

Please cite the Published Version

Kumar, A, Kumar, A, Bashir, AK, Rashid, M, Ambeth Kumar, VD and Kharel, R (2022) Distance based pattern driven mining for outlier detection in high dimensional big dataset. ACM Transactions on Management Information Systems, 13 (1). p. 8. ISSN 2158-656X

DOI: <https://doi.org/10.1145/3469891>

Publisher: Association for Computing Machinery

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/630983/>

Additional Information: © ACM 2021. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in ACM Transactions on Management Information Systems, <http://dx.doi.org/10.1145/3469891>

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Chest X Ray and Cough Sample based Deep Learning Framework for Accurate Diagnosis of COVID-19

Santosh Kumar^a, Rishab Nagar^a, Saumya Bhatnagar^a, Ramesh Vaddi^b, Sachin Kumar Gupta^c, Mamoon Rashid^{d,e*}, Ali Kashif Bashir^f, Tamim Alkhalifah^g

^aDepartment of Computer Science and Engineering, International Institute of Information Technology-Naya Raipur, Raipur, Chhattisgarh, India-493661,

^bDepartment of Electronics and Communication Engineering, School of Engineering and Applied Sciences, SRM University, Amaravati, Guntur, AP, India-522240,

^cSchool of Electronics and Communication Engineering, Shri Mata Vaishno Devi University, Katra, India,

^dDepartment of Computer Engineering, Faculty of Science and Technology, Vishwakarma University, Pune, India,

^eVishwakarma University Research Center of Excellence for Health Informatics, Pune, India,

^fDepartment of Computing and Mathematics, Manchester Metropolitan University, Manchester, UK,

^gDepartment of computer science, College of Science and Arts in Ar Rass, Qassim University, Saudi Arabia,

Abstract

All witnessed the terrible effects of the COVID-19 pandemic on the health and work lives of the population across the world. It is hard to diagnose all infected people in real time since the conventional medical diagnosis of COVID-19 patients takes a couple of days for accurate diagnosis results. In this paper, a novel learning framework is proposed for the early diagnosis of COVID-19 patients using hybrid deep fusion learning models. The proposed framework performs early classification of patients based on collected samples of chest X-ray images and Coswara cough (sound) samples of possibly infected people. The captured cough samples are pre-processed using speech signal processing techniques and Mel frequency cepstral coefficient features are extracted using deep convolutional neural networks. Finally, the proposed system fuses extracted features to provide 98.70% and 82.7% based on Chest-X ray images and cough (audio) samples for early diagnosis using the weighted sum-rule fusion method.

Keywords: Multimodal, COVID-19, Deep learning, Chest X-Ray, Cough-Breathing sounds, Segmentation

1. Introduction

The CoVID-19 pandemic had unprecedented global social and economic impact cite1. With more than 50 million infected patient cases and more than 500,000 deaths as of mid-July 2020, the pandemic continues to spread globally with coronavirus new strain without signs of abating. It is still increasing the number of infected people worldwide. Disruptive technologies and

Email address: santosh@iiitnr.edu.in, rishabh18101@iiitnr.edu.in, saumya18100@iiitnr.edu.in, ramesh.v@srmmap.edu.in, sachin.gupta@smvdu.ac.in, mamoon873@gmail.com, dr.alikashif.b@ieee.org, tkhliefh@qu.edu.sa (Santosh Kumar^a, Rishab Nagar^a, Saumya Bhatnagar^a, Ramesh Vaddi^b, Sachin Kumar Gupta^c, Mamoon Rashid^{d,e*}, Ali Kashif Bashir^f, Tamim Alkhalifah^g)

Preprint submitted to Computers & Electrical Engineering

September 5, 2022

emerging computing algorithms are being used by various organizations and public union states to find efficient solutions for early patient diagnosis[1].

Several interdisciplinary researchers and clinical testing organizations have deployed diagnosis models for patient tracing, testing, and treatment strategies that have successfully crushed the pandemic situation and facilitated medical services around the world [2]. These clinical testing methods include polymerase Chain Reaction (PCR) and other procedures that rapidly make millions to billions of copies of specific Deoxyribonucleic acid (DNA) samples. It is widely used to identify COVID-19 patients through the use of biological Macromolecule Ribonucleic Acid (m-RNA), a nasopharyngeal, oropharyngeal swab of infected people [3].

By detecting unique sequences of virus RNA, the real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) method is also used to confirm COVID-19 infected cases. While the (rRT-PCR) testing system is the current gold standard cite4. It is insufficient for controlling the pandemic for several reasons, including, but not limited to, (1) limited availability of COVID-19 testing due to huge geographical and temporal factors, (2) the demand for in-person visits to a hospital, clinical lab, or mobile lab center, (3) The real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) method is also used to confirm COVID-19 infected cases by detecting unique sequences of virus RNAg.

To address these issues, several researchers [4] [5] proposed a model based on the possibility of using widespread and simple chest X-ray (CXR) imaging for early diagnosis and screening of COVID-19 patients, attracting considerable interest from both clinical procedures and AI-based early diagnosis techniques. Several researchers cite9-cite12 proposed and tested frameworks for screening patients based on various statistical modeling techniques such as epidemiological-based methods and real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) method is also used to confirm COVID-19 infected cases by detecting unique sequences of virus RNA [6].

The unimodal model processes accumulated sample data about susceptible, infected, recovered, and deceased populations to provide significant statistical measures based on a patient database for early classification (e.g., positive, negative, and inclusive) of COVID-19-infected patients. Due to the proper availability of the annotated COVID-19 database, these models require only statistical measures for further data processing and produce inaccurate results. Furthermore, the performance of these frameworks and models is poor, which is hampered by the lack of properly labeled datasets for accurate classification of the COVID-19 case. It is a significant difficulty for early classification. As a result, statistical learning-based unimodal frameworks cannot produce meaningful results for early diagnosis.Hence, there is a need to design and develop multimodal learning frameworks to process the massive number of databases for the early diagnosis of infected patients using deep multimodal techniques.[4].

Deep multimodal learning frameworks are pulling the eyes of many researchers nowadays. These techniques help us provide better solutions to fight this virus within-population screening, medical help, notification, and infection control suggestions [7]. These emerging techniques improve the planning, treatment, and reported outcomes of this pandemic. The deep multimodal learning frameworks are highly used to represent better extracted discriminatory feature sets from multiple heterogeneous databases of COVID-19 patients. The multimodal framework processes the captured data through different deep learning techniques and extracts features from each separate dataset. The measured scores from unimodal datasets can be integrated using deep multimodal fusion methods for accurate classification. The multimodal frameworks provide the predict mortality risk by adequately analyzing these collected datasets of the patients [8].

1.1. Motivation

The second wave of the COVID-19 pandemic has resulted in an increase require expensive kits including testing, that are not always easily obtainable, working to develop or undeveloped countries are also still working hard to improve their detection performance [8]. As a consequence, low-cost, easily available, and dependable pre-screening tests are critical for identifying and diagnosing COVID-19 and limiting local emergence and spread of infection [9]. To provide solutions for early diagnosis of COVID-19 cases, numerous researchers have contributed to significantly alleviating early diagnostic procedures for COVID-19 patients using deep learning and machine learning techniques [10] [11].

In this paper, we address the following issue: *How can COVID-19 patients be diagnosed early?* We proposed a novel multimodal learning framework based on chest X-ray images, cough (sound) samples, and facial expressions using deep learning techniques to solve this problem. The framework contains a diagnostic model based on chest X-rays and a model based on cough for early COVID-19 detection. In addition to these models, we used the fatigue detection model to detect patient fatigue based on facial expressions captured in video. The model uses machine learning techniques to detect the patient's fatigue level after classifying the patient's facial emotions. Cough (audio) samples are known to behave energetically in lower frequencies for statistical analysis of cough samples to predict various respiratory diseases. As a result, COVID-19 analysis based on cough (audio) sample plays the important role in addressing COVID-19 problems and further investigates the distinctness of path morphological alterations in the human respiratory system induced by post-COVID-19 infections compared to other earlier COVID-19 and other respiratory infections [20]-[22]. As a result, we used cough samples to make an early diagnosis of COVID-19. The proposed framework uses speech signal processing-based techniques to remove noises from cough (audio) samples collected from COVID-19 patients. The proposed cough model computes the Mel spectrogram of the processed cough samples for better representation and analysis of cough samples. The discriminatory features are extracted from the spectrograms and classified using a convolutional neural network (CNN) model.

The multimodal framework integrates the chest -X-ray-based diagnostic model to perform early diagnosis of COVID-19 based on captured chest computed tomography (CT) images using the deep learning-based U-NET learning model. The CT scan-based diagnostic has been recognized as an informative tool for diagnosing critical diseases. The chest X-ray model performs the extractions of multiple types of features (such as radiomic features and handcrafted features) for the classification of COVID-19 and Non-COVID-19 using deep learning techniques.

1.2. Contribution

The major contributions of this work are as follows:

1. We proposed a novel multimodal framework that consists of a chest-X-ray imaging model, and cough (voice) sample model for early diagnosis of COVID-19 using deep multi-modal fusion and weighted sum rule fusion techniques.
2. In the proposed framework, a chest X-ray-based model is used to extract texture and holistic features from chest X-ray image datasets using deep U-Net-based learning techniques and the Darknet model to segment the images for better representation of extracted features using structured latent representation learning techniques. It provides robustness, generalization, and stability to the proposed framework.
3. To improve the performance of the proposed framework, a cough diagnostic model is proposed to extract the features from the cough samples of infected patients, because the

cough (audio) sample of patients has distinct latent features from distinct respiratory syndromes. The extracted discriminant Mel frequency Cepstral Coefficients (MFCCs) features of cough samples are extracted by speech signal processing techniques and convolutional neural network to train a cough diagnostic learning model for performing the preliminary diagnosis solely based on the cough to differentiate COVID-19 cough from non-COVID-19 cough sample.

4. The proposed framework integrates (fuse) the extracted features from chest-X-ray images, cough (voice) samples, and facial expression images using the weighted sum-rule fusion technique to perform accurate classification for early diagnosis of patients.
5. The experimental results of the existing handcrafted texture feature descriptor techniques and holistic feature extraction along with the representation methods are used to evaluate the performance analysis of the proposed framework for patient diagnosis.

The remaining part of the paper is organized as follows: Section 2 provides a literature review of the chest and cough-based methods for early diagnosis of patients. Section 3 shows a proposed framework for early classification using deep learning methods and the weighted sum rule fusion method. The experimental result and analysis are illustrated in Section4. The comparative analysis of the proposed framework is given with the existing method for cough and chest X-ray-based methods. Finally, the conclusion and future directions are given in section5.

Table 1: Existing works for COVID-19 detection using machine learning techniques based on chest-X ray images and cough(audio) samples.

Study	Year	R	RSS	NS	NR	PS	TM	Per(%)
[3]	2020	CXR	-	247	DL	FC	-	92.5
[4]	2019	CXR	125	200	DNET	ML	SP	98%
[5]	2020	X-ray	230	DL	CON	SP	SP	90%
[6]	2022	B	BB	220	MFCCS	DL	Low-D	70%
[7]	2022	Sound	200	Chest	HSR	SFs	CNN	70%
[15]	2020	cough	4352	SP	MFCC	DL	CNN	80%
[16]	2020	Web	Cough	COD	5320	DL	R2	97.10%
[19]	2020	X-Ray	30	R1	CNN	DL	SP	95.4%
[21]	2021	XCXR	-	200	CNN	CNET	SP	92.4
[23]	2020	SP	Cough	CO	3621	ST	R-18	AUC(0.72)
[24]	2021	ES	B	COV10	10	2DFT	Inv3	80%

Abbreviation: FC=(FC)-DenseNet103, CON=COVIDX-N, R1=ResNet50, S=SFT

SVM=Support vector machine, SP= Speech processing, T= dataset consists of 4352 unique people collected from the web app + 2261 unique people from the Android app+4352 and 5634 samples. Cough=Crowd sourced Respiratory Sound Data, SF=Shape chest Features, 2DFT = Two-dimensional (2D) Fourier transformation, COVID10= [5 COVID19 5 healthy 10], ES=Electronic stethoscope, BB=120 COVID+100 healthy) B=Breathing, Sp= Smartphone app, TM= Trained Model, Per.= Performance(%) R=Recordings Source, RSS =Respiratory Sound, NS=Number of Subjects, NR=Number of Recordings, PS=Pre-processing Steps, ST=Short-term magnitude spectrogram, CO = COVID-19 1620 healthy, Mel=Mel-frequency cepstral coefficients (MFCC), COD=2660 COVID-19 2660 healthy, CODD=114 COVID-19, 1388 healthy, MFCC= Spectrogram Mel spectrum, power spectrum Tona, Mel-frequency cepstral coefficients (MFCC), ENCNN=Ensemble CNN, Total= (COVID-19 34516 samples+ 2100(1500 (chest X-ray positive+600(Chest X-ray: Negative)

2. Literature Work

In this section, the literature work is divided into different subsections describing each part of the system's literature survey separately. The summary of the existing work for early detection of COVID-19 based on chest X-ray images and cough samples is shown in Table 1.

2.1. Chest-X-Ray Modality Based System

Various works done for COVID-19 detection using chest X-Ray or similar inputs are mentioned in this subsection. Linda et al. [1] proposed COVID-net for prediction of the covid case and reported 93.3% accuracy. Soares et al. [2] proposed a system using three learning architectures Xception, ResNet, and VGG-16 for classification of the infected patients based on Chest-X-ray images. These three models are used as the base and optimize only the fully connected layers. The accuracies claimed for Xception is 95.90%, for Resnet is 94.60%, and for VGG-16, the accuracy is 97.30%. Nevertheless, there is a significant overfitting issue in Exception and Resnet while low overfitting in VGG-16. Yujin et al [4] have also used a backup algorithm of lung segmentation to get better results with 88.90% with segmentation and 79.80% without segmentation using the FC-Dense Net learning model.

Tulin et al. [4] have made the classifier without transfer learning and getting good accuracy without over-fitting. They have implemented two kinds of classifiers binary for COVID-19 and no-findings and tertiary for COVID-19, Pneumonia, and no-findings. However, they have only one algorithm means they are directly putting the chest x-ray images into the classifier if there was a different algorithm like segmentation that can make the results more trustworthy.

2.2. Cough based diagnostic Model

In this subsection, cough-based COVID-19 diagnosis has been reported in the literature. Imran et al. [7] collected cough samples from COVID-19, bronchitis, and pertussis patients along with healthy individuals and created an Artificial Intelligent (AI) engine/system.

In a similar direction, Laguarta et al. [8] collected variable length, cough audio recordings with bit-rate 16KBs. It consists of 2660 COVID-19 positive samples with a 1:10 ratio of positive to control subjects. Each sample is split into 6seconds audio chunks and processed with an MFCCs feature extraction technique. The proposed model consists of 4 biomarkers, namely: (1) changes in vocal cords, (2) changes in lungs and (3) respiratory tracts, muscular degradation, and (4) changes in sentiment/mood able to classify with an accuracy of 97.10% [20]-[25]. Based on overall observation, we conclude that the availability of the open source datasets of chest X-ray images is not efficient to perform the early diagnosis and accurate prediction of COVID-19 patients using machine learning techniques. The available open-source cough and chest X-ray datasets are limited and not available in proper annotated with appropriately labeled datasets for training various ML-based models for fast diagnosis and accurate prediction of COVID-19. For achieving a better analysis of different subject studies, there is a need to acquire chest-X ray images as well as cough sample datasets of the same subjects (patients) for performing early diagnosis of COVID-19 and post-COVID-19 symptoms using deep learning techniques.

3. Proposed Framework

In this section, the work of the proposed multimodal framework is illustrated. The framework takes input data from (1) chest-X-ray images, (2) cough the (voice/speech) samples, and (3) facial expression (fatigue-based detection model) of infected users (shown in Fig. 1).

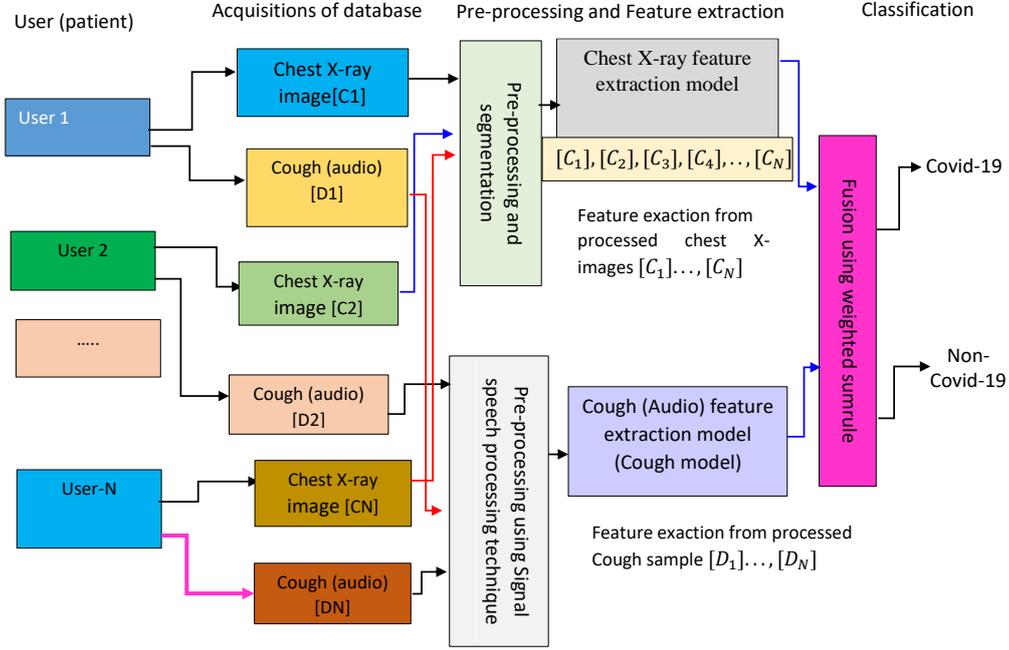


Figure 1: Illustrates working of the proposed multimodal framework for the diagnosis of COVID-19.

Chest-X-ray images are scanned by X-rays (Optima IGS 330). The facial expression of the patient is captured through videos by the surveillance system. The system collects cough (speech) data using the microphone of the mobile phone of the users. While performing alphabet pronunciation, the user also speaks out the alphabet/vowels, which are being selected. It attempts to build a diagnostic tool for COVID-19 based on respiratory, cough, and speech sounds. The participants need to record breathing sounds, cough sounds, sustained sound of vowel sounds, and a counting exercise that takes around 5-7 minutes. No personally identifiable data is collected from the participants. After that, the system performs the pre-processing and feature extraction from different models for accurate prediction of COVID-19. The extracted features are fused using the weighted sum-rule fusion method. A brief description of the model is given in the next subsections.

Table 2: Database of chest X-ray images.

Dataset	Size	Used model
A	468	Used to make COVID-19 class data
B	5,860	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 [12], 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

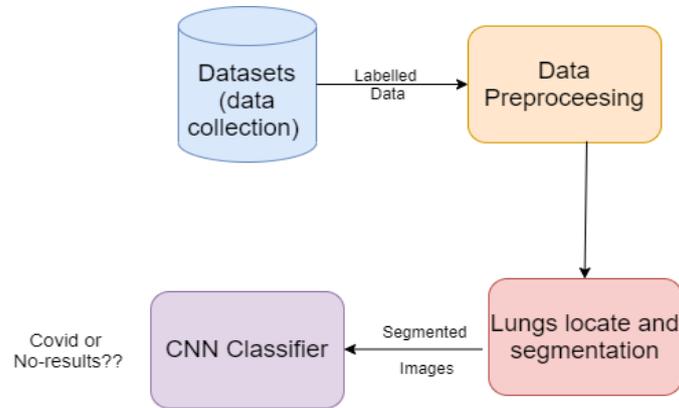


Figure 2: Illustrates working of pipeline chest X-ray imaging-based model.

3.1. Chest X-Ray Based COVID-19 Detection Model

The chest X-ray image database is collected from different sources for extracting rich and distinct information for early classification. The primary objective is to perform early diagnosis for COVID-19 patient and community-acquired pneumonia for characterizing the relationships between multiple types of discriminatory features from captured chest X-ray images and these diseases, which caters to a possible pipeline for early diagnosis of COVID-19 using deep learning techniques [9]. In the proposed multimodal system for COVID-19 detection, a deep learning model is used to classify infected and on-infected persons based on the collected database.

The Chest-X-ray-based working model consists of several steps: (1) database description and pre-processing, (2) segmentation of images, (3) feature extraction and representation, and (4) classification (shown in Fig. 2). The chest X-ray classification model classifies given input chest X-ray images into two classes: COVID-19 and non-COVID-19, it is a binary classification task which is reported the final result comes as COVID-19 or non-COVID-19. It is found that as increase the number of classes, the system may get less accurate. have divided the whole work of the chest X-ray model into several distinct patches using U-net based segmentation technique. The working of the model is shown in Fig. 3. Table 2 shows the different used chest X-ray image databases for the classification of COVID-19 using deep learning techniques.

3.2. Data Collection

We have collected the COVID-19 chest X-ray images and cough samples from open datasets [10]. Specifically, details on the X-ray images database are summarized in Table 2. The major significance between the prepared and open cough sample datasets is given as follows:

Various datasets are used which are available at open-source platforms from GitHub and Kaggle. For the COVID-19 detection using chest X-ray images, the data sets are mentioned in Table 2.

1. All four data sets have different usage in the model. The first chest x-ray image data set is IEEE-8023 which consists of 180 chest x-ray images of COVID-19 positive X-rays which are used to make the data for COVID-19 class.
2. Pneumonia dataset from Kaggle: It contains 5863 CXR images of pneumonia and normal patients used to make data for the non-COVID-19 class.

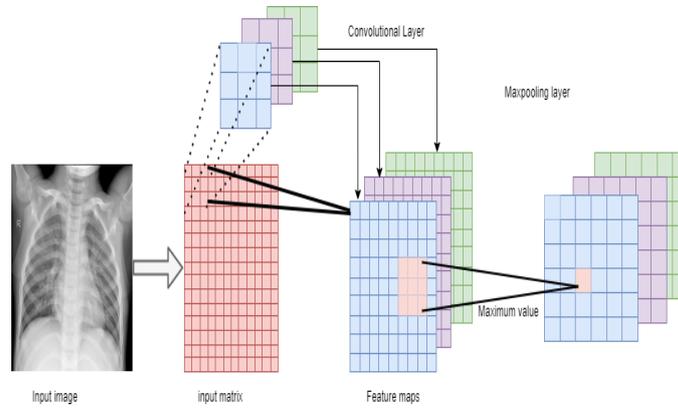


Figure 3: shows segmentation model based on U-net learning architecture

3. Shenzhen dataset and Montgomery dataset: have two image datasets, namely (1) Shenzhen dataset and (2) Montgomery dataset. Both datasets are used for obtaining information about lung segmentation due to lung mask availability for these datasets. Shenzhen dataset is used to train the segmentation model, and Montgomery is used to validating the lung segmentation model. There are 180 X-ray images involved in this research study. For COVID-19 positive, a total of 180 Chest-X-images are taken from GitHub. For Non-COVID class, the normal, and pneumonia chest X-ray images are considered from Kaggle. These two datasets are used to augment the classification of COVID-19 and non-COVID patients.
4. After data collection, the database is pre-processed using image processing techniques to remove noises and other artifacts from images before providing it as input to the Chest-X-ray-based classification model. As mentioned in Fig. 2, the next step is lung segmentation using deep U-Net learning architecture to find the region of interest for extracting features to train the proposed model.

3.3. Segmentation of Chest-X-ray Image

The main objective of the segmentation process taken into the collected data is to archive proper extraction of discriminatory information from these chest X-rays database that plays an essential role in the classification of COVID-19 and No-findings. To divide the X-ray image into distinct partitions, deep U-Net learning architecture is used to extract dilabeledatory information from these regions for classification purposes.

The dataset is segmented to mark lungs and regions of interest to allocate the COlabeling-fected parts in the given chest X-Ray image. The segmentation algorithm helps to localize the lungs, which is the essential part of studying chest X-ray images. To improve the accuracy of the proposed system deep learning U-Net-based segmentation model is used. trained the U-Net model on the Shenzhen hospital X-rays dataset and validate it on the Montgomery Country dataset. Since the region of interest in this model is to lungs part in a chest X-ray, the Montgomery and Shenzhen dataset are used with the lungs masks and train the model with the masks to get the lungs region as output (shown in Fig. 4).

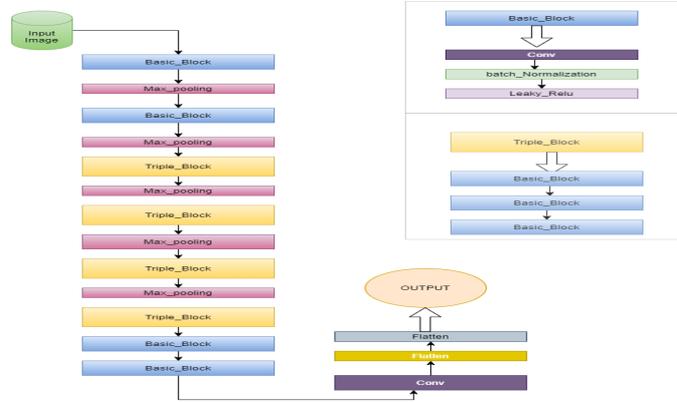


Figure 4: Depicts the architecture of the proposed classification model.

3.4. Classification

This step builds a classification model using the extra-labeled X-ray image feature vectors. The proposed system employs a deep learning model U-Net learning model based convolutional neural network (CNN) to classify COVID-19 and non-COVID-19 patients. The basic architecture of U-Net learning-based CNN is shown in Fig. 4 which contains convolution, pooling, and fully connected layers.

For the Darknet-19 model classification model, we have found it more rational to use existing work as a starting point rather than scratch. Darknet-19 is a well-accepted model in deep learning which has 19 convolutional and 5-max-pool layers. Each convolutional layer has a different number of filters. The number of filters increased as we go deeper into the architecture. We are using fewer layers in the proposed system architecture since non-availability of the sufficient chest and cough databases are a major challenging problem for training of the proposed model. To solve these problems, the proposed model is trained based on segmented images by generated mask. The proposed system used the Darknet model directly (without segmentation) to solve the overfitting problem. However, the overfitting issue is achieved due to the high parameters used in this work. This proposed system has used 17 convolutional layers with batch normalization method and leaky ReLU nonlinear activation function to train the model over-collected image database to solve the overfitting issue.

The batch normalization method is used to make the proposed segmentation model more stable and to make inputs in standard form. Batch normalization is a regularization that is easy to implement and compatible with many other models and training algorithms. The batch normalization method mitigates generalization error and allows dropout to be omitted because of the noises in the estimate of the statistics used to normalize each hyperparameters and variable.

The leakyRelu is a variation of Rectified Linear Unit (ReLU), which is used to fix the dying ReLU problem. In leakyRelu, instead of the function becoming zero for $(x \leq 0)$, it has a slight negative slope, nearly 0.01, which prevents the dying ReLU. We used leakyRelu activation in the proposed architecture. An activation function takes a number and performs a mathematical operation. Based on the output of the activation function, it decided which neuron has to activate. The pooling layer reduces feature maps' dimension using down-sampling by summarizing the features using the downsampling technique. It uses the maximum value from each cluster of

neurons at the last layer since it is a binary classification, so the proposed system has two classes, COVID-19 and non-COVID-19 prediction, for early diagnosis.

For the compilation of the model, Adam optimizer technique is used to update the weights and binary cross-entropy as loss functions. The fully connected layer measures the scoring probability of the output classes. These layers can be repeated according to the system requirement.

The CNN model uses the backpropagation technique to learn the suitable connection weights by computing the losses in the scores. Figure 4 shows the architecture of the proposed classification model. This model consists of 17 convolutional layers, 16 batch normalization, 5 max-pooling layers, 1 Flatten, and 1 linear normalization, 5 max-pooling layers, 1 Flatten, and 1 linear. To better understand the architecture, have divided it into blocks, the Basic Block, which consists of one convolutional layer followed by batch normalization and leaky ReLU. Triple Block is the combination of basic block three times, one after the other. Finally, the flattened layer and linear layer I used to predict the output of the early diagnosis. To better understand the architecture, it is divided into blocks, the basic block, consists of one convolutional layer followed by batch normalization and leaky ReLU. The triple block is the combination of basic blocks three times, one after the other.

3.5. Cough Sample-Based Covid-19 Diagnosis

The noticeable symptoms of COVID-19-infected patients have cough and breathing difficulties. When these breathing and cough samples are analyzed using machine learning techniques, it is envisioned that the respiratory sounds of patients give useful insights, enabling an early classification framework/method of patients as a diagnostic tool's design using deep learning techniques.

3.5.1. Cough Sample collection

For the detection of COVID-19 patients, the cough, breathing, and speech sound samples of individual patients (speaker/users) are considered for the training of the proposed framework. have used several datasets from different sources available in open-source platforms. The coswara cough sample database is used and provided by IISc Bangalore, India under the Coswara [9] project for diagnosis of COVID-19 patients. In the Coswara cough database, audio recordings are collected via worldwide crowd-sourcing using the b applications from different speakers. 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 normal) along with metadata information (shown in Table 2).

The Coughvid [10], DetectNow [10] and Virufy datasets are publicly available. Each month's audio recordings of cough samples are recollected in a separate folder and compressed along with its metadata information in multiple files. The files are then uncompressed to get the extracted recordings. The cough sample consists of audio recordings in WAV format at 44.1KHz. The cough audio sample has recordings from COVID-19 and non-COVID-19 people. The size of the cough audio sample was 5000 samples (i.e., 5000= 500 subjects (200 COVID-19+150(healthy)+150 (non-COVID-19 patient)× 10 samples of each people. We have chosen the COVID-19 positive samples and the negative samples randomly chosen for balanced distribution. Figure 5 shows the spectrogram of the cough sample based on different respiratory characteristics of the patient. Figure 6 shows the spectrogram of the cough sample based on different respiratory systems.

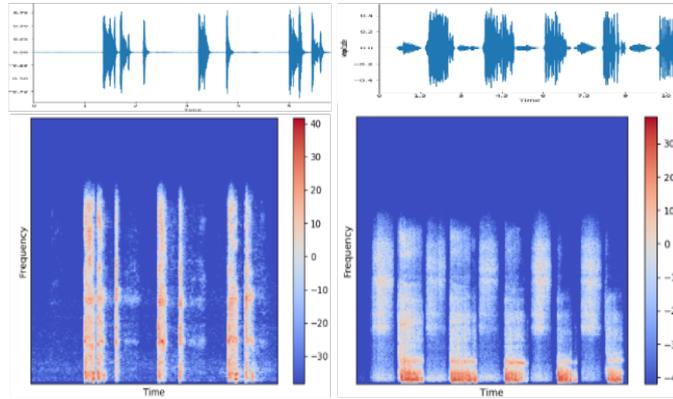


Figure 5: Working prototype for the cough-based model.

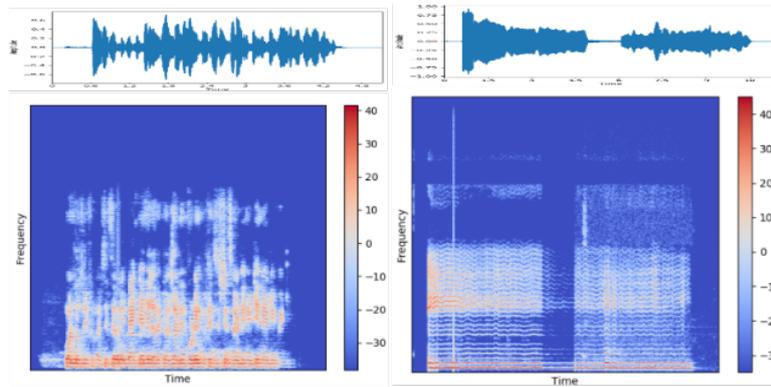


Figure 6: Shows the spectrogram of cough (voice) based on pronounced vowels (a, e, i, o, u) by patients.

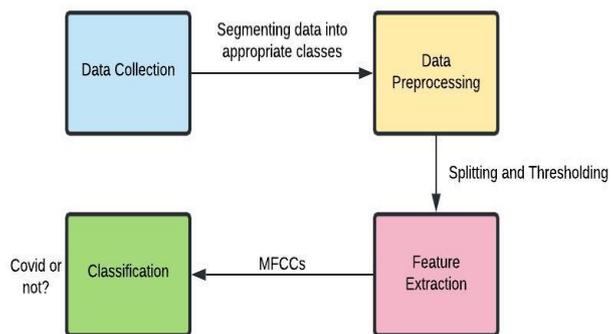


Figure 7: Working prototype for the cough-based model.

Algorithm 1: Preprocessing step

1. **Initialization:** A variable-length audio signal of sample rate 44.1KHz as input audio signal.
 2. Downsampling of the recorded cough sounds into the mono channel using the down-sampling technique at 16 kHz.
 3. **Thresholding:** Passed through an amplitude envelope of threshold value 120 to trim off background noise.
 4. **Hamming Window based splitting:**The recorded cough samples are in variable length audio signal. The samples are divided into small audio frames or chunks of four seconds for better analysis.
-

3.5.2. Pre-processing step

The audio recording datasets of patients consist of cough, breathing, phonation, and counting samples. These datasets are manually segmented into binary classes, namely, positive and negative. 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 are resegmented as negative. We used Feature extraction is a dimensional reduction process for statistical analysis of cough samples. The Mel-Frequency Cepstral Coefficients (MFCCs) features are extracted from the segmented signals using the hamming window method. The MFCC of the cough speech signal is used as a feature for classification purposes. The pre-processing algorithmic step for the cough sample is given in Algorithm 1. The short-time spectrum analysis of processed cough samples is achieved by using the short-time spectrum analysis method (shown in Algorithm 2). down-sampled the recordings at a 16kHz sample rate. The working of the prototype model is shown in Fig. 7. The computation of MFCCs features is shown in Fig. 8.

The proposed system takes cough samples (speech data) using the users' mobile phone microphones. After that, the Discrete Fourier Transform (DFT) technique is a spectrogram that gets the signal in the time interval from the frequency domain after dividing the speech signal into a small number of speech frames. We computed the power spectrum that is obtained for mapping it onto the Mel scale using triangular filters. Figure 9 shows the power spectrum of a speech input volume. Next, log outputs are found using the Discrete Cosine Transform (DCT) technique. Finally, delta (Δ) and delta-delta (Δ^2) coefficients are calculated as follows: Let us consider the MFCC of a window frame (t) is denoted by C_t . The delta coefficient (ΔC_t) is computed as (shown in Eq. 5), where $[i]$ is a (Δ) window and is usually set to 6 to 10 frames, as the consumer devices' speech input may have different signal duration, the MFCCs feature vector will also be of various lengths. Therefore, the proposed system normalizes the feature vector by constructing MFCCs with no sound for shorter signals. A detailed description of feature extraction is given in Algorithm 2.

3.5.3. Feature extraction

Coughing is one of the most important respiratory symptoms for early diagnosis of critical dis of different statistical information in lower level frequency with significant spectrogram coefficients or vectors. Therefore, we considered the cough (audio) signal for the representation of the sequence of spectral vectors for COVID-19 diagnosis. This representation of cough samples in the time vs. frequency representation is called a spectrogram with a high energy

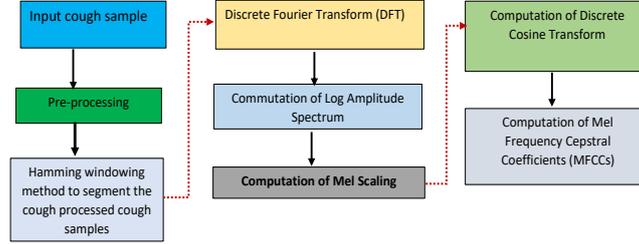


Figure 8: Extraction of MFCCs from the cough (voice) samples.

Algorithm 2: Short time spectrum analysis of samples.

1. Calculation of Power spectrum
2. Initialization: Let $y(n)$ = input cough (voice signal) samples, $y[n] = x[n]-ax[n]$.
3. Apply convolution (\otimes) and correlation methods on cough (voice signal). Consider a vector H and G of size $M \times N$. The linear convolution of H and G is the size of S of size $K = M + N$ as given below:

$$S_K = \sum_{i=0}^K H_{[K-i]} \otimes G_K, \text{ (where, } 0 \leq K \leq 1) \quad (1)$$

Where H_m and $G_n=0$ if $(m, n) \geq M, N$.

4. Perform the circular convolution of H(M) and G(N) with (For example, M, N=128) and measure the linear correlation of H(M) and G(N).
 5. Evaluate linear auto-correlation of H(M) and G(N) and used short-time spectrum analysis-effects window size and shape of the window of input speech.
 6. **Segmentation using Hamming Window:** We used the Hamming Window ($W_j(n)$) (20 msec data) to reduce the ripple and give a more accurate idea of the voice signal's frequency spectrum.
 7. **Discrete Fourier transform (DFT) analysis:**
 8. Apply 512 point DFT to get S_K
 9. **Frame the cough signal into short frames:** For each frame, the spectrogram is estimated as the power spectrum, and we applied the Mel filter bank method and perform logarithm operation of filter bank energies.
 10. **Computation of Discrete Cosine Transformation (DCT):** We have taken DCT of the log filter bank energies. The DCT coefficients 2-13 are considered and the remaining are discarded. The Mel scale frequency is measured as: $M(f) = 1125 \ln(1+f/700)$
 11. **Short-time analysis:**
 12. Short-time spectrum of the user speech signal.
 13. Plot log spectrum for voiced and unvoiced segments.
 14. Obtain spectrograms of a short utterance.
 15. Observe features of WB and NB spectrograms.
-

band for analysis of multi-pulmonary level cough signals for early diagnosis of COVID-19. The MFCCs feature of the Intrinsic Mode Functions (IMFs) and windowed cough signal samples are extracted from cough spectrograms to classify COVID-19 and non-COVID-19 using the convolutional neural network (CNN) model. Because MFCCs features consist of discriminatory features which depict the human vocal respiratory characteristics that vary with time. Therefore, the discriminatory features are extracted from the cough sample together with (Δ) first-order derivatives, and second-order ($\Delta\Delta$) derivatives on measured short-term power spectrogram of the cough sound sample for early diagnosis. The extracted MFCC features from the prepared cough audio datasets resulted in a 25000×36 feature matrix. Then, the feature reduction method is used to select the significant values from the extracted features using the Principal Component Analysis (PCA) method. The computation of MFCCs features is illustrated in Algorithm 3.

Algorithm 3: Feature extraction

1. Initialization: Let us consider the cough (voice signal) $y(n)$ as the input data.

$$y[n] = x[n] - ax[n] \quad (2)$$

2. Processing and filtering noise of the input signal $y[n]$ (shown in Eq. 3):

$$y[n] = \frac{1}{n} \sum_{i=0}^{n-1} y[n-i] \quad (3)$$

3. **Segmentation using Hamming Window:** The hamming window reduces the ripple and provides an accurate voice signal's frequency spectrum.
4. **Mel frequency cepstral coefficients features** Extraction of MFCCs features.
5. **Frame the signal into short frames:** For each frame calculate the spectrogram from the estimated log power spectrum. Apply the mel filter bank and take the logarithm of all filter bank energies.
6. **Computation of Discrete Cosine Transformation (DCT):** We computed the DCT of the log filter bank energies and keep DCT coefficients 2-13, discarding the rest.
7. The formula for converting from frequency to Mel scale is (shown in eq.4):

$$M(f) = 1125 \ln(1 + f/700) \quad (4)$$

The performance of speech voice-based classification tasks is greatly enhanced by adding time derivatives to the basic static MFCC parameters of the cough samples. This is popularly known as delta coefficients (delta (Δ) and delta-delta ($\Delta\Delta$) coefficients). It describes the vocal tract system in the human body based on computed acceleration values of the vocal cough system. Therefore, we have measured the derivatives to analyze MFCCs features and are computed using the following regression formula. The detailed description is given step by step procedures as follows (shown in Eqn. 6):

1. The Discrete Fourier Transform (DFT) technique is applied to get the signal in the time interval from the frequency domain after dividing the speech signal into a small number of speech frames.

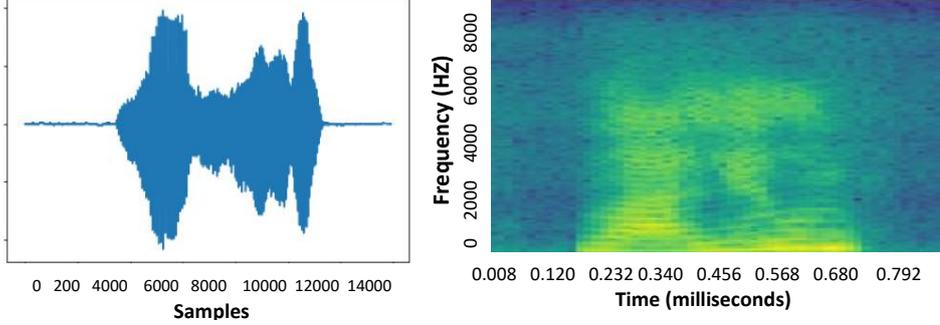


Figure 9: Illustration of cough for volume and its power spectrum.

2. Ultimately, the power spectrum has been obtained for mapping it onto the Mel scale. Fig. 9 depicts the power spectrum of a cough audio input volume.
3. Next, log outputs are found using the Discrete Cosine Transform technique.
4. Finally, delta (Δ) and delta-delta ($\Delta\Delta$) coefficients are calculated as follows:
5. Let us consider the MFCC of a window frame (t) is denoted by C_t . The delta coefficient (Δ_t) is computed as (shown in Eq. 5):

$$\Delta_t = \frac{\sum_{i=1}^I i \times (C_{t+i} - C_{t-i})}{2 \times \sum_{i=1}^I i^2} \quad (5)$$

Where, I is a (δ) window is usually set to 6 to 10 frames, as the consumer devices' speech input may have different signal duration, the MFCCs feature vectors will also be of various lengths. Therefore, the proposed system normalizes the feature vector by constructing MFCCs with no sound for shorter signals.

$$\Delta_t = \frac{\sum_{i=1}^I i \times (C_{t+i} - C_{t-i})}{2 \times \sum_{i=1}^I i^2} \quad (6)$$

3.5.4. Classification

This step builds a classification model using the extracted MFCCs feature vector with delta and delta-delta. The proposed system employs a deep CNN to classify the speech signal. The basic architecture of the CNN model contains convolution, pooling, and fully connected layers (shown in Fig. 8). The CNN model starts with an input layer that takes a spectrogram of the obtained features (i.e., feature map) and passes it to the convolution layer. The convolution layer uses a ReLU activation function to generate feature maps. As the input, the MFCC features are used for training the CNN learning model. It consists of two blocks of layers. Each block layer comprises two convolutional layers followed by a 2×2 max-pooling layer, batch normalization layer, and a 0.20 dropout to prevent the model from overfitting problems (shown in Fig. 9).

Next, the pooling layer reduces the dimension of feature maps using down-sampling by summarizing the features. The basic architecture of CNN is shown in Fig.12. The batch normal-

ization layer is used to standardize the input layer and stabilize the learning process. The convolutional layers in the first block use a kernel size of 5×5 with 32 filters, whereas, the second block uses a kernel size of 3×3 with 32 filters each in both convolutional layers. They learned complex features from these four convolutional layers that are flattened and then passed to a fully connected layer of 256 neurons followed by a dropout layer of a dropout rate of 0.30 to prevent overfitting. Finally, the output layer with two neurons and softmax () activation functions is used to classify COVID-19 and non-COVID-19 case patients for the given input.

3.6. Weighted Sum-Rule Fusion Method

The weighted sum-rule fusion method is used to integrate the matching scores from the chest X-ray detection model and cough sample-based model to perform an accurate prediction of infected people. The combined results of the two models provide a more robust and reliable result. The fatigue detection model is used to analyze the facial expressions of infected people for doctors so that the physical appearance of the patient can be taken into consideration for the diagnosis of symptoms of people. The weighted sum-rule fusion method integrates score probabilities (P^*) of chest-X-rays and cough models. The fusion of the two models makes the result more accurate for early diagnosis of patients (shown in Fig. 9). The individual weight for each chest X-ray model and cough sample model is given by (Eq.9). The confusion matrix is shown in Table 3. It shows predicated actual positive and negative classes (shown in Eq. 7).

$$P^* = \frac{C_{10} - C_{00}}{(C_{10} - C_{00}) + (C_{01} - C_{11})} \quad (7)$$

Next, assigned the individual weights (W_1) and (W_2) to the chest X-ray diagnostic and cough sample based model to combine the accuracy for accurate production of COVID-19. The proposed multimodal framework performs fusion of both models to calculate the mean weighted accuracy for accurate prediction of COVID-19 and non-COVID-19. The computation of mean weighted accuracy is given by Eq. 8.

$$Accuracy_{Avg} = \frac{W_1 + W_2 \times S_2}{W_1 + W_2} \quad (8)$$

Where, [S1] and [S2] are the matching scores of models. This weighted average accuracy of the proposed framework provides robust and more reliable results, about whether the patient can be COVID-19 positive or not for early diagnosis of COVID-19.

4. Experimental Results and Discussions

In this section, the experimental results are evaluated based on the different data sets such as chest X-ray, cough, and facial image data set sample-based multimodal learning framework. The proposed system is trained with 566 images with lung masks using U-Net-based deep learning architecture. The segmentation model provided the input segmented images to the classification model of chest X-rays with a validation accuracy of 98.46%. The input chest X-ray image, actual masks, and output predicted mask is highlighted in Fig. 10. The graphs of training and validation losses and accuracy with respect to the number of epochs for the segmentation model are shown in Fig. 11.

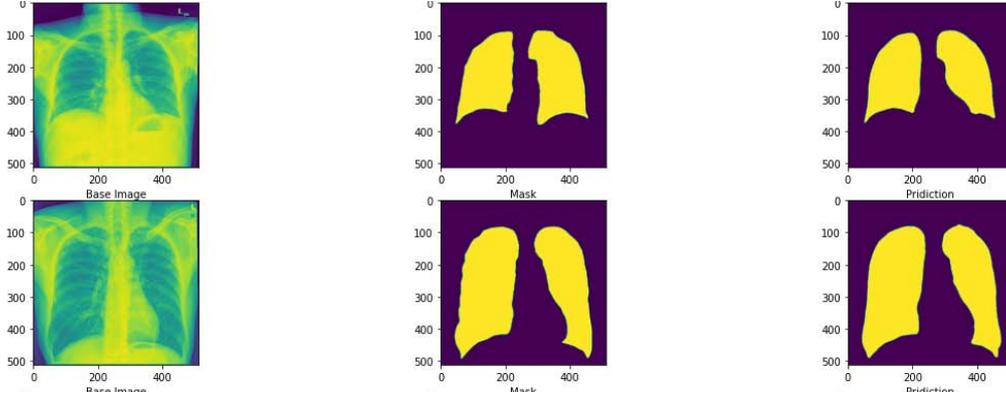


Figure 10: Shows base image, mask, and predicted mask

4.1. Classification Model

The segmented data is used as input for the proposed multimodal-based classification model. For training the model 180 images, for COVID-19 positive and 720 images for the non-COVID class. Out of these 900 images, the 20% and 80% split of the total databases is used for validation and training sets of the proposed model. We have validated the proposed framework using a 5-fold cross-validation technique. It is given 20% of total images for validation and the remaining 80% of total images for the proposed model's training. The automatic hyper-parameter selection method is used to tune the proposed framework for early diagnosis of COVID-19 cases. The proposed framework reduces the need to understand the selection strategy for the training of the models. The automatic selection method of the hyperparameters is highly computationally costly. Figure 11 shows the loss and accuracy curves of the segmentation model.

To evaluate the performance of the classification methods, we evaluated the following indicators based on confusion matrix-based measures (shown in Eq. 9-Eq. 13):

$$Accuracy(A) = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$Precision(P) = \frac{TP}{TP + F} \quad (10)$$

$$Recall(R) = \frac{TP}{TP + FN} \quad (11)$$

$$F1 - measure = \frac{2(P+R)}{P+R} \quad (12)$$

$$Specificity = \frac{TN}{TN + FP} \quad (13)$$

where TP is defined as True Positive, FP is False Positive, TN is True Negative, and FN is 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 3 and Table 4. The proposed multimodal framework for identifying COVID-19 and non-COVID-19 discoveries uses 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

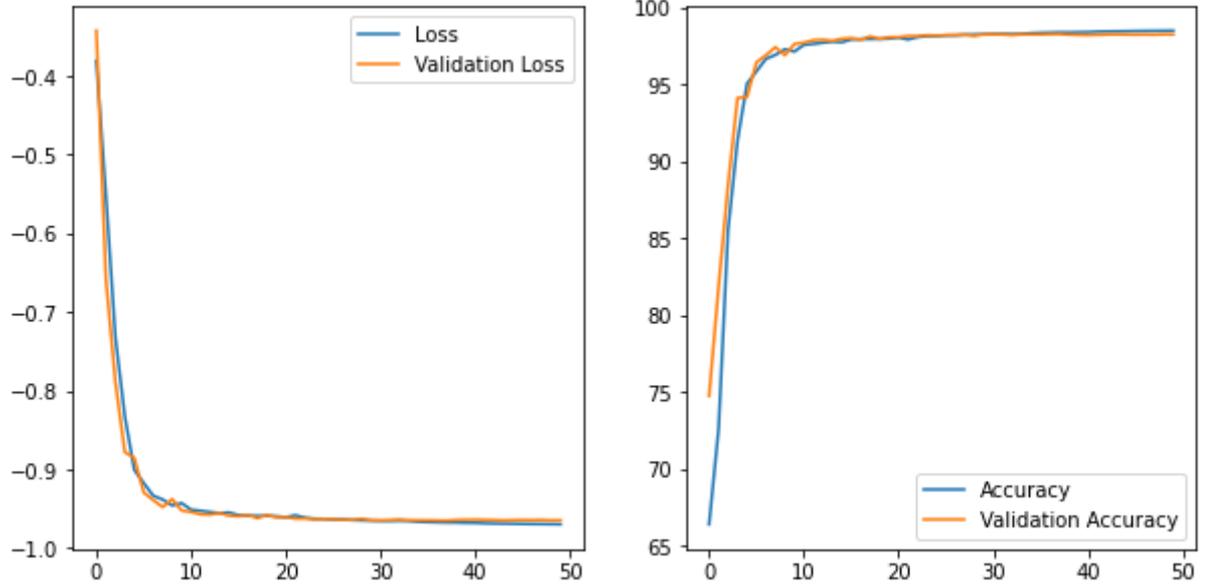


Figure 11: Loss and accuracy curves of the segmentation model

Table 3: 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

image database for training and validation of the proposed model, respectively, and for measuring system performance. A 5-fold cross-validation procedure is used to validate the suggested model. Figure 12 shows the confusion matrix-based measure for chest X-ray image and cough-based diagnosis.

To obtain higher performance of the proposed framework, the chest-X ray images are divided into different distinct regions of interest using the U-net model. Figure 13 depicts the segmentation of chest X-ray images using the proposed model. The confusion matrix of the proposed method is shown in Fig. 14 (a) for chest X-ray images and in Fig. 14(b) for cough audio signals. We used 20% of the total Chest-x-ray images for validation and the remaining 80% for the suggested model training scheme (see Fig. 15 shows fold 3 and fold-4 accuracy).

4.2. Cough Sound Analysis

The experimental results demonstrate that the proposed cough detection algorithm classifies the given cough samples into two classes: COVID-19 positive and non-COVID-19 (no findings) events with an overall accuracy of 82.30% (average accuracy is shown in Tables 4 and 5, respectively).

Table 4: Performance of proposed cough model based on fold cross-validation.

Fold	Sensitivity	Specificity	Precision	Accuracy	F1
1	0.9279	.9641	0.9538	0.9867	0.8792
2	0.9722	0.9861	0.9459	0.9833	0.9589
3	0.8657	0.91736	0.9367	0.9479	0.9746
4	0.9484	0.97288	0.9289	0.9576	0.9689
5	0.89594	0.94686	0.9266	0.9386	0.9979

Table 5: Performance of proposed cough model based on fold cross-validation.

Fold	Sensitivity	Specificity	Precision	Accuracy	F1
1	0.9279	.9641	0.9538	0.9867	0.8792
2	0.8131	0.8214	0.8131	0.8230	0.8131

The accuracy-based error graph based on computed average (mean) loss versus epochs of the neural network-based model is shown in Fig. 15. Figs. 16 and 17 show the confusion matrix for cough-based diagnosis and the accuracy of the proposed framework with the segmentation method on Chest-X-ray images. Figure 18 illustrates the average loss of the proposed model for the classification of symptoms with the segmentation method.

The L2 norm-based regularization technique is used to reduce the overfitting problem. It is also called the ridge regression method which adds a squared magnitude of coefficients as a penalty term to the loss function. The L2 regularization technique is used to remove the overfitting issues.

Moreover, various hyper-parameters like dropout rate, activation functions, and the number of hidden layers, learning rates of deep neural network-based models have also been performed to solve the overfitting problem. The number of epochs versus the decay of model loss has been investigated to rule out the possibility of overfitting [12] [13]. These performance metrics are based on a confusion matrix (shown in Table 4, Table 6, and Fig. 16).

Table 6: Accuracy(%) of the proposed Framework on selected weights.

Model Type	Weight	Accuracy
X-Ray Based Detection	0.80	0.983
Cough-based Detection	0.52	0.827

After calculating the accuracy of the individual models, our next task was to calculate the average accuracy of the two models i.e., chest X-ray-based detection-based model and Cough based detection model. The fusion of the two model's accuracy will increase the reliability and robustness of the overall framework. Since both the models are trained and tested using separate data sets, combining the models was a difficult task. The weighted average method is used for calculating the average accuracy. Some weight was assigned to both the models and using these weights calculated average accuracy. For calculating the weights, have used the confusion matrix of each model. The weights assigned to both models are. For calculating the weighted mean of the accuracy have used the formula 12 and 13. Therefore, the classification accuracy of the framework is 90.50%.

4.3. Comparison with the existing methods

In this section, the comparative study of existing approaches for COVID-19 detection is illustrated in Table 7 and Table 8. Each learning-based modality in multimodal which are chest



Figure 12: Segmentation of chest X-ray images using the proposed model.

		Actual	
		Covid-19	Non-Covid-19
Actual	Covid-19	36	0
	Non-Covid-19	6	138
		Predicted	
		Covid-19	Non-Covid-19

(a)

		Actual	
		Covid-19	Non-Covid-19
Actual	Covid-19	35	1
	Non-Covid-19	2	142
		Predicted	
		Covid-19	Non-Covid-19

(b)

Figure 13: Confusion Matrix features (a) chest X-ray images and (b) cough based diagnosis.

		Confusion matrix	
		Covid-19	Non-Covid-19
Actual	Covid-19	34	2
	Non-Covid-19	0	144
		Predicted	
		Covid-19	Non-Covid-19

(a)

		Confusion matrix	
		Covid-19	Non-Covid-19
Actual	Covid-19	36	0
	Non-Covid-19	0	144
		Predicted	
		Covid-19	Non-Covid-19

(b)

Figure 14: Confusion Matrix for (a) chest X-ray images and (b) cough based diagnosis.

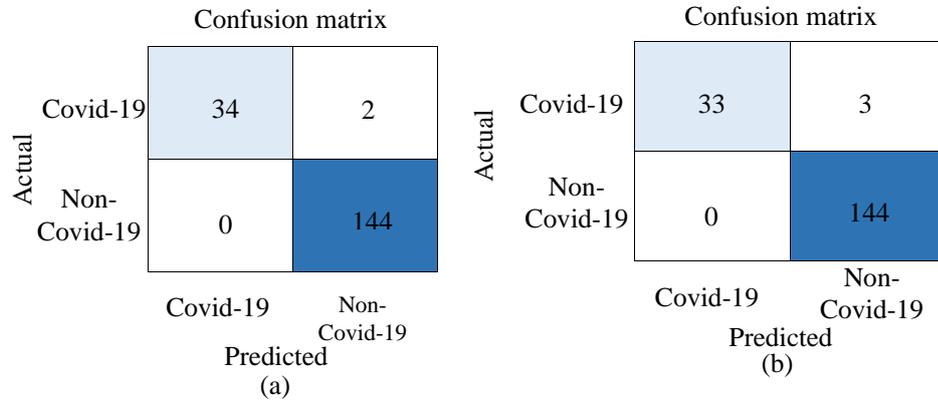


Figure 15: (a) shows the accuracy of fold 3 and (b) fold 4 for CXR classification.

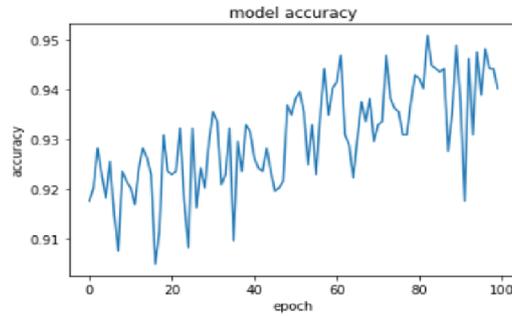


Figure 16: Shows model accuracy vs epoch for COVID-19 classification based on CXR images.

Table 7: Comparisons with existing techniques on cough for COVID-19 detection.

References	dataset	Technique used	Performance(%)	Remark
[17]	chest-CT	Res-A	86.7%	Takes more time, less accuracy
[18]	Chest CT	MI	82.90%	Less dataset
[7]	Chest-X ray	DTL	92.1%	Overfitting problem
[14]	CC	ED	77.1%	Less accuracy
Prop.	Coswara	CNN	82.70%	Less processing time

Abbreviation: Ref.=Reference, Prop.=Proposed, DLT= DLT-based classifier, ED=Ensembles DNN, TL=Transfer Learning with VGGish, chest-CT=Chest CT Scan ,RES-A=ResNet and Location Attention,MI=M-Inception techniques, CNN=Convolutional neural Network , Coswara = Coswara cough audio database, CC=Coswara/ Coughvid

X-ray-based, fatigue-based, and cough sound-based is taken into consideration to validate the accuracy of the proposed multimodal-based framework with current state of art methods that are discussed.

To demonstrate the importance of our chest X-ray model there is a comparison with various works done so far in the early diagnosis of COVID-19 field. To demonstrate the importance of the dataset, the experiments of different X-ray models are used for COVID-19 diagnosis based on different data sets and the current state of the arts. The comparative analysis of the existing methods for chest X-rays and cough samples is shown in Table 8. It mentioned various works and proposed works with the type of data used and results. The comparative analysis of our

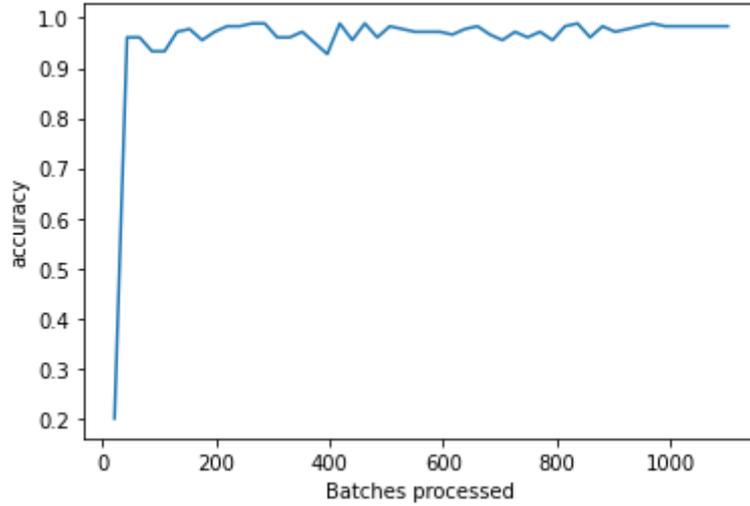


Figure 17: Depicts accuracy of the proposed framework with segmentation method on Chest-X-ray images

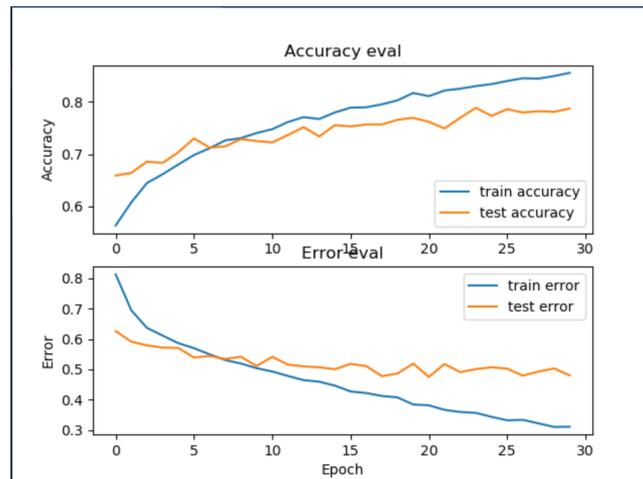


Figure 18: Illustrates the losses of the CXR classification model with segmentation method

work with various work done so far in this field they have compared the experiments of different cough-based models of COVID-19 diagnosis based on different datasets. In Table 8, the results are mentioned based on various existing works and proposed works with the type of data used and results.

5. Conclusion and Future Directions

The proposed multimodal framework performs the fusion of two models using the weighted sum-rule fusion technique for the classification of COVID-19 patients and the average accuracy

Table 8: Existing work for early detection of COVID-19 detection using ML techniques based on chest X and coughing samples.

Study	Year	R	RSS	NS	NR	PS	TM	Per(%)
[21]	2021	XCXR	-	200	CNN	CNET	SP	92.4
[19]	2020	X-Ray	30	R1	CNN	DL	SP	95.4%
[4]	2019	CXR	125	200	DNET	ML	SP	98%
[5]	2020	X-ray	230	DL	CON	SP	SP	90%
[7]	2022	Sound	200	Chest	HSR	SFs	CNN	70%
[15]	2020	cough	4352	SP	MFCC	DL	CNN	80%
[16]	2020	Web	Cough	COD	5320	DL	R2	97.10%
[23]	2020	SP	Cough	CO	3621	ST	R-18	AUC(0.72)
[24]	2021	ES	B	COV10	10	2DFT	Inv3	80%
[16]	2022	B	BB	220	MFCCS	DL	Low-D	70%
Prop.	2022	CT+C	Total	(98.33+82.7%)				

achieved in models. For chest X-ray classification, an accuracy of 98.33% is obtained and the cough sound model achieved 82.7% accuracy for the classification of patients. Based on overall observations, the final accuracy of the proposed framework is 92.03% for the early diagnosis of patients. The proposed system can be used remotely at different places, especially where lack no medical facilities, COVID-19 detection centers, monitoring systems, and other diagnosis centers. The major shortcomings and limitations faced for developing the multimodal framework for early diagnosis of the patient are: (1) there is no availability of a sufficient database of Chest-X ray and cough (audio) samples of the same patients for training the proposed models for early diagnosis of COVID-19 cases which can be COVID-19 positive class. In open source datasets, proper label and annotated datasets of chest X-ray and cough samples are available individually, however, these datasets consist of the minimum number of samples with proper labels. The non-COVID-19 case, mild infection cases, and other detection of severity levels, (2) there is always room for improvement for analysis for Cough (audio) samples-based diagnosis using deep multimodal fusion techniques. This work can also be improved by bringing new disruptive technologies and methods for fast computation of early prediction of COVID-19 cases at different levels. The future directions for the proposed multimodal are illustrated as follows: (1) The databases will be increased to evaluate results, (2) a smart system will be developed based on the Android application for early diagnosis of patients and other users of the medical services

Acknowledgement

The researchers would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project.

Conflicts of Interest:

The authors declare no conflict of interest.

References

- [1] Alsalem, Mohammed Assim, Rawia Mohammed, Osamah Shihab Albahri, Aws Alaa Zaidan, Abdullah Hussein Alamoodi, Kareem Dawood, Alhamzah Alnoor et al. "Rise of multiattribute decision-making in combating COVID-19: A systematic review of the state-of-the-art literature." *International Journal of Intelligent Systems* (2021).

- [2] Ghayvat, Hemant, et al. AI-enabled radiologist in the loop: novel AI-based framework to augment radiologist performance for COVID-19 chest CT medical image annotation and classification from pneumonia. *Neural Computing and Applications* (2022): 1-19.
- [3] Oh, Yujin, Sangjoon Park, and Jong Chul Ye. "Deep learning COVID-19 features on cxr using limited training data sets." *IEEE Transactions on Medical Imaging* (2020).
- [4] Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O. and Acharya, U.R., 2020. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine*, 121, p.103792.
- [5] 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.
- [6] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, et al., Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia, 2020 arXiv preprint arXiv:200209334.
- [7] Imran A, Posokhova I, Qureshi HN, Masood U, Riaz S, Ali K, John CN, Hussain I, Nabeel M, AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app, *Informatics in Medicine Unlocked* (2020).
- [8] Laguarta, Jordi, FerranHueto, and Brian Subirana. "COVID-19 artificial intelligence diagnosis using only cough recordings." *IEEE Open Journal of Engineering in Medicine and Biology* 1 (2020): 275-281.
- [9] Sharma, Neeraj, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Prasanta Kumar Ghosh, and Sriram Ganapathy. "Coswara—A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis." arXiv preprint arXiv:2005.10548 (2020).
- [10] 2020. Detect Now. source: <https://detect-now.org/>, no te= Accessed on 2021-02-12, time 5:30PM.
- [11] 2020. coughvid. <https://coughvid.epfl.ch/about/> accessed time 4: 30am date 25-03-2022.
- [12] Dong, Shunjie, Qianqian Yang, Yu Fu, Mei Tian, and Cheng Zhuo. "RCoNet: Deformable Mutual Information Maximization and High-order Uncertainty-aware Learning for Robust COVID-19 Detection." *IEEE Transactions on Neural Networks and Learning Systems* (2021).
- [13] Zhang, Yichi, Qingcheng Liao, Lin Yuan, He Zhu, Jiezhen Xing, and Jicong Zhang. "Exploiting shared knowledge from non-covid lesions for annotation-efficient COVID-19 ct lung infection segmentation." *IEEE Journal of Biomedical and Health Informatics* (2021).
- [14] Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough Audio Samples, accessed on 25-05-2021 time 11.45 am.
- [15] Chloe Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. "Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data." *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining(KDD 2020)*
- [16] J. Laguarta, F. Hueto and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis using only Cough Recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, doi: 10.1109/OJEMB.2020.3026928.
- [17] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, et al., Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia, 2020 arXiv preprint arXiv:200209334.
- [18] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Xu, B. (2020). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). medRxiv
- [19] Svetnik, Vladimir, Ting Wang, Christopher Tong, Andy Liaw, Robert P. Sheridan, and Qinghua Song. "Boosting: An ensemble learning tool for compound classification and QSAR modeling" *.Journal of chemical information and modeling* 45, no. 3 (2005): 786-799.
- [20] Subramani, Prabu, and Parameshachari BD. "Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients." *Personal and ubiquitous computing* (2021): 1-14.
- [21] Suppakitjanusant, P., Sungkanuparph, S., Wongsin, T., Virapongsiri, S., Kasemkosin, N., Chailurkit, L. and Ongphiphadhanakul, B., 2021. Identifying individuals with recent COVID-19 through voice classification using deep learning. *Scientific Reports*, 11(1), pp.1-7.
- [22] R. Pramono, A. Imtiaz, and E. Rodriguez-Villegas, "Automatic identification of cough events from acoustic signals," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 2019, 07 2019, pp. 217–220.
- [23] J. Laguarta, F. Hueto, and B. Subirana, "COVID-19 artificial intelligence diagnosis using only cough recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 275–281, 2020.
- [24] Bagad P, Dalmia A, Doshi J, Nagrani A, Bhamare P, Mahale A, et al. Cough against COVID: Evidence of COVID-19 signature in cough sounds. arXiv preprint arXiv:200908790. 2020.
- [25] Sait, U., KV, G.L., Shivakumar, S., Kumar, T., Bhaumik, R., Prajapati, S., Bhalla, K. and Chakrapani, A., 2021. A deep-learning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images. *Applied Soft Computing*, 109, p.107522.

Author Biography

Santosh Kumar received his Ph.D. in Computer Science and Engineering from IIT (B.H.U.), Varanasi, India, in 2017. He is an assistant professor in the Dept. of CSE, IIIT Naya Raipur, India.

Rishabh Nagar has completed his B.Tech. in CSE from the Department of Computer Science and Engineering (CSE), IIIT Naya Raipur in 2022. His area of interest includes machine learning, Deep Learning, and computer vision.

Saumya Bhatnagar has completed his B.Tech. in CSE from the Department of Computer Science and Engineering (CSE), IIIT-Naya Raipur in 2022. His area of interest includes machine learning, deep learning, and IoT.

Sachin Kumar Gupta received his B.Tech in ECE from NIT Raipur in 2008 and his Ph.D. from IIT BHU in 2016. He is working as an assistant professor in the school of ECE in SMVDU Jammu, India.

Ramesh Vaddi is working as an Associate Professor with the Department of Electronics and Communication Engineering, SRM University, Andhra Pradesh, India. His area of interest includes Energy Efficient Circuit Design with Post-CMOS Devices, Hardware Security for IoT, VLSI Accelerators, and In-memory Computing architecture for AI-based models and systems.

Mamoon Rashid is currently working as an Assistant Professor in the Department of Computer Engineering, Vishwakarma University, Pune, India. He is also serving as a Director, Research Center of Excellence for Healthcare Informatics, Vishwakarma University, Pune, India. He received his Ph.D. from Punjabi University, Patiala, India in the field of Medical Informatics. His research interests include Big Data Analytics, Machine Learning, Medical Imaging, Image Processing.

Ali Kashif Bashir is currently working as a Reader of Networks and Security at the Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom. He is serving as the Editor-in-chief of the IEEE Technology, Policy and Ethics NEWSLETTER. He is leading many conferences as a chair (program, publicity, and track) and had organized workshops in flagship conferences like IEEE Infocom, IEEE Globecom, IEEE Mobicom, etc. He has given more than 35 invited and keynote talks at international conferences across the globe.

Tamim Alkhalifah received the master's degree in computer science from Swansea University and the Ph.D. degree in computer science from Flinders University, Adelaide, Australia, in 2018. He is currently working as an Assistant Professor with the Department of Computer, College of Science and Arts, Qassim University, Ar Rass, Saudi Arabia. He has published several articles in the IT field. His primary research interests include e-technologies, mobile development, mobile learning, and gamification.