Gradient Boosting Machine and Efficient Combination of Features for Speech-Based Detection of COVID-19

Tusar Kanti Dash, *Senior Member, IEEE*, Chinmay Chakraborty, *Senior Member, IEEE*, Satyajit Mahapatra, and Ganapati Panda, *Life Senior Member, IEEE*

***Abstract*—In recent times, speech-based automatic disease detection systems have shown several promising results in biomedical and life science applications, especially in the case of respiratory diseases. It provides a quick, cost-effective, reliable, and non-invasive potential alternative detection option for COVID-19 in the ongoing pandemic scenario since the subject's voice can be remotely recorded and sent for further analysis. The existing COVID-19 detection methods including RT-PCR, and chest X-ray tests are not only costlier but also require the involvement of a trained technician. The present paper proposes a novel speech-based respiratory disease detection scheme for COVID-19 and Asthma using the Gradient Boosting Machine-based classifier. From the recorded speech samples, the spectral, cepstral, and periodicity features, as well as spectral descriptors, are computed and then homogeneously fused to obtain relevant statistical features. These features are subsequently used as inputs to the Gradient Boosting Machine. The various performance matrices of the proposed model have been obtained using thirteen sound categories' speech data collected from more than 50 countries using five standard datasets for accurate diagnosis of respiratory diseases including COVID-19. The overall average accuracy achieved by the proposed model using the stratified k-fold cross-validation test is above 97%. The analysis of various performance matrices demonstrates that under the current pandemic scenario, the proposed COVID-19 detection scheme can be gainfully employed by physicians.**

***Index Terms*—COVID-19 detection, LightGBM, speech classification, feature fusion, health informatics.**

I. Introduction

RECENT developments in speech signal processing have shown numerous clinical applications for non-invasive diagnosis of diseases which helps in effective remote health monitoring and remote healthcare facilities [1], [2], [3], [4]. In the current Coronavirus Disease 2019 (COVID-19) pandemic scenario, this speech-based remote health monitoring system can play a crucial role. According to the World Health Organization data, more than 579 million people have suffered including six million deaths reported till August 8, 2022, due to COVID-19 [5]. The standard and reliable test of COVID-19 is the Reverse transcription-polymerase chain reaction test (RT-PCR) test which is expensive (US$125 per test package, and over $15,000 to set up a processing lab) and also time-consuming (4–6 hours of processing time, and a turn-around of 2–4 days, including shipping) [6]. To deal with this challenging situation, there is a huge requirement for large-scale testing for isolating infected individuals and contact tracing [7]. Under this scenario, speech-based COVID-19 detection (CD) is one of the simplest, safest as well cost-effective methods [8].

Several temporal and spectral acoustic features of subjects have been used as inputs to a random forest model for the classification of speech into nine categories such as shallow and deep breathing, shallow and heavy cough, sustained vowel phonation (/o/, /e/, /a/), and normal and fast counting [9]. Detection accuracy of 66.74 % is reported in this study. In [10], respiratory sounds such as cough and breathing have been employed to classify COVID-19 from asthma using 733-dimensional features including 477-dimensional handcrafted features and 256-dimensional VGGNet-based features. The Logistic Regression-based classifier is used to provide an area under the receiver operator characteristic curve (ROC-AUC) of above 80%. The CD from online available speech data has been carried out using phoneme level analysis, Mel filter bank features, and the SVM classifier. It is reported that an accuracy of 88.6% is achieved from a limited number of 19 speakers [11]. An automated machine learning-based COVID-19 classification model is developed using glottal, prosodic, and spectral features from short-duration speech segments [12]. The proposed model yields a classification accuracy of 80%. Modified cepstral features are extracted from two speech databases and fed to the support vector machine (SVM) classifiers for CD and maximum

Manuscript received 15 April 2022; revised 23 June 2022 and 20 July 2022; accepted 6 August 2022. Date of publication 10 August 2022; date of current version 7 November 2022. *(Corresponding author: Chinmay Chakraborty.)*

Tusar Kanti Dash and Ganapati Panda are with Electronics and Communications Engineering, C V Raman Global University, Bhubaneswar 752054, India (e-mail: tusarkantidash@gmail.com; ganapati.panda@gmail.com).

Chinmay Chakraborty is with Electronics and Communication Engineering, Birla Institute of Technology, Mesra 835215, India (e-mail: cchakrabarty@bitmesra.ac.in).

Satyajit Mahapatra is with the School of Electrical and Electronics Engineering, VIT Bhopal University, Bhopal 466114, India (e-mail: satyajit6243@gmail.com).

Digital Object Identifier 10.1109/JBHI.2022.3197910

accuracy of 85% is obtained [13]. Transfer learning-based deep neural network classifiers are used for CD for cough, breath, and speech with a ROC-AUC of 0.982, 0.942, and 0.923 respectively [14]. Several machine learning-based algorithms are analyzed for the mobile health solutions of CD and it is observed that the SVM technique provides the highest accuracy of 97% for the Coswara database [15]. A mobile application is developed for CD by combining the symptoms checker with voice, breath, and cough signals for robust performance on openly sourced and noisy data sets by using deep CNN and gradient boosting [16].

Even though several speech-based CD methods have been proposed, there is still scope for improvement in terms of detection accuracy, computational complexity as well as testing on multiple datasets in different categories of speech. As the early CD is essential, the higher and more reliable accuracy of detection is very important which would drastically reduce the spread and medical emergency of the detection. Additionally, many researchers have focused on using chest X-rays for CD using several image processing techniques [17], [18], [19], [20], [21]. Although it achieved superior performance in terms of accuracy but acquisition of chest X-rays is a cumbersome task. A physical visit, a well-trained technician for successful data acquisition, and a medical practitioner are all required. In light of these considerations, the current research focuses on the development of an improved CD system based on speech. For efficient extraction of information from the speech samples, an effective combination of speech features is used in this paper along with Light Gradient Boosting Machine which was proposed by Microsoft in 2016 [22]. It provides improved training performance requiring minimum memory, and parallel processing ability as well as handling large-scale data compared to the traditional machine learning algorithms. In recent years, it has been employed for genomics data analysis [23], speech processing [16], image processing [24], arrhythmia detection [25], and others. Because of the associated advantages, the gradient boosting technique is chosen in the current implementation to achieve better classification performance. The main research contributions of the paper are listed below:

- Application of intelligent preprocessing techniques to bring the speech quality of the different real-life recorded speech to equal acoustic levels.
- Extraction of spectral, cepstral, and periodicity features at frame level for efficient combination of high dimensional relevant audio features at sample level to accurately detect several respiratory diseases including COVID-19 and Asthma.
- Development of Gradient Boosting Machine as a classifier and comparison of the detection performance matrices of the proposed method with those obtained from the standard methods using five datasets in thirteen different categories.
- Assessment of the generalization ability of the proposed model which can be presented as a clinical application method wherein the model is trained with a large number of speech samples from the cough category of multiple datasets. Later it can predict the condition of the patient from his/her cough sound.

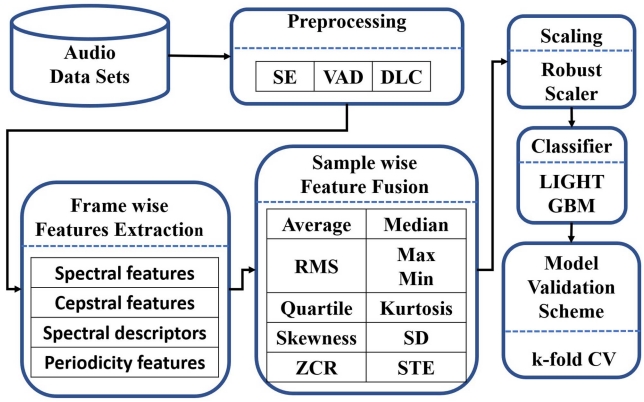

Fig. 1. Block diagram of the proposed speech-based COVID-19 detection scheme.

The paper is organized into four sections with Section I dealing with the introduction, literature review, motivations, and objectives of the investigation. The details of the materials and methods employed are dealt with in section II. Section III contains an analysis of results, and contributions in terms of research findings. The outcome of the research, limitations, and future research scope are presented in section IV.

II. Material and Methods

The block diagram of the proposed speech-based COVID-19 detection scheme is presented in Fig. 1 consisting of the following steps: dataset collection, preprocessing and features extraction, scaling of features, classification model training, and validation, and performance evaluation.

A. Datasets

Five datasets have been used to evaluate the performance of the suggested model in this study. These are: Coswara (Dataset-1) [9], Crowdsourced respiratory by the University of Cambridge (Dataset-2) [10], Virufy (Dataset-3) [26], recorded interviews from online platforms in telephone quality speech (Dataset-4) [11], Coughvid (Dataset-5) [7]. Out of these, data set-2 is used for both binary (COVID-19 positive, and healthy) and multi-class classification (COVID-19 positive, Asthma positive, and healthy) whereas datasets-1,3,4,5 are used for the binary classification task. These datasets contain speech samples of subjects from more than 50 countries. The dataset preparation follows a standard technique as shown in Fig. 2. Due to the deadly spreading nature of the COVID-19, the speech samples are recorded for most of the speech datasets in the online mode either by using mobile or web-based applications [7], [9], [10], [11], [26]. Along with the audio samples the COVID-19 status, location, gender, age, and the health conditions of the patients are also stored. The brief details of these five datasets are listed in Table I. A total of 4178 speech samples have been used in the simulation study.

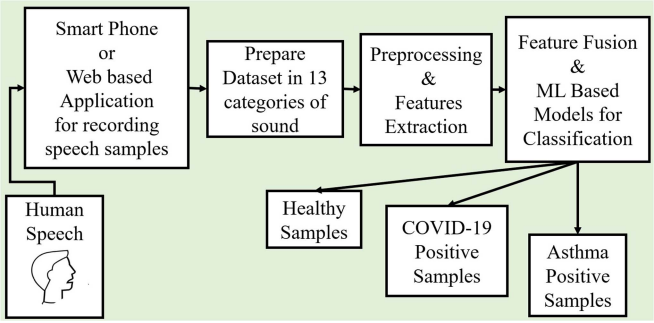

Fig. 2. Flowchart for the dataset preparation and classification.

TABLE I
DETAILS OF THE FIVE EXPERIMENTAL DATASETS USED IN THE SIMULATION

Name of Dataset	Categories	Number of speech samples in each class
Dataset-1 [?]	Breathing-deep	50 N + 47 P
	Breathing-shallow	49 N + 47 P
	Cough-heavy	50 N + 47 P
	Cough-shallow	50 N + 47 P
	Counting-Fast	17N + 42 P
	Counting-Normal	17N + 42 P
	Vowel-/o/	50N + 47 P
	Vowel-/e/	50N + 47 P
	Vowel-/a/	50N + 47 P
Dataset-2 [?]	Breathing	64N + 46P + 167 AP
	Cough	200N +47 P + 112 AP
Dataset-3 [?]	Cough	73 N + 48 P
Dataset-4 [?]	Spoken Sentence	237 N + 465 P
Dataset-5 [?]	Cough	1155 N + 1155 P

* The classes are named as COVID-19 Positive (P), COVID-19 Negative or healthy (N), and Asthma Positive (AP)

Complete details of these datasets are given in supplementary information S1.

B. Preprocessing

Speech preprocessing is critical to the overall success of developing a robust and efficient speech recognition system [27]. When speech is recorded by different users in different environments, then the speech quality varies drastically in one category within the dataset as well as across different datasets [28]. The background noise level significantly affects the overall performance of the speech recognition system [29], [30]. For highly non-stationary situations, the noise level is computed using the noise estimation algorithm [31]. To evaluate the effect of preprocessing, the variation in noise level and coefficient of variation are plotted in Figures 3 and 4 for two cases before and after preprocessing. The coefficient of variation measures the variation in the noise level by calculating the ratio between the standard deviation and mean of the estimated noise levels for one class [32]. For the noise level estimation, the cough category sound is used for dataset-1,2,3,5 and complete sentence sounds for dataset-4. The steps involved in preprocessing are mentioned below.

1) Low Pass Filtering: The sampling frequency of speech signals is different for different datasets. However, significant information is found within the 8 kHz bandwidth [33]. It is also

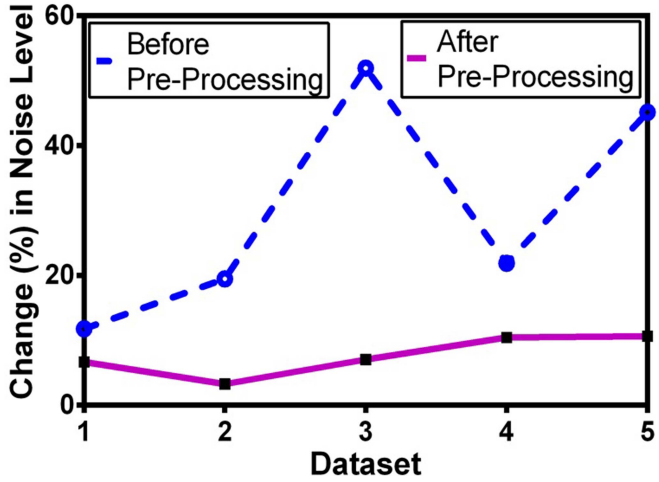

Fig. 3. Change in the noise level and between positive and negative class.

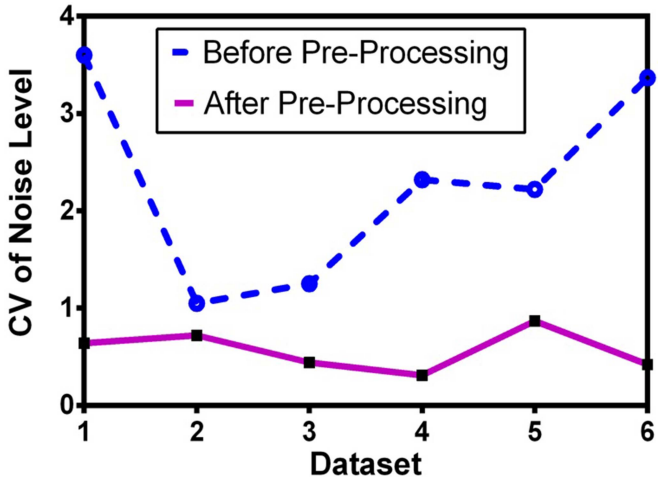

Fig. 4. Change in Coefficient of variation (CV) of noise level between positive and negative class.

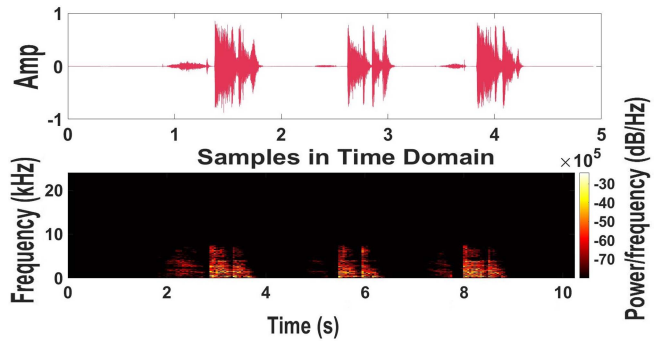

Fig. 5. Spectrogram of the cough signal from dataset-2.

evident from Fig. 5, where the time-frequency representation of one cough signal of dataset-2 is plotted using the spectrogram. To remove the unwanted signal components which are not associated with human speech, all the audio signals are passed through a low pass filter of 10 kHz. To maintain a uniform sampling rate and to extract the same number of features for

each frame, all speech signals are resampled at the maximum available sampling frequency (48 kHz) of all the datasets.

2) Speech Enhancement: The multi-band spectral subtraction approach has been employed to denoise the speech samples of all five datasets [34]. This is a simple and effective method for denoising signals affected by colored noises where spectral subtraction is performed separately at different frequency bands.

3) Voice Activity Detection and Dynamic Level Control: To separate the voiced frames from the unvoiced frames, a simple short-term energy-based voice activity detection (VAD) algorithm is used. The voiced frames are then passed through a Dynamic Level Controller (DLC). It is made up of an expander and a compressor, with the expander boosting low signal levels and the compressor lowering peak levels [35].

C. Features Extraction

In this section, the details of the audio features extraction techniques used in the investigation are dealt with. At the frame and sample levels, numerous audio features are extracted in the frequency, structural, statistical, and temporal domains. The complete recording of a single user in one category comprises one sample, while a frame is a subset of the entire audio data found in a sample. Considering there is 'n' number of frames present in each sample, the details of the frame-level features are described below. The features are named as f(serial number of the feature) such as f1 to f5701.

- Spectral Features — The speech signal is a non-stationary signal but the properties remain constant over fixed time intervals of 10–30 ms. The short-time spectral features are obtained by converting the time domain signal into the frequency domain by applying different Transform techniques. These features provide information about spectral information which plays an important role in speech recognition [36]. In this work, the hamming window is chosen as it provides less spectral leakage and the side lobes of this window are lower than the others [37]. A window size of 25 msec duration with 50% overlapping between two successive frames has been considered. The spectral features extracted are: Linear Spectrum (n×512), Mel Spectrum (n×32), Bark Spectrum (n×32), and Equivalent Rectangular Bandwidth (ERB) Spectrum (n×44). Therefore, the total dimension of spectral features is (n×620).
- Cepstral Features — The cepstral features help in extracting relevant speech information for speech emotion recognition tasks by using filter banks based on human speech perception [13]. The cepstral features are Mel-frequency cepstral coefficients (MFCC), MFCC Delta, MFCC Delta Delta, Gammatone cepstral coefficients (GTCC), GTCC Delta, GTCC Delta Delta, each of dimension (n×13). Therefore, the total dimension of cepstral features is (n×78)
- Spectral Descriptors — These features extract statistical information from the lengthy spectral features. These features are widely used in speaker, music, mood recognition, and classification tasks [38]. The spectral descriptors used are: Centroid, Crest, Decrease, Entropy, Flatness, Flux, Kurtosis, Roll-off Point, Skewness, Slope, and Spread, each having dimension (n×1). The total dimension of spectral descriptors is (n×11).
- Periodicity Features —These features provide important time-domain information of speech which helps in monaural speech analysis [39]. The features used are: Pitch (n×1), and Harmonic Ratio (n×1).

For this purpose, MATLAB-based audioFeatureExtractor is used [40], [41]. The fusion of spectral features, cepstral features, spectral descriptors, and periodicity features yields an n* 712-dimensional feature vector for each speech sample. As the frame numbers vary for each sample, so training in machine learning becomes difficult. Therefore, in this work, the statistical measures are computed at the sample level and it provides a fixed length of features for each sample. To extract statistical distributions at the sample level, several statistical features are extracted from the frame-level features [10]. The sample level features are: mean (f1:f712), median (f713:f1424), RMS (root-mean-square) (f1425:f2136), maximum (f2137:f2848), minimum (f2849:f3560), quartile (1st and 3 rd quartile, interquartile range) (f3561:f3563), standard deviation (SD) (f3564:f4275), skewness (f4276:f4987), kurtosis (f4988:f5699) of all frame-level features. Also, the Zero crossing rate (ZCR) (f5700), and Short-time energy (STE) (f5701) are calculated sample-wise. Each combined feature vector is the concatenation of the sample level features and it is a 5701-dimensional feature vector. Outliers in the high-dimensional feature vector can have an impact on the learning algorithm's performance. As a result, feature scaling is an important preprocessing step. The robust scaler removes the median and scales the data according to the quantile range, removing outliers from the features [42].

D. LightGBM (LGM)

The LGM is an effective gradient boosting decision tree with gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) to increase computational efficiency without affecting the accuracy [22]. The steps involved in LGM modeling are: (i) defining the loss function, (ii) performing the GOSS sampling, and identification of the optimal segmentation point using a histogram-based algorithm, (iii) calculation of feature dimension by the EFB method, (iv) performing the leaf-wise algorithm to combine the samples to fit residuals, and (v) splitting the nodes based on the objective function and generate a decision tree.

Let us consider **X** as the input feature vector and **Y** as the class labels. The aim of LGM is to determine the approximation function $\widehat{F}(x)$ so that the loss function $(L(y, F(x)))$ gets minimized [43].

$$\widehat{F}(x) = \underset{F}{argmin}\ E_{xy}\left[L(y, F(x))\right] \tag{1}$$

The final LGM model ($F_M(X)$) is formed using M decision trees such that

$$F_M(X) = \sum_{m=1}^{M} F_m(X) \tag{2}$$

The LGM is trained in an additive form at step m and can be expressed as:

$$\tau_m = \sum_{i-1}^{n} L(y_i\ ,\ F_{m-1}(x_i) + F_m(x_i)) \cong \sum_{i=1}^{n} (g_i F_m(x_i))\ + \frac{1}{2} h_i F_m^2(x_i)) \tag{3}$$

Where, g_i and h_i represent the first and second-order gradient statistics of the loss function. By denoting the sample set I_j of leaf $j(1 \le j \le J)$(3) can be written as:

$$\tau_m = \sum_{j=1}^{J} \left(\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right) \tag{4}$$

The optimal leaf weight scores of each leaf node (w_j^*) is calculated as:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{5}$$

Let, I_L and I_R are the sample sets of the left and right branches, respectively. The leaf weight, the regular penalty factor, λ is used as a smoothing parameter in calculating gain in the process of splitting points. The objective function after adding the split is then calculated as

$$G = \frac{1}{2} \left(\left| \frac{\left(\sum_{i \in I_L} g_i\right)^2}{\left(\sum_{i \in I_L} h_i + \lambda\right)} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\left(\sum_{i \in I_R} h_i + \lambda\right)} + \frac{\left(\sum_{i \in I} g_i\right)^2}{\left(\sum_{i \in I} h_i + \lambda\right)} \right| \right) \tag{6}$$

In the conventional gradient boosting technique, the tree grows horizontally, while in LGM the tree grows vertically which makes it an efficient tool for processing large-scale data and features [43]. The GOSS technique of LGM effectively selects the input features with larger gradients and removes the features with smaller gradient values. This works as feature reduction in the current implementation where the input feature size is relatively higher and thereby, it increases the efficiency of the detection model.

III. Results and Discussions

The performance of the proposed model is assessed for two tasks, (I) binary classification task to predict the speech samples as COVID-19 positive or negative, and (II) multiclass classification task to predict COVID-19 positive, Asthma positive, and healthy speech samples. To perform this, the speech samples are passed through the additional preprocessing blocks such as low pass filtering, speech enhancement, voice activity detection, and dynamic level control. Then a total of 5701 features are extracted from each sample. Here, the preprocessing block is a part of the feature extraction. These features are combined with an LGM classifier and three baseline classifiers such as Random Forest (RF) [9], SVM [10], [11], and K-Nearest Neighbor (KNN) [44] used for the speech classification task. For the development of the classification model five-fold stratified cross-validation scheme is employed. Standard performance measures as reported in [45] such as Classification Accuracy (CA), F-2 Score (F-2), Precision (PR), Recall (RC), and area under the curve (AUC), are employed in this study. The details of the performance measures are described in supplementary information S2. Grid search is used to find the optimal parameters of the classifiers. These parameters are listed in supplementary information S3.

TABLE II
PERFORMANCE COMPARISON FOR DATASET-1 USING 5701 FEATURE VECTOR FOR BINARY CLASSIFICATION

Category	Evaluation Measures	LGM	SVM	RF	KNN
Breathing Deep (D-1)	CA	**0.969**	0.557	**0.969**	0.691
	F-2	**0.969**	0.502	**0.969**	0.687
	PR	**0.969**	0.641	**0.969**	0.697
	RC	**0.969**	0.557	**0.969**	0.691
	AUC	**0.968**	0.543	**0.968**	0.687
Breathing Shallow (D-1)	CA	**0.99**	0.5	0.948	0.604
	F-2	**0.99**	0.421	0.948	0.602
	PR	**0.99**	0.258	0.95	0.606
	RC	**0.99**	0.5	0.948	0.604
	AUC	**0.989**	0.489	0.947	0.602
Cough Heavy (D-1)	CA	**0.979**	0.598	0.969	0.773
	F-2	**0.979**	0.555	0.969	0.773
	PR	**0.979**	0.695	0.969	0.773
	RC	**0.979**	0.598	0.969	0.773
	AUC	**0.979**	0.586	0.968	0.772
Cough Shallow (D-1)	CA	**0.979**	0.701	0.969	0.701
	F-2	**0.979**	0.684	0.969	0.696
	PR	**0.98**	0.756	0.971	0.719
	RC	**0.979**	0.701	0.969	0.701
	AUC	**0.978**	0.693	0.968	0.704
Vowel-/a/ (D-1)	CA	**0.99**	0.505	0.959	0.557
	F-2	**0.99**	0.427	0.958	0.538
	PR	**0.99**	0.263	0.962	0.565
	RC	**0.99**	0.505	0.959	0.557
	AUC	**0.99**	0.49	0.957	0.548
Vowel-/e/ (D-1)	CA	**0.99**	0.515	0.959	0.742
	F-2	**0.99**	0.434	0.959	0.737
	PR	**0.99**	0.266	0.959	0.759
	RC	**0.99**	0.515	0.959	0.742
	AUC	**0.99**	0.5	0.958	0.737
Vowel-/o/ (D-1)	CA	0.969	0.866	**0.979**	0.68
	F-2	0.969	0.866	**0.979**	0.678
	PR	0.969	0.866	**0.979**	0.685
	RC	0.969	0.866	**0.979**	0.68
	AUC	0.969	0.866	**0.979**	0.677
Counting Normal (D-1)	CA	**0.966**	0.712	**0.966**	0.712
	F-2	**0.966**	0.659	**0.966**	0.687
	PR	**0.966**	0.507	**0.966**	0.668
	RC	**0.966**	0.712	**0.966**	0.712
	AUC	**0.958**	0.5	**0.958**	0.552
Counting Fast (D-1)	CA	**0.949**	0.712	0.932	0.644
	F-2	**0.949**	0.669	0.932	0.607
	PR	**0.949**	0.656	0.932	0.492
	RC	**0.949**	0.712	0.932	0.644
	AUC	**0.929**	0.517	0.917	0.452

A. Performance Evaluation as a Binary Classification Task

The comparative study between the performance of LGM, SVM, RF, and KNN classifiers for binary classification task are presented in Tables II and III. The LGM classifier provides an average accuracy of 0.978, an F-2 Score of 0.979, and an AUC of

TABLE III
PERFORMANCE COMPARISON FOR DATASET-2,3,4,5 USING 5701 FEATURE VECTOR FOR BINARY CLASSIFICATION

Category (Dataset)	Evaluation Measures	LGM	SVM	RF	KNN
Cough (D-2)	CA	**0.992**	0.966	**0.992**	0.983
	F2	**0.992**	0.966	**0.992**	0.982
	PR	**0.992**	0.966	**0.992**	0.960
	RC	**0.992**	0.966	**0.992**	0.957
	AUC	**0.986**	0.955	0.980	0.940
Breathing (D-2)	CA	**0.982**	0.909	0.964	0.818
	F2	**0.982**	0.909	0.964	0.815
	PR	**0.982**	0.91	0.964	0.824
	RC	**0.982**	0.909	0.964	0.818
	AUC	**0.981**	0.9	0.959	0.797
Cough (D-3)	CA	**0.969**	0.965	0.942	0.793
	F2	**0.968**	0.965	0.941	0.781
	PR	**0.969**	0.965	0.947	0.823
	RC	**0.969**	0.965	0.942	0.793
	AUC	**0.961**	0.960	0.927	0.746
Sentence (D-4)	CA	**0.992**	0.991	**0.992**	0.928
	F2	**0.992**	0.991	**0.992**	0.928
	PR	**0.993**	0.992	**0.993**	0.928
	RC	**0.992**	0.991	**0.992**	0.928
	AUC	**0.999**	0.988	**0.999**	0.985
Cough (D-5)	CA	**0.998**	0.993	**0.998**	0.921
	F2	**0.998**	0.993	**0.998**	0.922
	PR	**0.998**	0.994	**0.998**	0.924
	RC	**0.998**	0.993	**0.998**	0.921
	AUC	**0.999**	0.985	**0.999**	0.971

TABLE IV
PERFORMANCE EVALUATION FOR DETECTION COVID-19 POSITIVE, NEGATIVE AND ASTHMA FROM DATASET-2 USING 5701 FEATURE VECTORS

Category	Evaluation Measures	LGM	SVM	RF	KNN
Cough	CA	**0.971**	0.947	0.943	0.915
	F-2	**0.971**	0.946	0.941	0.915
	PR	**0.973**	0.949	0.946	0.914
	RC	**0.971**	0.947	0.943	0.915
	AUC	**0.991**	0.969	0.985	0.974
Breathing	CA	**0.981**	0.89	0.963	0.854
	F-2	**0.981**	0.889	0.963	0.851
	PR	**0.983**	0.893	0.965	0.854
	RC	**0.981**	0.89	0.963	0.854
	AUC	**0.999**	0.949	0.994	0.918

0.976 across all the categories in the five datasets. The average accuracy, F-2 Score, and AUC of the SVM classifier are 0.749, 0.717, and 0.712, respectively. Similarly, for the RF classifier, the average accuracy, F-2 score, and AUC are found to be 0.967, 0.966, and 0.963, respectively. For the KNN classifier, the values are 0.753, 0.745, and 0.728. The results show that the LGM classifier performs better on the high-dimensional features than the SVM, RF, and KNN classifiers.

B. Performance Evaluation as a Three-Class Classification Task

To further evaluate the prediction ability of the classifiers, an assessment of multi-class data has been carried out for dataset-2 contains samples of COVID-19 positive, Asthma positive, and healthy in the cough and breathing sound categories. The results are listed in Table IV. It is observed that the performance of the LGM classifier is superior in all the performance measures as compared to the SVM, RF, and KNN classifiers respectively. The ROC curves are two-dimensional plots that provide the relative

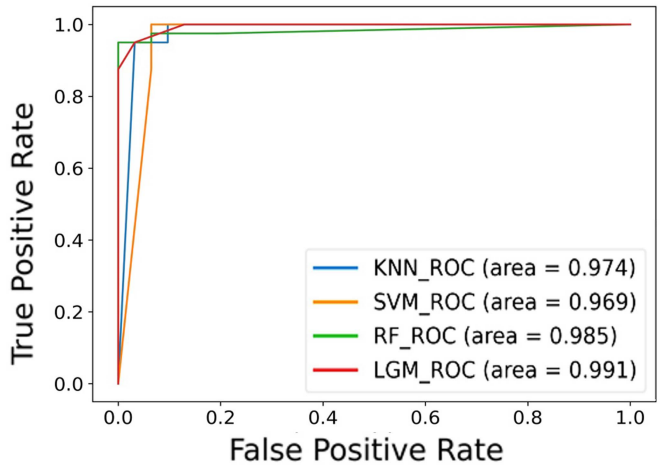

Fig. 6. Comparison of ROC curves of different classifiers for multiclass classification in cough category of dataset-2.

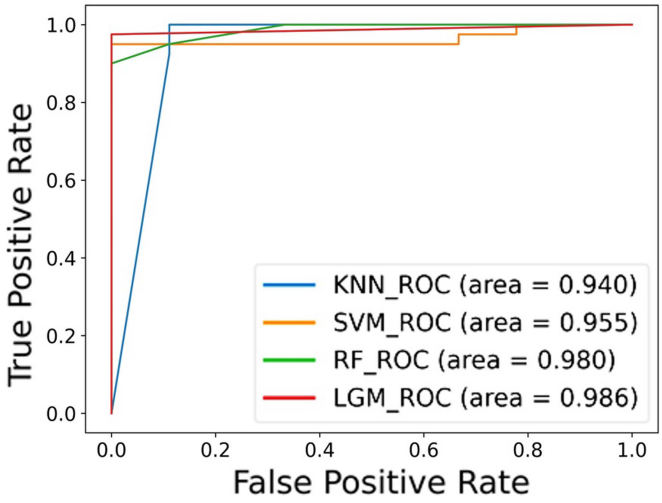

Fig. 7. Comparison of ROC curves of different classifiers for binary classification in cough category of dataset-2.

trade-offs between the true positive and false-positive rates [45]. The ROC curves of dataset-2 in the cough category (binary and multi-class) are shown in Fig. 6 and Fig. 7 respectively. The proposed approach has a high true true-positive rate and a low false false-positive rate, according to the ROC curves. The AUC of the proposed model is 0.99, which is better in comparison to the RF, SVM, and KNN models. The proposed features with the additional preprocessing provide better results compared to standard features and classifiers.

C. Comparison With Baseline Models and Combined Datasets

A comparative analysis of the proposed model over the existing methods used in the five datasets are shown in Table V. The Improvement in the detection performance is mentioned in the last column. It is observed that the proposed model shows consistent performance across all the datasets as well as in the

TABLE V
COMPARATIVE ANALYSIS OF THE OVERALL DETECTION OF PERFORMANCE OF EACH OF THE DATASETS

Classification task	Name of the dataset	Existing method	LGM	improvement
Binary	Dataset-1 (cough)	0.67	0.98	0.30
	Dataset-2 (cough)	0.80	0.97	0.17
	Dataset-3 (cough)	0.73	0.98	0.25
	Dataset-4 (spoken sentence)	0.88	0.98	0.09
	Dataset-5 (cough)	0.77	0.98	0.20
Multiclass	Dataset-2 (cough)	0.82	0.97	0.15

TABLE VI
PERFORMANCE COMPARISON FOR COMBINED DATASET (COUGH CATEGORY) USING 5701 FEATURE VECTOR

Evaluation Measures	LGM	SVM	RF	KNN
CA	**0.983**	0.922	0.971	0.858
F-2	**0.983**	0.922	0.971	0.858
PR	**0.983**	0.924	0.971	0.858
RC	**0.983**	0.922	0.971	0.858
AUC	**0.989**	0.972	0.982	0.932

combined dataset. There is approximately 30%, 15%, 25%, 9%, and 20% minimum improvement in CD performance for datasets 1,2,3,4,5. For the assessment of the generalization ability of the proposed model, a combined dataset is prepared with the speech signals in the cough category from datasets 1,2,3,5. In the combined dataset, there is a total of 1528 samples from the healthy category, while 1344 samples are from the COVID-19 positive category. The performance of all four methods is evaluated and the results are listed in Table VI. It is observed that the proposed model shows the highest accuracy of 0.983 over the other three standard models.

The minimum CD performance of the proposed method is approximately 97 % across all sound categories, databases, and CV schemes. The proposed approach has a high true-positive rate and a low false-positive rate, according to the ROC curves. The AUC of the proposed model is 0.99, which is better in comparison to the RF, SVM, and KNN models. The proposed features with the additional preprocessing provide better results compared to standard features and classifiers.

D. Statistical Analysis of Classifier Models

The statistical analysis of the comparison of the performance of the LGM model with the standard machine learning-based models SVM, RF, and KNN over five datasets is listed in Table VII. For this purpose, the t-statistic value between the two classifiers is computed as mentioned in (7).

$$t = \frac{c_1 - c_2}{\sqrt{v_1^2 + v_2^2}} \tag{7}$$

Where, the mean and variance of the 5-fold classification accuracy of classifier 1 and classifier 2 are denoted as c_1, c_2, and v_1^2, v_2^2 respectively [46]. Most of the t-values in Table VII are positive, which indicates the superior performance of the proposed model over the standard machine learning-based models.The above classification tasks, comparative, and statistical analysis results reveal the effectiveness of the proposed model with preprocessing and an efficient combination of audio features. The main

TABLE VII
COMPARISON OF T-STATISTIC OF PROPOSED MODEL WITH STANDARD ML-BASED MODELS

Dataset	LGM vs SVM	LGM vs RF	LGM vs KNN
Breathing Deep (D-1)	2.95	0	1.90
Breathing Shallow (D-1)	4.01	0.96	4.02
Cough Heavy (D-1)	3.88	0.14	1.52
Cough Shallow (D-1)	3.50	0.21	1.64
Vowel-/a/(D-1)	6.09	0.31	2.71
Vowel-/e/(D-1)	6.02	0.36	2.50
Vowel-/o/(D-1)	1.52	-0.14	2.46
Counting Normal (D-1)	1.84	0	3.50
Counting Fast (D-1)	1.51	0.19	3.17
Cough (D-2)	1.11	0	0.25
Breathing (D-2)	1.69	0.36	1.63
Cough (D-3)	0.09	0.35	2.27
Sentence (D-4)	0.17	0	4.46
Cough (D-5)	1.38	0	8.35

reason for this is the use of various signal processing techniques such as low pass filtering, speech enhancement, voice activity detection, and dynamic level control have substantially helped in reducing the effects of various environments while recording the speech signal of subjects. Secondly, The use of feature fusion-based statistical features evaluated from frame-level speech signal to the LGM classifier has yielded enhanced detection accuracy which is a minimum of 9% more than that obtained by the reported standard methods. The detection model has been observed to be robust as it offers a consistent detection performance of 97% while testing with five different speech datasets.

IV. CONCLUSION

In the current study, a non-invasive and effective respiratory disease detection scheme is developed and tested for COVID-19 and Asthma. The major contributions of the investigation are the use of improved preprocessing techniques, an effective combination of spectral, cepstral, and periodicity features along with the implementation of gradient boosting machines for robust and consistent performance across multiple datasets. The proposed model can be used for early and fast automatic diagnosis of COVID-19 without the subject visiting a hospital as well as without the assistance of a medical professional. However, it is suggested that the detection scheme by the use of the proposed intelligent model can be verified by the medical professional before a prescription is initiated. It may be noted that the proposed detection scheme involves more computations and training time. There is still room to improve the method's computing complexity for faster implementations. The effective preprocessing techniques, as well as the combination of audio features can be further implemented and tested for other speech recognition tasks including emotion recognition, Parkinson's disease, and heart disease detection.

ACKNOWLEDGMENT

The authors express their gratitude to Professor Cecilia Mascolo, Department of Computer Science and Technology and Chancellor, Master, and Scholar of the University of Cambridge for sharing the speech database of COVID-19 [10].

REFERENCES

[1] I. Trancoso, J. Correia, F. Teixeira, B. Raj, and A. Abad, "Analysing speech for clinical applications," in *Proc. Int. Conf. Stat. Lang. Speech Process.*, 2018, pp. 3–6.

[2] C. Chakraborty and A. N. Abougreen, "Intelligent Internet of Things and advanced machine learning techniques for COVID-19," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 7, no. 26, pp. e1–e1, 2021.

[3] K. Rezaee, H. G. Zadeh, C. Chakraborty, M. R. Khosravi, and G. Jeon, "Smart visual sensing for overcrowding in COVID-19 infected cities using modified deep transfer learning," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2022.3174160.

[4] F. Piccialli, V. D. Somma, F. Giampaolo, S. Cuomo, and G. Fortino, "A survey on deep learning in medicine: Why, how and when?," *Inf. Fusion*, vol. 66, pp. 111–137, 2021.

[5] "World Health Organization/Diseases/Coronavirus disease (COVID-19)," Accessed: Aug. 8, 2022. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[6] K. Ramdas, A. Darzi, and S. Jain, "'Test, re-test, re-test': Using inaccurate tests to greatly increase the accuracy of COVID-19 testing," *Nat. Med.*, vol. 26, no. 6, pp. 810–811, 2020.

[7] L. Orlandic, T. Teijeiro, and D. Atienza, "The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms," *Sci. Data, Nature Publishing Group*, vol. 8, no. 1, pp. 1–10, 2021. [Online]. Available: https://doi.org/10.1038/s41597-021-00937-4

[8] J. Han et al., "An early study on intelligent analysis of speech under COVID-19: Severity, sleep quality, fatigue, and anxiety," in *Proc. Interspeech*, 2020, pp. 4946–4950. [Online]. Available: https://doi.org/10.21437/Interspeech.2020--2223

[9] N. Sharma et al., "Coswara A database of breathing, cough, and voice sounds for COVID-19 diagnosis," in *Proc. Interspeech*, 2020, pp. 4811–4815. [Online]. Available: https://doi.org/10.21437/Interspeech.2020--2768

[10] C. Brown et al., "Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 3474–3484. [Online]. Available: https://doi.org/10.1145/3394486.3412865

[11] K. V. S. Ritwik, S. B. Kalluri, and D. Vijayasenan, "COVID-19 patient detection from telephone quality speech data," 2020, *arXiv:2011.04299*. [Online]. Available: http://arxiv.org/abs/2011.04299

[12] B. Stasak, Z. Huang, S. Razavi, D. Joachim, and J. Epps, "Automatic detection of COVID-19 based on short-duration acoustic smartphone speech analysis," *J. Healthcare Informat. Res.*, vol. 5, no. 2, pp. 201–217, 2021.

[13] T. K. Dash, S. Mishra, G. Panda, and S. C. Satapathy, "Detection of COVID-19 from speech signal using bio-inspired based cepstral features," *Pattern Recognit.*, vol. 117, 2021, Art. no. 107999. [Online]. Available: https://doi.org/10.1016/j.patcog.2021.107999

[14] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features," *Comput. Biol. Med.*, vol. 141, 2022, Art. no. 105153.

[15] L. Verde, G. D. Pietro, A. Ghoneim, M. Alrashoud, K. N. Al-Mutib, and G. Sannino, "Exploring the use of artificial intelligence techniques to detect the presence of coronavirus COVID-19 through speech and voice analysis," *IEEE Access*, vol. 9, pp. 65750–65757, 2021.

[16] A. Ponomarchuk et al., "Project achoo: A practical model and application for COVID-19 detection from recordings of breath, voice, and cough," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 2, pp. 175–187, Feb. 2022.

[17] L. Sun et al., "Adaptive feature selection guided deep forest for covid-19 classification with chest CT," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2798–2805, Oct. 2020.

[18] V. Ravi, H. Narasimhan, C. Chakraborty, and T. D. Pham, "Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images," *Multimedia Syst.*, vol. 28, no. 4, pp. 1401–1415, 2021.

[19] H. K. Bhuyan, C. Chakraborty, Y. Shelke, and S. K. Pani, "COVID-19 diagnosis system by deep learning approaches," *Expert Syst.*, vol. 39, no. 3, 2022, Art. no. e12776.

[20] A. Gumaei et al., "A decision-level fusion method for COVID-19 patient health prediction," *Big Data Res.*, vol. 27, 2022, Art. no. 100287.

[21] K. K. Wong, G. Fortino, and D. Abbott, "Deep learning-based cardiovascular image diagnosis: A promising challenge," *Future Gener. Comput. Syst.*, vol. 110, pp. 802–811, 2020.

[22] G. Ke et al., "Lightgbm: A highly efficient gradient boosting decision tree," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 3149–3157, 2017.

[23] C. Chen, Q. Zhang, Q. Ma, and B. Yu, "LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion," *Chemometrics Intell. Lab. Syst.*, vol. 191, pp. 54–64, 2019.

[24] Z. Kostic and A. Jevremovic, "What image features boost housing market predictions?," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1904–1916, Jul. 2020.

[25] D. Wang, Q. Meng, D. Chen, H. Zhang, and L. Xu, "Automatic detection of arrhythmia based on multi-resolution representation of ECG signal," *Sensors*, vol. 20, no. 6, 2020, Art. no. 1579.

[26] G. Chaudhari et al., "Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough," 2020, *arXiv:2011.13320*.

[27] A. Keerio, B. K. Mitra, P. Birch, R. Young, and C. Chatwin, "On preprocessing of speech signals," *Int. J. Signal Process.*, vol. 5, no. 3, pp. 216–222, 2009.

[28] C. Nass and K. M. Lee, "Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction," *J. Exp. Psychol.: Appl.*, vol. 7, no. 3, pp. 171–181, 2001.

[29] C. G. L. Prell and O. H. Clavier, "Effects of noise on speech recognition: Challenges for communication by service members," *Hear. Res.*, vol. 349, pp. 76–89, 2017.

[30] T. K. Dash, S. S. Solanki, and G. Panda, "Improved phase aware speech enhancement using bio-inspired and ANN techniques," *Analog Integr. Circuits Signal Process.*, vol. 102, no. 3, pp. 465–477, 2020.

[31] S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Commun.*, vol. 48, no. 2, pp. 220–231, 2006.

[32] L. D. Shriberg, J. R. Green, T. F. Campbell, J. L. Mcsweeny, and A. R. Scheer, "A diagnostic marker for childhood apraxia of speech: The coefficient of variation ratio," *Clin. Linguistics Phonetics*, vol. 17, no. 7, pp. 575–595, 2003.

[33] O. C. Ai, M. Hariharan, S. Yaacob, and L. S. Chee, "Classification of speech dysfluencies with MFCC and LPCC features," *Expert Syst. With Appl.*, vol. 39, no. 2, pp. 2157–2165, 2012.

[34] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. 4, pp. 44164–44164.

[35] D. Giannoulis, M. Massberg, and J. D. Reiss, "Digital dynamic range compressor design A tutorial and analysis," *J. Audio Eng. Soc.*, vol. 60, no. 6, pp. 399–408, 2012.

[36] D. Paul, M. Pal, and G. Saha, "Spectral features for synthetic speech detection," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 4, pp. 605–617, Jun. 2017.

[37] K. Kumar, R. K. Aggarwal, and A. Jain, "A hindi speech recognition system for connected words using HTK," *Int. J. Comput. Syst. Eng.*, vol. 1, no. 1, pp. 25–32, 2012.

[38] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis, "A speech/music discriminator of radio recordings based on dynamic programming and bayesian networks," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 846–857, Aug. 2008.

[39] Z. Chen and V. Hohmann, "Online monaural speech enhancement based on periodicity analysis and a priori SNR estimation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 1904–1916, Nov. 2015.

[40] "MATLAB and audio toolbox," The MathWorks, Inc., Natick, MA, USA.

[41] "Matlab: AudioFeatureExtractor function," 2020. [Online]. Available: https://it.mathworks.com/help/audio/ref/audiofeatureextractor.html/

[42] A. F. Vermeulen, "Industrial machine learning," in *Using Artificial Intelligence as a Transformational Disruptor*. Berkeley, CA, USA: A Press, 2020, pp. 137–180. [Online]. Available: https://doi.org/10.1007/978-1-4842-5316-8

[43] X. Sun, M. Liu, and Z. Sima, "A novel cryptocurrency price trend forecasting model based on LightGBM," *Finance Res. Lett.*, vol. 32, 2020, Art. no. 101084.

[44] C. Pao, C. Yeh, and Lin, "A comparative study of different weighting schemes on KNN-based emotion recognition in mandarin speech," in *Proc. Int. Conf. Intell. Comput.*, 2007, pp. 997–1005.

[45] J. Lever, M. Krzywinski, and N. Altman, "Points of significance: Model selection and overfitting," *Nature Methods*, vol. 13, no. 9, pp. 703–705, 2016.

[46] T. T. Wong, "Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets," *Pattern Recognit.*, vol. 65, pp. 97–107, 2017.