database. It includes CNN with ReLU+SVM and MLP+Bayesin methods. The framework provides classification accuracy of 96.57% and F1 measure of 95.30%, which are higher than the MLP+Bayesin method (accuracy 76.02%, F1 measure 83.29%). Moreover, we used MLP+K-NN, LR, and LSTM techniques to diagnose and accurately predict COVID-19 infection. The MLP+K-NN technique provides higher classification accuracy (93.5%), and F1-measure (89.69%) than LR (accuracy 68.30%, F1 measure 72.88%), and LSTM methods (accuracy 75.5%, F1 measure 73.62%).

The Linear Discriminant Analysis (LDA) technique measured optimal features based on within-class and between-class cough samples because the number of overlapped frames of segmented voice is similar in datasets. The F1 score of the proposed method is 98.9% accuracy which is greater than other methods. The LR technique is used for classifying cough sample classes based on extracted features of 120 MFCCs. LR techniques provide specificity, sensitivity, accuracy, and F1 measures are 67%, 91%, 94%, 75.5%, and 73.6%. Other classifiers such as LSTM, CNN+SVM techniques are used to evaluate performance while trained on the Coswara dataset and evaluated on the Sarcos and Coswara cough voice datasets. LSTM technique provides 63% specificity, 74% sensitivity, 68.3% accuracy, and 72.9% F1-measure.

To evaluate the performance of the proposed framework, we used SVM classifiers, CNN, LSTM+LDA, Resnet-50 network, and LSTM+SFS, LDA, and LSTM for classifying the chest X-ray samples for accurate prediction of COVID-19 and non-findings. Table 4 shows that the proposed method provides high accuracy (95.91%) for diagnosis of COVID-19 and non-COVID-19 cases based on subspace feature selection (SFS) method. We used the SFS technique to select a discriminatory set of features from the extracted features to train the proposed framework. These classifiers reported better accuracy in comparison to the other classifiers based on different feature selection techniques. For example, based on selected features using SFS, the LSTM+SFS method achieved 93.75% accuracy for the classification of COVID-19 patients.

Table 4
Average performance of proposed chest X-ray model of 5 fold cross validation results.

Method	Feature set	SP	ST	A	AC
CNN + SVM	120	61.95	85.90	73.02	75.50
LSTM + LDA	150	73.95	75.89	73.78	77.86
Resnet50	200	57.86	93	74.58	74.89
LSTM + SFS	250	91.75	98.80	92.45	90.75
Proposed	350	97.90	96.65	95.91	94.75

MFCC = MFCC Features, SP = Specificity, ST = Sensitivity, A = Accuracy (%), AC = Area Under Curve.

Table 5
Accuracy of the proposed chest X-ray model based on COVIDGR-1.0 dataset.

Class	Precision	Recall	F1	Accuracy	Support
COVID-19	0.967	0.989	0.971	0.9890	86
No-findings	0.972	0.969	0.966	0.975	86
Accuracy	–	–	–	0.97.7	172
Macro avg	0.9654	0.96285	0.967	0.9677	172
Weighted avg	0.9893	0.97985	0.976	0.9876	172

Table 6
Statistical significance analysis of lung areas based on measured intensity value of lung images.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.546	0.0589	n/a	n/a	n/a	n/a
TB	0.532	0.043	A	n/a	n/a	n/a
BI	0.558	0.042	n/a	C	n/a	n/a
VI	0.509	0.047	B	n/a	C	n/a
CO-19	0.504	0.051	C	B	C	n/a

BI = Bacteria infection, VI = Viral infection, TB = Tuberculosis, CO-19 = COVID-19.

3.3. Evaluation based on COVIDGR-1.0 sample dataset

Due to the high deviations between various executions, five different five-fold cross-validations are conducted in all the experimental setups. We have used 80% of the COVIDGR-1.0 chest X-ray image dataset for training, and the remaining 20% samples are used for testing. A random 10% of each training sample set is used to validate the proposed model to determine when to stop the training process. We conducted the data-augmentation techniques in each experiment, and good samples are selected carefully. All results are illustrated using the average values and the standard deviation of the 50 epoch executions. The experimental results are shown in Table 5.

3.3.1. Performance analysis on COVIDGR 1.0 dataset

We tested the pre-trained weights of the proposed framework using Cohen (also known as COVIDx) on the COVIDGR-1.0 dataset for accurate classification of COVID-19 patients. The overall accuracy of the proposed framework is 0.977%, with a total support of 172. The macro average accuracy of the proposed framework is 0.9654% (precision), and 0.96285% (recall). F1-score is 0.9669% accuracy is the 0.9677% with total support of 172. The weighted average accuracy of the proposed is 0.9893%, 0.97985%, 0.9756%, 0.9876%, and support of 172 (see Table 5).

3.4. Statistical analysis

In this work, we computed statistical measures over the chest X-ray image sample (as shown in Tables 6–8). The following standard measures of the biomarkers from chest X-rays image analysis are highlighted:

1. **Measuring the morphology structure:** Morphology structure analysis method is used to segment chest and lung area. As

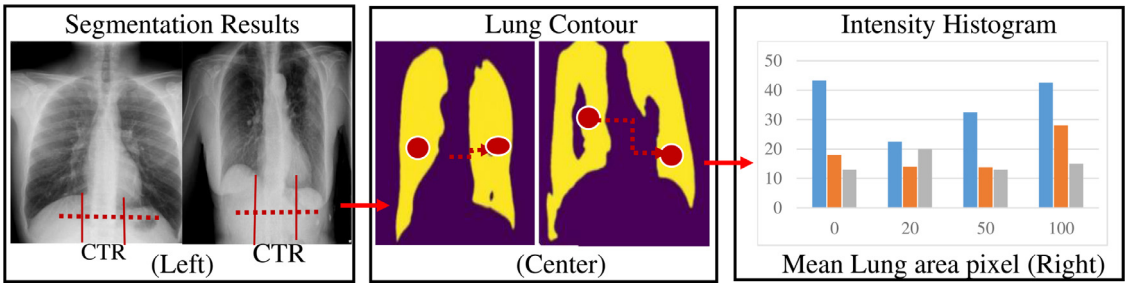

Fig. 11. (left) Lung segmentation result, (center) detection of contour in segmented images, and (right) scatter histogram plot of lung areas pixels.

Table 7
Statistical significance of lung areas based on measured mean intensity value of lung images.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.158	0.027	n/a	n/a	n/a	n/a
TB	0.135	0.017	n/a	n/a	n/a	n/a
BI	0.143	0.020	n/a	n/a	n/a	n/a
VI	0.163	0.025	C	C	B	n/a
CO-19	0.161	0.022	C	C	C	A

Table 8
Lung intensity based variance statistics.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.139	0.017	n/a	n/a	n/a	n/a
TB	0.136	0.016	n/a	n/a	n/a	n/a
BI	0.145	0.019	A	B	n/a	n/a
VI	0.165	0.022	C	C	C	n/a
CO-19	0.163	0.022	C	C	C	n/a

shown in Figs. 11(left) and 11(center), it was evaluated on different classes.

2. **Computation of Mean lung intensity:** We measured the statistical second-order (mean value, standard deviation, variance) to validate significant statistical measures of the segmented region of lung region values (see Fig. 11(right)).
3. **Standard deviation based Analysis:** We evaluated the second-order statistic measure (standard deviation) from the segmented chest X-ray images. The gray level intensity-based histogram is measured from the segmented lung pixels. The selected regions are considered from the segmented images. The black double-headed arrow is shown in Fig. 11(right).
4. **Cardiothoracic Ratio (CTR):** CTR is calculated as the ratio between the maximal transverse cardiac diameter and the maximal internal thoracic diameter annotated images; it is a widely used marker to diagnose cardiomegaly. It is suggested that if the cardiothoracic border in COVID-19 CXR [4,5] is obscured by rounding opacity or consolidation, a clear off-average CTR value can be used as an anomaly warning.
5. **Kolmogorov Smirnov test:** We performed the Kolmogorov Smirnov test [20] to measure the normal distribution of potential biomarkers candidates, because the cardiothoracic boundary becomes blurred by rounded opacities or consolidation in of the segmented chest X-ray images. We used the non normally distributed variables to perform the Wilcoxon performance measure method [21] to check the significance of the chest images using the rank testing method. The rank is computed to compare segmentation method performance with identical data size, and the rank-sum method using the Wilcoxon rank test compares COVID-19 candidates to those of other labels with different data sizes.
6. To provide significant statistical measures for computed features from segmented regions, we used the statistical significance (SS) levels for computed values which are indicated for p-test $p(A) \leq 0.05$, $p(B) \leq 0.01$, and $p(C)$ for $p \leq 0.001$ and F-measures.

Table 9
Variance statistics analysis based on inter- and intra-class segmented lung areas.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.446	0.051	n/a	n/a	n/a	n/a
TB	0.476	0.078	A	n/a	n/a	n/a
BI	0.472	0.074	n/a	n/a	n/a	n/a
VI	0.502	0.064	C	A	n/a	n/a
CO-19	0.499	0.068	C	C	A	n/a

C: infection of COVID-19.

Table 6 shows the corresponding results using chest X-rays images based on the statistical analysis method. It shows SS measures for different segmented images of the chest part of COVID-19 patients. We have computed the SS measures for the scatter plot of the histogram, as shown in Fig. 11 (centre). It depicts the major overlap regions between labelled chest image classes.

In the following, we summarize the analysis of lung areas intensity variance based on pneumonia dataset and COVID-19 chest X-ray images.Standard deviation is measured from the pixel intensity of each segmented lung area. The computed values are shown in Table 6. We concluded that variance value is higher for COVID-19 cases while viral cases have higher variance values than other labelled classes with SS ($p \leq 0.001$). To investigate the impact of scanning protocol on statistics, we analyzed by excluding AP Supine radiographs from the entire chest X-ray image dataset with documented patient information. The AP Supine protocol is used as a substituted method to standard chest Anterior Posterior (AP) or Posterior Anterior (PA) radiographs method that depends on patient condition since AP Supine protocol is not common in standard cases. We performed analysis based on obtained results from Tables 7, 8. The results are compared by different classification methods, as shown in Tables 6, 7, and 9, respectively. Based on overall observations, statistical measures method highlights differences in mean values and standard deviation values from computed histogram responses of segmented lung images. Based on these measures, the proposed framework classifies COVID-19 cases, tuberculosis (TB), bacterial infection and other viral infections.

Based on significant measures, the computed accuracy illustrates that the statistical mean and STD values of COVID-19 and viral classes show significant differences were based on highly intensity-variable characteristics in segmented lung areas. This is because we did not considered the dynamic changes in to scanning protocol. Based on overall observation, the classification between COVID-19 and other viral cases shows a significant difference. Despite statistical differences between the COVID-19 cases and other disease classes, the selected parameters (confidence intervals) are $p < 0.001$ for Normal and TB, $p < 0.05$ for Bacteria, broad overlaps between several classes.

Table 10
Weighted Sum Rule fusion method based Accuracy for classification of COVID-19 patients.

Modality	Weight	Mean accuracy (S_i)	W_{sm} score
Chest X ray	0.54	98.67	53.35
Cough audio	0.46	86.53	39.80

Table 11
Weighted sum rule fusion method based accuracy for classification of COVID-19 patients.

Model	W	A	WS
M1	0.54	98.37%	53.35
M2	0.46	90.53%	41.64
Fused Accuracy (%)	–	–	94.99

M1: Chest X ray, M2: Cough (audio), W=Weight ((W_i), A = Accuracy (S_i), WS = Weighted sum fusion score.

The method performs the classification of bacteria infection, tuberculosis, COVID-19, and other viral infection based on extracted features from lung regions. The statistical significance of extracted features is evaluated using a statistical model (as reported in Tables 10 and 11). Based on the statistical analysis of potential extracted features from chest X-ray images that signify essential changes in the pixel intensity distribution of the segmented lung images for accurate classification. The accurate distribution revealed that distinct patterns within the segmented lung area could be changed. Therefore, these features are the most effective in the early diagnosis and accurate prediction based on CXR.

3.5. Analysis of inter- and intra-class lung images

Based on extracted features from intra-class and inter-class segmented images, we computed local (shape features) and global pixel intensity distribution of chest X-ray images for accurate prediction for early diagnosis of COVID-19 patients. Our goal is to compute optimal discriminant between inter-class and intra-class of lung image database. The mean intensity for each lung imaging patch is calculated using the computed values. We computed the STD of every patch of lung pictures referred to as intra-patch intensity distribution for early identification of COVID-19 cases, and we denoted these patches as the inter-patch intensity distribution.

The distribution of inter-patch intensity of the unified COVID-19 and other viral infection class showed that the lower intensity values ($p \leq 0.001$ for all) of other classes are highly intensity-variant characteristics and showed it as a large error bar. The accuracy is measured based on selected different Regions of Interest (RoI) of lung images and computed histogram responses of each ROI for analysis of inter-patch and intra-patch lung images. In analyzing intra-patch intensity-based pixel distribution, we have concluded that there are no differences compared to the normal class ($p \geq 0.05$). Based on overall observations, the intra-class image patches based on pixel variance represent local texture information for classifying chest X-rays.

We computed the global features and multi-focal intensity from the segmented images, and the change can be discriminating features for COVID-19 diagnosis. It is statistically significant with a strong correlation to chest X image analysis, as shown in Table 7.

3.5.1. Theoretical and analytical model: weighted sum-rule fusion method

In this section we devise a theoretical and analytical model for early diagnosis and accurate classification of COVID-19 cases based on the weighted sum rule fusion method. To the best of our knowledge, there are no available fusion models and techniques based on chest X-rays images and cough (audio) data fusion for early diagnosis of COVID-19 cases in the literature.

The proposed multimodal framework performs data fusion from the chest X-rays and cough datasets obtained from different subjects samples.

The fusion-based accuracy is calculated as average accuracy of the two models i.e., chest X-ray based detection based model (M_1) and Cough-based detection model (M_2) using the weighted sum-rule method for both models. The weighted sum-rule-based fusion method is used to give more weight to the model, which is less susceptible to errors. Referring to the same analogy, we calculated the mean deviation value on the 5-fold cross-validation accuracy for both X-ray and cough sound models, as mean deviation directly corresponds to the system consistency. Therefore, we observe that the mean deviation (Md) value gives an idea of the method error rate.

For calculating the weights, we have used the confusion matrix of each model (see Tables 9 and 11).

$$Md = \frac{1}{n}\Sigma_{n=1}^{i}|x_i - S| \tag{17}$$

The similarity measures S_1 and S_2 are calculated as:

$$S_1 = \sum_{i=1}^{5}\frac{chestX - rayaccuracy(A_i)}{5} \tag{18}$$

$$S_2 = \sum_{i=1}^{5}\frac{Cough(A_i)}{5} \tag{19}$$

Also, the computation of weights for each models are illustrated as follows:

$$W_2 = \frac{Md_1 \times Md_2}{sum(Md_1 + Md_2)} = 0.46 \tag{20}$$

$$W_1 = 1 - W_2 = 0.54 \tag{21}$$

W_2 and W_1 are assigned weight for the model.

Finally, the multimodal framework fusion-based classification accuracy is computed as follows:

$$F_s = sum(W_{sm}Score) = 93.1538\% \tag{22}$$

3.6. Analysis of ensemble learning classification models

The primary objective to use the customized ensemble learning based classification models is to improve overall performance of the multimodal framework over individual cough and chest X diagnostic model. The multimodal framework integrates the ensemble learning based classification models to mitigate the spread or dispersion of the predictions and model performance for early diagnosis and accurate prediction of COVID-19.

We used a two-level stacking-based ensemble method consisting of the first base and second levels. The base-level constructs from ensemble methods include decision trees, random forests, logistic regression, boosted and bagged learning models using majority voting method. The predictions of the base level ensemble models are fed into the second level as inputs of feature sets for accurate prediction of COVID-19 cases based on the chest X-ray images. Therefore, ensemble classifier with 350 trees was used to classify participants positively and negatively to COVID-19 cases. Alternatively, the bagging model is configured with 350 decision trees with a maximum depth equal to 4 (see in Fig. 12). Finally, AdaBoost.M1 algorithm is used for adaptive boosting with 350 decision stumps, i.e., one-level trees with two leaf nodes. The second level in the ERLX model is an extreme gradient boosting (XGBoost) classifier. XGBoost is an ensemble algorithm based on decision trees and a gradient boosting framework.

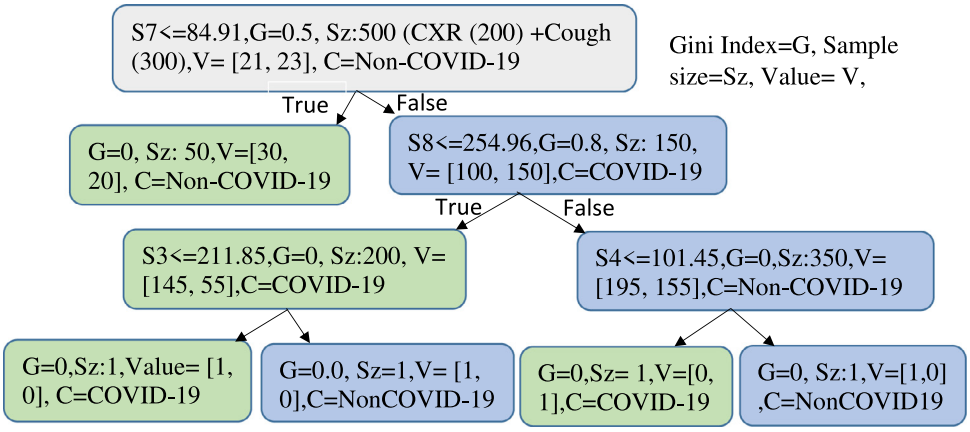

Fig. 12. RF classifier model.

Table 12
Comparison of multimodal system with other techniques.

Ref.	Dataset	Technique	Accuracy
[3]	CO	ML	66.74%
[14]	CXR	DTL	92.10%
[16]	CC	ED	77.10%
[22]	COV	S-Dnet	53.35%
[24]	COV	ML	68.64%
Prop.	CXR+CC	DL	94.99%

COV: COVIDGR-1.0, CXR: Chest X ray database, CC: cough sample, CO: Coswara cough samples, ML: Machine Learning techniques.

Table 13
Comparison of the performance of proposed system.

Ref.	Dataset	Technique	Accuracy
[28]	COVIDx 1.0	COD	92.60%
[14]	COVIDx 1.0	COP	93.70%
[13]	COVIDx 2.0	DK	87.02%
[22]	COVIDx 2.0	VG	91.35%
[24]	COV	CS	$76.18 \pm 2.70\%$
Prop.	CXR + CC	DL	94.99%

COD = COVIDx 1.0 + COVIDNet, COP = COVIDx 1.0 + COVID-CAPS (a capsule network-based model), CT = Chest X-ray8 dataset, Dk = DarkCovidNet + 5FVC Class: No-finding, COVID-19, Pneumonia), COV = COVIDGR-1.0, CS = COVID-SDNet, VG = VGG-19 + DK-161.

4. Discussion

We have developed a multimodal framework for early diagnosis of CoVID-19 patients based on extracted features from chest-X ray image and cough(audio) sample databases by applying DL and ensemble classification techniques. Hence, we have illustrated the contribution of various levels of predictors including statistical ML techniques, DL techniques and fusion techniques for early diagnosis. The proposed framework provided evaluation on chest-X ray based model and cough based model on individual chest X-rays and cough datasets. To improve the overall performance of the proposed framework, both chest-X ray based and cough based model are integrated to fuse the individual accuracy based on extracted features from each datasets using weighted sum rule fusion method. The overall accuracy is reported as 94.99% for early classification of COVID-19 cases. The existing ML techniques showed an adequate predictive ability with confusion matrices values for classification (shown in Tables 3–5). Apart from the evaluation on existing methods, we evaluated the performances and significance of proposed multimodal framework on different DL classification models.

We have computed results based on experimental setup on various combinations of public datasets and benchmark settings (shown in Table 13). The author [28] showed that COVIDNet provided accuracy of 92.60%, with 9.0% sensitivity in normal class, 90% in Non-findings and 87.10% in COVID-19. In [14], COVID-CAPS model achieved an accuracy of 93.70%, sensitivity of 90%, and specificity of 95.80%. These results look too remarkable compared to expert radiologist sensitivity, 69% [24]. This can be explained because the used dataset is biased to severe COVID-19 cases [22] due to statistically unreliable data and several authors evaluated results without cross-validation testing. Our proposed framework is better than these models [14,28] based on different evaluation matrices Table 12) and Table 13, respectively.

Finally, we compared the performance of the proposed multimodal framework with existing methods for early diagnosis and accurate prediction of COVID-19. To the best of our knowledge, most systems use CXR image used to evaluate the detection of COVID-19 [22]. This makes a direct comparison between the proposed system and the existing ones unfair. To make the comparison fair, we only compare the proposed system with the existing work that used CXR and cough samples to predict COVID-19 accurately (shown in Table 12).

5. Conclusions

In this work, a novel multimodal framework is proposed to predict COVID-19 patients accurately. The framework extracts feature from chest X-ray images and cough audio databases using DL techniques. The MFCC features are extracted from the Sarcos cough (audio) and Coswara sample database and classified by logistic regression, LSTM, CNN + SVM, and MLP techniques. The framework used U-Net and Darknet architectures to extract the chest X-ray database features [13]. DL techniques are used to measure the performance of the proposed framework, including CNN with ReLU + SVM and MLP + Bayesin methods. The frameworks provides classification accuracy of 96.57% and F1 measure of 95.30%, which are higher than the MLP + Bayesin method (accuracy 76.02%, F1-measure 83.29%). Moreover, we used MLP + K-NN, LR, and LSTM techniques to diagnose and accurately predict COVID-19 infection. The MLP + K-NN technique provides higher classification accuracy (93.5%), and F1-measure (89.69%) than LR (accuracy 68.30%, F1-measure 72.88%), and LSTM methods (accuracy 75.5%, F1 measure 73.62%).

We used the weighted sum rule fusion-based method for early diagnosis of COVID-19 patients with the accuracy of 94.99%. The

performance of the proposed framework is evaluated on discriminatory features using CNN, LSTM using LDA, Resnet-50 network, and LSTM + SFS. The overall accuracy of the framework is 98.90% which is higher than other methods based on Sarcos cough samples.

The performance proposed model will improve based on different database using deep multimodal fusion techniques. We will also develop intelligent devices for the early diagnosis of noncommunicable diseases in rural and remote areas worldwide.

Ethical approval

The authors declare that no ethical approval was required for this study.

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

Acknowledgements

Funding: This work was supported by the Science Foundation Ireland, co-funded by the European Regional Development Fund under Grant no. SFI/12/RC/2289_P2 and by the Italian MIUR, PRIN 2017 Project Fluidware" (CUP H24I17000070001) and PRIN 2020 Project "COMMON-WEARS". The work is also carried out within the frame of the SPINE Body of Knowledge (SPINE-BoK, https://projects.dimes.unical.it/spine-bok).

References

[1] F. Hu, H. Mingfang, J. Sun, X. Zhang, J. Liu, An analysis model of diagnosis and treatment for COVID-19 pandemic based on medical information fusion, Inf. Fusion 73 (2021) 11–21.
[2] D. Now, https://detect-now.org/, last accessed on jan, 2022,
[3] N. Sharma, K. Prashant, K. Rohit, R. Shreyas, S.R. Chetupalli, P.K. Ghosh, S. Ganapathy, Coswara–a database of breathing, cough, and voice sounds for COVID-19 diagnosis, 2020, arXiv preprint arXiv:2005.10548.
[4] WHO, Coronavirus disease 2019 (COVID-19) Situation Report - 35, WHO(2020).
[5] Y. Pathak, P.K. Shukla, K.V. Arya, Deep bidirectional classification model for COVID-19 disease infected patients, IEEE/ACM Trans. Comput. Biol. Bioinform. 18 (4) (2021) 1234–1241.
[6] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, M. Kaur, Classification of the COVID-19 infected patients using densenet201 based deep transfer learning, J. Biomol. Struct. Dyn. 39 (15) (2020) 5682–5689.
[7] D. Toppenberg-Pejcic, J. Noyes, T. Allen, N. Alexander, M. Vanderford, G. Gamhewage, Emergency risk communication: lessons learned from a rapid review of recent gray literature on Ebola, Zika, and yellow fever, Health Commun. 34 (4) (2018) 437–455.
[8] WHO, Advice for public, WHO Int., 2020.
[9] S.H. Alsamhi, B. Lee, M. Guizani, N. Kumar, Y. Qiao, X. Liu, Blockchain for decentralized multi-drone to combat COVID-19, 2021, arXiv:2102.00969.
[10] A. Windmon, M. Minakshi, P. Bharti, S. Chellappan, M. Johansson, B.A. Jenkins, P.R. Athilingam, Tussiswatch: a smart-phone system to identify cough episodes as early symptoms of chronic obstructive pulmonary disease and congestive heart failure, IEEE J. Biomed. Health Inform. 23 (4) (2018) 1566–1573.
[11] S.H. Alsamhi, B. Lee, M. Guizani, N. Kumar, Y. Qiao, X. Liu, Blockchain for decentralized multi-drone to combat COVID-19 and future pandemics: framework and proposed solutions, Trans. Emerg. Telecommun. Technol. (2021) 4255.
[12] S.H. Alsamhi, B. Lee, Blockchain-empowered multi-robot collaboration to fight COVID-19 and future pandemics, IEEE Access 9 (2020) 44173–44197.
[13] I. Afyouni, Z. Al, R. Aghbari, A. Razack, Multi-feature, multi-modal, and multi--source social event detection: a comprehensive survey, Inf. Fusion 79 (2022) 279–308.
[14] A.T. Porter, A Path-specific Approach to SEIR Modeling, University of Iowa, 2012 Ph.D. Thesis.
[15] Coronavirus Map, John Hopkins University, 17 March, 2020,
[16] J.T. Wu, K. Leung, G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, Lancet 395 (10225) (2020) 689–697.
[17] A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 228–233.
[18] L. Zhan, Y. Liu, A clarke transformation-based DFT phasor and frequency algorithm for wide frequency range, IEEE Trans. Smart Grid 9 (2016) 67–77.
[19] C.D. Vente, Automated COVID-19 grading with convolutional neural networks in computed tomography scans: a systematic comparison, IEEE Trans. Artif. Intell. 3 (2) (2021) 129–138.
[20] J. Wang, L. Gu, L. Yang, Oracle-efficient estimation for functional data error distribution with simultaneous confidence band, Comput. Stat. Data Anal. 167 (2022).
[21] Y. Oh, S. Park, J.C. Ye, Deep learning COVID-19 features on CXR using limited training data sets, IEEE Trans. Med. Imaging 39 (8) (2020) 2688–2700.
[22] S. Tabik, A. Gez-Ros, J. L. Martn-Rodrguez, I. Sevillano-Garca, M. Rey-Area, D. Charte, E. Guirado, J.L. Surez, J. Luengo, M.A. Valero-Gonzlez, P. Garca-Villanova, E. Olmedo-Snchez, F. Herrera, COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images, IEEE J. Biomed. Health Inform. 24 (12) (2020) 3595–3605.
[23] J.P. Cohen, P. Morrison, L. Dao, COVID-19 image data collection, arXiv Prepr. arXiv:2003.11597 2020.
[24] S. Tabik, A. Gmez-Ros, J.L. Martn-Rodrguez, I. Sevillano-Garca, M. Rey-Area, D. Charte, E. Guirado, et al., COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images, IEEE J. Biomed. Health Inform. 24 (12) (2020) 3595–3605.
[25] L. Garg, E. Chukwu, N. Nasser, C. Chakraborty, G. Garg, Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model, IEEE Access 8 (2020) 159402–159414.
[26] L. Garg, C. Chakraborty, S. Mahmoudi, V.S. Sohmen, Healthcare Informatics for Fighting COVID-19 and Future Epidemics, Springer International Publishing, 2022.
[27] A. Anand, A.K. Singh, Dual watermarking for security of COVID-19 patient record, IEEE Trans. Dependable Secure Comput. (2022) 1–9, doi:10.1109/TDSC.2022.3144657.
[28] L. Wang, Z.Q. Lin, A. Wong, Covid-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images, Sci. Rep. 10 (1) (2020) 1–12.
[29] R. Sahal, S.H. Alsamhi, K.N. Brown, D. O'Shea, B. Alouffi, Blockchain-based digital twins collaboration for smart pandemic alerting: decentralized COVID-19 pandemic alerting use case, Computational Intelligence and Neuroscience 2022 (2022) 1–14 Hindawi.
[30] A. Aggarwal, et al., COVID-19 risk prediction for diabetic patients using fuzzy inference system and machine learning approaches, J. Healthc. Eng. 2022 (2022) 1–10.