Cluster Computing
https://doi.org/10.1007/s10586-018-2177-0

Iterative Conditional Entropy Kalman filter (ICEKF) for noise reduction and Neuro Optimized Emotional Classifier (NOEC)

R. Kumuthaveni[1] · E. Chandra[2]

Received: 8 January 2018 / Revised: 10 February 2018 / Accepted: 13 February 2018

Abstract

Emotion has a most important aspect in terms of interactions among the humans and this would become ideal for human emotions to get mechanically identified by the machines and primarily for enhancing the communication among the human–machine. In the recent work Enhanced Bat algorithm with Simulated Annealing (EBSA) are introduced for solving emotion recognition problem. Here the removal of noises from the speech samples and reduction in the number of speech features becomes very difficult task which reduces the accuracy of the classifier. To solve this problem this research work involves detection of emotions from speech which stimulates machines understanding human behavioral tasks namely reasoning, decision making and interaction. EBSA is used in the previous system to identify the happy, sad and neutral emotions from speech input. The performance of the previous system has been decreased due to recognition accuracy and feature selection. Improved Artificial Bee Colony (IABC) with Neuro Optimized Emotional Classifier (NOEC) solves this issue in the proposed system. The Iterative Conditional Entropy Kalman filtering (ICEKF) is initially processed to effectively filter the noisy features from the inputted speech data. Mel Frequency Cepstrum Coefficient (MFCC), pitch, energy, intensity and formants are extracted as speech features. Every extracted feature is maintained in the database and annotated along with their emotional class label. IABC algorithm chooses the feature optimally, which in turn employs the best fitness function values. From the optimally selected dataset, the NOEC is processed. Emotions can be identified from the Tamil news speech dataset with the help of the supervised machine learning technique, which demands the training set (collection of emotional speech recordings). Every recording or sample in the dataset is named with the emotional class and they are indicated as n-dimensional vector of spectrum coefficients which in turn is extracted from the Tamil news speech dataset. This dataset is collected from real time via using the search engine sites like Google, YouTube, twitter etc. By implementing IABC with NOEC classification process, the work segregates the emotional classes such as happy, sad, anger, fear and neutral emotions perfectly. From the experimental verification, it is confirmed that the proposed method IABC with NOEC gives better performances with respect to accuracy, precision, recall and f-measure values.

Keywords Speech enhancement · Feature extraction · Feature selection · Iterative Conditional Entropy Kalman filtering (ICEKF) · Speech emotion recognition · Improved Artificial Bee Colony (IABC) algorithm · Neuro Optimized Emotional Classifier (NOEC) algorithm

1 Introduction

Human emotions are acquired from speech, facial expression, gesture, and so forth. Information obtained regarding these emotions from human speech gives a natural and intuitive interface for speech emotion recognition systems. Detection of human emotions from speech namely speech emotion recognition demonstrates as one among the most promising modalities in speech processing and automatic human emotion recognition. Application of speech emotion

✉ R. Kumuthaveni
kumuthaveni@gmail.com

[1] Dr. SNS. Rajalakshmi College of Arts & Science, Coimbatore 641049, Tamil Nadu, India

[2] Department of Computer Science, Bharathiar University, Coimbatore 641046, Tamil Nadu, India

Published online: 24 February 2018

recognition extends in the areas of healthcare, psychology, cognitive sciences, e-learning, monitoring, entertainment, law and marketing. While communicating, an emotion in the speech was carried from one communicator to another [1]. This reveals that the emotional state of a speaker easily trigger an interlocutor emotional state, which results in modification of the speech style or tone. With the help of this process, emotional states were changed and mutually shaped.

Emotion distinguishes the condition of mind, such as joy, anger, love and hate derived from natural instinctive states of mind [2]. Emotions appear in various shapes and in various representations. Signals about the emotions can be provided by the body reactions such as speech, facial expression, walking type etc. Emotions directly trigger human's function of job routine in a positive or negative manner. Recognition of optimistic emotions assist in encouragement of better performance of humans and also on the other end helps to prevent them from strange emotions supporting cynically to a situation. Various systems have been established to recognize the emotions from the speech signal. An analysis is done regarding the speech emotion recognition, which, in turn, depends on the earlier methodologies with various classifiers for the emotion recognition. Here the classifiers work is to discriminate the emotions like anger, happiness, sadness, fear, neutral state, etc. The emotional speech samples are collected from Tamil language particularly from Tamil news channels and added to the database of the speech emotion recognition system and the features were drawn-out from these samples. The features are energy, pitch, Mel Frequency Cepstrum Coefficient (MFCC). According to the extracted features, classification performance is done [3].

The characteristics of the speech signal stay constant for a short period, since it is slowly time varying signal. The characteristics of the signal vary to give back various speech sounds while speaking, when it is analyzed over long duration. A spectral representation helps in a substitutive way to distinguish the speech signal and to represent the information in-corporate with the sound. Formant features were elucidated as adaptive non uniform samples of the signal spectrum which are pin-pointed in the resonance frequencies of the vocal tract and generally occurs to have a signal-to-noise ratios. According to the phonemes and the position of the window along the phoneme [4], the count and location of these frequencies along the frequency axis may vary. For enhancing the short-time spectral coefficients of a noisy speech signal, a Minimum Mean-Square Error (MMSE) estimator is utilized. This research work, deduce the analytical solutions, for computing the Discrete Fourier Transform (DFT) coefficients in the MMSE sense when the previous density function's probability of the clean speech DFT coefficients can be formulated by a complex Laplace or by a tedious bilateral Gamma density [5]. The probability density function of the noise DFT coefficients may be formulated either by a difficult Gaussian or by a complex Laplacian density. According to the Gaussian assumption, MMSE short-time spectral amplitude estimator algorithms are compared, and on these super gaussian densities improve the signal-to-noise ratio.

A neural network is explained in [6], which is very powerful model for the classification of speech signals. Few highly simplified models can identify the small set of words. By the pre-processing technique, the performance of the neural networks is affected highly. It is noticed that the MFCC is a highly authentic tool for pre-processing stage. The major objective of this work is to introduce a new Improved Artificial Bee Colony (IABC) with Neuro Optimized Emotional Classifier (NOEC) for increasing the results of speech emotion recognition. With the help of these coefficients, the acquired output is highly effective. Substantial output is accomplished by the deployment of both the multilayer feed forward and radial basis function neural network with the back propagation algorithm while utilizing MFCC. The emotions like happy, normal anger, fear and sad were recognized by extracting the MFCC, Pitch (PS-ZCPA), intensity, formants and energy features. With the help of Improved Artificial Bee Colony (IABC) algorithm significant feature were chosen. By enforcing the combined PSO with ANN (NOEC) approach, the classification process is done, which assist to give the perfect output for classification process in the provided speech input.

2 Related work

A subband Kalman filtering scheme [7], which signifies the auditory masking properties for single channel speech enhancement. It tries to accomplish the extreme quality enhanced speech by enhancing the trade-off among the speech distortion and noise reduction. The Kalman filtering in the sub band rather than the full-band domain leads to drop in complication and performance enhancement. A novel approach is used to integrate the masking threshold with subband Kalman filtering, through which the prediction of the noise variance is employed in the Kalman filtering process and every subband is changed, based on the masking threshold. The optimal configuration for this scheme is analyzed. This approach tends to a good output when distinguished with the full-band and the conventional subband Kalman filtering methods. It provides that the scheme exceeds with respect to the objective as well as subjective measures, through intensive simulations.

A two-level classifier with a pitch-based gender identification method is proposed in [8], in order to defeat the difficulty of MFCC-based gender classification. The first-stage classified the gender, while the pitch represents the gender of the speaker perfectly, by utilizing the threshold-based decision rule. For undetermined speakers or tedious cases, the second-stage Gaussian Mixture Model (GMM) classifier is exploited, which gives satisfying accuracy values. In [9], proposed a subject independent and multi-style speech recognition which assist for enhancing the emotional and stressed state recognition system. It helps many applications like Human Computer Interaction (HCI), health care system, investigating criminal problems. Blue Cross and Blue Shield of Florida (BSBCFs) features were drawn-out from both speech and glottal waveforms and they are joined with the Inter speech 2010 features, in order to enhance the rate of the subject independent emotion recognition. To minimize the dimension of the Inter speech 2010 feature set, MCFS feature selection algorithm is enforced. To identify the perfect feature set for three different databases, Particle Swarm Optimization (PSO) and Hybrid BBO_PSO optimization were employed.

The technique of feature extraction from the spectrogram image of sound signals for automatic sound recognition [10], which explains the technique in an audio surveillance application and then computes the performance with the help of four common multiclass Support Vector Machine (SVM) classification techniques, one-against-all, one-against-one, decision directed acyclic graph, and adaptive directed acyclic graph. Experimentation is continued with the help of an audio database with 10 sound classes, and each has multiple subclasses with intra class diversity and interclass similarity with respect to signal properties.

ABC algorithm is proposed [11], which, in turn, used for data dimension reduction on classification issues. Then to choose the optimal subset from the original high-dimensional data, it is used. For fitness evaluation within the ABC framework, k-NN method is implemented. To generate an efficient dimension reduction method, ABC and k-NN were chained and wrapped together. Three kinds of bee varieties were utilized in ABC, they are employed bees, onlooker bees, and scout bees. This method enforces ABC wrapping with a k-NN classifier. To compute the fitness value of the ABC food sources, k-NN is used. For choosing the optimal subset of features, accuracy is used as a measure. From the gene expression analysis result, it is proved that the ABC–kNN method can effectively decrease the data dimension while controlling the extreme classification accuracy. An Artificial Neural Network (ANN) [12] is proposed for recognizing the flow pattern, but with the preprocessing stage with the help of natural logarithmic normalization. This pre-processing step assists in normalizing huge data range and to minimize the overlap among the flow patterns. Then with the help of dimensionless inputs, the validity of the model is lengthened, in order to execute the horizontal pipes of different diameters, liquid densities and viscosities. The idea is formalized by constructing and verifying the model with the assistance of the experimental data as well as well-known multiphase flow models.

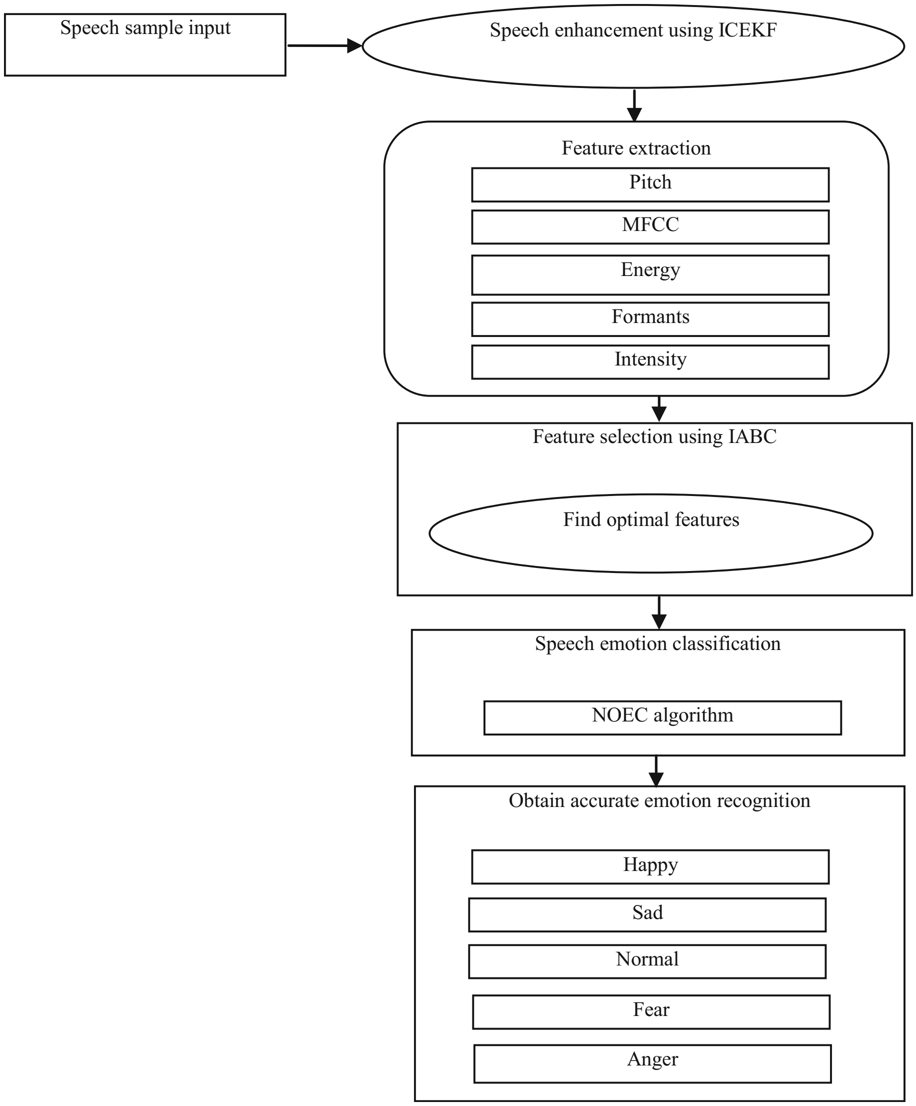
3 Proposed methodology

Enhanced Bat algorithm with Simulated Annealing (EBSA) was employed in the preceding work for speech emotion classification. The features like pitch, MFCC and energy were extracted and consumed to enhance the speech emotion recognition accuracy. However, in the feature extraction, the formant and intensity features weren't conceived. The feature selection was not indulged in the previous methodology. So, the effectiveness of the entire system was not to the satisfaction. Iterative Conditional Entropy Kalman filtering is applied in speech input to enhance the speech by clearing the noisy features. The Improved Artificial Bee Colony (IABC) algorithm is specifically suggested for feature selection in the proposed work to opt the significant features from extracted features and classification is brought-in through a Neuro Optimized Emotional Classifier (NOEC) to improve the entire system's performance. Further emotions intended for categorization in this projected method are fear and anger along with happy, sad and neutral features. In order to manage the feature's count and to give more accuracy in speech emotion recognition, suggested system is adequate. The primary contribution of the suggested system are: Iterative Conditional Entropy Kalman filtering for speech enhancement, feature extraction employing pitch, MFCC, energy, formants and intensity features, feature selection applying IABC optimization algorithm and speech emotion classification is performed with a NOEC. The Fig. 1 explains the entire block diagram of the proposed system.

3.1 Speech enhancement by Iterative Conditional Entropy Kalman Filtering (ICEKF)

The speech enhancement is the first module in the suggested architecture and it is determined as an examination of the input speech signal by Kalman filtering approach. This algorithm includes two steps: prediction and measurement update. The speech is enhanced by deducing the equation and covariance error. With the help of this filter, the accuracy of the classification for the provided input speech dataset is improved. The noise features are also predicted, according to the well-established Kalman filter. By exploiting the accessible data, prediction is done while updating the estimation process of the coefficients of the Kalman filter while no update is done when the data is

Fig. 1 Overall block diagram of the proposed system



missing for Kalman filter [13]. At the same time, consuming mean of the data, updating is done with the responsibility even for missing data, x(s) = a(x) + v(x), s = 1,2,…n, x(s) and a(x) denote the discrete time samples of noisy speech and clean speech correspondingly, and v(x) is a white Gaussian noise with variance. The clean speech signal is generally modeled as an Auto Regressive AR process, and approximated as the output of an all-pole linear system driven by an excitation signal, which is considered to be a zero-mean white Gaussian process with variance.

Algorithm

(1) Prediction

$$A_{s|s-1} = FA_{s-1|s-1} \tag{1}$$

$$P_{s|s-1} = FP_{s-1|s-1}F^{s} + Q \tag{2}$$

(2) Measurement update If the noise is occurring

(3) Update filter

$$\begin{cases} K_s = P_{t|t-1}H^T\left(HP_{t|t-1}H^T + R\right)^{-1} \\ A_{s|s} = A_{s|s-1} + K_s\left(B_s - HA_{s|s-1}\right) \\ P_{s|s} = \left(I - K_sH\right)P_{s|s-1} \end{cases} \tag{3}$$

(4) Filter

$$\begin{cases} F_s = P_{s|s}F^T p_{s+1|s}^{p-1} \\ A_{s|s} = A_{s|s} + F_s\left(A_{s+1|T} - A_{s+1|s}\right) \\ P_{s|T} = P_{s|s} + F_s\left(P_{s+1|T} - P_{s+1|s}\right)L_s^T \end{cases} \tag{4}$$

(5) The covariance matrix is estimated by using given below formula

$$\begin{cases} A_{s|j} = E(A_s|A_j) \\ P_{s|j} = E[(A_s - A_{s|j})^T], \end{cases} \tag{5}$$

where the A_s is the state vector, which has the interest for the system (e.g., position, velocity, heading). F is the state transition matrix which enforces the effect of every system state parameter. P_s is covariance matrix, H^T is the transformation matrix that maps the state vector parameters into the measurement domain and K_s is Kalman filter. The Q is noise speech features and F_s is filtered output once after minimizing the noise. The details from the predictions and measurements were joined to give a perfect estimation about the location of the train and it is utilized to substitute the noise values by foreseeing the values in the provided speech dataset [14].

In the Kalman filter, measurement update is still nonlinear and is specified as the Maximum A Posteriori (MAP) estimate according to following Eq. (6)

$$A_{s|s} = \underset{A_s}{\operatorname{argmax}}\, p(A_s|Y_s) \tag{6}$$

The posterior is proportional to the product of the likelihood and the prior.

$$p(A_s|Y_s) \propto p(y_s|A_S)p(A_s|Y_{s-1}) \tag{7}$$

$$\propto \exp -\frac{1}{2}((y_s - h(a_s))^T R_s^{-1}(y_s - h(A_s)) + \left(A_{s|s-1} - A_s\right)^T P^{-1}_{(s|s-1)}\left(A_{s|s-1} - A_s\right) \tag{8}$$

where terms not depending on A_s have been dropped. Since the exponential function is monotone and rising being capable to minimize the negative log instead which gives

$$A_{s|s} = \underset{A}{\operatorname{argmin}} \frac{1}{2}(y_s - h(A))^T R_s^{-1}(y_s - h(s)) + \frac{1}{2}\left(A_{(s|s-1)} - A\right)^T P^{-1}_{(s|s-1)}\left(A_{(s|s-1)} - A\right) \tag{9}$$

$$= \underset{A}{\operatorname{argmin}} \frac{1}{2} r^T(A)r(A) \tag{10}$$

where

$$r(A) = \begin{bmatrix} R_s^{-\frac{1}{2}}(y - h(A)) \\ P^{-\frac{1}{2}}_{(s|s-1)}\left(A_{(s|s-1)} - A\right) \end{bmatrix} \tag{11}$$

Make a fall of time dependence on the variable A. The former expression in the objective function is equivalent to the unidentified sensor noise and the later expression function is equivalent to the importance of the prior to the estimate. In addition, the optimization problem must send back a covariance approximation. There is no common technique for resolving (6) and the nonlinear character entails so as to iterative methods are required to attain an approximate estimate [14]. The Newton step p is the key to the equation

$$\nabla^2 V(A)p = -\nabla V(A) \tag{12}$$

Opening from a first guess ×0, Newton's system iterates the following equations

$$A_{i+1} = A_i - \left(\nabla^2 V(A)\right)^{-1} \nabla V(A_i) \tag{13}$$

where $\nabla^2 V(A)$ and $\nabla V(A_i)$ are the Hessian and the gradient of the cost function, correspondingly. Consider that if $V(A)$ is quadratic flowingly (12) gives the least in a one step. Gauss–Newton can be used while the minimization problem is on nonlinear least-squares form as in (6). After that the gradient has an easy representation.

$$\nabla V(A) = J^T(A)r(A) \text{ where } J(A) = \left.\frac{\partial r(x)}{\partial x}\right|_{x=s} \tag{14}$$

and the Hessian is approximated as

$$\nabla^2 V(A) \approx J(A)^T J(A) \tag{15}$$

neglecting the necessity for second-order derivatives. The GN step follows as

$$A_{i+1} = A_i - \left(J_i^T J_i\right)^{-1} - J_i^T r_i \tag{16}$$

where $J_i^T = \mathrm{J(A_i)}$ and $\mathrm{r_i = r(A_i)}$ are establish to ease notation. Furthermore, the Jacobian estimates to

$$J_i = -\begin{bmatrix} R_s^{-\frac{1}{2}} H_i \\ P^{-\frac{1}{2}}_{(s|s-1)} \end{bmatrix} \text{where } H_i = \left.\frac{\partial h(x)}{\partial x}\right|_{x=s_i} \tag{17}$$

is the measurement Jacobian. In the Kalman filtering method, consider that this work is to introduce and define an entropy function to update the state of the kalman filter. The entropy of a discrete random variable A with signal and Y with signal, the conditional entropy of Y given A is defined as the weighted sum of H(Y|A) for each possible value of A, using p(a) as the weights.

$$P_{s|j} = H\left(Y_j|A_s\right) = \sum_{A \in a} p(a_s)H(Y|A = a) \tag{18}$$

$$= -\sum_{a \in A} p(a) \sum_{y \in Y} p\left(y_j|a_s\right) \log p\left(y_j|a_s\right) \tag{19}$$

In the step 5 the covariance matrix is estimated by using the above mentioned formula (18). So the proposed work, Iterative Conditional Entropy Kalman filtering (ICEKF) is for noise removal.

3.2 Feature extraction

The following features such as the PS-ZCPA, MFCC, energy, Formants and Intensity features are extracted from the speech denoised samples.

3.2.1 PS-ZCPA, MFCC and energy

The pitch-synchronized peaks were utilized in the PS-ZCPA method to extract the given input's features. The ZCPA method calculates the spectrum. To reduce the effect of noise in the performance, and to define the value of n for the speech signals with different SNR for the optimum output, a noise extraction procedure is applied in the suggested PS-ZCPA method. In each and every channel output, the noise level is checked; hence the temporal structure is authentic, even in high frequency bands. The better parametric representation of acoustic signals is extracted and it is desirable to generate the best recognition performance. For the next phase, the effectiveness of the current phase is desirable, since it affects its nature. Based on human hearing, MFCC works, which can't comprehend the frequencies over 1 kHz. For instance, the energy features can be drawn-out effectively from the provided input, which is applied to raise the speech emotion recognition concert [15].

3.2.2 Formants

The formant features were employed for indicating the spectral information in this module. Moreover, this feature gives more instruction regarding the provided input emotions. With the help of formant features, the provided input speech samples can be effectively identified, because it drew-out the appropriate features from the vocal region, which, in turn, assist to recognize the emotions according to the pitch effects. For establishing the different speech systems, cepstral coefficients were utilized as a standard front end features, nevertheless, they perform poorly with noisy or real life speech. Hence, the supplementary features along with basic cepstral coefficients were important to manage the real life speech. The higher amplitude regions of a spectrum like formants were comparatively less impacted by the noise. In [16] drawn-out spectral sub-band centroids from high amplitude regions of the spectrum were used for noisy speech recognition. As supplementary features to cepstral features, the formant parameters were employed in this research work. The current cepstral features uses only amplitude (energy) information from the speech power spectrum, on the other hand, the formant features uses the frequency details as well.

Formant tracks indicates the order of vocal tract shapes, therefore, formant analysis uses their strength, location and bandwidth assist to drew-out the vocal tract which is connected to the particular emotion information from a speech signal. Formants are the reverberations of the vocal tract. Their location and bandwidth prediction is significant in various applications. A technique implemented often for formant frequency includes determination of resonant peaks from the filter coefficients, which is acquired by LPC analysis. The formant parameters are either driven by peak picking on the filter response curve or by rectifying the roots of the equation $\mathrm{A(z)} = 0$, when the prediction polynomial A(z) is computed. In order to compute the respective formant frequency and bandwidth, every pair of complex root is utilized. The difficulty of the polynomial root finding techniques excludes them as approach frequently, for formant estimation. A number of standard references on speech processing give the following transformation from complex root pairs $z = r_0 e^{\pm\theta_0}$ and sampling frequency $\mathrm{f_s}$ to formant frequency and F 3 dB bandwidth B

$$F = \frac{f_s}{2\pi}\theta_0\mathrm{Hz} \tag{20}$$

$$B = \frac{f_s}{2\pi}\ln\mathrm{r}_0\mathrm{Hz} \tag{21}$$

Let us consider that the prediction polynomial is responsive to a second order all pole system, it shows that the formant frequency and 3 dB bandwidth are

$$\hat{F} = \frac{f_s}{2\pi}\arccos\left[\cos\theta_0\frac{\left(r_0^2+1\right)}{2r_0}\right] \tag{22}$$

$$\hat{B} = \frac{f_s}{2\pi}|\arccos(x_1) - \arccos(x_2)| \tag{23}$$

where

$$x_1 = \frac{(1+r_0^2)}{2r_0}\cos\theta_0 + \frac{(1+r_0^2)}{2r_0}\sin\theta_0 \tag{24}$$

$$x_2 = \frac{(1+r_0^2)}{2r_0}\cos\theta_0 - \frac{(1+r_0^2)}{2r_0}\sin\theta_0 \tag{25}$$

$$\left|\hat{F} - F\right| = \frac{f_s}{4\pi}k^2|\cot\theta_0|\left|1 + k + O\left(k^2\right)\right| \tag{26}$$

$$\left|\hat{B} - B\right| = \frac{f_s}{3\pi}k^3\left|1 + 3\cot^2\theta_0\right||1 + O(k)| \tag{27}$$

This work, enforce the first two formant frequencies for emotion recognition and indicate them as $f_F = [F1, F2]'$, where prime indicates vector transpose.

$$\Delta f_F = \frac{\sum_{k=-2}^{2} k f_F[n+k]}{\sum_{k=-2}^{2} k^2} \tag{28}$$

where $f_F[n]$ is the formant feature vector at time frame n and k number of values. The above equations computation represents perfect agreement among the two methods of measuring the formants and the bandwidth, when the respective root is near the unit circle. In this session, the formant estimation is done for emotional speech classification. In this process, the five emotional class speech samples experiments were passed via a filter, in order to clear-away the undesirable frequency components from the signal. The significant features were drawn-out more effectively and it improves the entire classification performance. Formant features linked with the normal spectral features enhances the emotion recognition performance of the systems.

3.2.3 Intensity

Intensity features per frame were computed on the absolute value of the FFT (Fast Fourier Transform) spectrum. They point to the spectral sum of the signal and spectral distribution in every sub-band, provided by

$$I(k) = \sum_{n=0}^{N/2} |FFT_k(n)| \tag{29}$$

$$D_i(k) = \frac{1}{I(k)} \sum_{n=L_i}^{H_i} |FFT_k(n)| \tag{30}$$

where k points to the frame, $I(k)$ is the intensity of the k-th frame and $D_i(k)$ is the intensity ratio of the i-th subband. L_i and H_i are correspondingly the lower and upper bounds of the i-th subband. Neutral or unemotional speech has a much narrower pitch range rather than the emotional speech, and it is recognized as the increased emotional intensity, the frequency and duration of pauses and stops normally recognized at the time of neutral speech were reduced [17].

3.3 Feature selection using Improved Artificial Bee Colony (IABC) algorithm

The optimal feature selection is done in this section for choosing the desirable feature from the earlier module, which is performed by IABC algorithm and it is utilized for raising the searching efficiency by minimizing the iterations count. Feature selection includes recognizing the feasible subset of the most of the helpful features in the provided dataset, which generate the compatible classification output as the overall dataset of the features. The honey bees were classified into three groups such as employed bees, onlooker bees and scout bees, in ABC algorithm. The employed bees count is identical to the onlooker bees. The food sources location and its quality details were collected by the employed bees, from these collected details, the onlooker bee search for food sources and it remains in the hive. From the deserted food sources, the scout bee search new food sources randomly. ABC solution is an iterative process, as like other population-based algorithms. A repetitive iteration of the three phases namely employed bee phase, onlooker bee phase and scout bee phase is necessary once after initializing the ABC parameters and swarm. The complete mechanism of ABC [18] is given in Fig. 2.

The exploration here indicates the capability to identify the global optimum by examining the different unknown areas in the solution search space and exploitation indicates the capability to identify improved solutions by executing the knowledge of the earlier solutions. In behavior, the exploration and exploitation challenge with one another, nevertheless, both capabilities should be graceful to accomplish good performance. The IABC mechanism is illustrated in Fig. 3 [19].

By utilizing the information given by a randomly selected potential solution inside the existing swarm, every potential solution updates itself in ABC. Here, a step size, defined as a linear combination of a random number $\Phi_{ij} \in [-1, 1]$, existing solution and a randomly selected solution were utilized. The quality of the updated solution functions depending on the step size. If the step size is

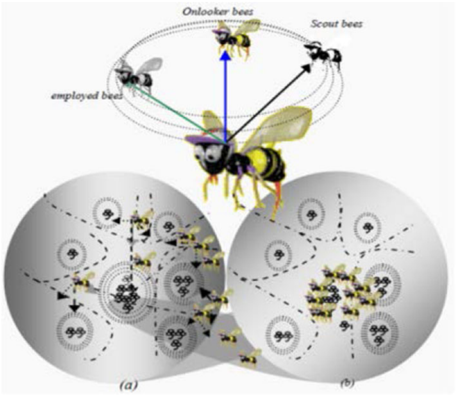


Fig. 2 General mechanism of ABC

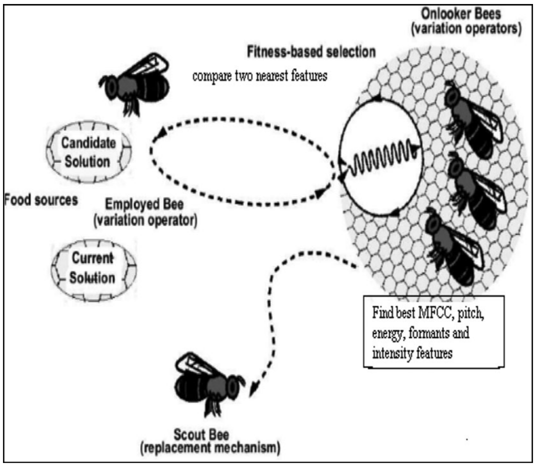


Fig. 3 Improved ABC algorithm process

excessively large, this occurs if the variations of current solution and randomly chosen solution is large enough with extreme absolute value of Φ_{ij}, then the updated solution can beat the true solution and if the step size is excessively small then the convergence rate of ABC may reduce desirably. Proper balances of the step sizes can compensate the ability of the exploration and exploitation of the ABC algorithm at the same time. But, this step has random component, so its balance can't be performed manually. The Improved solution update strategy based on the fitness of the solution, the exploration and exploitation can be balanced. In the general ABC, the food sources are updated, as shown in Eq. (31).

$$v_{ij} = x_{ij} + \Phi_{ij}(x_{ij} - x_{kj}) \tag{31}$$

where v_{ij} is defined as the changes made on enhancing the parameter j, that is, x_{ij} is changed. Indices j and k are random variables. Changes to the present solution (food source) are done by employed bees in the employed bee phase, which works upon the details of the individual experience and the new solution fitness value. If the fitness value of the new solution is higher than the old solution, then the bee updates the position to the new solution and old one will be rejected. The position update equation for ith candidate in this phase is

$$v_{ij} = \begin{cases} x_{ij} + \Phi_{ij}\left(x_{ij} - \Phi_{kj}\right), & \mathit{if}\ r_j \leq \mathit{modified\ rate} \\ x_{ij}, & \mathit{otherwise} \end{cases} \tag{32}$$

where $k \in \{1, 2, \ldots SN\}$ and $\{1, 2, \ldots D\}$ are randomly selected indices. k Should vary from i. Φ_{ij} is a random number between [− 1, 1]. To enhance the exploitation, it is required to consider the details of the global best solution, in order to train the search of candidate solutions, the solution search equation is examined by Eq. (32) and it is improved as follows:

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right) + (2.0 - \mathrm{prob}_i) \times (x_{bestj} - x_{ij}) \tag{33}$$

The fitness based self adaptive mutation method is included in the general ABC, to progress the capacity of the exploitation. The perturbation in the solution is based on the fitness of the solution, that occurs in the proposed approach. It is clear that the number of update in the dimensions of the ith solution works on prob$_i$ and it implies nothing but a function of fitness. This technique is based on the suggestion that the perturbation will be high for low fit solutions as for that the value of prob$_i$ will be low as long as the perturbation in high fit solutions will be low because of the high value of prob$_i$. It is pictured that the global optima should be closer to the better fit solutions and if the perturbation of better solutions is high, then we have an option of skipping the true solutions because of its huge step size. So, the step sizes, which were correspondingly based on the perturbations in the solution, are ewer for good solutions and it is high for worst solutions and it is answerable for exploration. So, in the proposed approach, the better solutions will be accomplished in the search space as long as the low fit solutions examine the search area.

For acquiring the optimal features, traditional ABC is utilized, but the convergence rate of ABC isn't desirable, since there occurs an inherent constraint and other composite functions, that result in a premature convergence. So to enhance the searching efficiency, it is examined to change the parameter which is utilized to compute the new candidate food sources. In this work, coefficients were distinguished as functions of the fitness in the search process of IABC.

Once v_{ij} is produced, the fitness value of the food source is obtained using

$$\mathit{Fitness}_i = \begin{cases} \dfrac{1}{1+f_i}, & \mathit{if}\ f_i \geqslant 0 \\ 1+f_i & \mathit{if}\ f_i < 0 \end{cases} \tag{34}$$

The coefficients ϕ_{ij} to change the search process for various feature selection as mentioned in Eq. (35). F itness(i) represents the fitness value of the features

$$\phi_{ij} = \frac{1}{(1 + \exp(-\mathrm{fitness\ (i)}))} \tag{35}$$

The ϕ_{ij} is modified using (35) and updated in (33) to improve the optimal features

Algorithm procedure

```
Cycle =1 , Maximum Cycle Number(MCN) =100, no of bees= depending on training samples, $Fitness_i = 95\%$

Initialize IABC parameters ($v_{ij}, \phi_{ij}, x_{ij}$, solution $x_i$, prob$_i$ and $j \in (1, D)$

        $x_{ij} = x_{min,j} + rand(0,1)(x_{max,j} - x_{min,j})$        (36)

Where i=1 „.., N,j=1 „. .,D, such that N is the number of food sources and D is the number of optimization parameters.

Evaluate the fitness of each individual feature

Repeat the process

{

    Construct employed bee phase for creating new food sources;

    For $j \in \{1\ to\ D\}$ do

    Compute neighborhood exploration using equation (33)

    Evaluate fitness function using (34)

    where $f_i$is a cost function. For increasing the issues, the cost function can be directly utilized as a fitness value.

    Compute the probability of feature solution

    Construct solution by onlooker bees

    The probability of an onlooker bee to select a food source to be explored is integrated to its fitness shown by given below equation

        $p_i = \frac{Fitness_i}{\sum_{n=1}^{F} Fitness_i}$        (37)

Select the features based on probability $p_i$

Through the values of exploration probabilities, the food sources were chosen by the onlooker bees.

Onlooker bees phase for updating the food sources works on their nectar amounts;

Determine the scout bee and the abandoned solution

Scout bee phase for identifying the new food sources in place of abandoned food sources;

Compute the best feature

Memorize the best optimal features

Cycle = cycle +1

End while

End for

Output the best solution found so far
```

From this algorithm, it is understood that the if value of prob$_i$ is high and then in the case of high fitness solution, then the step size solution will be less So, it is clear that there occurs high chance of high fitness solution to navigate in its neighborhood when distinguished to the low fitness solution and so, a better solution will be accomplished in the search area. Alternatively, our solutions will exploit or explore the search area according to the likelihood of the fitness function.

The IABC algorithm is enforced to choose the optimal features. Modifications in this ABC algorithm, is applied to precede its local search ability and also to utilize to raise the speed of the convergence quickly. The IABC algorithm is very simple, robust and it is a population based stochastic optimization algorithm. In ABC algorithm, the solution of the optimization is indicated by the food source and the quality of the solution is indicated by the nectar amount of the source (fitness). In 1st step of the algorithm, the locations for the food source were generated without any order. Employed bees will define the food source neighborhood in its memory and then it exchanges the details with the onlookers inside the hive and this onlooker choose one of the food sources. Onlookers choose a food source inside the neighborhood of the food sources which is selected by them. Employed bee which has deserted source will become a scout and it begins to search for new food source randomly. The IABC algorithm is utilized for minimizing the repetitions optimally and moreover it will raise the searching ability when distinguished with the previous algorithms. It will choose the best features according to the greatest fitness value.

3.4 Speech recognition using NOEC algorithm

A Hybrid Particle Swarm Optimization-Artificial Neural Network (HPSO-ANN) is brought in the proposed system, to enhance the speech recognition results of the entire system. The fear and anger emotions aren't considered in the previous system [15], even if it consumes much time for implementation and also feature selection process isn't performed. In order to rectify these problems, NOEC is proposed in this work, when compared to EBSA approach this is much faster. Here, fear and anger emotions were conceived and computed much faster with the help of fast neuron layers. In training as well as testing results, NOEC has merits of greater accuracy. The learning and testing [20] are the two primary phases in ANN are illustrated in Fig. 4. The former one will adjust and change the neural network weights in response to the training input patterns which is represented at the input layer. In learning phase, the optimal size for the speech dataset like the total number of layers, the number of hidden units in the middle layers, and number of units in the input and output layers with respect to accuracy on a test set, and the training algorithm were employed, it will perform the extraction process from a noisy training set in order to accomplish good speech enhancement. The target of learning is to reduce the cost function according to the error signal $e_i(t)$, in terms of parameters (weights), so, the actual response of every output neuron in the network approaches the target response [21, 22]. MSE is the common case which is exploited in cost function, which is determined as a mean-square value of the sum squared error:

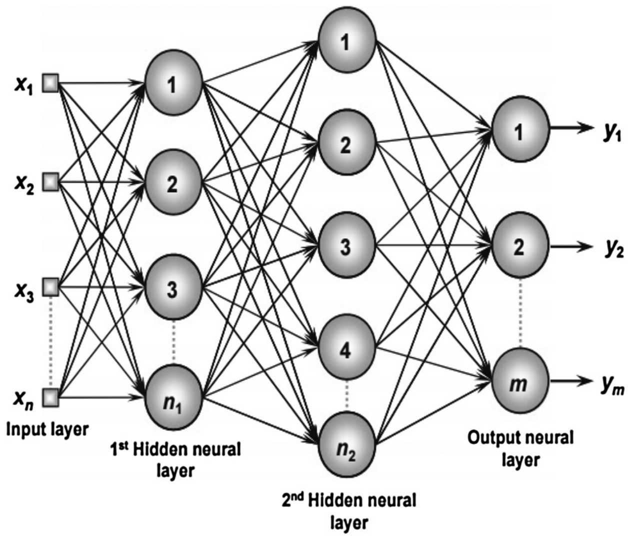


Fig. 4 General structure of ANN

$$J = E\left[\frac{1}{2}\sum_i (e_i(t))^2\right] \tag{38}$$

$$= E\left[\frac{1}{2}\sum_i d_i(t) - y_i(t))^2\right] \tag{39}$$

where E is the statistical expectation operator and the summation is outer layer of the entire neurons. With the help of the desired signal $di(t)$ only, the adaptation of weights is done generally. It is declared that a new signal $d_i(t) + n_i(t)$ can be utilized as an appropriate signal for output neuron I rather than utilizing the original desired signal $d_i(t)$, where $n_i(t)$ is a noise term. This noise term is considered to be a white Gaussian noise, independent of both the input signals $x_k(t)$ and the desired signals $d_i(t)$. To identify the optimal solution, it is required in each and every ANN modeling to be crosschecked. In task like prediction, classification, time series projection, etc., ANN approach is enforced, which, in turn, demands various structures and features to be analyzed.

This is established as generalizations of mathematical models of biological nervous systems, once after bringing-in the simplified neurons, a first interest in neural networks.

The artificial neurons, or simply neurons or a node is a basic processing element of neural networks. The effect of the synapses were indicated as connection weights that adjust the effect of the associated input signals, and the nonlinear characteristic shown by neurons is indicated by a transfer function, this happens in the simplified mathematical model. The computation of the weighted sum of the input signals is done by the neuron impulse, which is then transformed by the transfer function. The learning capability of an artificial neuron is accomplished by adapting the weights in accordance to select the learning algorithm. The neuron output signal O is provided by the following relationship:

$$o = f\left(\sum_{j=1}^{n} w_j x_j\right) \tag{40}$$

where w_j is the weight vector, x_j is input of neurons. Three types of neuron layers are there in the basic architecture, they are: input, hidden, and output layers. The signal flow is from input to output units, in the feed-forward networks, which is closely in a feed-forward direction. The data processing is continued over various layers of units.

A neural network has to be configured, were the application of a set of inputs generates an appropriate set of outputs. With the help of the prior knowledge, the strength of ANN set the weights clearly. One more thing is to guide the neural network by nourishing its teaching patterns and allowing it to modify its weights based on the learning rules. The learning situations in the neural networks is divided into supervised learning, here, an input vector is represented at the inputs together with a set of appropriate responses, one for every node, at the output layer. A forward pass is performed, and it is required to identify the errors or conflicts among the appropriate and actual response for every node. The ANN has to be optimized because some features cannot be recognized and accuracy of the emotion classification is reduced considerably. To avoid these issues, the learning of the ANN is improvised by using PSO and accuracy is increased by training most informative features. Some noise or other randomness is included to the training data, to produce a strong and trustworthy network, and also to acquire the network familiarized with noise and natural variability in real data. Unreliable and unpredictable network will be developed if the training data is poor. For a prefixed number of epochs or when the output error reduced below a specific error threshold, the NOEC is trained. The network shouldn't be over trained. Figure 5 illustrates the NOEC architecture diagram with PSO algorithm.

The network become adjusted in the learning the samples from the training set extremely, if it over-trained, so it leads to inaccurate samples outside the training set. The training speed is decreased desirably, when the ANN has demerits at the time of dynamic behavior. To rectify this issue, the NOEC approach are be enhanced with respect to neuron selection. The network is inadequate to learn the relationships amidst the data and the maximized error, because of the few hidden neurons. The Fig. 6 shown the NOEC algorithm procedure

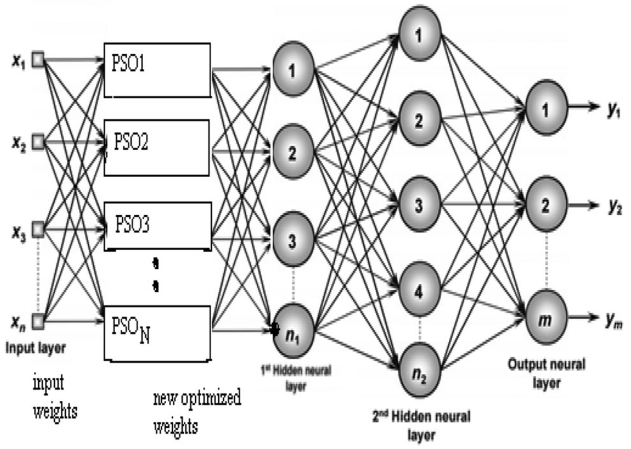


Fig. 5 Improved Artificial Neural Network (IANN)

Huge hidden neuron assures proper learning and the network has the ability to perfectly foresee the data, which has been trained. The NOEC collects the emotion features with MFCC, energy, pitch, intensity and formants. By implementing the IABC optimally, the individual best threshold value is computed. With the help of the best objective function from the total input features, the appropriate features were chosen. These features were provided to the NOEC to proceed with the training and testing to divide more accurate emotions like sad, happy, normal, fear and anger. The hidden layers drew-out the associated features and rapidly calculate to divide the best output in output layer.

PSO algorithm starts by initializing a random swarm of M particles for weight optimization in ANN. The fitness of every particle is computed based on the fitness function, at every epoch. The algorithm collects and evenly substitute the best previous position of every particle (pbest$_i$, i = 1,2,…,M) as well as an individual best particle (gbest). In PSO, the velocity and position of every particle were substituted by phase. The optimization-based speech classification, required to determine the weight values of ANN. The average error admits the noisy speech signal and the estimated noise signal in every frame are utilized as the cost function. Fitness particles have less cost function values.

$$J_i = \frac{1}{N+1}\sum_{k=0}^{N}[d(k) - y_i(k)]^2 \tag{41}$$

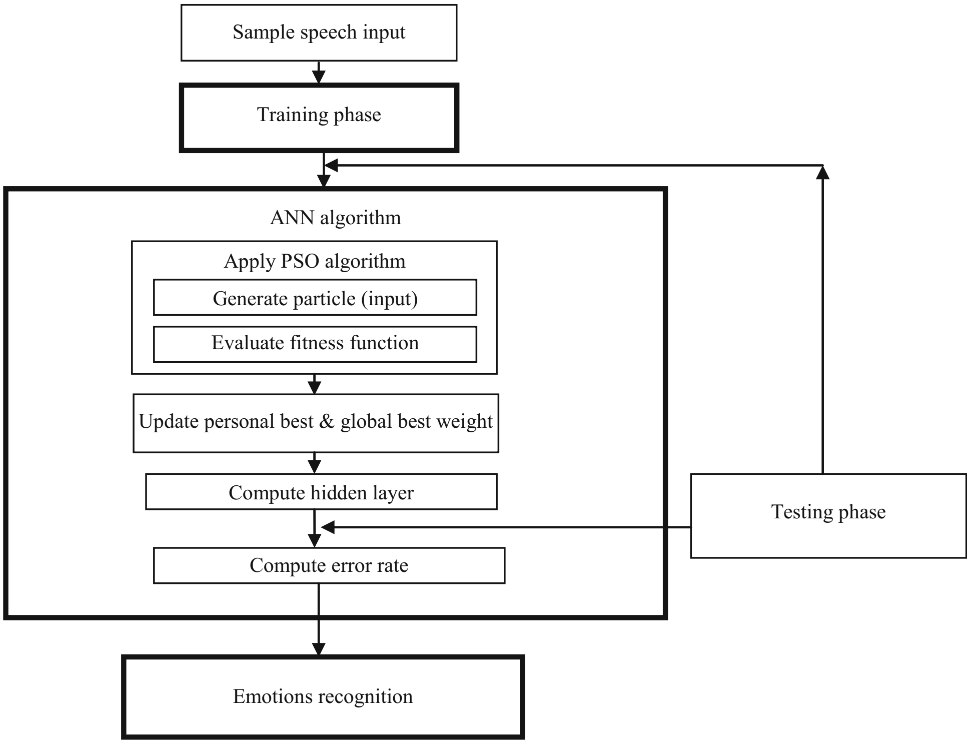


Fig. 6 Neuro Optimized Emotional Classifier (NOEC)

where N is the speech sample's count in every frame, and y(k) is the output. When J_i is minimum, then the parameters denotes the best estimation and generate lower error rate. If the error values between the training and testing is lesser means then the weight values are created higher. The position of every particle in the swarm is a candidate for the coefficients, in the PSO-based optimization speech classification. After certain iterations, the optimal feature (emotion feature) is computed based on the position vector of the best (global) particle in the swarm (gbest). Then, these features were divided more accurately by utilizing the optimized values. The Particle Swarm optimization (PSO) is brought-in this work, to manage the smoothing features and non recognition features optimally. With the help of PSO particles, the ANN learning process is enhanced. The best fitness value is computed to update the local best and global best values in the PSO algorithm, which concentrates on enhancing the entire emotion classification accuracy. PSO algorithm, improves the ANN architecture, which will raise the optimal emotion recognition results by maximizing the convergence speed. Develop an initial ANN, which has three layers, i.e., an input = 150, an output = 150, and a hidden layer = 300. For every training pattern

1. Enforce the input speech features to the network
2. Train the network on the training set till the error is almost constant $e_{th} = 0.1$ for a specific number of training epochs $\tau = 100$ that is specified by the user.
3. Assume the five emotions for input nodes and hidden nodes
4. Compute the output for each neuron from the input layer, by utilizing the hidden layer(s), to the output layer
5. Pitch, MFCC, energy, formants and intensity feature values as input pattern for the neural network
6. Calculate the error of the ANN according to the validation set.
7. If the error is identified to be unacceptable (i.e., too large), then consider that the ANN has irrelevant architecture, and go to the next step.
8. Else, stop the training process until it reaches $e_{th} = 0.1$. The error E is computed based to the following equations.

$$E(w, v) = \frac{1}{2}\sum_{i=1}^{k}\sum_{p=1}^{C}(S_{pi} - t_{pi})^2 \tag{42}$$

where k the pattern's count and C is is the number of output nodes. t_{pi} and S_{pi} are the target and actual outputs for the ith pattern of the pth output node. The actual output S_{pi} is computed based on thefollowing equation.

$$S_{pi} = \sigma(\sum_{m=1}^{h}\delta((x_i)^T w_m)v_m) \tag{43}$$

Here h is the hidden node's count in the network, xi is an n-dimensional input pattern, $i = 1, 2, \ldots k$,w_m is weights for the arcs connecting the input layer and the m-th hidden node, m = 1, 2, …, h, v_m weights for the arcs connecting the m-th hidden node and the output layer, σ is activation function and δ is the hidden layer hyperbolic tangent function.

9. Add one hidden node to the hidden layer. Randomly initialize the weights of the newly included node and proceed to training process.
10. Apply PSO algorithm

    10.1. Create the particle swarm (speech input) with random position and velocity

    10.2. Compute the fitness function by utilizing the given below equation

$$v_i(t+1) = w \times v_i(t) + (c_1 \times \text{rand} \times (P\text{best}(t) - x_i(t))) + (c_2 \times \text{rand} \times (G\text{best}(t) - x_i(t))) \quad (44)$$

    where, w is weight that is ranged from 0 to 1, v_i is particle velocity, $c1, c2$ is speeding factors, Pbest: best value of particle i, x_i ith particle of swarm, Gbest is best value that one of swarm particles reach.

    10.3. The new fitness of particle is calculated as follows

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (45)$$

    10.4. The PSO creates best values of ANN learning rates along with weight value among the input and hidden nodes

    10.5. If the features aren't identified by ANN then

    10.6. The neurons checks error values

    10.7. Enforce the weight adjustments

    10.8. The training and testing of NOEC utilize the features like formants, MFCC, pitch, energy and intensity

    10.9. PSO identify the features optimally by using the best fitness value

    10.10. Update the global best solutions

11. Classify speech output emotions as happy, sad, normal, fear and anger based on above extracted features
12. Emotions are classified more accurately

4 Experimental results

The dataset is collected from regional Tamil news. The audio samples range is around range of 100-150 samples and from various news channels, it is assumed. The samples are in the formats of '.wav' format and the time duration is 2 min per sample. The different voices were conceived with emotions like happy, sad, fear, anger and normal. The proposed NOEC algorithm, exactly acquire the emotion speech recognition. The performance metrics were assumed as accuracy, specificity, sensitivity, precision, recall and f-measure metrics were evaluated by existing approaches like EBSA, SVM, KNN and proposed NOEC algorithm.

4.1 Peak signal to noise ratio (PSNR)

It is the ratio between the utmost possible powers to the power of corrupting noise is given as Peak Signal to Noise Ratio. It considerably has an effect on the fidelity of its representation. It can be also observed that it is the logarithmic function of peak value of image and Mean Square Error (MSE).

$$\text{PSNR} = 10 \log_{10}\left(\text{MAX}_\text{i}^2/\text{MSE}\right) \quad (46)$$

4.2 Mean square error (MSE)

Mean Square Error (MSE) of an estimator is to quantify the difference between an estimator and the true value of the quantity being estimated.

$$\text{MSE} = \frac{1}{\text{mn}} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \quad (47)$$

The Table 1 shows the comparison of preprocessing methods with respect to PSNR and MSE

From the above Fig. 7, it can be observed that the comparison metric is computed among the pre-processing methods with respect to PSNR metric. In x-axis the algorithms are considered and in y-axis the PSNR value is considered. The proposed ICEKF algorithm provides higher PSNR value of 40.95 dB which is 3.39 and 5.32 dB higher when compared to IKF and KF methods respectively. The result confirms that the proposed system gains greater speech recognition results.

From the above Fig. 8, it can be observed that the comparison metric is computed among the pre-processing methods with respect to MSE metric. In x-axis the algorithms are considered and in y-axis the MSE value is considered. The proposed ICEKF algorithm provides lesser MSE value of 0.25 which is 0.13 and 0.17 lesser when compared to IKF and KF methods respectively. The result confirms that the proposed system gains greater speech recognition results. The Table 2 shows the comparison of existing and proposed system with respect to accuracy, error rate, precision, recall and f-measure

Table 1 Comparison of preprocessing methods

Performance metrics	Preprocessing methods		
	KF	IKF	ICEKF
PSNR(dB)	35.63	37.56	40.95
MSE	0.42	0.38	0.25

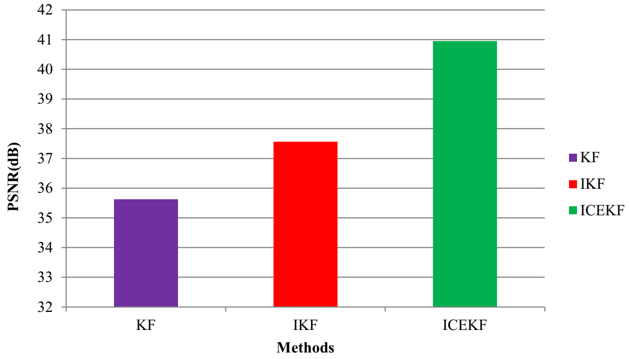


Fig. 7 PSNR comparison verses methods

4.3 Accuracy

Accuracy is defined as the complete correctness of the model and is evaluated as the total actual classification parameters $(T_p + T_n)$ which is classified by the sum of the classification parameters $(T_p + T_n + F_p + F_n)$. The accuracy is computed as like :

$$\text{Accuracy} = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)} \tag{48}$$

where T_P is known as the amount of correct predictions that an instance is negative, T_n is called the amount of incorrect predictions that an instance is positive, F_p is known as the amount of incorrect of predictions that an instance negative, and F_n is known the amount of correct predictions that an instance is positive.

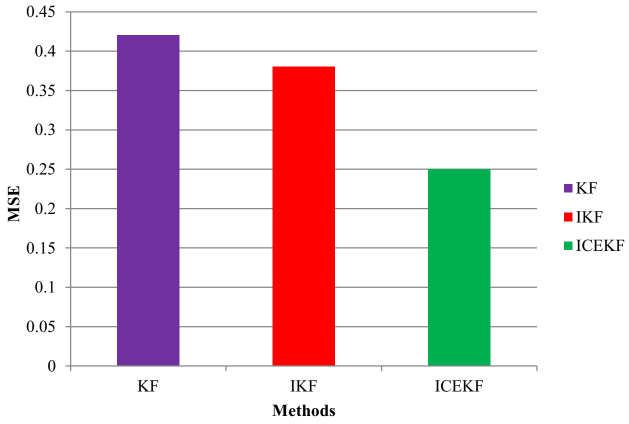


Fig. 8 MSE comparison versus methods

From the above Fig. 9, it can be observed that the comparison metric is computed among the existing and proposed method with respect to accuracy. In x-axis the algorithms are considered and in y-axis the accuracy value is considered. The existing methods gives lower accuracy in fact, the proposed system gives higher accuracy for the provided speech sample input. The proposed feature selection IABC algorithm chooses best speech features. These features are then enforced in NOEC training and testing phase to generate appropriate emotions. The result confirms that the proposed system gains greater speech recognition results with the help of IABC with NOEC algorithm. Hence, the proposed IABC with NOEC algorithm is higher than the previous the EBSA, and ANN algorithms.

From the above Fig. 10, it can be observed that the error rate metric is computed among the existing and proposed methods. The result confirms that the proposed system gains lesser error rate when compared to other methods.

4.4 Precision

Precision is explained as the ratio of the true positives opposite to both true positives and false positives result for imposition and real features. It is distinct as given below

$$\text{Precision}(P) = \frac{T_p}{T_p + F_p} \tag{49}$$

From the above Fig. 11, it can be observed that the comparison metric is computed among the existing and proposed method with respect to precision. In x-axis the algorithms are considered and in y-axis the precision value is considered. The existing methods gives lower precision in fact, the proposed system gives higher precision for the provided speech sample input. The proposed system has greater precision because the feature selection process occurs with the help of IABC algorithm. It chooses the optimal speech features from the extracted features like MFCC, energy, pitch, formants and intensity. These features are then enforced in NOEC training and testing phase to generate appropriate emotions. The result confirms that the proposed system gains greater speech recognition results with the help of IABC with NOEC algorithm. Hence, the proposed IABC with NOEC algorithm is higher than the previous the EBSA, and ANN algorithms.

4.5 Recall

Recall value is evaluated on the root of the data retrieval at true positive forecast, false negative. Normally, it can be decided as follows,

Table 2 Comparison of existing and proposed algorithms

Performance metrics	ABC			IABC		
	EBSA	ANN	NOEC	EBSA	ANN	NOEC
Precision	80.322	82.68	85.873	83.56	85.92	90.41
Recall	93.349	93.89	94.46	94.53	95.63	95.86
F-measure	86.83	88.28	90.16	89.04	90.77	93.13
Accuracy	86.67	87.41	89.94	88.95	90.52	92.18
Error rate	13.33	12.59	10.06	11.05	9.48	7.82

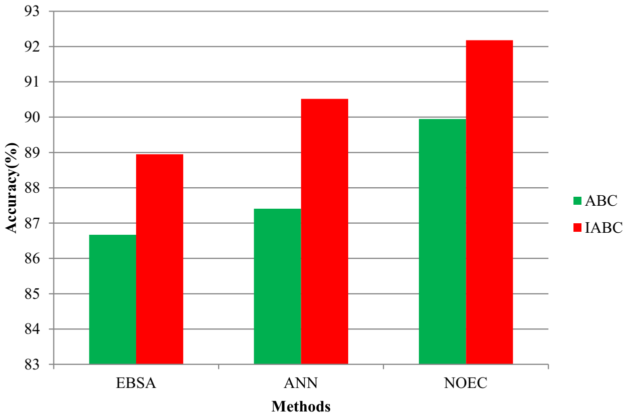


Fig. 9 Accuracy comparison

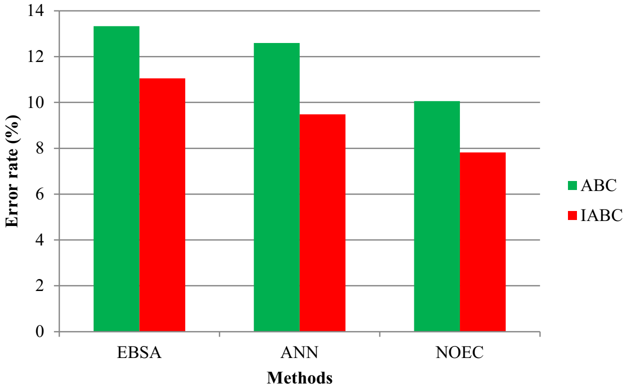


Fig. 10 Error rate comparison

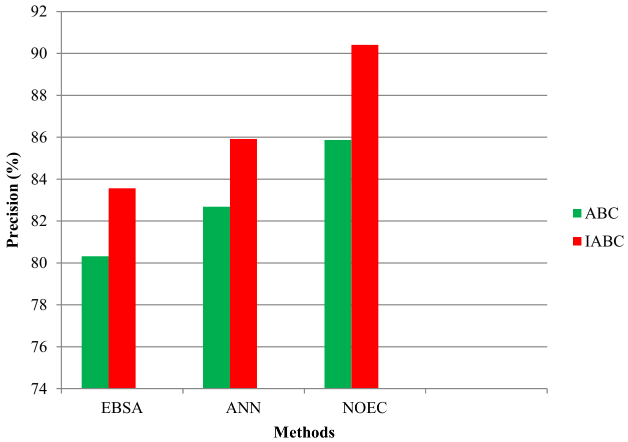


Fig. 11 Precision comparison

$$Recall(R) = \frac{T_p}{T_p + F_n} \tag{50}$$

From the above Fig. 11, it can be observed that the comparison metric is computed among the existing and proposed method with respect to precision. In x-axis the algorithms are considered and in y-axis the recall value is considered. The existing methods gives lower precision in fact, the proposed system gives higher precison for the provided speech sample input. Similarly the proposed system has greater recall because the feature selection process occurs with the help of IABC algorithm is illustrated in Fig. 12. It chooses the optimal speech features from the extracted features like MFCC, energy, pitch, formants and intensity. These features are then enforced in NOEC training and testing phase to generate appropriate emotions. The result confirms that the proposed system gains greater speech recognition results with the help of IABC with NOEC algorithm. Hence, the proposed IABC with NOEC algorithm is higher recall than the previous the EBSA, and ANN algorithms.

4.6 F-measure

It is a measure of an accuracy of the test. It consider both the precision p and the recall r of the test to compute the score.

From the above Fig. 13, it can be observed that the comparison metric is computed among the existing and proposed method with respect to f-measure. In x-axis the algorithms are considered and in y-axis the f-measure value is considered. The existing methods give lower f-measure in fact; the proposed system gives higher f-measure for the provided speech sample input. The proposed system has greater f-measure because the feature selection process occurs with the help of IABC algorithm. It chooses the optimal speech features from the extracted features like

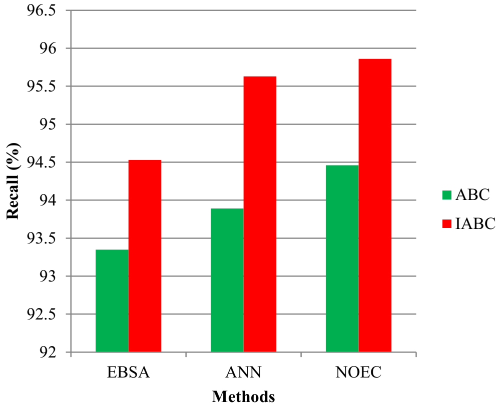


Fig. 12 Recall comparison

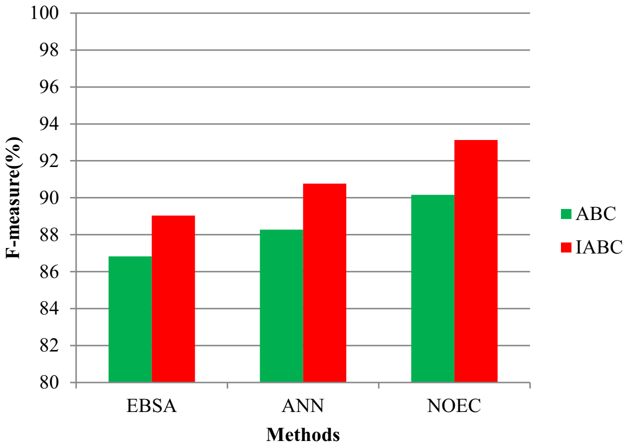


Fig. 13 F-measure comparison

MFCC, energy, pitch, formants and intensity. These features are then enforced in NOEC training and testing phase to generate appropriate emotions. The result confirms that the proposed system gains greater speech recognition results with the help of IABC with NOEC algorithm. Hence, the proposed IABC with NOEC algorithm is higher than the previous the EBSA, and ANN algorithms. The proposed NOEC with IABC classifier provides higher accuracy results of 92.18%, which is 1.66 and 3.23% higher when compared to ANN and EBSA methods respectively. The proposed NOEC with IABC classifier provides higher f-measure results of 90.52%, which is 2.36 and 4.09% higher when compared to ANN and EBSA methods respectively. The proposed NOEC with IABC classifier provides higher recall results of 95.86%, which is 0.23 and 1.33% higher when compared to ANN and EBSA methods respectively. The proposed NOEC with IABC classifier provides higher precision results of 90.41%, which is 4.49 and 6.81% higher when compared to ANN and EBSA methods respectively.

5 Conclusion and future work

Accurate emotion recognition systems were important for the advancement of human behavioral informatics and in the design of effective human machine interaction systems. In this work, IABC with NOEC approach is proposed to enhance the speech emotion recognition more efficiently. The proposed system has modules like speech enhancement, feature extraction, feature selection and speech recognition. The Iterative Conditional Entropy Kalman filtering (ICEKF) performs the first process of speech enhancement; it is utilized to maximize the emotion classification accuracy desirably. Then the feature extraction is done with the help of MFCC, pitch, energy, intensity and formants. The feature selection is desirable, which is included to improve the entire system performance optimally. IABC algorithm is enforced to choose significant features by producing the best fitness function values. These features are considered into NOEC classification phase, and it has training and testing phase to construct an accurate model which gives superior neural network performance. The highest value of the emotion recognition accuracy is for male and female together by utilizing the entire metrics. The accuracy enhanced to maximize the dataset size. This method examines the effect of various features on the performance of the emotion recognition system. The NOEC approach is utilized to identify five emotions like happy, sad, fear, normal and anger emotions more accurately. The result proves that the proposed system gives greater classification results with respect to greater accuracy, precision, recall and f-measure metrics by utilizing IABC with NOEC approach.

References

1. Schuller, B., Gerhard, R., Manfred, L.: Hidden Markov model-based speech emotion recognition. Multimed. Expo. ICME'03. In: Proceedings. 2003 International Conference on. vol. 1. IEEE (2003)
2. Schuller, B., Gerhard, R., Manfred, L.: Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. Acoustics, Speech, and Signal Processing, 2004. In: Proceedings.(ICASSP'04). IEEE International Conference on, vol. 1. IEEE (2004)
3. Ingale, A.B., Chaudhari, D.S.: Speech emotion recognition. Int. J. Soft Comput. Eng. (IJSCE) **2**(1), 235–238 (2012)
4. Anila, R., Revathy, A.: Emotion recognition using continuous density HMM. Communications and Signal Processing (ICCSP), 2015 International Conference on. IEEE (2015)
5. Martin, Rainer: Speech enhancement based on minimum mean-square error estimation and supergaussian priors. IEEE Trans. Speech Audio Process. **13**(5), 845–856 (2005)

6. Tsenov, G.T., Mladenov, V.M.: Speech recognition using neural networks. In: Neural Network Applications in Electrical Engineering (NEUREL), 2010 10th Symposium on. IEEE (2010)
7. You, Chang Huai, Koh, Soo Ngee, Rahardja, S.: Subband Kalman filtering incorporating masking properties for noisy speech signal. Speech Commun. **49**(7), 558–573 (2007)
8. Hu, Y., Wu, D., Nucci, A.: Pitch-based gender identification with two-stage classification. Secur. Commun. Netw. **5**(2), 211–225 (2012)
9. Yogesh, C.K., et al.: Hybrid BBO_PSO and higher order spectral features for emotion and stress recognition from natural speech. Appl. Soft Comput. **56**, 217–232 (2017)
10. Sharan, R.V., Moir, T.J.: Noise robust audio surveillance using reduced spectrogram image feature and one-against-all SVM. Neurocomputing **158**, 90–99 (2015)
11. Prasartvit, Thananan, Banharnsakun, Anan, Kaewkamnerdpong, Boonserm, Achalakul, Tiranee: Reducing bio–informatics data dimension with ABC–KNN. Neurocomputing **116**, 367–381 (2013)
12. Al-Naser, Mustafa, Elshafei, Moustafa, Al-Sarkhi, Abdelsalam: Artificial neural network application for multiphase flow patterns detection: a new approach. J. Petrol. Sci. Eng. **145**, 548–564 (2016)
13. Arai, Kohei: Recovering method of missing data based on proposed improved Kalman filter when time series of mean data is known. Int. J. Adv. Res. Artif. Intell. **2**(7), 18–23 (2013)
14. Skoglund, M.A., Hendeby, G., Axehill, D.:Extended Kalman filter modifications based on an optimization view point. 18th International Conference on Information Fusion (Fusion), pp 1856–1861 (2015)
15. Kumuthaveni, R., Chandra, E.: An enhanced bat algorithm with simulated annealing method for speech emotion recognition. Int. J. Adv. Res. Dyn. Control Syst. **1**, 125–138 (2017)
16. Chen, J., et al.: Recognition of noisy speech using dynamic spectral subband centroids. IEEE Signal Process. Lett. **11**(2), 258–261 (2004)
17. Bozkurt, E., et al.: Formant position based weighted spectral features for emotion recognition. Speech Commun. **53**(9), 1186–1197 (2011)
18. Morrison, Donn, Wang, Ruili, de Silva, L.C.: Ensemble methods for spoken emotion recognition in call-centres. Speech Commun. **49**(2), 98–112 (2007)
19. Mogaka, L., Murage, D.K., Saulo, M.J.: Rotating Machine based power optimization and prioritization using the artificial bee colony algorithm. In: Proceedings of Sustainable Research and Innovation Conference (2016)
20. Mezura-Montes, Efrén, Cetina-Domínguez, Omar: Empirical analysis of a improved artificial bee colony for constrained numerical optimization. Appl. Math. Comput. **218**(22), 10943–10973 (2012)
21. Badri, Lubna: Development of neural networks for noise reduction. Int. Arab J. Inf. Technol. **7**(3), 289–294 (2010)
22. Dorronsoro, J., López, V., Cruz, C., Sigüenza, J.: Autoassociative neural networks and noise filtering. IEEE Trans. Signal Process. **51**(5), 1431–1438 (2003)

R. Kumuthaveni Assistant Professor in Department of Computer Applications, Dr. SNS.Rajalakshmi College of Arts & Science, Coimbatore. She received B.Sc degree from Bharathiar University, Coimbatore in 2001. She was awarded with M.C.A from IGNOU (Indira Gandhi National Open University, Delhi) in 2007. She obtained her M.Phil from Vinayaka Missions University, Salem in 2008. She has 10 years of teaching experience and pursuing Ph.D as part-time research scholar in the same institution. She has qualified SET (State Eligibility Test for Lectureship) in April 2017. Her area of interest lies in Speech and Emotion recognition and published 3 papers in International journals and presented 4 papers in national and international conferences in that area.

E. Chandra Professor and Head in the Department of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India. She has more than 22 years of teaching experience and 20 years of Research Experience. Her Area of Specialization includes Neural Networks, Speech Recognition System. She has produced 6 Ph.D scholars. She has authored more than 72 papers published in refereed International journals, presented 45 papers in national and International Conferences and published 2 books. She has obtained funding projects from UGC in the field of speech signal processing. . She is an active member of CSI, Life member of Society of Statistics and Computer Applications. She is the Board of Studies Member for various affiliated Institutions, Member of various Professional Societies, Inspection Commission member for various colleges and selection committee member of different institutions. She is the Coordinator for RUSA scheme from Bharathiar University and Reviewer for International Journals.