Pattern Recognition 117 (2021) 107999

Contents lists available at ScienceDirect

Pattern Recognition

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

Detection of COVID-19 from speech signal using bio-inspired based cepstral features

Tusar Kanti Dash[a], Soumya Mishra[a], Ganapati Panda[a], Suresh Chandra Satapathy[b,*]

[a] Electronics and Telecommunications Engineering, C V Raman Global University, Bhubaneswar, India
[b] School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India

ARTICLE INFO

Article history:
Received 10 December 2020
Revised 9 April 2021
Accepted 18 April 2021
Available online 24 April 2021

Keywords:
Bio-inspired computing
COVID19
Speech signal

ABSTRACT

The early detection of COVID-19 is a challenging task due to its deadly spreading nature and existing fear in minds of people. Speech-based detection can be one of the safest tools for this purpose as the voice of the suspected can be easily recorded. The Mel Frequency Cepstral Coefficient (MFCC) analysis of speech signal is one of the oldest but potential analysis tools. The performance of this analysis mainly depends on the use of conversion between normal frequency scale to perceptual frequency scale and the frequency range of the filters used. Traditionally, in speech recognition, these values are fixed. But the characteristics of speech signals vary from disease to disease. In the case of detection of COVID-19, mainly the coughing sounds are used whose bandwidth and properties are quite different from the complete speech signal. By exploiting these properties the efficiency of the COVID-19 detection can be improved. To achieve this objective the frequency range and the conversion scale of frequencies have been suitably optimized. Further to enhance the accuracy of detection performance, speech enhancement has been carried out before extraction of features. By implementing these two concepts a new feature called COVID-19 Coefficient (C-19CC) is developed in this paper. Finally, the performance of these features has been compared.

1. Introduction

Coronavirus disease 19 (COVID- 19) which exhibits acute respiratory syndrome is a deadly viral infection. As reported, it has started in Wuhan, China in 2019 and has affected the whole world [1]. As per the report of the World Health Organization, more than a hundred million people have suffered till 7th March 2021 out of which more than 2.5 million deaths have been reported [2]. The social distancing of 1.6 m to 3 m is recommended to control the rapid spreading of COVID-19 cases [3]. It is observed from the experiences of the medical practitioners that rather than the deadly nature of the virus, its fear of stigma is stopping people from going to medical laboratories for testing purposes [4]. Under such circumstances, it has become a huge challenge for developing an appropriate method for the early detection of this disease. It is a fact that the speech-based detection of COVID-19 is a simpler and safer approach for this purpose [5]. In this section, a review of related literature is carried out in two parts: speech based COVID-19 detection and speech recognition using MFCC features.

1.1. Literature review

In this section, the literature review has been carried out in two parts: speech based COVID-19 detection and use of MFCC features based speech recognition.

1.1.1. Review on COVID-19 detection using speech signals

Speech analysis is one of the important methods used for the detection of parkinson [6] alzheimer, asthma [7]. In the recent past, attempts have been made in the area of speech based COVID-19 analysis and diagnosis. The details in terms of databases, feature extraction methods, and performance analysis have been presented. A crowdsourced data set of respiratory sounds has been prepared using coughs and breathing sounds for detecting COVID-19. Several audio features such as speech time duration, onset, tempo, period, RMS energy, spectral centroid, Roll-Off frequency, Zero-crossing, MFCC have been used as inputs to classification methods such as Logistic Regression, Gradient Boosting Trees, and Support Vector Machines (SVM) for the classification task. It is reported that a maximum accuracy of 80% in Receiver Operator Characteristics Area Under Curve (AUC) [8].

A review of Artificial Intelligence based methods used for COVID-19 detection is presented in [9]. It explains multi-modal

[*] Corresponding author.
E-mail address: suresh.satapathyfcs@kiit.ac.in (S.C. Satapathy).

https://doi.org/10.1016/j.patcog.2021.107999
0031-3203/© 2021 Elsevier Ltd. All rights reserved.

approach using audio, text, and image for achieving better detection results. The log Mel spectra of a speech signal are mapped with the respiratory sensors to train the neural network-based models. A sensitivity of 91.2% for breathing based detection and a mean absolute error of 1.01 breaths per minute have been reported using the proposed methods. In another paper, the health condition of COVID-19 patients is categorized into four types with respect to the severity of illness, sleep quality, fatigue, and anxiety [10]. Audio dataset has been collected from twenty females and thirty-two males COVID19 patients from two hospitals in Wuhan, China during March 20, – 26, 2020. Two acoustic feature sets from the computational paralinguistics challenge and extended Geneva minimalistic acoustic parameter sets have been used as inputs to SVM to achieve an average classification accuracy of 69%.

1.1.2. Review on optimization in MFCC features

The cepstral analysis is one of the oldest and popular signal analysis methods which finds applications in the speech signal and mechanical systems [11]. The MFCC based features are very popular and effective for speech recognition, music information retrieval, speech evaluation parameters, etc. An optimization of MFCC features is achieved by reducing the feature space using Linear Discriminant Analysis Fisher's F-ratio [12]. The use of these features achieves faster convergence of the ANN model and improvement in recognition accuracy at different SNR levels. The source recognition of Cell-Phone is carried out using the optimization of different cepstral coefficients such as Mel, linear, and Bark frequency [13]. The use of minimum and maximum frequencies of MFCC and cepstral variance normalization has enhanced the identification rate to 96.85%. With an objective to minimize the dissimilarity between the perceptual and feature domain distortions, modified MFCC based features are proposed in [14]. These simple feature vectors provide improved speech recognition performance in noisy as well as clean conditions. In another paper, the mean and standard deviation of the feature space including MFCC are optimized using genetic algorithm, and differential evolution [15]. These improved features are used for the punjabi language speaker recognition.

An analysis of different frequency bands has been carried out using the F-ratio method for speech unit classification and it is observed that 1 kHz to 3 kHz frequency range to be provided more emphasis and accordingly an optimization of features using the F-ratio scale is proposed in [16]. A significant reduction in the sentence error rate is reported by using the proposed feature optimization technique. The central and side frequency parameters of MFCC filter banks have been optimized using particle swarm optimization and genetic algorithm [17]. These optimized features are then applied in the Hindi vowel recognition using the Hidden Markov model and Multilayer perceptrons under different noise conditions. A hybrid approach is proposed by taking gammatone and mel frequency cepstral coefficients with PCA, and multi tapered method using differential evolution, and Hidden Markov model (HMM) based classifiers for robust recognition of Punjabi speech under different noise conditions [18]. The frequency range of the filter banks of MFCC is optimized for emotion recognition using two databases [19]. This method has improved the speaker-independent emotion recognition accuracy by 15% for the Assamese database and 25% for the Berlin database.

1.2. Motivation

It is observed from the literature review that early detection of COVID-19 from speech data is a challenging and timely research area [20–22]. For the remote online based COVID-19 detection from the speech signal, the patients have to use mobile applications in the real-life noisy environment. Investigation in this direction has not been fully explored particularly in the selection of proper audio features of COVID-19 patients for detection purposes. The nature of the speech signals used for the analysis of COVID-19 are mainly the breathing and cough sounds and hence it is quite different from the speech signal comprising complete sentences. The MFCC features have also been used in COVID-19 detection in [23], but these features are directly used without any modification or improvement in the features and hence the detection performance is poor. Thus, in the present COVID-19 scenario, there is a huge requirement to develop a better effective tool for improved detection from a safe distance and remotely recorded speech. Hence, there is a need to find and identify improved features which are expected to improve the detection accuracy of the classifier. The focus of the current investigation is to develop potential features from the speech data for facilitating the classifier to yield higher accuracy of detection. In addition, speech enhancement is required to be carried out for extraction of the proper audio features [24]. It is further observed that there exists a huge class imbalance present in the speech data available on the online platforms [8,23]. This class imbalance affects the overall training and testing performance of classifiers and hence this issue needs to be addressed and resolved. These problems have been identified during the literature review and taken up in this paper.

There are different techniques used for achieving higher accuracy in classification and prediction tasks out of which the deep learning-based techniques are preferred if the data size is high and features extraction and selection are difficult [25,26]. On the other hand, ML based methods use extracted features whereas CNN extracts the appropriate features through a series of convolution operations. Further DL methods involve more time for classification [27–29]. In the present problem, two speech datasets [8,23] with categorically less data are available and the feature optimization is the target. Though Cepstral analysis is an old method, still it is quite popular, effective and a lot of recent articles still employ this feature [30,31]. Therefore, for the present problem, Cepstral optimised features are obtained by a bio-inspired technique and used for classification purposes.

1.3. Research objectives

Based on the motivation of research arising out of the literature review, the problem has been formulated with the following research objectives. Thus, the research objectives of the paper are:

1. To analyze the cepstral features used in speech recognition and to suitably optimizing the conversion scale in the frequency domain, and frequency range of filter banks using the bio-inspired technique to achieve better COVID-19 detection.
2. To identify the best possible sound patterns during coughing, breathing, and voiced sounds to detect COVID-19.
3. To employ the adaptive synthetic sampling approach for achieving efficient training for class imbalance in the database and to facilitate proper classification using SVM.
4. To compare and analyze the detection performance of the proposed cepstral feature-based classifier with that obtained from other reported results.

1.4. Organization of the paper

Based on the objectives of the research the organization of the paper proceeds as follows. The introduction, literature review, motivation, research objectives are presented in Section I. In section II, a detailed review of the related work on the theme of the problem is presented. The salient characteristics of data obtained from the standard database of COVID-19 are provided in Section III. In Section IV, the proposed methodology is presented in detail. The

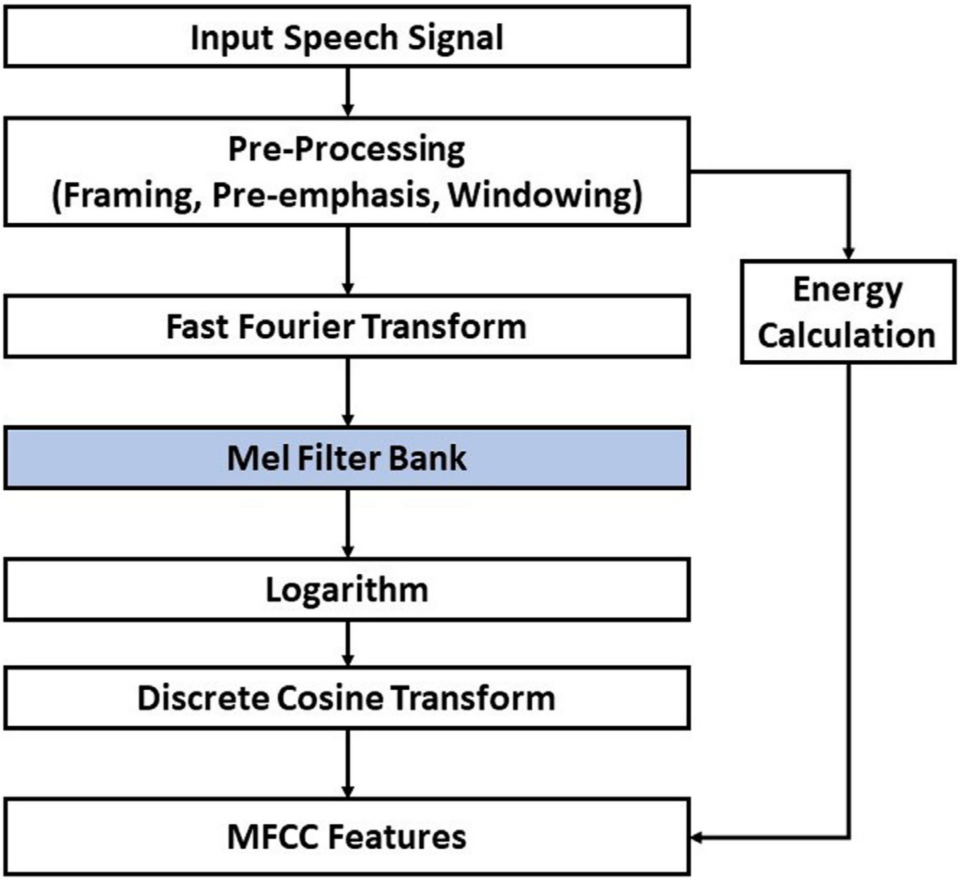

Fig. 1. Steps for the MFCC feature extraction.

simulation-based experiment using two standard data sets and discussion of results as well as the contribution of the paper is presented in Section V. Finally, the conclusion of the paper and scope for future work are dealt with in Section VI.

2. Related works

The cepstral analysis is one the oldest and popular signal analysis methods used in various applications like speech signal processing, mechanical engineering problems, and analysis of multiple inputs, multiple output systems, etc [11]. The MFCCs are very popular and effective features in speech recognition, music information retrieval, speech evaluation parameters. It is normally calculated using the following steps [32].

- Step-1 — Apply the pre-processing like windowing, framing to the input signal
- Step-2 — Calculate the energy of the frame
- Step-3 — Find the Discrete Fourier transform using the FFT method
- Step-4 — Apply the Mel filter bank by mapping the power spectrum into the mel scale and by using triangular overlapping windows.
- Step-5 — FInd the logarithm of Step 4
- Step-6 — Apply the Discrete cosine transform
- Step-7 — Combine Energy and other features from step-6 to get MFCC features

These steps are also shown in Fig. 1. Step-4 deals with the use of the Mel filter bank. There are several variations of these filter banks reported in the speech processing such as triangular filter bank using mel-scale, human factor scale, Bark-scale, ERB-scale, and Gammatone filter bank. Depending on these filter banks the cepstral coefficients are named accordingly. The basic steps used for the filter bank design are shown in Fig. 2 [33]. In this case, the Perceptual scale means frequency scale based on the human perception of sound like the mel scale. The details of the four types of features are discussed in the sequel.

2.1. Triangular filter bank using mel-scale (TFBCC-M features)

The steps involved in the design of the filter bank for extraction of TFBCC-M features [33] are explained in this section. The steps are as follows:

- Convert linear scale frequency of DFT (f) to Mel scale (f_m) using Equation (1).

$$f_m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

- The linear scale lower and higher cut-off frequencies (f_l and f_h) are converted into melscale (fm_l and fm_h) respectively. Now, the center frequencies $\left(f_{m_{c\,p}}\right)$ of each filter are calculated using Equation (2).

$$f_{m_{c\,p}} = f_{ml} + p \cdot \left(\frac{f_{mh} - f_{ml}}{P + 1}\right) \tag{2}$$

where p = 1, 2, 3,..., P-1. P is the number of Mel filters.

- The center frequencies $\left(f_{m_{c\,p}}\right)$ of pth filter band is to be converted to linear scale using $\left(f_{c\,p}\right)$given in Equation (3).

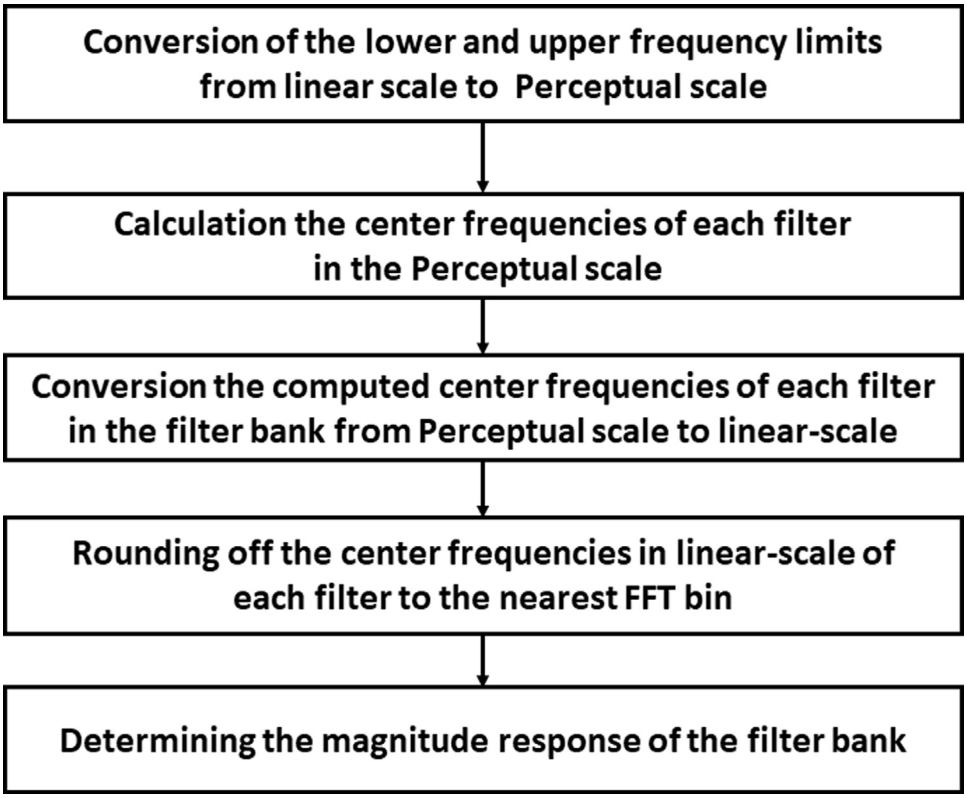
Conversion of the lower and upper frequency limits from linear scale to Perceptual scale
Calculation the center frequencies of each filter in the Perceptual scale
Conversion the computed center frequencies of each filter in the filter bank from Perceptual scale to linear-scale
Rounding off the center frequencies in linear-scale of each filter to the nearest FFT bin
Determining the magnitude response of the filter bank

Fig. 2. Steps for the Filter bank computation.

$$f_{c\,p} = 700 \cdot \left(10^{\frac{f_{mc\,p}}{2595}} - 1\right) \tag{3}$$

- The magnitude response of each of the filters $H_p(k)$ in the mel filter bank is calculated using Equation (4).

$$H_p(k) = \begin{cases} 0, & if\ k < f_r(p-1) \\ \frac{k - f_r(p-1)}{f_r(p) - f_r(p-1)}, & if\ f_r(p-1) \le k \le f_r(p) \\ \frac{f_r(p+1) - k}{f_r(p+1) - f_r(p)}, & if\ f_r(p) \le k \le f_r(p+1) \\ 0, & if\ k > f_r(p+1) \end{cases} \tag{4}$$

Where,k is the frequency domain index of FFT, $f_r(p)$ is the rounded center frequencies $\left(\left(f_{c\,p}\right)\right)$ and it is calculated using Equation (5) and fs is the sampling frequency of the speech signal.

$$f_r(p) = (length\ of\ FFT\ block + 1) \cdot (f_{c\,p} / f_s) \tag{5}$$

2.2. Triangular filter bank using bark scale (TFBCC-B features)

The calculation of TFBCC-B features is similar to the TFBCC-M features with only one difference [33]. Instead of conversion between linear scale frequency of DFT and mel scale (as mentioned in Equations (1) and (3)), the conversion is between linear scale frequency of DFT (f) and Bark scale (f_b)as mentioned in Equations (6) and (7).

$$f_b = 6\ \sin h^{-1}\left(\frac{f}{600}\right) \tag{6}$$

$$f_{c\,p} = 600\ \sin h\left(\frac{f_{b_{c\,p}}}{6}\right) \tag{7}$$

where $\left(f_{b_{c\,p}}\right)$ is the center frequency in the Bark scale.

2.3. Triangular filter bank using human factor scale (TFBCC-H features)

The difference between the designing of filter bank for TFBCC-M and TFBCC-H features lies in the calculation of the critical bandwidth using the ERB approximation [34].

- The center frequencies $\left(f_{c_p}\right)$ of the first (p=1) and last (p=P) filters are computed as mentioned in Equation (8)

$$f_{c_p} = \frac{1}{2} \cdot \left(-\beta + \sqrt{\beta^2 - 4\gamma}\right) \tag{8}$$

where p is the index of filter and P is the maximum number of filters. For the first filter

The $\left(f_{c_p}\right)$ is obtained usingEquation (9) and Table 1.

$$\beta = \frac{\bar{\beta} - \hat{\beta}}{\bar{\alpha} - \hat{\alpha}}\ ,\ \gamma = \frac{\bar{\gamma} - \hat{\gamma}}{\bar{\alpha} - \hat{\alpha}} \tag{9}$$

Table 1
Variables used in TFBCC-H features calculation.

parameter	value
$\bar{\alpha}$	0.00000623
$\bar{\beta}$	0.09339
$\bar{\gamma}$	28.52
$\hat{j}$	0 (for the first filter)
$\hat{j}$	1 (for the Pth filter)
f_p	f_l (for the first filter)
f_p	f_h (for the Pth filter)

$$\begin{aligned}\hat{\alpha} &= (-1)^{\hat{j}}\left(\frac{0.5}{700+f_p}\right)\\ \hat{\beta} &= (-1)^{\hat{j}}\left(\frac{700}{700+f_p}\right)\\ \hat{\gamma} &= (-1)^{\hat{j}+1}(0.5\ f_p)\left(1+\frac{700}{700+f_p}\right)\end{aligned} \tag{10}$$

- After obtaining the values of $\left(f_{c_p}\right)$, these are converted from linear-scale to mel-scale as mentioned in Equation (11), where f'_{c_1} and f'_{cP} are the center frequencies of the first and last filter in the mel-scale computed using Equation (1).

$$f'_{c_p{=}}f'_{c_1} + (i-1)\left(\frac{f'_{cP}-f'_{c_1}}{P-1}\right) \tag{11}$$

- Convert the computed center frequencies of all the filters from mel-scale to linear-scale using Equation (3).
- Calculate the lower and upper limit of the frequency of each of the filters in the filter bank using Equations (12), (13) and (14).

$$f_{lp} \ = -(700+ERB_p) + \sqrt{(700+ERB_p)^2 + f_{c_p}\left(f_{c_p}+1400\right)} \tag{12}$$

$$f_{hp} = f_{hp} + 2\cdot ERB_p \tag{13}$$

$$ERB_p = \bar{\alpha}\ \ {f^2}_{c_p} + \bar{\beta}\ \ f_{c_p} + \gamma \tag{14}$$

- Round off the center frequencies and magnitude response calculation of each filter as is done in case of TFBCC-M features using Equations (4) and (5).

2.4. Triangular filter bank using ERB scale (TFBCC-E features)

The calculation of TFBCC-E features is similar to that of TFBCC-H features with only one change [33]. Instead of conversion between linear scale frequency of DFT and mel scale (as mentioned in Equations (11), (1) and (3)), the conversion is made between linear scale frequency of DFT (f) and ERB scale (f_e)as mentioned in Equations (15) and (16).

$$f_e = 24.7(0.00437\cdot\ f\ +1) \tag{15}$$

$$f_{c_p} = 228.72\left(\left(\frac{f_{e_{cp}}}{24.7}\right)-1\right) \tag{16}$$

2.5. Parameters for optimization

It is observed from the detailed review of the existing Cepstral coefficients that the two parameters such as the conversion of frequency from linear scale to perceptual scale and the selection of cut-off frequencies are the determining factor for the performance of the cepstral analysis. This issue is addressed in the literature [13–19]. This problem has been further investigated in Section 5.2.

3. Database preparation

Two Speech databases are used in the simulation-based experiments carried out in Section 5. The brief details of the databases used are discussed in this section.

3.1. Coswara database

The Coswara database (DB-1) has been developed by the Indian Institute of Science, India in the year 2020 [23]. A web application is used for collecting audio samples for diagnosing COVID-19 prevalence using breath, cough, and voice sounds. The audio files are arranged in groupings of different respiratory indicators such as shallow and deep breathing, shallow and heavy cough, continuous vowel pronunciation /a/, /e/, and /o/, counting normal and fast. This database also contains additional information in terms of age, gender, demography, existing health history, and the existence of chronic health preconditions. The volunteering members have individually contributed to multiple sound clips for multiple segments. The maximum duration of the sound clips for the individual segment is approximately 15 seconds duration with the sampling frequency as 41 kHz or 48 kHz. This data set covers 570 participants and each participant has contributed 9 audio files pertaining to various categories. Cumulatively, this data set comprises 3470 clean, 1055 noisy, and the remaining are highly degraded sound samples.

3.2. Crowdsourced respiratory sound data

The Crowdsourced Respiratory Sound Data (DB-2) has been developed by Cambridge University, the UK in the year 2020 [8]. The Android and web-based sound application is used for capturing speech/audio samples for detecting Corona Virus disease prevalence using breath, and cough sounds. For capturing the required sounds, the volunteers are prompted to follow instructions for coughing and breathing a couple of times along with reading phases. Finally, the users are asked whether they have clinically tested COVID positive so far. Since the project employs two applications to collect the data, the indicating words 'web' and 'android' are frequently used to distinguish between the recorded audio samples. Also, while naming the audio files, 'no cough' and 'with cough' designate a volunteer's report of a condition to dry or wet cough, while 'nosymp' means the volunteer showed no signs at that time. The selected audio.wav files have been arranged in groupings under categories of cough, breath, and asthma. The maximum duration of the sound clips for the individual segment recorded by the android application varies from 8 seconds to 20 seconds duration approximately with a sampling frequency of a maximum 48 kHz. Similarly, the maximum duration of the sound clips for the individual segment recorded by Web application varied from 11 seconds to 24 seconds duration approximately with the sampling frequency of maximum of 48 kHz samples. This dataset consists of 4352 unique users from the web app and 2261 unique users from the Android app. Out of these, total of 235 users are declared COVID-19 positive.

In the two databases, the signals are recorded at 44.1 kHz and 48.1 kHz sampling rates. It is observed from the literature [35] the most of the latent features are within 8 kHz bandwidth and the Hence pre-processing is done by downsampling the speech signal to 16 kHz.

4. Proposed methodology

The details of the proposed method are discussed in this Section. The methodology is depicted in the block diagram form (Fig. 3) and the need and operation of each block are explained. The algorithm is divided into two parts: configuration and application [36]. The configuration part deals with the preparation of a clean balanced dataset, and the application part explains the extraction and use of the proposed C-19CC features. In the configuration stage, the two available datasets are analysed and converted into a labeled balanced dataset by using the Adaptive

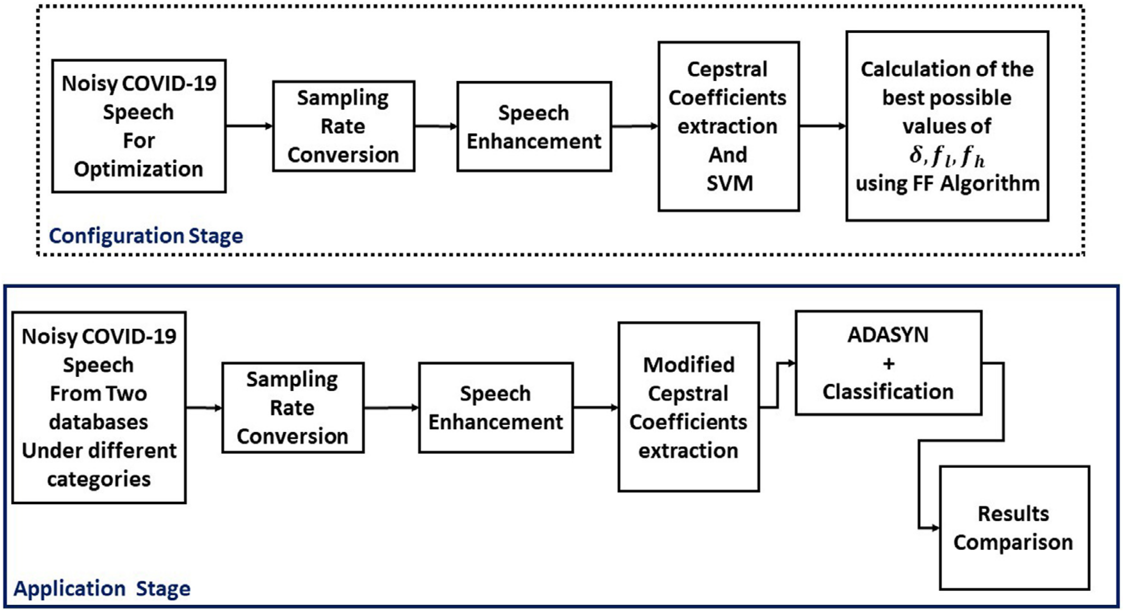
Noisy COVID-19 Speech For Optimization
Sampling Rate Conversion
Speech Enhancement
Cepstral Coefficients extraction And SVM
Calculation of the best possible values of δ, f_l, f_h using FF Algorithm
Configuration Stage
Noisy COVID-19 Speech From Two databases Under different categories
Sampling Rate Conversion
Speech Enhancement
Modified Cepstral Coefficients extraction
ADASYN + Classification
Results Comparison
Application Stage

Fig. 3. Block Diagram of Proposed detection model of COVID-19.

Synthetic Sampling Approach [37] which is applied to deal with meant for Imbalanced Learning method. To transform the dataset with a uniform single sampling rate, the sampling rate conversion strategy is applied. Subsequently, the Speech Enhancement algorithm is employed for the reduction of noise present in the data. In the configuration stage, the calculation of the best possible values of the Cepstral features is carried out from the pre-processed dataset. Finally, the previously configured set of C-19CC features are extracted from speech databases. The classifier is then trained using these extracted features for the identification of the appropriate class.

4.1. Fire-Fly optimization algorithm

Based on the flashing pattern of fireflies, an efficient optimization algorithm known as FF Algorithm has been developed in the last decade. It has the advantages of swarm intelligence and also it has other advantages as compared to the standard swarm intelligence based algorithms due to its automatic subdivision and the ability of dealing with multi-modality [38]. This section deals with a brief discussion on the operation and implementation of the FF algorithm.

The basic concept of the FF algorithm is based on the attraction and attacking principle of the firefly species. They produce short and rhythmic flashes and the attractiveness of a firefly is calculated by its brightness (light intensity). This principle is modeled as the objective function. The attractiveness is governed by light intensity variation with distance and the absorption coefficient and expressed in Equation (17).

$$I = \frac{I_s}{r^2}, \quad I = I_o\, e^{-\gamma r} \tag{17}$$

The explanations of the parameters used in the FF Algorithm are listed in Table 2.

By combining the two factors of Equation (17) the instantaneous intensity can be expressed asEquation (18)

$$I = I_o\, e^{-\gamma\, r^2} \tag{18}$$

Similarly, the attractiveness (β) is represented as Equation (19).

$$\beta = \beta_o\, e^{-\gamma\, r^2} \tag{19}$$

The distance between any two fireflies i and j at position $\mathbf{x}_i$ and $\mathbf{x}_j$ is computed as the Cartesian distance given in Equation (20).

$$r_{ij} = \|\mathbf{x_i} - \mathbf{x_j}\| = \sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2} \tag{20}$$

Where $x_{i,k}$ is the kth component of the spatial coordinate $\mathbf{x}_i$ of the i th firefly. The movement of a firefly i attracted to another brighter firefly j is expressed as Equation (21), where the second term denotes the attraction and the third term is for inserting randomization. For most cases the value of $\beta_o = 1$ and $\alpha \in [0, 1]$. The speed of the convergence and the overall effectiveness of the FF algorithm depends upon the parameter γ which denotes the variation of the attractiveness. The value of γ varies from 0.1 to 10. When any ith firefly is attracted by a brighter (more attractive) firefly j, then its movement is expressed as in Equation (21).

$$\mathbf{x_i} = \mathbf{x_i} + \beta_o\, e^{-\gamma\, r_{i,j}^2}\,(\mathbf{x_j} - \mathbf{x_i}) + \vartheta\, \varepsilon_i \tag{21}$$

The flow chart of the FF algorithm is shown in Fig. 4. In the FF algorithm, the whole population can be easily subdivided into subgroups based on the attraction principle. These subgroups help to find the local optimum solutions and correspondingly the global optimum. This concept of subdivision allows the fireflies to get the optima simultaneously if the population size is sufficiently [39]. Also, it has been proved that the FF algorithm works better than the traditional Particle Swarm Optimization and Genetic Algorithm in terms of both efficiency and success rate [39]. Recently several speech processing based optimizations have been implemented successfully using FF algorithms [40]. Hence, the FF algorithm has been chosen for finding the best possible cepstral features to be used for the COVID-19 classification.

4.2. Classification

The classification is an important task for the detection of COVID-19. Support Vector Machines (SVM) are simple but potential classifiers having lower computational complexity providing higher classification accuracy as compared to other non-linear classifiers [41]. It is further observed from the literature that for the several speech signal based classifications, SVM with the Gaussian kernel is one of the effective classifiers due to its overall good per-

Table 2
Parameters used in the FF Algorithm.

Name of Parameter	Details
I	light intensity at any particular instant
I_s	light intensity at the source
I_o	Original light intensity used in the calculation of absorption effect
γ	light absorption coefficient
r	distance
β	attractiveness
β_o	attractiveness at distance r = 0
ϑ	randomization parameter
$\boldsymbol{\varepsilon}$	vector of uniformly distributed zero mean random numbers in the range of – 0.5 to 0.5

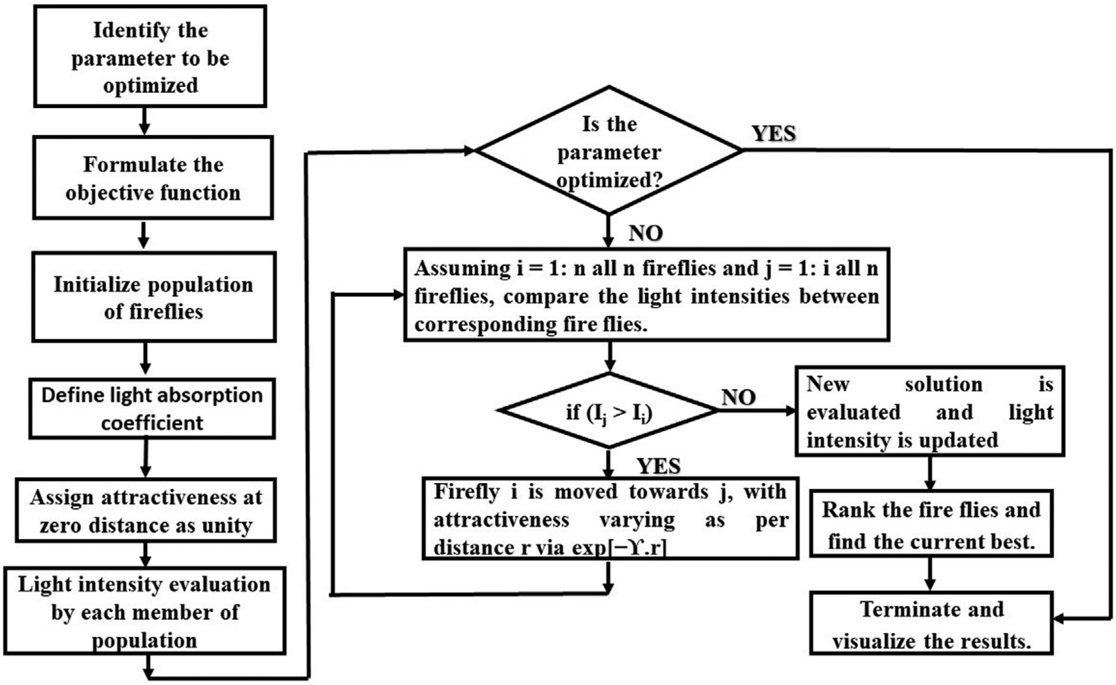
Identify the parameter to be optimized
Formulate the objective function
Initialize population of fireflies
Define light absorption coefficient
Assign attractiveness at zero distance as unity
Light intensity evaluation by each member of population
Is the parameter optimized?
YES
NO
Assuming i = 1: n all n fireflies and j = 1: i all n fireflies, compare the light intensities between corresponding fire flies.
if (I_j > I_i)
NO
YES
New solution is evaluated and light intensity is updated
Firefly i is moved towards j, with attractiveness varying as per distance r via exp[−ϒ.r]
Rank the fire flies and find the current best.
Terminate and visualize the results.

Fig. 4. Flow Chart of Fire Fly Algorithm.

formance and requirement of the optimization of fewer parameters associated with penalty and kernel parameters [42]. The cross-validation process is used for the calculation of the accuracy of the classifier. In this case, the input data is divided into two parts such as training and testing. The validation accuracy is calculated from the unknown testing part of the data set. In the present case, 80% of the data is used for training and the remaining 20% is used for testing and a five-fold cross-validation scheme is also used. The simulation study is carried out in the MATLAB Platform and the validation accuracy is calculated using Equation (22).

$$V\text{alidation Accuracy} = 1 - (k - \text{fold Loss}) \tag{22}$$

In the optimization process, the cost function is considered as a minimization problem, the k-fold loss is taken as the objective function. The range of k-foldloss is from 0 to 1 and the lower the value the better is the overall performance of the classifier.

4.3. Problem formulation

The problem dealt with in this paper is to calculate the optimum values of conversion between linear to mel scale and higher and lower cut-off frequencies in of the filters related Equations (1), (2) and (3). These equations are re-written here with the proper variables that needs to be optimized. The conversion of the linear scale frequency of DFT (f) to Mel scale (f_m) is written with a variable factor δ in theEquation (23). This idea of optimizing $\delta$for classification is inspired a similar implementation of the Optimization in Automatic Speech Recognition [14]. In normal MFCC calculations, the value of δ is taken as 7.

$$f_m = 2595 \cdot \log_{10}\left(1 + \frac{f}{(\delta \times 100)}\right) \tag{23}$$

Correspondingly, the conversion of mel scale to linear scale is also written with the variable δ.

$$f_{c\,p} = (\delta \times 100) \cdot \left[\left(10^{\frac{fm_{c\,p}}{2595}}\right) - 1\right] \tag{24}$$

It is also observed from the literature that the linear scale lower and higher cut-off frequencies (f_l and f_h) and correspondingly the melscale lower and higher cut-off frequencies (fm_l and fm_h) in Equation (2) can also be optimally chosen for improvement in the overall classification accuracy [13].

4.4. Need for finding optimum values of δ., fl and fh using FF algorithm

The effect of the change in magnitude of the parameters on the accuracy of classification is dealt in this section. In Fig. 5, the effects of the change in the value of δ.and maximum frequency of the speech signal on the validation accuracy are shown. It is noticed from these two figures that the selection of the best possible value of the δ.to achieve the highest validation accuracy is

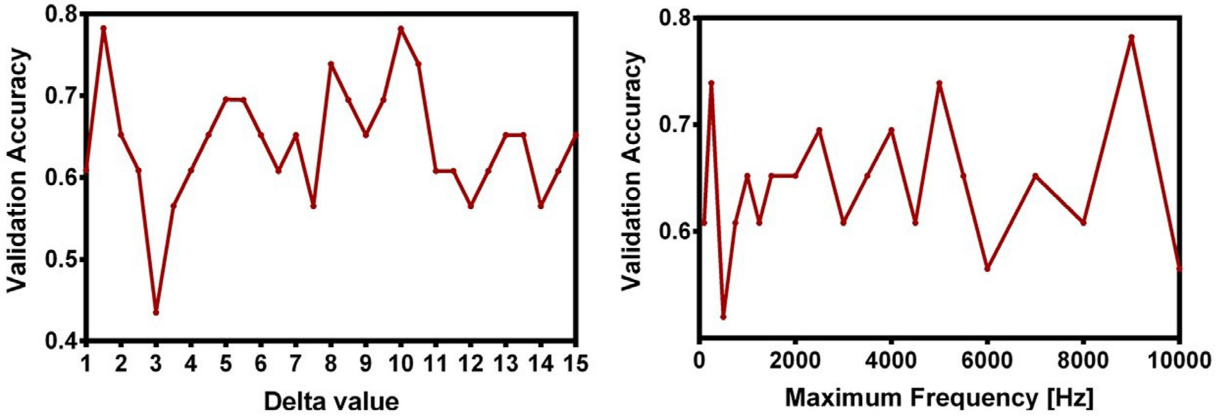

Fig. 5. Effect of change in the Delta value and maximum frequency on the Validation Accuracy.

not straight forward and hence can not be selected using the empirical calculations. Due to the random nature in the pattern of these graphs, to select the best parameters to achieve the maximum validation accuracy, the bio-inspired optimization technique is employed. To fulfill this requirement, the FF algorithm is chosen in this paper.

The objective cost function is to minimize the k-fold loss. It is defined as

$$Objective\ Function\ =\ \min_{\delta,\ f_l,\ f_h}\{\ k-fold\ loss\ \} \tag{25}$$

The k-fold loss of the classifier depends on the values of δ, fl and fh. Hence, these three variables (δ, fl and fh) are chosen for optimization. The k-fold loss needs to be minimized by suitably optimizing by using nature inspired technique. The range of k-fold loss lies between 0 and 1. The goal is to calculate the best possible values of these variables to obtain the lowest possible value of the k-fold loss at the output of the SVM classifier.

4.5. Speech enhancement

Speech enhancement is a process used to denoise the noisy speech signal and to improve the overall quality and intelligibility of the denoised speech. It is widely used in hearing aids, speech communications, and speech recognition tasks. Recently, the phase spectrum compensation based speech enhancement has been proposed [43] and its performance has been shown to be improved using the bio-inspired and ANN techniques [40]. This algorithm has been employed in this paper in the speech enhancement part. This algorithm is based on the concept of the use of proper scaling factor in the phase spectrum compensation. The Flow chart of this algorithm is shown in Fig. 6.

4.6. Adaptive synthetic sampling approach for imbalanced learning

Classification based on Imbalanced learning is a challenging task in the fields of machine learning and data mining. One of the effective approaches to handle such a problem is called Adaptive Synthetic Sampling Approach for Imbalanced Learning (ADASYN). It solves the problem of the imbalanced classification problem by generating new data from the minority class (synthetic data). This is achieved by reducing the bias of the class imbalance, and gradually changing the classification decision boundary [37].

The motivation behind using the ADASYN algorithm in the proposed COVID-19 detection problem due to its encouraging performance in speech recognition (vowel) in [37], where the ratio between the number of the minority to majority samples is 90:900. Similarly, in the present case, the ratio of minority to majority class in Database-1 and Databse-2 are above 30% to 70% which are unbalanced case.

The steps of the or the ADASYN algorithm are expressed below.

- Step-1 — Calculate the degree of class imbalance (d) from the given training data
- Step-2 — Generate the number of new data samples (synthetic data = G) for the minority class
- Step-3 — Calculate the r_i from the K nearest neighbors calculation and normalize it. (where $r_i = \frac{\triangle_i}{K}$, where $\triangle_i$ is the number of examples in the K nearest neighbors of $\boldsymbol{x_i}$ (feature vector of the ith sample) that belong to the majority class)
- Step-4 — Calculate the number of synthetic data samples that need to be generated for each minority sample
- Step-5 — Generate the synthetic data using Equation (26) for the loop varying from 1 to g_i

$$s_i = x_i + (x_{zi} - x_i) \times \lambda \tag{26}$$

where g_i is the number of synthetic data that need to be generated for each minority sample, λ is a random number between 0 and 1, and $\boldsymbol{x_{zi}}$ is the random selection of one minority data sample from the K-nearest neighbors. The important parameter which makes ADASYN algorithm better than other similar technique [44] is the use of a density distribution which automatically calculates the number of synthetic samples that need to be generated for each minority data sample.

5. Experiment

In this section the details of the simulation based experiments carried out and various results are obtained. The different steps in simulation study snd the corresponding flowchart is shown in Fig. 3

Step-1 — A balanced labeled data set is prepared with 200 speech samples.

Step-2 — The sampling rate is converted to 16 kHz for all the speech samples.

Step-3 — Speech Enhancement principle is applied to remove the unwanted noise components from the data set.

Step-4 — The best possible values of δ, fl and fh in the MFCC implementation are obtained which helps to yield the lowest possible value of k-fold loss of the SVM classifier using the enhanced speech data set.

Step-5 — The optimized values of δ, fl and fh obtained from the Step-4 are applied and the corresponding modified cepstral features of the remaining speech samples of database-1 and 2 are obtained.

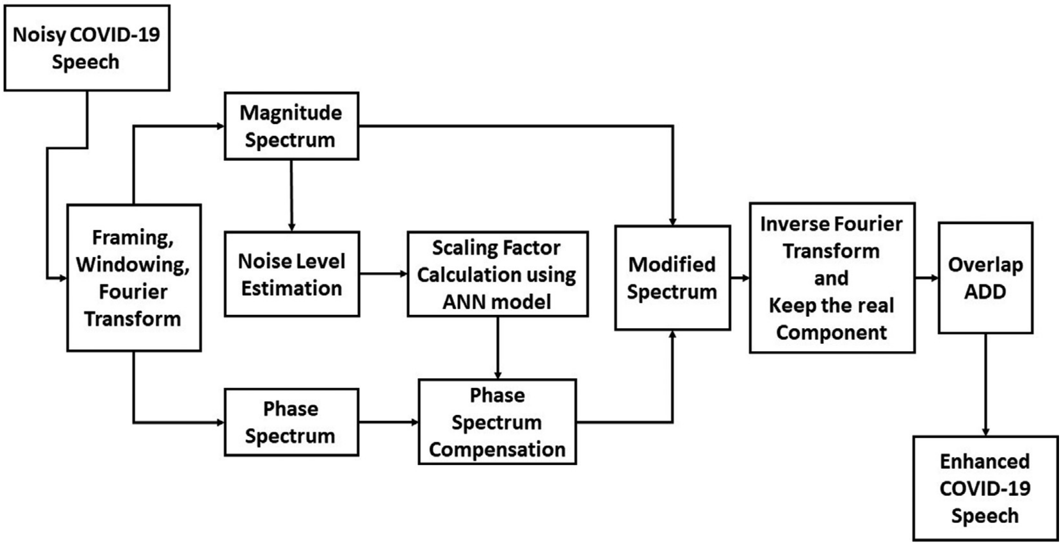
Noisy COVID-19 Speech
Framing, Windowing, Fourier Transform
Magnitude Spectrum
Noise Level Estimation
Scaling Factor Calculation using ANN model
Phase Spectrum
Phase Spectrum Compensation
Modified Spectrum
Inverse Fourier Transform and Keep the real Component
Overlap ADD
Enhanced COVID-19 Speech

Fig. 6. Flow Chart of Improved phase aware speech enhancement Scheme.

Step-6 - The various performance measures of classification using the optimized features are obtained from the SVM classifier.

5.1. Performance evaluation measures

5.1.1. Sensitivity

The performance of any classifiers is commonly evaluated by using either a numeric metric (accuracy), or a graphical representation of performance (receiver operating characteristic (ROC) curve). A commonly used classification metric, the Confusion matrix is calculated from the four values such as: TP (True Positive), FP (False Positive), FN (False Negative), and TN (True Negative) [45]. For the medical diagnosis, the FN plays a crucial role because it shows the class of patients who suffer from the COVID-19 disease but the classifier has falsely predicted them to be healthy [46]. But the FP has less significance in COVID-19 as the patient can go for the second round of tests to confirm. But the COVID-19 positive patients should not have false interpretation as if they are negative. To effectively find the FNs, the Recall (Sensitivity) value is used. It is calculated as

$$Sensitivity = \frac{TP}{TP + FN} \tag{27}$$

5.1.2. F-βScore

Traditionally, F-βScore is another performance measure of classification accuracy. The βvalue indicates whether the evaluation importance would be to be given to FP or FN. It is known from the medical field that identification of FN is more significant than FP and βvalue is taken as 2 for evaluation of classification accuracy [47]. The F-βScore and F-2 Score are computed as

$$\begin{gathered} F_\beta = \left(1 + \beta^2\right) \times \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall} \\ Precision = \frac{TP}{TP + FP} \text{ and } Recall = \frac{TP}{TP + FN} \\ F_2 = \frac{5 \times Precision \times Recall}{(4 \times Precision) + Recall} \end{gathered} \tag{28}$$

5.1.3. Kruskal-Wallis tests

Kruskal-Wallis tests are widely used to check the suitability of the input features to be used in the classification task. It is a non-parametric evaluation and no assumption is made about any prior distribution of the input data [6]. The test is based on statistical parameter H defined in Equation (29), where the total number of samples including all the classes is M, the number of samples in the jth class is m_j, and the sum of ranks of the jth class is R_j, and N is the number samples in the independent group.

$$H = \left(\frac{12}{M(M+1)} \sum_{j=1}^{N} \frac{R_j^2}{m_j}\right) - 3(M+1) \tag{29}$$

5.2. Performance evaluation using TFBCC-M features

In this Section, the performance analysis using the existing MFCC features (TFBCC-M features) is carried out. For this purpose, 50 uniformly distributed random are selected from the COVID-19 positive and negative subjects in each of cough (C-1), breathing (B-1), and voiced (V-1) from database-1 and cough (C-2), breathing (B-2) from database-2. The 13 MFCC feature vectors are extracted and classified using the SVM classifier using five-fold cross validation. The performance is evaluated in terms of the validation accuracy, sensitivity, F-2 score, and Kruskal-Wallis tests using two databases. The results are plotted in Fig. 7. The Receiver Operating characteristics (ROC) and Area Under Curves (AUC) are obtained and plotted in Fig. 8.

It has been observed that the overall performance of the TFBCC-M features are not satisfactory in the detection of COVID-19 and also it is noticed that the cough sounds are providing consistent performance compared to other categories of sounds. Therefore, to improve the performance of the TFBCC-M features the cough sounds are used for determining the optimum values of δ, f_l and f_h by using FF Algorithm.

5.3. finding the optimum values of δ, f$_l$ and f$_h$ using FF algorithm

The relationship between the best cost (k fold loss) and the number of iterations obtained from the simulation study is shown in Fig. 9. The associated parameters of the FF algorithm used in the simulation study are $\gamma = 0.2$, $\beta_0 = 1$, $\vartheta = 0.98$ and the population size is assumed to be 50. These parameters have been chosen based on trial and error which provides the least k-fold loss at the output of the classifier. The objective function to be minimized is given in Equation (25). The three attributes δ, f_l and f_h are associated as a member of the population of the FF algorithm. The light intensity value affects the k-fold loss for a given range of δ,

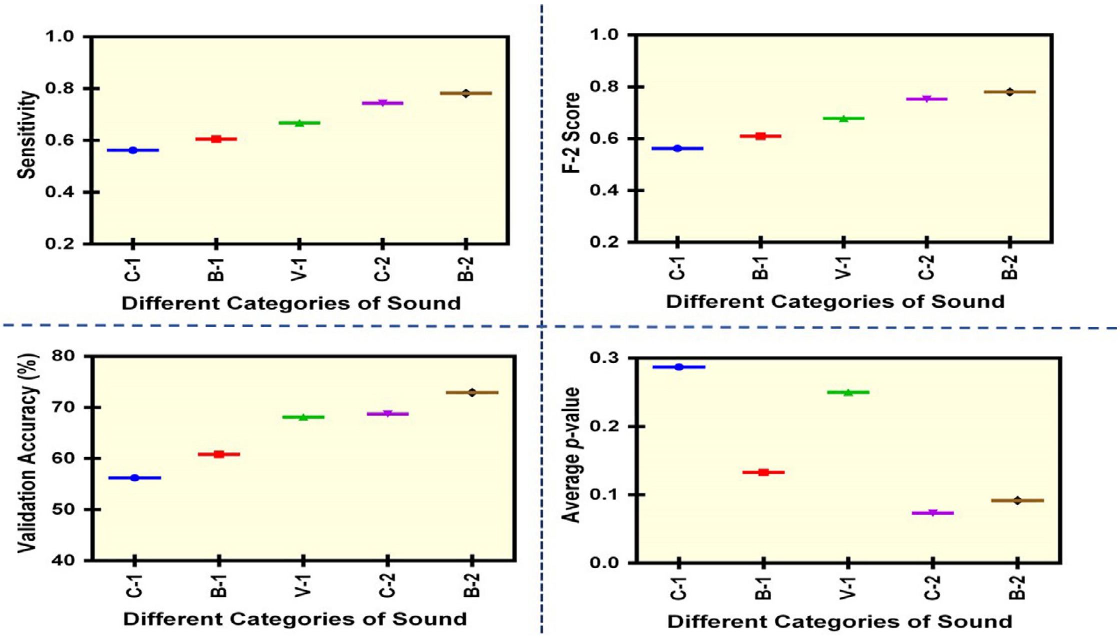

Fig. 7. Performance analysis of the TFBCC-M features.

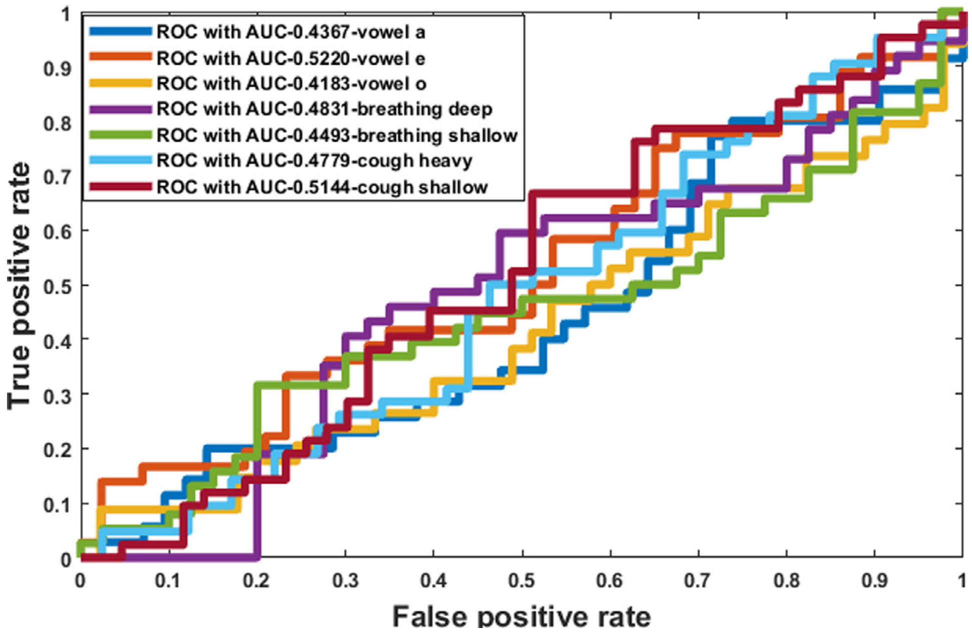

Fig. 8. Comparison of the performance of standard MFCC features for databse-1.

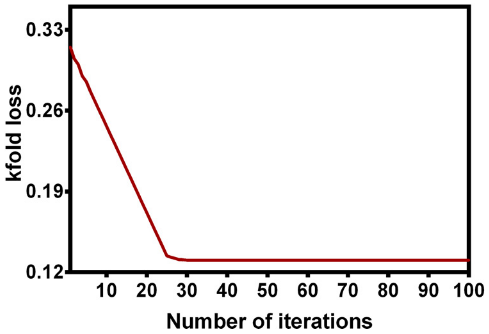

Fig. 9. The learning characteristics obtained during the optimization of FF algorithm.

f_l and f_h. The attractiveness is governed by light intensity variation with distance and the absorption coefficient. It is expressed in Equation (18). The minimization of k-fold loss continues using the FF algorithm until it attains the best possible minimum value. After obtaining the satisfactory convergence, the best member of the population provides the optimized values of δ, f_l and f_h.

5.4. Discussion

In this section, the detection performance using the proposed Cepstral Coefficients (C-19CC) is compared with another seven types of audio features such as TFBCC-M features (T-M), TFBCC-B (T-B) features, TFBCC-H (T-H) features, TFBCC-E features (T-E) [33], DWT based MFCC Features (D-M) [48], TQWT based MFCC Features (T-M) [49], and Temporal and Spectral acoustic features (T-S)

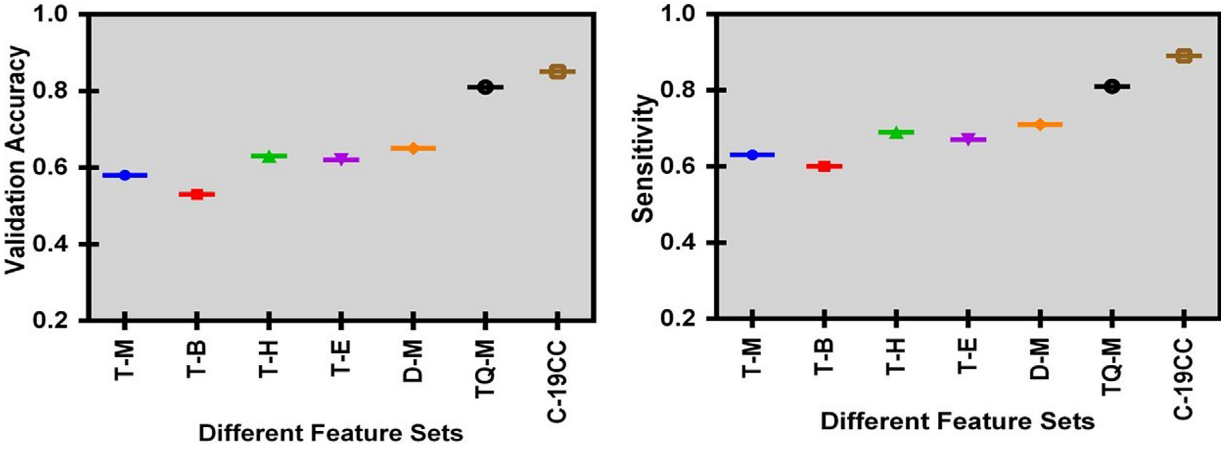

Fig. 10. Comparison of the validation accuracy and sensitivity using different feature sets.

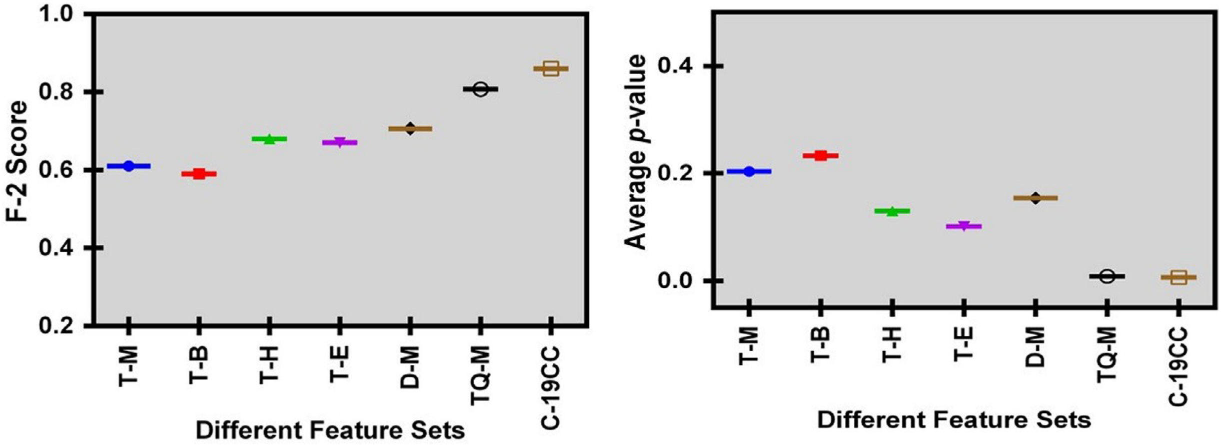

Fig. 11. Comparison of the F-2 score and average p-value using different feature sets.

[23] for cough sounds. The effectiveness of the proposed C-19CC features is illustrated using different measures such as validation accuracy, sensitivity, F-2 score and Kruskal-Wallis tests for the two databases.

The better classification performance demonstrates the efficacy of the proposed method. The optimization of MFCC features has been previously used in the area of spoken language recognition [13] speaker-independent emotion recognition [18], source recognition of Cell-Phones [19]. In this current research work, better-optimized MFCC features are obtained and used for efficient detection of COVID-19. An important advantage of using C-19CC features as the problem of feature selection or reduction is avoided. In traditional speech recognition tasks, several spectral, cepstral, temporal, and wavelet features are obtained and combined to form the desired feature vector which is a tedious and time-consuming process. On the other hand, the proposed C-19CC features are easy to generate and potential in performance. In Figs. 10 and 11, the comparison of various performance measures using different feature sets is made for the combined cough category of sounds using the relevant data of the two databases. It is observed that the sensitivity is high because of the lower FN values. This is very important in the case of medical data classification. The same is also evident from the plot of the F-2 Score and validation accuracy. To further justify the effectiveness of the proposed technique, the p-value comparative analysis is carried out for standard MFCC and C-19 CC features as inputs. The average comparative performance plots are presented in Figure 10. The average p-value is observed to be quite low. The lower the p-value, the better is the effectiveness of that feature. The average p-values obtained for all the 13 coefficients of C-19CC are quite low and thus are suitable features for the COVID-19 classification. The validation accuracy of different audio features is plotted in Fig. 10. It is noticed that the C-19CC features-based model outperforms the other input feature-based models. This justifies the effectiveness of the suggested features. The F-2 scores of the individual categories of both databases are listed in Table 3 using the two databases.

It is observed from Table 3 that in the category of cough sounds, the proposed C-19 CC provides the highest F-2 scores of 0.851 and 0.741 for the database-2 and database-1 respectively. While in the Vowel category, the E sound TQ-M features perform the best and the /e/ vowel is providing superior performance compred to /o/ and /a/. Similarly, for the counting case, the T-S features exhibit the best performance and in the breathing category, the T-M and D-M features are better than the others. The validation accuracy is also found to be the highest for the category of cough sound using the proposed C-19 CC features. Considering all the categories and databases, the cough sound is found to be the best one which provides the highest accuracy of detection as well as the highest F-2 score employing the C-19CC features. The acoustic properties of the coughing signal are different from the speech signal in terms of bandwidth and the way of perception of sound. The application of the additional speech enhancement block has further improved the detection performance of the proposed model. Additionally, the use of the ADASYN tool for removing the class imbalance in both the databases has improved the classification performance because it helps in identifying the better features.

Table 3
Performance Comparison of the different categories of sounds using 8 different feature sets.

Category of Sound	Evaluation Measures	Features							
		T-M	T-B	T-H	T-E	D-M	TQ-M	T-S	C-19CC
Vowel-/o/ (DB-1)	F-2 Score	0.547	0.527	0.604	0.515	0.577	**0.685**	0.674	0.619
	Accuracy	0.590	0.485	0.611	0.532	0.552	**0.692**	0.656	0.626
Vowel-/e/ (DB-1)	F-2 Score	0.587	0.543	0.535	0.528	0.597	**0.704**	0.645	0.645
	Accuracy	0.596	0.577	0.559	0.515	0.656	**0.682**	0.674	0.663
Vowel-/a/ (DB-1)	F-2 Score	0.543	0.523	0.523	0.423	**0.603**	0.595	0.598	0.595
	Accuracy	0.509	0.469	0.502	0.414	0.582	**0.593**	0.554	0.575
Counting-Fast (DB-1)	F-2 Score	0.681	0.641	0.671	0.611	0.691	0.701	**0.728**	0.708
	Accuracy	0.636	0.652	0.616	0.61	0.684	0.721	**0.735**	0.699
Counting-Normal (DB-1)	F-2 Score	0.543	0.525	0.532	0.456	0.564	0.597	**0.713**	0.713
	Accuracy	0.537	0.567	0.519	0.468	0.555	0.558	**0.725**	0.714
Cough-Shallow (DB-1)	F-2 Score	0.608	0.546	0.535	0.454	0.583	0.612	0.701	**0.729**
	Accuracy	0.592	0.567	0.574	0.502	0.576	0.614	0.685	**0.741**
Cough-Heavy (DB-1)	F-2 Score	0.612	0.532	0.562	0.462	0.632	0.662	0.698	**0.711**
	Accuracy	0.586	0.496	0.492	0.499	0.592	0.666	0.659	**0.723**
Breathing-Deep (DB-1)	F-2 Score	**0.594**	0.529	0.502	0.466	0.576	0.546	0.577	0.557
	Accuracy	**0.623**	0.529	0.506	0.439	0.586	0.564	0.562	0.607
Breathing-Shallow (DB-1)	F-2 Score	0.611	0.481	0.475	0.402	**0.615**	0.561	0.556	0.536
	Accuracy	0.586	0.497	0.486	0.496	**0.622**	0.601	0.563	0.534
Cough (DB-2)	F-2 Score	0.735	0.691	0.715	0.695	0.742	0.776	0.751	**0.851**
	Accuracy	0.717	0.651	0.718	0.701	0.759	0.772	0.792	**0.857**
Breathing (DB-2)	F-2 Score	0.751	0.746	0.722	0.719	0.697	0.742	**0.806**	0.732
	Accuracy	0.794	0.788	0.685	0.711	0.694	0.728	**0.815**	0.736

6. Conclusion

The detection of COVID-19 using speech signal can serve as an important cost-effective tool as it does involve any complicated medical test. This approach can easily diagnose the preliminary condition of a patient even without visiting a hospital and without the help of any medical staff as it serves as an automatic detection tool. In this paper, a new audio feature called C-19CC is proposed and used for detection of COVID-19 in this paper and the performance of the method is tested using two standard speech databases. The proposed model has been demonstrated to be superior to other existing speech based COVID-19 detection model reported in the literature. However, it is suggested that the detection accuracy need to be ascertained by appropriate medical experts. The performance can be further be increased by combining the new C-19CC with other temporal and statistical features. The proposed method and the combination of features can also be applied for detection of other speech related diseases.

The proposed C-19CC features are based on the selection of the best possible conversion scale and frequency range of the Cepstral filter bank by using the bio-inspired technique. This is achieved by Identification of the appropriate sound patterns to efficiently detect COVID-19 and application of the speech enhancement schemes for the improvement of the classification performance. In this paper, a simple SVM based classifier is used for detection purpose. However, the classification accuracy can further be improved by using deep learning-based techniques.

The attributes given in the dataset are breath, cough and voiced vowel sounds. Moreover, the analysis can be extended to study the phonetic relevance and identification of phonemic grouping of speech based COVID-19 detection. For this study there is a requirement of preparation of the phonetically balanced dataset of COVID-19. The optimization method of the filter bank parameters can also be extended to different mechanical applications of cepstral analysis, where the properties of the input signal is quite different from that of standard human speech signals. There is a scope for further work to reduce computational complexities associated with this method so that it may be suitable for the real-life application using FPGA[8].

Declaration of Competing Interest

All authors declare that there is no conflict of interest in this work.

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 of the Old Schools Trinity Lane, Cambridge CB2 1TN, UK for sharing the speech database of COVID19 sound App of the paper published in ACM KDD [8].

References

[1] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 Infection: origin, transmission, and characteristics of human coronaviruses, J. Adv. Res. (2020).

[2] WHO Coronavirus Disease (COVID-19) Dashboard Data, https://covid19.who.int/.

[3] C. Sun, Z. Zhai, The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission, Sustainable cities and society 62 (2020) 102390.

[4] More than virus, fear of stigma is stopping people from getting tested: Doctors, 2020, web edition, https://www.newindianexpress.com/states/karnataka/2020/aug/06/more-than-virus-fear-of-stigma-is-stopping-people-from-getting-tested-doctors-2179656.html.

[5] J. Han, K. Qian, M. Song, Z. Yang, Z. Ren, S. Liu, J. Liu, H. Zheng, W. Ji, T. Koike, An early study on intelligent analysis of speech under COVID-19: severity, sleep quality, fatigue, and anxiety, arXiv preprint arXiv:2005.00096 (2020).

[6] B. Karan, S.S. Sahu, J.R. Orozco-Arroyave, K. Mahto, Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech, Biomed. Signal Process. Control 61 (2020) 102050.

[7] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P.H. Robert, Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease, Alzheimer's &amp; Dementia: Diagnosis, Assessment &amp; Disease Monitoring 1 (1) (2015) 112–124.

[8] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, C. Mascolo, Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data, arXiv preprint arXiv:2006.05919 (2020).

[9] G. Deshpande, B. Schuller, An overview on audio, signal, speech, &amp; language processing for COVID-19, arXiv preprint arXiv:2005.08579 (2020).

[10] J. Han, K. Qian, M. Song, Z. Yang, Z. Ren, S. Liu, J. Liu, H. Zheng, W. Ji, T. Koike, An early study on intelligent analysis of speech under COVID-19: severity, sleep quality, fatigue, and anxiety, arXiv preprint arXiv:2005.00096 (2020).

[11] A.V. Oppenheim, R.W. Schafer, From frequency to quefrency: a history of the cepstrum, IEEE Signal Process. Mag. 21 (5) (2004) 95–106.
[12] K.A. Sheela, K.S. Prasad, Linear discriminant analysis F-Ratio for optimization of TESPAR &amp; MFCC features for speaker recognition., J. Multimed. 2 (6) (2007).
[13] C. Hanilçi, F. Ertas, Optimizing acoustic features for source cell-phone recognition using speech signals, in: Proceedings of the first ACM workshop on Information hiding and multimedia security, 2013, pp. 141–148.
[14] S. Chatterjee, W.B. Kleijn, Auditory model-based design and optimization of feature vectors for automatic speech recognition, IEEE Trans. Audio Speech Lang. Process. 19 (6) (2010) 1813–1825.
[15] V. Kadyan, A. Mantri, R.K. Aggarwal, A heterogeneous speech feature vectors generation approach with hybrid hmm classifiers, Int. J. Speech Technol. 20 (4) (2017) 761–769.
[16] Y. Sun, Y. Zhou, Q. Zhao, Y. Yan, Acoustic feature optimization based on F-ratio for robust speech recognition, IEICE Trans. Inf. Syst. 93 (9) (2010) 2417–2430.
[17] R.K. Aggarwal, M. Dave, Filterbank optimization for robust ASR using GA and PSO, Int. J. Speech Technol. 15 (2) (2012) 191–201.
[18] V. Kadyan, A. Mantri, R.K. Aggarwal, Improved filter bank on multitaper framework for robust punjabi-ASR system, Int. J. Speech Technol. 23 (1) (2020) 87–100.
[19] H. Kou, W. Shang, I. Lane, J. Chong, Optimized MFCC feature extraction on GPU, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 7130–7134.
[20] Z. Wang, Y. Xiao, Y. Li, J. Zhang, F. Lu, M. Hou, X. Liu, Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays, Pattern Recognit. 110 (2021) 107613.
[21] A. Oulefki, S. Agaian, T. Trongtirakul, A.K. Laouar, Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, Pattern Recognit. (2020) 107747.
[22] N. Dey, V. Rajinikanth, S.J. Fong, M.S. Kaiser, M. Mahmud, Social group optimization-assisted Kapur's entropy and morphological segmentation for automated detection of COVID-19 infection from computed tomography images, Cognit. Comput. 12 (5) (2020) 1011–1023.
[23] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S.R. Chetupalli, P.K. Ghosh, S. Ganapathy, Coswara-A database of breathing, cough, and voice sounds for COVID-19 diagnosis, arXiv preprint arXiv:2005.10548 (2020).
[24] C.H. You, M.A. Bin, Spectral-domain speech enhancement for speech recognition, Speech Commun. 94 (2017) 30–41.
[25] L. Wang, C. Zhang, J. Liu, Deep Learning Defense Method Against Adversarial Attacks, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 3667–3672.
[26] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, Q. Sun, Deep learning for image-based cancer detection and diagnosis- A survey, Pattern Recognit. 83 (2018) 134–149.
[27] G. Zhong, L.-N. Wang, X. Ling, J. Dong, An overview on data representation learning: from traditional feature learning to recent deep learning, The Journal of Finance and Data Science 2 (4) (2016) 265–278.
[28] Y.-D. Zhang, S.C. Satapathy, S. Liu, G.-R. Li, A five-layer deep convolutional neural network with stochastic pooling for chest CT-based COVID-19 diagnosis, Mach. Vis. Appl. 32 (1) (2021) 1–13.
[29] S. Ahuja, B.K. Panigrahi, N. Dey, V. Rajinikanth, T.K. Gandhi, Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices, Applied Intelligence 51 (1) (2021) 571–585.
[30] P.S. Sujitha, G.K. Pebbili, Cepstral analysis of voice in young adults, Journal of Voice (2020).
[31] E. Benmalek, J. Elmhamdi, A. Jilbab, Multiclass classification of Parkinson's disease using cepstral analysis, Int. J. Speech Technol. 21 (1) (2018) 39–49.
[32] E.S. Doc, Speech processing, transmission and quality aspects (STQ); distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithms, ETSI ES 202 (050) (2002) v1.
[33] N. Sugan, N.S.S. Srinivas, L.S. Kumar, M.K. Nath, A. Kanhe, Speech emotion recognition using cepstral features extracted with novel triangular filter banks based on bark and ERB frequency scales, Digit. Signal Process. (2020) 102763.
[34] N. Sugan, N.S. Srinivas, N. Kar, L.S. Kumar, M.K. Nath, A. Kanhe, Performance comparison of different cepstral features for speech emotion recognition, in: 2018 International CET Conference on Control, Communication, and Computing (IC4), 2018, pp. 266–271.
[35] B. Karan, S.S. Sahu, K. Mahto, Parkinson disease prediction using intrinsic mode function based features from speech signal, Biocybernetics and Biomedical Engineering 40 (1) (2020) 249–264.
[36] N. Strisciuglio, M. Vento, N. Petkov, Learning representations of sound using trainable COPE feature extractors, Pattern Recognit. 92 (2019) 25–36.
[37] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, in: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), 2008, pp. 1322–1328.
[38] X.-S. Yang, Firefly algorithms for multimodal optimization, in: International symposium on stochastic algorithms, 2009, pp. 169–178.
[39] X.-S. Yang, X. He, Firefly algorithm: recent advances and applications, International journal of swarm intelligence 1 (1) (2013) 36–50.
[40] T.K. Dash, S.S. Solanki, G. Panda, Improved phase aware speech enhancement using bio-inspired and ANN techniques, Analog Integr. Circuits Signal Process. (2019) 1–13.
[41] L. Auria, R.A. Moro, Support vector machines (SVM) as a technique for solvency analysis (2008).
[42] B. Karan, S.S. Sahu, J.R. Orozco-Arroyave, K. Mahto, Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech, Biomed. Signal Process Control 61 (2020) 102050.
[43] A.P. Stark, K.K. Wójcicki, J.G. Lyons, K.K. Paliwal, Noise driven short-time phase spectrum compensation procedure for speech enhancement, Ninth annual conference of the international speech communication association, 2008.
[44] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284.
[45] J. Lever, M. Krzywinski, N. Altman, Points of significance: classification evaluation., 2016, (????).
[46] S.A. Hardwick, I.W. Deveson, T.R. Mercer, Reference standards for next-generation sequencing, Nat. Rev. Genet. 18 (8) (2017) 473.
[47] D. Devarriya, C. Gulati, V. Mansharamani, A. Sakalle, A. Bhardwaj, Unbalanced breast cancer data classification using novel fitness functions in genetic programming, Expert Syst. Appl. 140 (2020) 112866.
[48] Z. Soumaya, B.D. Taoufiq, B. Nsiri, A. Abdelkrim, Diagnosis of Parkinson disease using the wavelet transform and MFCC and SVM classifier, in: 2019 4th World Conference on Complex Systems (WCCS), 2019, pp. 1–6.
[49] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, Appl. Soft Comput. 74 (2019) 255–263.

Dr. Tusar Kanti Dash has received Ph.D. from Birla Institute of Technology, Mesra India in 2020 in the area of Speech Processing. He has got his B.E. Degree in Electronics & Telecommunication Engineering from the Utkal University, India in 2003, and M.Tech. Degree in Electronics & Communications Engineering from IIT, Kharagpur, India in 2013. He is working at the Department of Electronics & Telecommunications Engineering, C.V. Raman Global University, Bhubaneswar, India as an Assistant Professor. His research interests include Audio Signal Processing and Evolutionary Computing.

Ms. Soumya Mishra has received B.Tech. and M.Tech. degree from Biju Patnaik University of Technology, India in Electronics and Communications Engineering in the year 2008 and 2011 respectively. She is working at C V Raman Global University, Bhubaneswar, India as an Assistant Professor in the Department of Electronics and Telecom Engineering. She is pursuing a Ph.D. at C.V. Raman Global University, India under the guidance of Dr. T K Dash and Dr. G. Panda. Her research interests include Speech & Audio Signal Processing and Wireless Communication.

Dr. Ganapati Panda did his postdoctoral research work at the University of Edinburgh, UK (19841986) and Ph.D. from IIT Kharagpur in 1981 in the area of Electronics and Communication Engineering. He has already guided 42 Ph.D.s in the field of Signal Processing, Communication, and Machine Learning. Currently, he is working as Professor and Research Advisor at C V Raman Global University, Bhubaneswar, India. Prof Panda is also the Professorial Fellow at the Indian Institute of Technology Bhubaneswar. Prior to this, he was working as Professor, Dean, and Deputy Director in the School of Electrical Sciences of IIT Bhubaneswar.

Dr Suresh Chandra Satapathy has received his PhD in Computer Science Engg from JNTUH, Hyderabad. He has done his MTech Computer Science from NIT, Rourkela. Presently he is Professor in Computer Science at KIIT Deemed to be University, Bhubaneswar, Odisha, India. He hs over 30 years of teaching experience. Machine Learning, Swarm Intelligence etc are his areas of research. He has published more than 150 research articles in various reputed journals and conferences. SGO and SLEO are his two evolutionary optimization algorithms developed by his collaborators.