



A novel deep fusion strategy for COVID-19 prediction using multimodality approach[☆]

Ankush Manocha^{*}, Munish Bhatia

Lovely Professional University, Phagwara, 144411, Punjab, India

ARTICLE INFO

Keywords:

Semi-supervised learning
Smart healthcare
Covid-19
Deep learning
Smart monitoring

ABSTRACT

Over the last two years, the novel coronavirus has become a significant threat to the health of the public, and numerous approaches are developed to determine the symptoms of COVID-19. To deal with the complex symptoms of COVID-19, a Deep Learning-assisted Multi-modal Data Analysis (DMDA) approach is introduced to determine COVID-19 symptoms by utilizing acoustic and image-based data. Furthermore, the classified events are forwarded to the proposed Dynamic Fusion Strategy (DFS) for confirming the health status of the individual. Initially, the performance of the proposed solution is evaluated on both acoustic and image-based samples and the proposed solution attains the maximum accuracy of 96.88% and 98.76%, respectively. Similarly, the DFS has achieved an overall symptom determination accuracy of 98.72% which is highly acceptable for decision-making. Moreover, the proposed solution shows high reliability with an accuracy of 95.64% even in absence of any one of the data modalities during testing.

1. Introduction

The novel COVID-19 scientifically known as SARS-CoV-2 was first announced by the end of the year 2019 in China and instantly became a pandemic in the entire world [1]. COVID-19 infects the lungs and affects the repository system of an individual which needs to detect early to reduce its severity. In COVID-19, shortness of breath, continuous dry cough, and infection in the lungs are very common symptoms in patients. The seriousness of this issue can be decreased (i.e. “flatten the curve”) by following some precautionary measures such as hand washing, social separation, and wearing a mask [2]. To monitor the status, several X-ray centers are established in the cities and villages. However, the availability of radiologists is considered a big issue. In this manner, the development of a smart monitoring framework is required for continuous monitoring and immediate decision-making. Modern data processing solutions such as deep learning (DL) and machine learning (ML) could likewise assist to slow the speed of spreading the virus and observing new cases by processing complex data in an efficient manner [3]. Therefore, the development of a preliminary self-monitoring solution became necessary to get immediate and considerable results.

Most commonly, the parameter of temperature is considered as a preparatory method to distinguish the individuals suffering from COVID-19 in public places. However, the symptom of temperature might be missing in asymptomatic patients which can be a significant reason for spreading the virus. To deal with this issue, acoustic signals can help to expand the utility of the monitoring solution by alleviating the spread of the virus. From the prior research, it can be analyzed that the acoustic signals of the individuals experiencing respiratory infirmities are having recognizable highlights. These highlights can be extracted from indicative vocal traits such as breathing, speech, and cough by utilizing different signal preprocessing techniques [4–6]. After feature extraction, the deep

[☆] This paper is for special section VSI-igdn. Reviews were processed by Guest Editor Dr. Hamid Reza Boveiri and recommended for publication.

^{*} Corresponding author.

E-mail addresses: ankushmanocha31@gmail.com (A. Manocha), munishbhatia90@gmail.com (M. Bhatia).

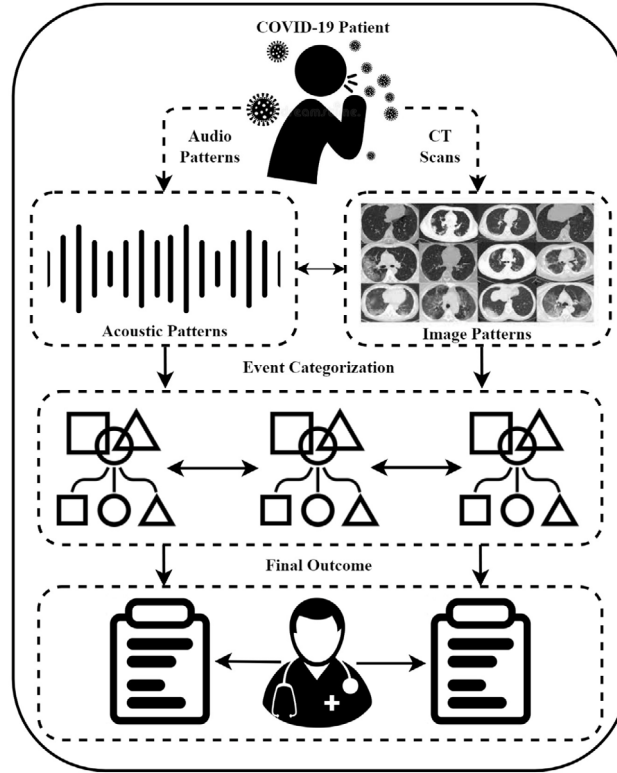


Fig. 1. The conceptual framework of the proposed solution.

learning approach can be evaluated on the extracted features to monitor the health status of the patient suffering from COVID-19 [7]. Hence, it is important to analyze the predefined vocal characteristics of individuals with mild symptoms of COVID-19 in a continuous manner [8]. In this manner, several studies have been proposed to monitor different symptoms of COVID-19 and a review of some of these methods with its limitation is summarized in Table 1.

As utilizing a single data modality can be a reason for event misclassification, multiple data modalities can be used to increase the probability of symptom determination with more reliability. In this manner, multimodal data fusion strategies can be used to combine different data types in a single solution to predict different symptoms related to COVID-19. Thus, the novelty of the proposed framework is shown in Fig. 1. The contribution of the proposed work is further divided into the following objectives.

1. To develop a semi-supervised monitoring framework that utilizes two different data modalities such as acoustic signals (breathing, speech, and cough) and images (CT scan) for the adequate analysis of COVID-19 symptoms in real-time.
2. To develop deep learning-based solutions namely *Acoustic Prediction* and *Image Prediction* for the determination of COVID-19 symptoms from the acoustic and image-based data samples, respectively.
3. To increase the event determination reliability, a decision tree classification approach-based dynamic fusion strategy is proposed to fuse the predicted outcomes concerning the proposed classification networks.
4. To justify the symptom determination performance of the proposed solution by comparing the calculated outcomes with the state-of-the-art approaches.

The remaining contents of an article are coordinated as follows. Section 2 is explaining some of the primary work related to the proposed study. The detailed working process of the developed approach is discussed in Section 3. The experimental evaluation related to the performance of the proposed solution is explained in Section 4. Lastly, the conclusive remarks related to the proposed work are discussed in Section 5.

2. Related work

In article [14], the authors examined the feasibility of using a model that assists experts in identifying COVID-19-positive patients. For the identification of COVID-19 patterns utilizing CT scans, a novel therapeutic hybrid DL model has been applied to meet this goal. The suggested paradigm consists of two primary stages: segmentation and identification. The suggested model for COVID-19 outperformed various existing literature techniques in terms of detection accuracy. In article [9], a Convolutional Neural Network

Table 1

A comparative analysis of COVID-19 determination methods.

Sr. No.	Reference	Data samples	Technique	Accuracy	Limitation
Acoustic data analysis	[5]	Cough	LSTM and SVM	88.2%	Only a single input modality was employed, resulting in adequate confidence levels (<100%). Utilizing supplementary modalities can enhance performance.
	[6]	Voice	Two-way ANOVA and Wilcoxon's rank-sum test	83.69%	Only a single input modality was employed, resulting in adequate confidence levels (<100%). Utilizing supplementary modalities can enhance performance.
	[7]	Cough	Transfer learning	92.85%	Utilizing two or three deep models with several layers for a single input modality enhances the implementation complexity. This necessitates a compact framework.
	[9]	Cough	CNN and MFCC	98.54%	Only a single input modality was employed, resulting in adequate confidence levels (<100%). Utilizing supplementary modalities can enhance performance.
Image data analysis	[10]	CT scans	AlexNet	AuC of 0.995	Due to the variety in symptoms across patients, a single modality of input was employed, which cannot be depended on. Additionally, such models are unfit for noisy situations.
	[11]	CT scans	Attention-based multiple instance learning	97.9%	3D imaging increases training and spatial complexity.
	[12]	X-rays	Patch-based CNN	91.9%	Only a single input modality was employed, resulting in adequate confidence levels (<100%). Utilizing supplementary modalities can enhance performance.
	[13]	X-rays	Fusion of ResNet-101 and ResNet-152	96.1%	Utilizing two or three deep models with several layers for a single input modality enhances the implementation complexity. This necessitates a compact framework.

(CNN) is proposed to detect the presence of COVID-19 in an individual by analyzing cough samples. The proposed architecture is designed by following the architecture of ResNet 50. The model was able to achieve 94.2% and 83.2% accuracy for specificity for symptomatic and asymptomatic subjects, respectively. In article [10], the process of feature extraction, feature selection, and feature classification is followed to determine the symptoms related to COVID-19 from CT scans. Three different techniques such as AlexNet, Guided-Whale Optimization Algorithm, and a machine learning-inspired classifier are used for feature extraction, selection, and classification, respectively. In this manner, the authors were able to achieve an accuracy of 0.995 for AUC under the curve. In article [11], a 3D instance learning approach is proposed to determine COVID-19 symptoms from CT scans. The authors collected a total of 460 images related to chest CT scans to train the model and achieved a prediction accuracy of 97.9%. In article [12], authors proposed a patch-based CNN approach to diagnose COVID-19 symptoms from CXR radiographs. With this approach, the authors were able to achieve considerable accuracy with respect to COVID-19-based symptom prediction. In article [13], the transfer learning approach was used by the authors to classify the symptoms into 3 categories such as (i) normal, (ii) COVID-19, and (iii) viral. After completing the process of training, the authors were able to achieve an accuracy of 96.1% from X-ray images. In article [15], authors utilized a DRL-based technique to address mask extraction problems. This innovative technique employs a modified Deep Q-Network to permit the mask detector to choose masks from the investigated picture. Based on COVID-19 computed tomography scans, authors utilized DRL masked extraction-based approaches to capture visual characteristics of COVID-19 affected regions and deliver a precise medical assessment, all while improving the pathogenic lab test and speeding up the process. In article [16], authors developed a COVID-19 diagnosis system using a convolutional neural network (CNN), stacked autoencoder, and deep neural network. In this approach, categorization is modified before the three CT imaging techniques are used to distinguish between normal and COVID-19 instances. For the training of the applied learning algorithm and the assessment of its ultimate performance, a large-scale, demanding CT image dataset was utilized. In article [17], authors suggested a technique for the categorization and early detection of COVID-19 utilizing X-ray pictures for image analysis. The evaluation findings indicate a high diagnostic accuracy ranging from 89.2% to 98.6%. The proposed technique for early screening and categorization of COVID-19 utilizing image processing on X-ray images has been demonstrated to be effective. [18] present a thorough analysis of the most current Deep Learning and Machine Learning approaches for COVID-19 diagnosis. The authors instruct the scientific community on the future expansion of machine learning for COVID-19 and stimulate their future studies. Authors in [19] gave an overview of the artificial intelligence (AI) approaches utilized in the COVID-19 study. Authors proposed a solution for capturing the healthcare data such as temperature by utilizing the sensors of cellphones [20]. Machine learning algorithms were proposed by [21] to detect COVID-19 cases. The process of training is dependent upon the cauterization of data from the user via an online survey that can be viewed from a smartphone. Following the outbreak of COVID-19, authors emphasized the necessity to create common standards for sharing information amongst smart cities in pandemics. For example, by utilizing data captured by thermal cameras located in smart cities, AI algorithms can be utilized to predict the status of the COVID-19 situation [22]. Authors developed an IoT-based method for detecting novel covid-19 cases. An inference system based on fuzzy rules was used to design the rules [23].

Limitations and research gaps: However, a major limitation of utilizing the single data modality in the proposed solutions for the determination of COVID-19 symptoms is observed in the above-discussed literature. By analyzing the sensitivity of the domain, high precision in the event prediction is required. The high scale of precision can be obtained by utilizing multimodal data that includes audio and image-based data such as breathing noises, cough, speech, and CT scans. Therefore, according to the best of our knowledge, no one has developed a comprehensive framework for identifying and monitoring the irregular events of the patients infected with the COVID-19 virus by utilizing several data modalities. In this manner, some of the research gaps are enlisted as follows;

1. Utilizing two or three deep models with several layers for a single input modality enhances the implementation complexity. This necessitates a compact framework.
2. Only a single input modality was employed, resulting in adequate confidence levels (<100%). Utilizing supplementary modalities can enhance performance.
3. The high scale of precision can be obtained by utilizing multimodal data that includes audio and CT scan image-based data such as breathing noises, cough, speech, and CT scans, but no work has been performed in this context.
4. Limited research work has been performed for identifying and monitoring the irregular events of the patients infected with the COVID-19 virus by utilizing several data modalities.

3. Proposed solution

In this work, a preliminary screening method is proposed to predict COVID-19 remotely without visiting a healthcare clinic. In the proposed solution, acoustic and image data is used to determine the symptoms of COVID-19 in an individual. Utilizing different data modalities in a single solution can help to expand the domain of health monitoring and attain maximum symptom determination accuracy. To determine the outcome, the output calculated by each network is forwarded to a machine learning classifier that can help to reduce the chances of event misclassification. The complete working of the proposed solution is presented in Fig. 2. The proposed architecture consists of three different modules as follows;

1. Data collection and preprocessing,
2. Semi-supervised multi-modal for COVID-19 symptom determination,
 - (a) Acoustic data analysis,
 - (b) Image data analysis,
3. Dynamic fusion strategy

Every possible parameter related to each module is explained ahead.

3.1. Data collection and preprocessing

In the proposed study, two types of data modalities such as acoustic signals and images-based data are used to determine the symptoms of a patient infected with COVID-19. The acoustic dataset is containing three types of audio signals belongs to cough, breath, and speech. On the other hand, the image-based dataset is containing CT scans belonging to the chest of the patients. A total of 289 cough-based samples is gathered from virufy¹ and coswara² GitHub repository. Similarly, a total of 276 and 502 samples belong to speech and breath audio samples, respectively. To deal with the situation of overfitting, different data augmentation techniques have been applied to the acoustic dataset to increase the number of samples. After performing the operation of data augmentation on the acoustic dataset, the updated number of audio samples are 428, 388, and 742 for cough, speech, and breath, respectively. Moreover, the augmented samples are categorized into positive and negative classes based on their type and frequency. The positive class represents irregular health events and the negative class represents normal events. In this manner, 404 number of cough samples belongs to the positive class and 24 cough samples belong to the negative class. Similarly, 352 speech samples belong to the positive, and 36 speech samples belong to the negative class. For breath, 706 samples belong to the positive class and 26 samples belong to the negative class. For the image-based dataset, CT scan images are collected and obtained from an open-source online platform. The chest images dataset [24] comprises 1252 coronavirus positive cases and 1229 typical examples.

Preprocessing: As the proposed solution is dealing with both acoustic and image-based events, preprocessing is required to deal with the overlapping events. In this manner, the hard-labeling preprocess technique named median filtering is utilized to preprocess the signals. Median filtering helps to convert the representation of an event $y_I(e)$ into a binary representation such as $\bar{y}_I(e)$ that determines whether a label is present ($\bar{y}_I(e) = 1$) or absent ($\bar{y}_I(e) = 0$). It also helps to increase the symptom determination stability of the model by removing noisy outputs such as single-frame output.

¹ Source: Virufy COVID-19 Open Cough Dataset, <https://github.com/virufy/virufy-data>.

² Source: Coswara-Data, <https://github.com/iiscleap/Coswara-Data>.

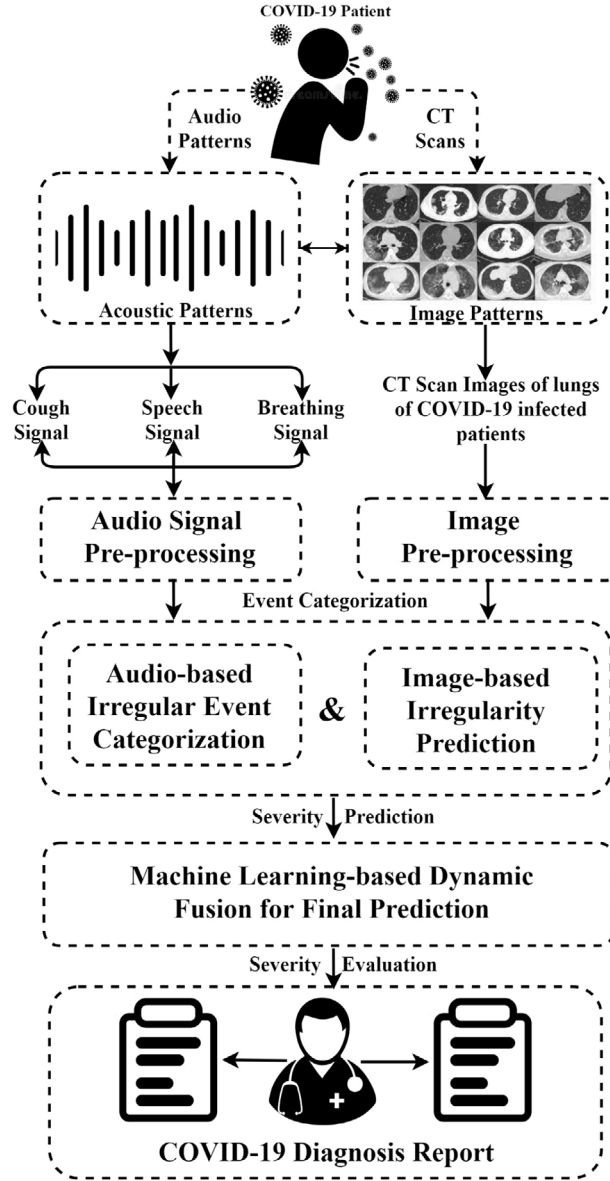


Fig. 2. The detailed flow of the proposed solution.

Definition 1. Let $Y = (y_1(e), \dots, y_i(e), \dots, y_T(e))$ be a probable succession of events denoted as e and θb defines the value of threshold. The categorized binary events with respect to its threshold value is denoted as $\bar{Y} = (\bar{y}_1(e), \dots, \bar{y}_i(e), \dots, \bar{y}_T(e))$.

Mathematically, the process of event categorization is represented as;

$$\bar{y}_i(e) = \begin{cases} 1, & \text{if } y_i(e) > \theta b \\ 0, & \text{Otherwise} \end{cases} \quad (1)$$

In the proposed study, the median filter calculates the middle point of each event with the window size γ within a sequence of events denoted as \bar{Y} . In this manner, a median filter eliminate any event with the duration of $\leq \lfloor \frac{\gamma}{2} \rfloor$ and combining the two events with distance of $\leq \lfloor \frac{\gamma}{2} \rfloor$.

3.2. Semi-supervised multi-modal for COVID-19 symptom determination

The complete process of COVID-19 symptom determination is divided into three modules such as acoustic signal processing, image processing, and multi-modal dynamic fusion. The targeted parameters acquired from the expected individuals can be provided

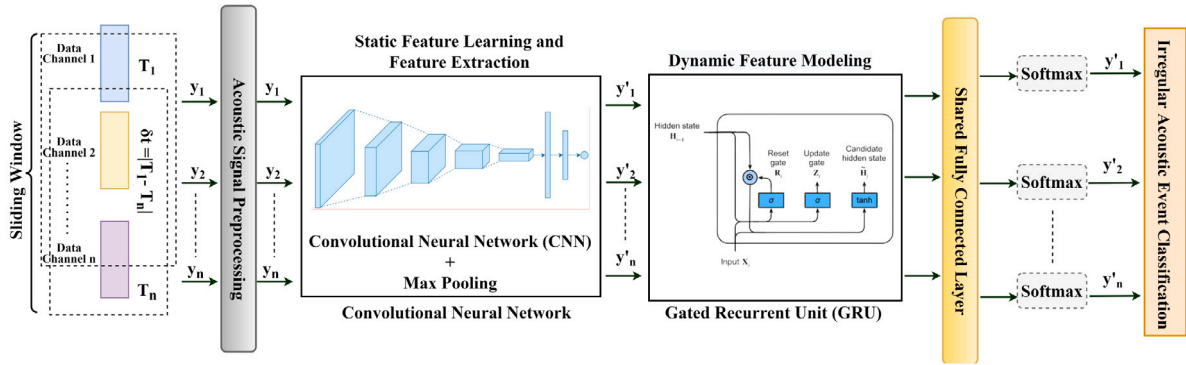


Fig. 3. Irregular acoustic event classification.

as input to each separate classifier to obtain expected predictions. Cough, speech (utterance of vowel 'a'), breathing-based acoustic signals, and CT scans are utilized to train the proposed *Acoustic_{Prediction}* and *Image_{Prediction}* model, respectively. At last, the proposed dynamic fusion strategy is responsible for combining the predicted outcomes of the *Acoustic_{Prediction}* model and *Image_{Prediction}* model to evaluate the final scale of positive probability.

3.2.1. Acoustic data analysis

To determine the COVID-19 symptom, raw acoustic signals are transferred as input to the proposed deep network. The Mel Frequency Cepstral Coefficients (MFCC) signals are utilized in the proposed study to analyze the targeted audio patterns. Here, the audio classification is done by performing two primary operations as illustrated in Fig. 3.

1. Feature extraction
2. Symptom determination

Feature extraction: In general, audio categorization begins with manual feature extraction, then proceeds on to a feature selection approach, and lastly to symptom determination. Another approach is to use raw audio waveforms in the proposed deep learning models for categorization. In our study, we leverage Mel Frequency Cepstral Coefficients (MFCC) that can be used directly for symptom determination with the help of the Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU). To determine the Cepstral Coefficients of sound signals, a cepstral investigation is carried out on the Mel Spectrum. As a result, these numbers are referred to as MFCCs and multi-dimensional MFCC feature extraction is done from the acoustic signals [25]. Firstly, the acquired acoustic signals are converted into the Mel scale to calculate higher acoustic resolution in the case of lower audio frequencies which helps to enhance symptom determination stability. The MFCC is deemed more realistic to depict these signals since effective acoustic information is always transmitted at lower frequencies. Therefore, it is expected to make the updation in the signal frequency that all the more intently reflect perceptible changes. There are various strategies for updating the scale of frequency to the Mel scale. In this study, the scale of frequency denoted as f is transformed into Mel scale m as:

$$m = 2595 \times \log_{10}\left(1 + \frac{f}{500}\right) \quad (2)$$

Symptom determination: In the proposed study, CNN and GRU techniques are acting as a backbone to identify the targeted symptoms as represented in Fig. 3. Moreover, temporal pooling and subsampling layers help the network to learn probabilities from the acoustic clips during the time of training. The proposed *Acoustic_{Prediction}* model consisting a total number of five CNN layers followed by the number of GRU cells. In particular, the proposed *Acoustic_{Prediction}* model uses two-fold convolution layers and three subsampling stages rather than five. Besides, the performance of the model significantly increases by utilizing the parameter of Activation Block Structure, Convolution, Batch-normalization, and LeakyReLU. A bidirectional GRU with 128 units handles the extracted features by the CNN. The motivation behind utilizing the layers of subsampling accompanies three reasons:

- Subsampling layers help to deal with disjoint symptom determination such as short length of acoustic frames, consecutive sequence of one-zero values $[0, 1, 0, 1, \dots] \rightarrow [0, 0, 0, 0, \dots]$.
- It helps to summarize all the small portions of the time sequences by reducing the steps of time in the Recurrent Neural Network (RNN).
- With the assistance of subsampling in the domain of time–frequency, dynamic frequency of time can be experienced by the proposed solution through the operation of L-norm pooling.

The process of subsampling defines the operations performed in the middle of the local kernels with the size of 2×2 in the domain of time–frequency. The process of subsampling is generally performed by utilizing max or average operators. L-norm subsampling

Table 2
The detail of parameters of the DFS.

Sr. No.	Parameters	Value
1.	Validated dataset X	$X'_1, X'_2, X'_3, \dots, X'_{200}$
2.	Test samples	$X_i = X'_a, X'_b, X'_c, X'_d$
3.	Sample of cough	X_a
4.	Sample of speech	X_b
5.	Sample of Breathing	X_c
6.	Sample of CT scan	X_d
7.	Test subject	n
8.	Models	k
9.	Looping variables	i, j
10.	Prediction vector	P
11.	Training-based prediction vector	P_x
12.	Test-based prediction vector	P_y
13.	DT classification model	DTC
14.	Number of trees	n
15.	Number of features	n
16.	Misclassified prediction vector	F

helps to achieve robustness by detecting the short and inconsistent events. In this manner, the process of L-norm subsampling opted in the proposed study is characterized as:

$$L_p(x) = \sqrt[p]{\sum_{x \in K} x^p} \quad (3)$$

where the value of L_p -norm subsampling is decided by calculating the interpolation in between the max and mean operations. The interpolated value of subsampling helps to preserve the temporal features of a signal and extricate the optimized features as similar to the max operation. The operation of subsampling is performed in three phases denoted as s_1 , s_2 , and s_3 . Each stage calculates the temporal features by calculating the component of s_j where $j \in [1-3]$. Generally, the frequency of time is subsampled by calculating the variable v with the value of 4. The dimension of frequency is decreased by the value of 4 at every one of the three stages. At last, the dimension of frequency is reduced by the value of 64. Furthermore, the operation of linear upsampling is included at the last boundary layer that re-establishes the original temporal resolution denoted as $\frac{T}{v} \rightarrow T$. In this manner, the $Acoustic_{prediction}$ model can detect the irregular COVID-19 symptoms from the short acoustic signals with a time frame of 20 ms at the resolution range of 50 Hz.

$Acoustic_{prediction}$ generates the probabilistic output $y_i(e) \in [0, 1]$ with respect to the input x_i at a specific time δt . By following the concept of temporal pooling, the sound clip-based calculated probabilities $y(e) \in [0, 1]$ are aggregated together to generate frame-based probability represented as $y(e) = agg_t \in T(yt(e))$. Furthermore, as temporal pooling is significantly contributing to performing the operation of event tagging and event localization, the linear softmax (LinSoft) function is utilized in the proposed study and represented as;

$$y(e) = \frac{\sum_i^T y_i(e)^2}{\sum_i^T y_i(e)} \quad (4)$$

LinSoft temporal pooling operation is responsible to calculate the average of weights in an automated manner. Therefore, the proposed $Acoustic_{prediction}$ model is producing two results:

1. Acoustic clip-based symptom determination $y(e)$ and
2. Sequential symptom determination $y_i(e)$.

As acoustic clips are utilized to train the model, the model parameters $y(e)$ are updated by aggregating the clip-level outputs and performing the operation of back-propagation. In this manner, the edge level results $y_i(e)$ are exclusively used for assessment.

3.2.2. Image data analysis

The area of Artificial Intelligence has been transformed with the emergence of deep learning and image processing techniques. For the precise determination of COVID-19 symptoms, a radiology-based imaging technique can be also used with deep learning. The noticeable information regarding the virus has been shown through CT scans. However, real-time classification methods necessitate an efficient model with a low network complexity. In this way, deep learning models like DenseNet, U-Net, Inception, ResNet, YOLO, MobileNet, and RCNN have been designed and widely utilized for image processing tasks in recent years. Below are the structure of these architectures.

1. In DenseNet, instead of addition, the output from the preceding layer is concatenated. The increased complexity resulting from a large number of layers and skip connections is the biggest limitation.
2. The U-Net architecture is a sort of CNN architecture that is mostly used for image segmentation. It comprises a concatenated encoder and decoder route with skip connections. During upsampling, these skip links can share the information globally.
3. Inception model is a CNN architecture-based deep network with 27 layers that are commonly used for image classification.

4. The residual and skip connections are used in the Resnet architecture. The aggregate of the output from the previous layers is utilized as the input for the subsequent layers in this case. This prevents the information from being lost or abstracted during the early stages of learning.
5. Object detection, classification, object localization, and segmentation are all performed using the YOLO architecture. YOLO (version 1), YOLO (version 2), and YOLO (version 3) performed much better than their earlier versions.
6. Darknet19 is used by item YOLOv2, and Darknet53 is used by item YOLOv3. These models are highly effective, but they are more complicated and take up more memory. Despite this, the little edition of YOLOv3 runs substantially quicker and uses less amount of memory. As a result, this model is considered one of the best models for a real-time system.
7. Region-based CNN and You Only Look Once (YOLO) models are commonly utilized for object detection. To check the presence of the object in a rectangular box, RCNN uses a collection of boxes called regions in the image, which results in a longer convergence time.

The proposed model is built using the architecture of YOLOv3 as a framework. The proposed model is developed by reducing the number of layers of the existing YOLOv3 tiny for real-time monitoring with considerable speed. To save the local information, an additional skip connection which is similar to the U-Net architecture is proposed. This makes it easier for the model to locate abnormal symptoms in images and appropriately identify them as depicted in Fig. 4. The proposed model has 12 convolution layers for the extraction of features and 7 max-pooling layers for dimensionality reduction. Every convolution is preceded by a ReLU activation function and the activations are batch normalized. After the feature maps are reduced into a feature vector, a softmax layer gives the outcome.

3.3. Dynamic fusion strategy

The Decision Tree (DT) approach is following the late fusion approach in the proposed Dynamic Fusion Strategy. The scale of randomness in features and data instances helps to enhance the symptom determination diversity of the state-of-the-art learners that helps to deal with the problem of overfitting. The unimodal-based calculated outcomes are forwarded to DT to determine the health severity. Even, DT can deal with missing information. This permits the framework to work even in the absence of any information source. Moreover, the wrongly predicted outcomes are additionally forwarded to the proposed dynamic fusion strategy for retraining. As the process of retraining unimodal increases the complexity of the framework, only the dynamic fusion module is retrained with the misclassified events. The detail of the parameters utilized in the proposed dynamic fusion strategy is presented in Table 2. The detailed solution of overall symptom determination through dynamic fusion strategy is explained in Algorithm 1.

Algorithm 1: Dynamic Fusion Strategy

Input: Validated dataset X_i

Output: The final prediction of Y with respect to COVID-19 symptoms

Step 1. Initialize: $k = 5$, $Estimators_n = 10$, $n = 250$, min samples split = 2

```

for i = 1 to n do
  for j = 1 to k do
     $P(i, j) = P(i, j) + \text{predict}(x_i, m_j)$ 
  end for
end for

```

Step 2. Splitting the dataset P into $P_x(80\%)$ and $P_y(20\%)$

Step 3. Building DFS with features max = sqrt(n)

Step 4. Train DFS with the training set P_x

Step 5. Dataset $Y =$ Testing P_y

```

for i = 1 to 20 do
  if
     $Y_i$  belongs to misclassification then
    Transfer  $P_i$  to Vector  $F$ 
  end if
end for

```

Step 6. Retraining of DFS with vectors F for the updation of trained parameters

Step 7. Return Y

Step 8. Exit

4. Performance evaluation

In this section, the performance of the proposed solution is determined by evaluating the audio-based and image-based irregular events individually. After evaluating the symptom determination performance of each model, the performance of the dynamic fusion strategy is evaluated for health severity symptom determination. Moreover, the scale of reliability and stability is also evaluated to analyze the efficiency of the proposed solution for real-world implementation. The detail of each experimentation is enlisted as:

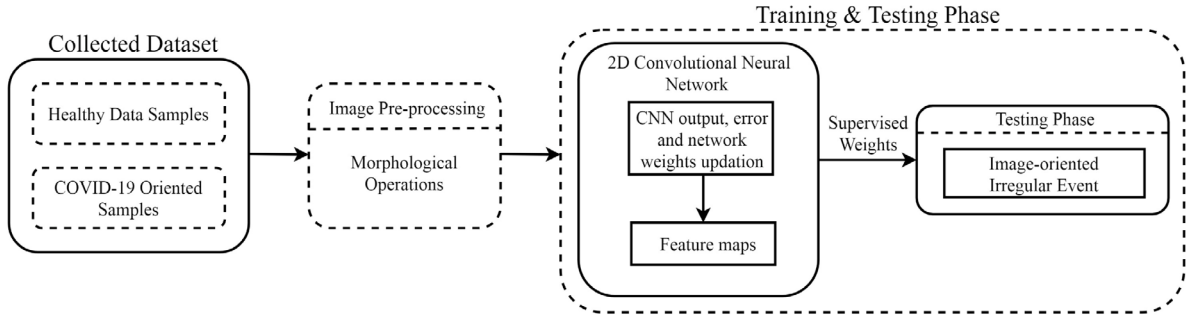


Fig. 4. Image-based irregular event determination.

Table 3
Hyperparameters of *Acoustic_{Prediction}* model.

Model	Hyperparameter	Range	Optimized value
Proposed	Rate of learning	[0.0003, 0.001, 0.01]	0.001
	Size of batch	[16, 32, 64]	32
	Number of epochs	[20, 40, 60, 80, 100]	80
	Sampling rate	[22050, 44100]	44100 Hz
	No. of MFCCs	[13, 26]	13
	Convolution layer	[3, 6, 7, 9, 13]	13
	No. of GRU cells	[32, 64, 128]	128
	Activation function	Fixed	ReLU
	Optimizer	Fixed	Adam
	Loss function	Fixed	Sparse categorical
	Train and test ratio	Fixed	80%–20%

1. Audio-based COVID-19 symptom analysis
2. Image-based COVID-19 symptom analysis
3. Dynamic Fusion Strategy-based overall health determination analysis
4. System reliability analysis
5. System stability analysis

4.1. Audio-based COVID-19 symptom analysis

The total duration of audio is utilized to determine the number of acoustic samples with the rate of the sample. Moreover, the function of segmentation is performed on the acoustic clips to enhance the size of the dataset. After performing the operation of segmentation, multiple samples have been extracted from each segment. At last, the final number of MFCC arrays are acquired by partitioning the acoustic samples based on the length of the hop. The extracted MFCC arrays are extricated and stored as a JSON file that is utilized to train the proposed solution. To extract the MFCC features from the acoustic samples, a librosa-based audio library is utilized. After extracting the features, the JSON file is converted into a 2D format which is fed as an input to train the model to classify the sounds related to the targeted COVID-19 symptoms. The data is divided into two 80% training and 20% testing groups. The proposed model is trained for 80 epochs and the performance is optimized by utilizing the Adam optimizer. In this manner, the proposed solution model is utilized for the symptom determination of COVID-19 from acoustic data samples. The optimal hyper-parameter of the proposed solution is presenting in Table 3.

The symptom determination stability of the proposed solution is achieved by selecting the optimized hyperparameters. The non-linear features are extracted by utilizing the ReLU activation function. Moreover, the issue of overfitting which is very common in deep learning is overcome by utilizing max-pooling layers. It has been analyzed that the process of training has got stuck by achieving the value of local minima over an optimized solution while the learning rate was excessively low. On the other hand, the model converged too quickly and generate non-optimized results when the learning rate was too high. The optimal value of the learning rate was selected by achieving the maximum training and testing accuracy on different models. It has been observed that the model has achieved the maximum symptom determination accuracy with a batch size of 32. Moreover, the model is trained on a different set of epochs such as 20, 40, 60, 80, and 100. From the calculated outcomes, it can be analyzed that the model has achieved the maximum symptom determination accuracy of up to 80 epochs and the model started to overfit by increasing the number of epochs. The calculated outcomes related to the training and validation are illustrated in Fig. 5.

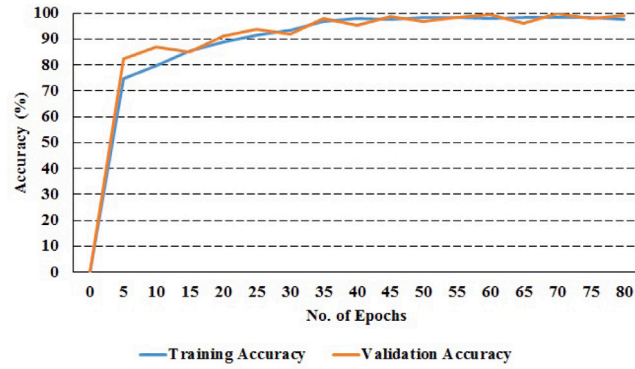
Fig. 5. Training and validation accuracy of *AcousticPrediction* model.Fig. 6. Training and validation loss of *AcousticPrediction* model.

Table 4

Fold-1 based cross validation.

Dataset	Cough	Speech	Breathing
Cough	97.12%	0.0%	0.78%
Speech	0.0%	100%	0.0%
Breathing	0.0%	0.0%	97.45%

Table 5

Fold-2 based cross validation.

Dataset	Cough	Speech	Breathing
Cough	95.24%	0.0%	0.98%
Speech	0.0%	99.05%	0.0%
Breathing	0.85%	0.0%	96.54%

The graph represented the training and validation accuracy of the proposed solution on 80 epochs. It can be analyzed that a significant accuracy has been recorded up to 80 epochs and the rate of loss is decreased. Moreover, Fig. 6 defines the continuous convergence by achieving the maximum rate of accuracy and limited loss at 80 epochs.

Irregular Event-based classification efficiency: In this subsection, all the event-based results and the rate of errors generated by the proposed solution are explained in detail. To evaluate the symptom determination performance, many performance measures such as “Accuracy”, “Precision”, “Recall”, and “F-measure” are calculated. In this manner, four measures “True Positive (TP)”, “False Positive (FP)”, “False Negative (FN)”, and “True Negative (TN)” are represented in the confusion matrix. To validate the performance, the k-fold cross-validation technique has opted where the dataset is divided into 5 folds. In this technique, each fold has got the chance to be a part of the training set, and the testing set and calculated outcomes are represented as follows:

As shown in Tables 4–8, a total of 27 events with respect to COVID-19 symptoms are misclassified by *AcousticPrediction*. From the total number of misclassified events, it can be analyzed that the events falsely predicted by the proposed model are very low as compared to the correctly identified events which are highly acceptable. Similarly, Table 9 represents the symptom determination performance of the proposed solution related to the other performance measures.

Table 6
Fold-3 based cross validation.

Dataset	Cough	Speech	Breathing
Cough	96.21%	0.0%	0.78%
Speech	0.0%	100%	0.0%
Breathing	1.2%	0.0%	95.36%

Table 7
Fold-4 based cross validation.

Dataset	Cough	Speech	Breathing
Cough	96.52%	0.0%	1.3%
Speech	0.0%	99.57%	0.0%
Breathing	0.81%	0.0%	94.36%

Table 8
Fold-5 based cross validation.

Dataset	Cough	Speech	Breathing
Cough	97.41%	0.0%	1.27%
Speech	0.0%	100%	0.0%
Breathing	0.81%	0.0%	96.63%

Table 9
Overall prediction accuracy.

Dataset	Accuracy	Precision	Recall	F-measure
Cough	97.12%	96.96%	94.11%	96.42%
Speech	96.08%	100%	92.87%	91.68%
Breathing	97.45%	100%	93.53%	97.18%
Mean	96.88%	98.99%	92.61%	95.09%

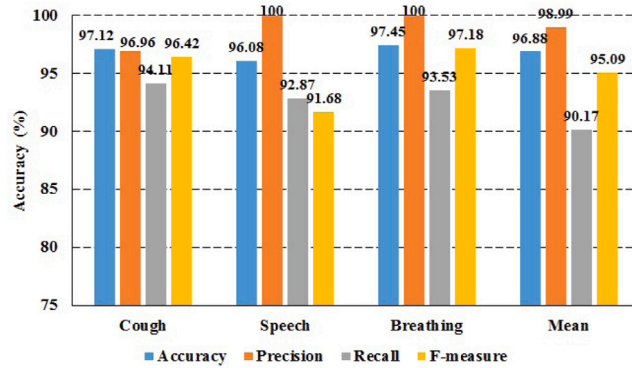


Fig. 7. Event-based symptom determination performance evaluation of *AcousticPrediction* model.

The outcomes shown in Table 9 depicted that the proposed *AcousticPrediction* model accomplished significant accuracy for the targeted symptoms-based event classes. From the calculated outcomes, it can be realized that the proposed solution can achieve the overall accuracy of 96.88% for symptom determination. Similarly, the proposed model has achieved 92.61%, 98.99%, and 95.09% of accuracy for recall, precision, and F-measure, respectively (see Fig. 7).

Comparative analysis: To justify the symptom determination performance of the proposed solution, the most popular acoustic event determination techniques such as CNN, RNN, and LSTM have opted. These approaches are widely used to characterize regular and irregular acoustic events. The calculated outcomes are compared with the state-of-the-art event determination approaches and presented in Table 10.

From the calculated outcomes presented in Table 10, it can be observed that the proposed model has achieved the highest accuracy in terms of Accuracy, Precision, Recall, and F-measure as compared to the CNN, RNN, and LSTM for cough dataset. For the speech dataset, all the models have achieved 100% of accuracy for precision.

Table 10
Comparison of the proposed solution with state-of-the-art models.

Dataset	Model	Accuracy	Precision	Recall	F-measure
Cough	Proposed	97.12%	96.96%	94.11%	95.14%
	CNN	92.57%	86.45%	88.12%	87.28%
	RNN	91.84%	90.90%	82.05%	86.25%
	LSTM	94.25%	92.52%	91.41%	90.48%
Speech	Proposed	96.08%	100%	92.87%	91.41%
	CNN	94.63%	100%	88.54%	90.43%
	RNN	93.91%	100%	85.65%	87.42%
	LSTM	92.98%	100%	89.74%	89.52%
Breathing	Proposed	97.45%	100%	93.53%	97.24%
	CNN	94.33%	95.65%	94.86%	93.25%
	RNN	92.73%	93.44%	92.88%	91.58%
	LSTM	93.56%	94.84%	92.98%	95.43%

Table 11
Hyperparameters of the proposed *Image_{Prediction}* model.

Model	Hyperparameter	Range	Optimized value
Proposed	Rate of learning	[0.0003, 0.001, 0.01]	0.001
	Size of batch	[16, 32, 64, 128]	32
	Epochs	[20, 40, 60, 80, 100]	80
	Convolution layer	[11, 12, 13]	12
	Activation function	Fixed	ReLU
	Optimizer function	Fixed	Adam
	Loss calculator	Fixed	Cross entropy
	Train and test ratio	Fixed	80%–20%

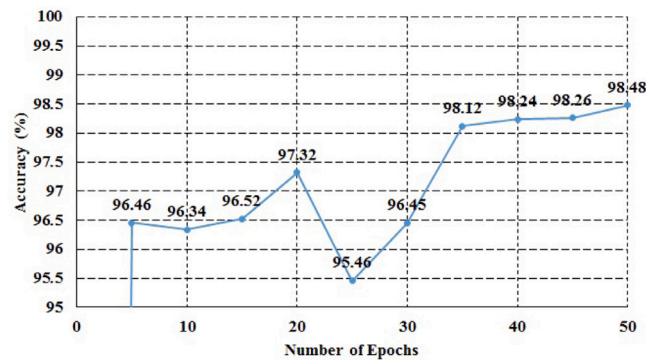


Fig. 8. Training accuracy graph of the proposed network.

4.2. Image-based COVID-19 symptom analysis

The proposed framework is implemented utilizing Pytorch and the fastai module. The training of the proposed solution has been done on 80% of data belonging to CT scan images with the pre-specified requirements. To optimize the process of training, the Adam optimizer function is utilized, and training of one of the models has been done on 80 epochs by setting the learning rate to 0.001. After completing the process of training, the test dataset is utilized to evaluate the symptom determination performance towards the infection in the lungs of the individual. Table 11 shows the optimum hyper-parameters of the proposed approach for image analysis. In addition, Fig. 8 illustrated the accuracy achieved by the proposed solution while training. Moreover, Fig. 9 explained the rate of loss generated by the proposed model.

From the readings illustrated in Figs. 8 and 9, it can be analyzed that the model achieves a satisfactory rate of convergence after 50 epochs of training. The proposed model has taken a total of 5 hours of training time.

Comparative Analysis: Some state-of-the-art models, such as ResNet-50, ResNet-18, ResNet-50, MobileNet-v2, U-Net, and DenseNet-121, are trained and evaluated to support the proposed solution's symptom determination performance. Table 12 shows the results of the calculations.

The proposed model achieved the accuracy of 98.76% for the CT scan dataset, as shown in Table 12. Although, the ResNet-50 has achieved higher recall and F-measure than the proposed model. However, ResNet-50 is having four times the number of layers as compared to the proposed solution. This demonstrates that the proposed methodology provides optimum performance by minimizing

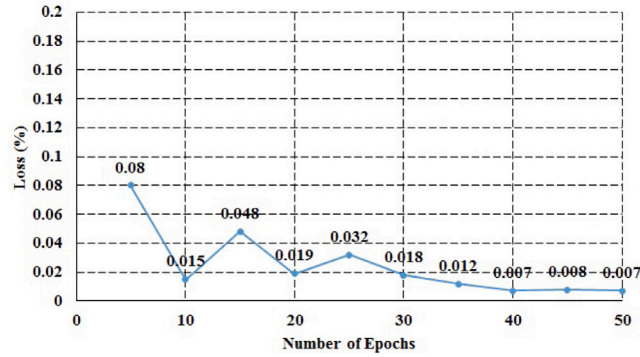


Fig. 9. Training loss graph of the proposed network.

Table 12

Comparison of the proposed model with state-of-the-arts models for image dataset.

Dataset	Framework	Accuracy	Precision	Recall	F-Measure	RoC
CT-Images	Proposed model	98.76%	100%	90.64%	94.55%	99.45%
	ResNet-50	98.33%	100%	96.67%	98.31%	98.90%
	DenseNet-121	96.80%	94.35%	78.95%	88.24%	96.51%
	MobileNetV2	96.80%	100%	86.67%	92.86%	97.64%
	ResNet-18	97.60%	100%	84.21%	91.43%	98.75%
	UNet	97.38%	97.90%	96.68%	97.29%	98.80%

Table 13

Hyperparameters of the dynamic fusion model.

Model	Hyperparameter	Range	Optimized value
Dynamic fusion	Number of decision tree	Fixed	10
	Splitted samples	Fixed	2
	Training ratio	Fixed	80
	Testing ratio	Fixed	20

the network's complexity. Similarly, Table 12 demonstrates the optimum performance of the proposed solution in terms of precision, recall, ROC, and F-measure.

4.3. Dynamic fusion strategy-based overall health determination analysis

In DFS, the correctly predicted image and acoustic events related to *Acoustic_{prediction}* and *Image_{prediction}* models are loaded. In the testing dataset, a total of 100 +ve and 100 -ve samples for breathing, speech, cough, and CT scan images are loaded. It has been observed that there are only 72 positive speech samples. To maximize the range of symptom determination, the technique of data augmentation is applied on 72 samples which helps to increase the total number of test samples. Each test sample is forwarded to its respective trained model as an input to acquire the unimodal outcomes. In this manner, the calculated outcomes with respect to each unimodal are divided into training and testing ratios with the size of 80% and 20%, respectively. The training and testing ratio is utilized by the DFS to calculate the outcome. The misclassified scores are subsequently appended to a vector of misclassification scores that have been used by the DFS for retraining. The performance of the proposed DFS is compared with the traditional Maxvoting Fusion approach. Table 13 shows the optimal hyperparameters of the proposed DFS approach. Table 14 shows the performance comparison of symptom determination.

The performance of the proposed DFS and the commonly known MaxVoting fusion approach is compared and the calculated outcomes are presented in Table 14. From the calculated outcomes, it can be observed that the MaxVoting classifier made inaccurate predictions in a rare test situation with wrong predictions from three unimodal models found. On the other hand, the proposed DFS approach efficiently handled the case and provided correct predictions. Due to the maximum accuracy of individual models, there was no occasion where the unimodal models produced inaccurate predictions at the same time. When the unimodal models make wrong predictions, the dynamic retraining approach changes the learned parameters of the DT model, increasing its accuracy to 98.74%.

4.4. System reliability analysis

It is an important metric for forecasting the framework's overall effectiveness. To assess the dependability of the suggested model, its findings are compared to those of state-of-the-art models. The accuracy level of the four models is determined by the change in

Table 14
Overall symptom determination accuracy.

Model	Accuracy	Precision	Recall	F-measure
Dynamic fusion strategy	98.72%	100%	94.87%	96.68%
MaxVoting fusion	95.54%	100%	93.53%	94.18%

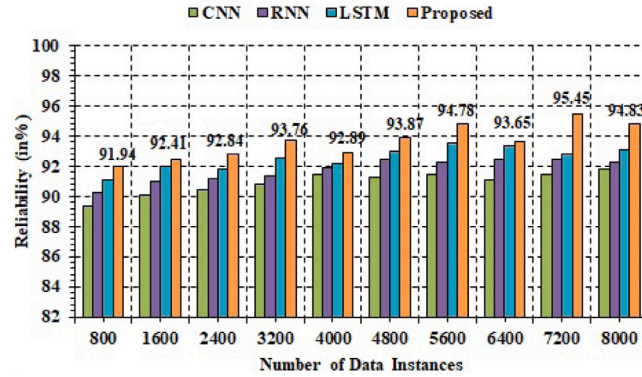


Fig. 10. Reliability analysis.

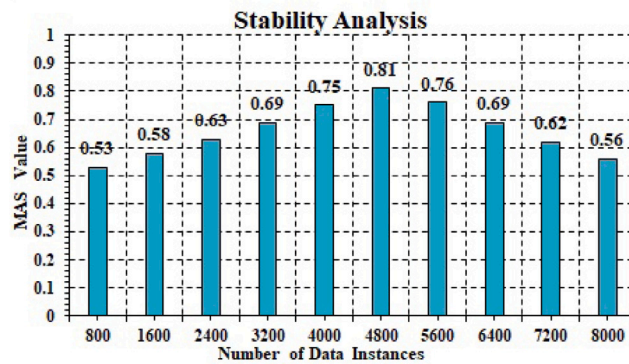


Fig. 11. Stability analysis.

dependability patterns. As the number of instances grows, so does the effectiveness of experimental simulation. Fig. 10 depicts the results of the simulation. As a consequence, when compared to CNN (91.37%), RNN (93.18%), and LSTM (92.14%), the provided model has the greatest reliability score (95.64%).

4.5. System stability analysis

It demonstrates the framework's long-term stability. Mean Absolute Shift (MAS) is a statistic for measuring the stability of a system. The MAS value ranges between 0 and 1, with 0 being the least stable system and 1 representing the most stable system. Fig. 11 depicts the results. The minimum and highest values for MAS are 0.53 and 0.81, respectively, with an average of 0.76. As a consequence, when compared to previous methodologies, the suggested model for forecasting COVID-19 is exceptionally accurate and efficient.

5. Discussion and limitation

The less number of data processing layers makes the proposed *AcousticPrediction* and *ImagePrediction* models less complex. In this manner, the optimized network helps to extract more meaningful features from the acoustic and image data that help to achieve considerable symptom determination accuracy. The overall accuracy of 96.88% and 98.76% with respect to *AcousticPrediction* and *ImagePrediction* network, respectively, defines the considerable symptom determination efficacy as compared to the state-of-the-art models. Similarly, DFS helps to achieve an accuracy of 98.72% as compared to the traditional approach which defines the advantage of the late fusion strategy as compared to early fusion. Moreover, the proposed approach has achieved a considerable symptom

determination reliability of 95.64% by achieving average stability of 0.76. However, a major limitation related to the size of the data is observed. It has been realized that the small size of acoustic data can be a reason for the limited accuracy of the *Acoustic Prediction* network that can be improved by increasing the size of the data.

6. Conclusion

As the surge of COVID-19 colossally affects almost every field such as engineers, researchers, students, and medical practitioners, virologists are continuously trying to find the solution to stop the spread of COVID-19. In this manner, a smart monitoring solution has been proposed for the determination of COVID-19 symptoms by utilizing multimodal samples. The proposed solution is driven by the two unimodal networks named *Acoustic Prediction* and *Image Prediction* for the determination of COVID-19 Symptoms from acoustic and image samples, respectively. Furthermore, the symptom determination outcomes of the unimodal networks are forwarded to the DFS to calculate the probabilistic outcome that helps to make a final decision towards the individuals having COVID-19 symptoms. It has been realized that the *Acoustic Prediction* and *Image Prediction* network has achieved 96.88% and 98.76% accuracy for the determination of COVID-19 symptoms from acoustic and image samples. Even, the small depth of the proposed *Acoustic Prediction* and *Image Prediction* models with the optimal number of 6 and 10 layers makes the deployment of the proposed framework simpler and easy for instant decision-making in real-time. Moreover, DFS has achieved more convincing outcomes for overall decision-making by achieving an accuracy of 98.72% as compared to the MaxVoting Fusion Strategy with an accuracy of 94.18%. Furthermore, the strategy of online retraining is also presented that helps to deal with the noisy environment. This makes the framework more dynamic with limited overhead. The limitation of the proposed study is that the proposed solution can only determine the symptoms of COVID-19. However, disorders related to lungs and pneumonia can also be considered as a future work of the proposed study. Also, the proposed solution can be combined with an IoT platform to achieve self-analysis by delivering a non-contacting diagnostic method that can effortlessly be accessed anywhere.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.compeleceng.2022.108274>.

Data availability

Data will be made available on request.

References

- [1] Ahamad MM, Aktar S, Rashed-Al-Mahfuz M, Uddin S, Liò P, Xu H, ... Moni MA. A machine learning model to identify early-stage symptoms of SARS-Cov-2 infected patients. *Expert Syst Appl* 2020;160:113661. <http://dx.doi.org/10.1016/j.eswa.2020.113661>.
- [2] Vedaei SS, Fotovvat A, Mohebbian MR, Rahman GM, Wahid KA, Babyn P, . Sami R. COVID-SAFE: an IoT-based system for automated health monitoring and surveillance in post-pandemic life. *IEEE Access* 2020;8:188538. <http://dx.doi.org/10.1109/ACCESS.2020.3030194>.
- [3] Hlaing PM, Nopparatjamjornas TR, Nopparatjamjornas S. Digital technology for preventative health care in Myanmar. *Digit Med* 2018;4(3):117. http://dx.doi.org/10.4103/digm.digm_25_18.
- [4] Pinkas G, Karny Y, Malachi A, Barkai G, Bachar G, Aharonson V. SARS-CoV-2 detection from voice. *IEEE Open J Eng Med Biol* 2020;1:268–74. <http://dx.doi.org/10.1109/OJEMB.2020.3026468>.
- [5] Vijayakumar DS, Sneha M. Low cost Covid-19 preliminary diagnosis utilizing cough samples and keenly intellectual deep learning approaches. *Alex Eng J* 2021;60(1):549–57. <http://dx.doi.org/10.1016/j.aej.2020.09.032>.
- [6] Asiaee M, Vahedian-Azimi A, Atashi SS, Keramatfar A, Nourbakhsh M. Voice quality evaluation in patients with COVID-19: An acoustic analysis. *J Voice* 2020;1–7. <http://dx.doi.org/10.1016/j.jvoice.2020.09.024>.
- [7] Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, . Nabeel M. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlocked* 2020;20:100378. <http://dx.doi.org/10.1016/j.imu.2020.100378>.
- [8] Quatieri TF, Talkar T, Palmer JS. A framework for biomarkers of COVID-19 based on coordination of speech-production subsystems. *IEEE Open J Eng Med Biol* 2020;1:203–6. <http://dx.doi.org/10.1109/OJEMB.2020.2998051>.
- [9] Laguarda J, Huetto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open J Eng Med Biol* 2020;1:275–81. <http://dx.doi.org/10.1109/OJEMB.2020.3026928>.
- [10] El-Kenawy ESM, Ibrahim A, Mirjalili S, Eid MM, Hussein SE. Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. *IEEE Access* 2020;8:179317–179335. <http://dx.doi.org/10.1109/ACCESS.2020.3028012>.
- [11] Han Z, Wei B, Hong Y, Li T, Cong J, Zhu X, . Zhang W. Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning. *IEEE Trans Med Imaging* 2020;39(8):2584–94. <http://dx.doi.org/10.1109/TMI.2020.2996256>.
- [12] Oh Y, Park S, Ye JC. Deep learning COVID-19 features on CXR using limited training data sets. *IEEE Trans Med Imaging* 2020;39(8):2688–700. <http://dx.doi.org/10.1109/TMI.2020.2993291>.
- [13] Wang N, Liu H, Xu C. Deep learning for the detection of COVID-19 using transfer learning and model integration. In: 2020 IEEE 10th international conference on electronics information and emergency communication (ICEIEC). 2020, p. 281–4. <http://dx.doi.org/10.1109/ICEIEC49280.2020.9152329>.
- [14] Ibrahim DA, Zebari DA, Mohammed HJ, Mohammed MA. Effective hybrid deep learning model for COVID 19 patterns identification using CT images. *Expert Syst* e13010. <http://dx.doi.org/10.1111/exsy.13010>.
- [15] Alloui M, Mohammed MA, Benameur N, Al-Khateeb B, Abdulkareem KH, Garcia-Zapirain B, . Maskeliūnas R. A multi-agent deep reinforcement learning approach for enhancement of COVID-19 CT image segmentation. *J Personalized Med* 2022;12(2):309. <http://dx.doi.org/10.3390/jpm12020309>.
- [16] Abdulkareem KH, Mostafa SA, Al-Qudsy ZN, Mohammed MA, Al-Waisy AS, Kadry S, . Nam Y. Automated system for identifying COVID-19 infections in computed tomography images using deep learning models. *J Healthc Eng* 2022;2022. <http://dx.doi.org/10.1155/2022/5329014>.

- [17] Hasoon JN, Fadel AH, Hameed RS, Mostafa SA, Khalaf BA, Mohammed MA, Nedoma J. COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images. *Results Phys* 2021;31:105045. <http://dx.doi.org/10.1016/j.rinp.2021.105045>.
- [18] Alyasseri ZAA, AL Betar MA, Doush IA, Awadallah MA, Abasi AK, Makhadmeh SN, . Zitar RA. Review on COVID 19 diagnosis models based on machine learning and deep learning approaches. *Expert Syst* 2022;39(3):e12759. <http://dx.doi.org/10.1111/exsy.12759>.
- [19] Nguyen TT, Nguyen QVH, Nguyen DT, Hsu EB, Yang S, Eklund P. Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions. 2020, <http://dx.doi.org/10.48550/arXiv.2008.07343>, arXiv preprint [arXiv:2008.07343](https://arxiv.org/abs/2008.07343).
- [20] Maghded HS, Ghafoor KZ, Sadiq AS, Curran K, Rawat DB, Rabie K. A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: design study. In: 2020 IEEE 21st international conference on information reuse and integration for data science (IRI). IEEE; 2020, p. 180–7. <http://dx.doi.org/10.1109/IRI49571.2020.00033>.
- [21] Rao ASS, Vazquez JA. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine. *Infect Control Hosp Epidemiol* 2020;41(7):826–30. <http://dx.doi.org/10.1017/ice.2020.61>.
- [22] Allam Z, Jones DS. On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. In: *Healthcare*, Vol. 8. Multidisciplinary digital publishing institute; 2020, p. 46. <http://dx.doi.org/10.3390/healthcare8010046>, No. 1.
- [23] Fatima SA, Hussain N, Balouch A, Rustam I, Saleem M, Asif M. IoT enabled smart monitoring of coronavirus empowered with fuzzy inference system. *Int J Adv Res Ideas Innov Technol* 2020;6(1):188–94.
- [24] Angelov P, Almeida Soares E. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv* 2020. <http://dx.doi.org/10.1101/2020.04.24.20078584>.
- [25] Jayachitra VP, Nivetha S, Nivetha R, Harini R. A cognitive IoT-based framework for effective diagnosis of COVID-19 using multimodal data. *Biomed Signal Process Control* 2021;70:102960. <http://dx.doi.org/10.1016/j.bspc.2021.102960>.

Ankush Manocha received a Ph.D. degree in computer science from Lovely Professional University, India. His current research interests include IoT, Computer Vision, Cyber-Physical Systems, Edge Computing, and Cloud Computing. He has authored several research papers, including in International journals, such as WILEY Transactions, IEEE, ELSEVIER, Taylor and Francis, and SPRINGER.

Munish Bhatia works as an Assistant Professor in the CSE Department of LPU, India. He has published in reputed journals like Wiley Transactions, IEEE, Elsevier, and Springer. Previously, he worked as Project Researcher at IIT MANDI under DBT, India Project titled FARMERZONE. He is a reviewer of numerous International Journals including NATURE, IEEE Transactions, ELSEVIER, SAGE, WILEY, Taylor and Francis.