How to Cite:
Manjutha, M., & Subashini, P. (2022). Survey on optimization algorithms in speech processing. *International Journal of Health Sciences*, *6*(S5), 2997–3017.
https://doi.org/10.53730/ijhs.v6nS5.9307

Survey on optimization algorithms in speech processing

Manjutha M
Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India
Email: manjutham@gmail.com

Subashini P
Professor, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India
Email: parthasarathysubashini@gmail.com

***Abstract*---**Nature is a tremendous provenance of resolving hard and complex problems that exist in the field of computer science because it reveals a very diverse, robust, dynamic, and interesting phenomenon. It constantly finds the optimal solution to resolve its problem which accomplishes exact equity among its element. Nature-inspired algorithms are the heuristic high-level procedure that interprets nature to solve the optimization problem which popularized in the new era of computation. The main objective of this paper is to evaluate the modern technology and enhancement in the nature-inspired algorithm, especially in the application of speech processing, speech recognition, and speech feature selection problems. This paper presents broad collections of global optimization algorithms which have been successfully applied to generate recognition systems that are integrated with metaheuristic algorithms.

***Keywords*---**Evolutionary Computation, Feature Selection, Fitness Function, Nature Inspired Algorithm, Optimization and Speech Recognition.

I. Introduction

In general speech recognition is an intelligence of machines that are developed to identify the spoken language words and phrase and translate it into the machine-understandable format. Hence, this process is widely known as Automatic Speech Recognition (ASR) system or Computer Speech Recognition (CSR) and Speech to Text conversion [32], [33]. The primary applications of the speech recognition framework revolve around innovations of machine learning and pattern

recognition approach [18], [19]. The speech recognition system is extremely difficult to execute by a machine because of the changes occurred in the individual speech which provides complex speech signals. Therefore, this complex signal must be processed through automatic speech signals and which are handled through an ASR system. There are many significant areas like automatic speech recognition, unconstrained speech, and robust speech recognition has been examined for the recent development of spoken language frameworks [12],[51]. Nowadays several users adapted to lightweight modern applications that assist real-time recognition with limited vocabulary [43]. The speech processing system uses a machine learning paradigm in which a classifier employs an appropriate learning process and the performance of the classifiers is strongly associated with significant features. Hence it is complex to select the salient features from the large set of feature vectors. Extracting significant features from the existing set of features will ease the dimensionality of data and sequentially it enhances the classifier performance and its accuracy during runtime [8] [13] [17]. From the existing set of features, choosing the most discriminative and appropriate set of features is a major challenge. The significant feature alone is selected from the available set of features to remove redundant and irrelevant features. The feature selection encompasses two major processes listed as follows [7]:

- Searching strategy that seeks the search space which chooses a features subset.
- Evaluation procedure that computes the feature subset quality that results in the best subset has to be selected.

Determining feature subsets for a large number of feature sets is considered a crucial problem. Therefore, a reliable searching strategy is mandatory to avoid estimating a huge amount of combinations in the entire feature subsets. For this reason, many of the authors proposed various searching strategies such as complete search, bidirectional search, sequential forward selection, and backward selection. Further, the search procedure includes two major approaches namely filter and wrapper. The filter search procedure classifies the features or subset of features independently of the classifier whereas the wrapper search procedure evaluates the subset of features by the classifier. In certain cases, uses an embedded method to avail benefits of both approaches [46], [49]. Spontaneously to enhance the searching strategy of significant speech features and the performance of classifiers, optimization algorithms have been presented and implemented by many researchers. Generally, an optimization algorithm is stochastic or deterministic in nature and is used to identify the best possible solution in the search space. This algorithm solves a wide range of problems existing in many applications such as image processing, video processing, robotics, job scheduling, telecommunications, agriculture, weather prediction, space research, vehicle routing, protein folding, data mining, speech, and signal processing. Hence, it is considered as an active research area in the new era [24].

Machine learning is the general algorithm with other applications and mathematical composition of the learning issues in speech processing based on model creation from statistical principles, observation, and prediction of data quantify misfit in the loss function, a relevant parameter for the developed models, predecessor for these constraints, standardizes the function which

evaluates the complexity of the constraints and for the huge number of a dataset. The speech and language processing encompasses labeled instances has words, phonemes, phrases, or even sentences and also for a specific speaker. The global objective function of the speech database is partially separable, depending upon the parameter of the model each data belongs to a single item which may be sentences, words, or frame of speech. Hence this application often uses optimization algorithms to avoid losses in data accuracy, performance, and strong convergences in the features. However, the metaheuristic algorithm issues in speech and language processing are non-convex in nature due to the complexity of the models in use, and the loss functions are designed to compute the actual recognition and classification errors. Optimization problem needs more computational efforts thus lean to be unsuccessful due to the enlargements of problem size. This is the reason behind the exploitation of nature-inspired optimization algorithms as the computationally dynamic approach. Optimization algorithm works based on the iteration enhancement of the single solution or population of solutions and predominantly places randomization and local search to resolve the optimization problem. From the heuristic design, the metaheuristic algorithm is designed which is appropriate to a wide set of optimization problems with negligible changes [38]. Many researchers introduced optimization algorithms that are widely used in various applications among that, the most commonly used algorithms such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony Optimization (ABC), Simulation Annealing (SA) Firefly algorithms and Genetic algorithm (GA).

This optimization algorithm uses the searching techniques and adopted the local search rather than global search during the whole process, hence it is a challenging task to attain near-optimal to optimal solutions. Therefore, in the modern world many drive from the computational intelligence community for establishing heuristic-based search algorithms to report speech recognition problems that target the global search algorithm by using local search appropriately. The optimization algorithm is a multi-agent system and addresses the problem of identifying the typical solutions in polynomial time. [25], [37], [44]. The metaheuristic algorithm general process is represented in Fig.1.

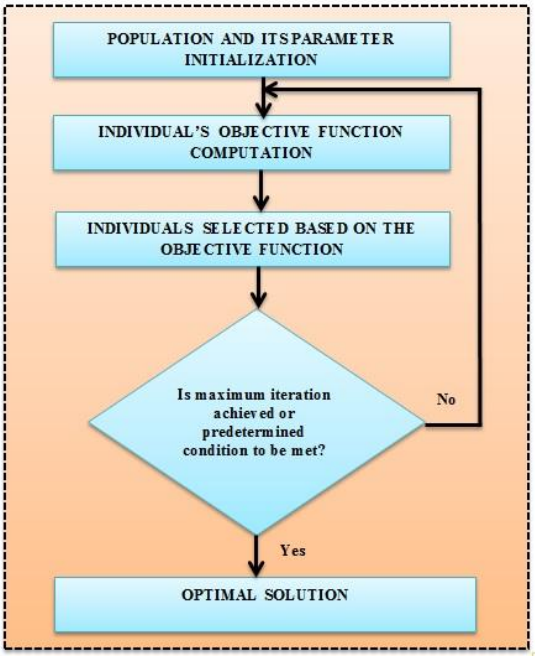

Fig.1 General Process of Metaheuristic Algorithm

Primarily optimization algorithm encompasses five major steps that are initialization of population and the parameter, computing objective function, choosing the best, pre-determined condition evaluated based on criteria and achieves optimal solution.

Step 1: In the initialization stage, the definite numbers of individuals are represented in the population array with the predetermined essential parameter values.

Step 2: For each iteration, individuals exist in the population array are randomly selected to compute the objective (fitness) using fitness functions.

Step 3: Depending on the cost value, individual best is chosen at the execution of each program cycle.

Step 4: This procedure continued to identify the global optimum as far as the predetermined conditions are to be attained or till reaches utmost number of iterations.

Step 5: The feasibility of identifying optimal solution i.e. maxima or minima primarily based on different aspects namely, parameters, objective function, characteristics of convergence (exploitation) and divergence (exploration), stagnation problem in local optima.

Eventually, to solve optimization issues in computational algorithm by using an accurate test function becomes challenging task. Because of this relevant objective function is selected based on the nature of dataset, search space dimension, scalability, modality, objectives, performance of the algorithm, and discrete features of the objective function. All over the world, many researchers proposed different techniques to enhance the speech processing applications embedded with nature-inspired algorithms has been elaborated in this paper.

This paper emphasizes the analysis of test functions, which are applied in many metaheuristic algorithms to resolve the challenges in speech recognition, feature selection, and classification problems. The organization of articles is as follows. Section 2 presents a detailed view of the speech recognition applications in various metaheuristic algorithms such as Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA). Section 3 discuss the various techniques and algorithm adapted to solve classification and recognition problems in speech processing applications. The conclusion and future extension of the research works are presented in Section 4.

II. Elaborated review on speech processing system

A. Dev et al (2010) proposed the robustness of speech front-end which is improved by using MFCC feature vector with three steps. First, the relative higher-order autocorrelation coefficients were obtained. Then the obtained speech signals magnitude spectrum is computed by FFT (Fast Fourier Transform) with respect to frequency and finally, the magnitude spectrum is converted into MFCC coefficients represented MFCCs mined from Differentiated Relative Higher Order Autocorrelation Sequence Spectrum (DRHOASS). The performance of recognition rate achieves MFCC, Autocorrelation MFCC (AMFCC), Relative Autocorrelation Sequence MFCC (RAS-MFCC) and DRHOASS-MFCC as 98.241%, 98.246%, 98.30% and 99.64% respectively [1].

In 2011, F.L. Huang introduced a dynamic approach for the Chinese independent speech recognition system, especially for small vocabulary size based on Hidden Markov Model (HMM). The speech signal was recorded by 4 native male and female speakers producing a total of 640 speech samples. The internal and external preliminary testing result achieves 89.6% and 77.5% respectively and also the precision rates of internal and external test attains an average of 92.7% and 83.8%. [14]

The most promising model developed is based on the human speech expression with different emotions like happiness, anger, and sadness. The salient feature including energy, pitch, and MFCC was computed to analyze the emotional state. In order to select the significant features Ant Colony Optimization is proposed which reduces the feature set by over 16.6% in a total of 300 iterations without losing the information [10].

I.E. Henawy et.al (2014) modeled an accurate speech recognition system for Arabic spoken digits which recognizes the human speech exactly from any speaker. This is due to the discriminative feature extraction techniques, such as MFCC, cepstrum features, formant frequencies, and time-domain short-time energy (STE). This feature set improves the recognition accuracy and results in the highest accuracy of 98% for MFCC based on HMM. Other features like cepstrum, formant frequency, and STE achieves a recognition rate of 94%, 95.5%, and 92% respectively. The outcome of the results suggests that MFCC with HMM provides a high recognition rate [23].

Czyzewski et.al (2003), proposed an intelligent algorithm to classify the Polish language in which Artificial Neural Network (ANN) and Rough set algorithm

achieves a classification accuracy of 73.25% and 91 % respectively [11]. Wisniewski et.al (2007) detected the continuous speech disorder by using Hidden Markov Model (HMM). The speech sample contains 24 disfluencies, MFCC features were extracted and HMM achieves 80% accuracy [55]. Ravikumar et.al (2008) introduced an objective assessment for stuttered speech. The database consists of 150 Standard English passages was uttered by 10 people. The MFCC feature and Dynamic Time Warping (DTW) feature were extracted which is further classified by using SVM and attains 94.35% accuracy [27].

Mahesha and Vinod (2012) introduced MFCC feature-based speech dysfluency classification using k-NN that produces average accuracy of 86.67% for stuttered speech and 93.34% for normal speech respectively [42]. The same author (2015) classifies the 30 speech disfluency data from the UCLASS database using SVM and GMM super vector providing 98.25% accuracy [41]. Genetic Algorithm, GMM, and PSO based feature selection is proposed using GMM for TIMIT database in that PSO based GMM feature reduces over 85%, that decreases the complexity of Automatic Speaker Verification System. The four types of human emotions feature extracted using pitch, energy, zero-crossing rate, and MFCC. These feature extracted and classification accuracy reached 81%, 78%, 76%, and 77% for anger, happy, sad, and neutral respectively [50].

A. Zabidi et.al (2010) implies healthy and unhealthy infant cry signal features were optimized MFCC parameter using PSO. The obtained optimized feature is classified by MLP which results in good classification accuracy [4]. In 2016 (Bright Kanisha and Ganesan Balarishnanan), the adaptive PSO is used to extract multiple speech features like normal speech signal and conventional speech signal, peak frequency modulation, MFCC, Tri spectral features, and DWT. The obtained all features were given as the population of APSO and it is further classified by multi SVM and the classifier achieves 97.8% using APSO [9].

Gražina Korvel et al, (2018) developed a cepstral and spectral phoneme framework using LPC and MFCC features. SVM and Naïve Bayes algorithm classifies the extracted features and achieves 90.8% and 95.2% as classification accuracy [29].
In 2019, Manjutha M et al., developed a classification system especially for Tamil stuttered and normal speech that classified using SVM and Naïve Bayes. The normal and stuttered speech feature was extracted by introducing computational algorithms PSO and SFO in the feature selection process. The Synergistic Fibroblast Optimization (SFO) produces 96.08% accuracy integrated with naive bays classifier compared to PSO [37].

From the above literature review, it is inferred that the speech processing application explored with nature-inspired algorithms in a different field. The overall implementation of global optimization algorithms like ACO, BPSO, PSO, and GA used by many researchers to solve the diverse speech processing problems is presented in Tables 1, 2, 3, and 4.

Table 1 Ant Colony optimization algorithm for various speech applications

Algorithm	Methodology	Applications of Speech Processing and Dataset	Adapted Techniques	Fitness Function	Results and Discussions
ACO, GA	To develop a text-independent speaker verification system an improved ant colony optimization algorithm was proposed for the speech database (Mehdi Hosseinzadeh Aghdam, 2012) [39]	Speaker Verification, TIMIT Database	MFCC, LPCC, GMM-Universal Background Model (UBM),	Equal Error Rate (EER)	Compared to GA, ACO increases the speech verification and reduces the complexity of ASV system
ACO	To reduce the input number of DNN, feature selection is done by using a hybrid of genetic algorithm (GA) and ant colony optimization (ACO) (Mansour Sheikhan, 2013) [40]	Speech Synthesis, Spoken Words	DNN, Root Mean Square Error (RMSE)	Mean square error (MSE)	The RMSE was reduced by using hybrid GA and ACO
ACO	The ACO is proposed to select speech features for automatic speech recognition (C.Poonkuzhali, et al., 2013) [10]	Feature selection, Speech Signals	MFCC	Features	Compared to 100 iterations, features get diminished 16.6% in ACO of 300 iteration
ACO	The ACO is proposed to solve speech dynamic time warping (Xing Wei, Xiaojin Yang, 2013) [56]	Speech recognition, Chinese pronunciation English data	LPCC, DTW	Path length	The traditional algorithm DTW recognition rate is 92.4% and the proposed ACO-DTW has better global search ability with is 96.75% recognition rate
ACO	Selection of optimal features by introducing computational and machine learning algorithms (M.Kalamani, S.Valarmathy et al., 2014) [28]	Speech Recognition, Quran sound	MFCC, DFT, SVM, HMM, Word Error Rate	Accuracy	ACO algorithm high performance is achieved by using feature selection of fuzzy rough set with immense recognition rate and minor WER.

ACO	In order to enhance the emotions of practical English, and improved quantum ACO is introduced. (Lihui DU and Yueguang Li, 2015) [30]	Emotion Recognition, English emotion speech database	SVM	Gaussian mutation	The recognition rate of English speech significantly improved by using quantum ACO
ACO	In order to extend the text-independent speaker verification system, multi-objective optimization is imported based on hybrid Ant and Bee colony. (J. Sirisha Devi and Srinivas Yarramalle, 2015) [25]	Speaker verification, BERLIN dataset	MFCC, Gaussian Mixture Model (GMM)	Feature subsets	The computational complexity of ASV get decreased using hybrid ACO and ABC which optimizes features over 85%
ACO	To develop a speech recognition system fuzzy-based ACO clustering is proposed. (Fooad Jalili, and Milad Jafari Barani, 2016) [16]	Speech recognition, Speech signal	Denoising Fuzzy model	Euclidean distance	Compared to the fuzzy system the fuzzy and ACO based method has less time complexity

Table 2 Binary Particle Swarm Optimization algorithm for several speech applications

Algorithm	Methodology	Speech Processing Applications and Dataset	Adapted Techniques	Fitness Function	Results and Discussions
Binary Particle Swarm optimization	The infant cry asphyxia is recognized and features were selected by using Binary Particle Swarm optimization (Azlee Zabidi, Wahidah Mansor et al., 2011) [4]	Speech Recognition, University of Milano–Bicocca offers asphyxia infant cry signals and the normal cry signals from Instituto Nacional de Astrofisica, Óptica	MFCC, PCA, multilayer perceptron (MLP) ANN	Features	MLP achieves the highest recognition using MFCC with 14 coefficients
Binary Particle Swarm optimization	Binary Particle Swarm Optimization is introduced for pathological voice features selection	Feature Selection, Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab	Multilayer Perceptron (MLP) neural network, Wavelet 2-D,	Accuracy	The pathological condition of the larynx is identified

	using neural network (Taciana A. Souza, Micael A. Souza, et al., 2015) [52]	produced the database	Haralick Texture Features		better than the healthy vocal folds

Table 3 Particle Swarm Optimization algorithm for various speech applications

Algorithm	Methodology	Speech Processing Applications and Dataset	Adapted Techniques	Fitness Function	Results and Discussions
PSO	PSO algorithm is proposed to optimize the SVM classifier to improve the classification accuracy (Xueying Zhang and Yueling Guo, 2009) [57]	speech recognition, Vocabulary words	MFCC, SVM, SNR	k cross validation accuracy	The optimal SVM achieves 97.43 % and conventional method reaches 96.48 for fifty words
PSO	A speech recognition system based on HMM is developed and optimized by using PSO (Negin Najkar, Farbod Razzazi and Hossein Sameti, 2010) [45]	Speech Recognition, TIMIT speech database includes isolated words and phone classifiers	HMM, MFCC	logarithmic domain over one of the possible paths treated as fitness function	The proposed segment ProbPSO method reduces error rate as 0.66% comparing to Viterbi error 0.87%
PSO and Modified PSO MPSO	A modified PSO is introduced to enhancement speech signal in the adaptive filter (Laleh Badri Asl and Vahid Majid Nezhad, 2010) [6]	Speech Enhancement, Speech Signal	MSE, SNR, Adaptive IIR filters	MSE function	The adaptive filter uses the MPSO algorithm that enhances the SNR ratio as 26.59 dB over PSO 25.13 dB
PSO+ mRMR	The human emotion state has been recognized using optimized features based on PSO integrated with mRMR (S Rajarajeswari, Shree Devi B N and Sushma G, et al., 2013) [50]	Emotion Recognition, Human Emotions such as happiness, anger, sad and normal	energy, pitch, MFCC, Gaussian Mixture Model (GMM), Minimum Redundancy Maximum Relevance (mRMR), SVM	Accuracy	The proposed PSO with mRMR optimal feature subset is selected by GMM techniques which reduces the complexity and load on GMM.

PSO	The speech glottal signal emotions were recognized by introducing PSO that enhances the selection of features (Hariharan Muthusamy, Kemal Polat and Sazali Yaacob, 2015) [22]	Emotion Recognition, Emotional speech like simulated, elicited, and natural.	MFCC, LPCCs, Gamma tone filter bank outputs (GTFBOs), Perceptual linear predictive (PLP) analysis, Timbral Texture Features (TTFs), SWT based Timbral Texture Features (SWT-TTFs), Relative wavelet packet energy and entropy features (RWPFs), ELMkernel classifier	Euclidean distance for PSO clustering for feature Enhancement and new fitness function Geometric mean used for the Feature Selection	The proposed PSO based speaker-independent system enhances the feature selection and improves the multi-class emotions
PSO	Particle swarm optimization-based feature is extracted to develop an automatic speech recognition system (G. C. Batista, et al., 2016) [17]	Speech Recognition, Brazilian Portuguese of the digits (0-9)	MFCC, DCT, SVM	Mel-cepstral coefficients	The proposed PSO algorithm reduces the processing time and computational load during the training of SVM
PSO	A new method of sound verification system is proposed using particle swarm optimization (Thaer M. Tax`ha, Hazem M. El-Bakry, et al., 2016) [53]	Speech Recognition, Human Voice is known as Urban Sound	DCT, FFT, SVM, k-fold cross-validation	Cost function	The proposed sound verification system has a reduced rate of time achieves 23.5% using PSO-SVM
PSO	The human active voice is identified and optimized by using PSO, especially for Tamil stuttered speech database (M. Manjutha and Dr.P.Subashini, 2017) [36]	Features Extraction, Real-time Tamil speech data	Zero crossing rate, Short-time energy, autocorrelation, and normalized autocorrelation	sphere	The PSOVAD improves the detection of active speech more efficiently than the conventional VAD method.
PSO	To develop a Hindi speech recognition database features were optimized by using PSO	Speech recognition, TIFR in Mumbai developed a	Mel frequency (MF) cepstral coefficient (MFCC), perceptual linear	Crossover Accuracy	The accuracy of the proposed Q-PSO enhances the

	(Mohit Dua, Rajesh Kumar Aggarwal, and Mantosh Biswas, 2020) [34]	Hindi speech database	prediction (PLP), and Gamma tone frequency (GF) cepstral coefficient (GFCC), SNR, HMM and PSO with crossover (C-PSO), and PSO with quadratic crossover (Q-PSO).GMM,		recognition system over PSO and C-PSO. The different dataset exists in between the range of 1% to 3%

Table 4 Genetic Algorithm and Particle Swarm Optimization algorithm for various speech applications

Algorithm	Methodology	Speech Processing Applications and Dataset	Adapted Techniques	Fitness Function	Results and Discussions
GA and PSO	In order to verify the speaker individuality, optimized feature selection has been introduced using PSO (Shahla Nemati and Mohammad Ehsan Basiri, 2010) [51]	Feature Selection, TIMIT corpora	MFCCs, LPCCs and GMM	feature subset	Without losing performance the PSO selects the required features than GA and it diminishes the feature vector size over 85% to avoid the complexity of the recognition system
GA and PSO	To develop a robust automatic speech recognition system Filter bank is optimized with GA and PSO (R.K. Aggarwal and M. Dave, 2012) [47]	Speech Recognition, Hindi Phonemes	MFCC, HMM and Multilayer perceptron (MLP)	HMM	The proposed filter bank achieves high accuracy and also the other filter bank achieved up to 4% better than the Mel-scale in various background noises
Chaos PSO, PCA-CPSO-ANN and PCA-GA-ANN,	A novel algorithm as Chaos particle swarm optimization is introduced to recognize the speaker using a backpropagation neural network. (Guowen Wang, Shixin Luo et al., 2013) [20]	Speech Recognition, TIMIT library	LPCC, MFCC, ANN, Principal Component Analysis (PCA),	Mean Square Error	The recognition rate of ANN, PCA-ANN, PCA-CPSO-ANN, and PCA-GA-ANN as 90%, 94%, 98%, and 98%. The optimal solution increased gradually for CPSO-BP

GA and PSO	Improved PSO algorithm based HMM is introduced to enhance the speech recognition system (Lokesh Selvaraj and Balakrishnan Ganesan, 2014) [31]	Speech Recognition, English spoken Words	MFCC, Peak, Pitch Spectrum, Mean, Standard Deviation, Minimum and Maximum, GA based Vector Quantization, NN, IPSO based HMM technique (IP-HMM)	1. Fitness function of GA is vector and the codebook 2. Fitness function of PSO-HMM	The proposed IPSO based HMM attains 97.14% recognition accuracy and NN, HMM, PSO achieves 88.58%, 78.33% and 84.34% respectively
GA, Adaptive GA PSO, Adaptive (APSO), and Harmony Search (HS)	An adaptive PSO based advanced feature has been extracted to build a speech recognition system (Bright Kanisha and Ganesan Balarishnanan, 2016) [9]	Speech features Extraction, Recorded English spoken Words	Peak Frequency Modulation, MFCC, Tri spectral features, DWT, SVM, ANN	maximum accuracy	The proposed APSO results 97.8% accuracy compared to the conventional SVM linear kernel function

III. Observations and Discussion

Nowadays, speech processing is essential for any mode of communication which plays a significant role in many applications and the basic functional model of speech processing systems includes acquiring speech data or formulating new speech database and pre-processing.

Further, the relevant information about the speech utterances is extracted to build a more perfect recognition or classification system. The speech recognition system or the classification system definitely finds the best match by tuning the utterances embedded with optimization algorithms. The basic model of speech processing is shown in Fig.2. which includes data acquisition, Pre-processing, feature extraction or feature selection, and classification or recognition.

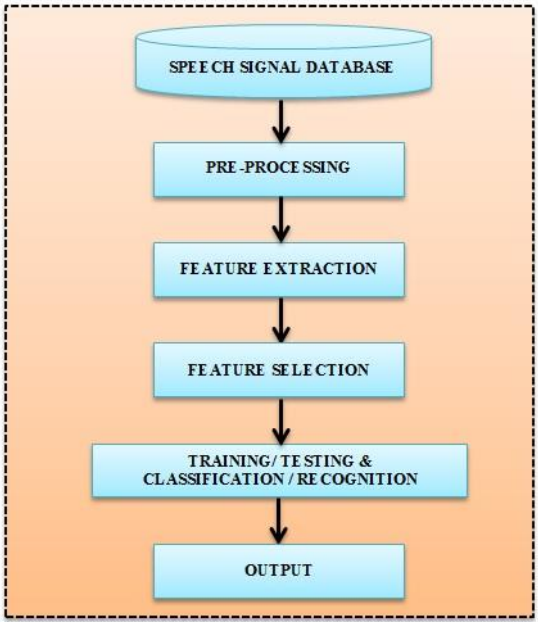

Fig 2 Speech processing Methodology

A. Speech Database

The speech database is created based on the speaker utterances recorded by microphone or using voice recorder or any other audio device in the form of audio waves. The captured speech signals were transformed into electrical signal and then into digital signal since the machine should acknowledge the signals. The proposed methodology efficiency is tested based on the existing database and real-time speech corpus in the native language such as English, Chinese, Brazilian Portuguese, Hindi, Arabic, and Tamil. These speech data consist of spoken words, phonemes, vocabularies, digits, and sentences.

A standard like TIMIT database, BERLIN dataset, and Hindi database developed by Tata Institute of Fundamental Research (TIFR) was used for the speech processing work. The speech emotions can be classified and recognized by using human emotions like happy, sad, angry, and normal. The Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab provide the asphyxia of infant cry signals obtained from the University of Milano–Bicocca and Instituto Nacional de Astrofisica, Óptica offers normal cry signal of infants. From the overall study, it is understood that most of the existing work is computed based on the TIMIT database i.e. “TIMIT acoustic-phonetic Continuous Speech Corpus” and for the human speech like happy, sad, angry, and normal database were widely used.

B. Speech Pre-processing

The speech data is pre-processed with sampling, pre-emphasis, filtering, framing, and windowing. During pre-processing, the amplitude of high-frequency bands that exist in the speech signal get increases and the amplitudes of the lower band get decreases in the adapted filtering techniques. The audio wave signals were

divided into sequential frames and the size of the frame was considered as 25 ms. Many types of windowing techniques are presented in speech processing such as Hann Window, Hamming Window, Blackman Window, Rectangular Window, Gaussian Window, Kaiser Window, Triangular Window, Flat Top Window, Welch Window, and Nuttall Window. Moreover, the conventional windowing techniques adapted by existing work are Hamming or Hanning windowing because the edges of each frame signal discontinuity were reduced.

C. Feature Extraction and Selection

In order to attain efficient accuracy, feature selection is considered as a significant process in speech processing from that relevant feature subsets are acquired for the classification. The performance enhancement of the acoustic data analysis depends only on the feature selection method [35].

Several researchers adopted classical feature extraction techniques like Mel Frequency Cepstral Coefficient (MFCC), Linear Predictive Cepstral Coefficient (LPCC), Linear Predictive Coding (LPC), Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), Independent Component Analysis (ICA), Perceptual linear predictive (PLP) analysis, Tri Spectral Features, Gamma tone Filter bank Outputs (GTFBOs), Timbral Texture Features (TTFs), Wavelet 2-D, Haralick Texture Features, Stationary Wavelet Transform (SWT) based Timbral Texture Features (SWT-TTFs) (includes Energy Entropy, Short-Time Energy, Zero-Crossing Rate, Spectral Rolloff, Spectral Centroid, and Spectral Flux), Relative Wavelet Packet Energy and Entropy features (RWPFs), Autocorrelation, Gamma tone Frequency (GF), Gamma tone Frequency Cepstral Coefficient (GFCC), DTW, Energy, Peak, Pitch Spectrum, Mean, Standard Deviation, Minimum and Maximum. Of these existing techniques widely used for extracting features are MFCC. The computational intelligence is another efficient algorithm that is used to select the optimal features based on the cepstral features, Equal Error Rate (EER), Feature subsets, accuracy and Mean Square Error that significantly improved the performance efficiency of speech recognition and classification system. Among the feature, the selection is applied to many speech applications like speech and speaker recognition, speech enhancement, voice analysis, and speech coding [50] which is represented in Table 5. It specifies the significant needs in the subset feature selection task and to improve the performance of classification and recognition system optimized feature selection plays a major part in the speech processing application.

Table 5 Optimized feature selection techniques for speech processing applications

Authors and Year	Metaheuristic Algorithm	Applications of Speech Processing
W Zha GK and Venayagamoorthy (2007)[54]	PSO	Speech Coding
LB Asl, and VM Nezhad (2010) [6]	PSO	Speech Enhancement
J.S. Lee, Park CH (2010) [26]	SA	Visual Speech Recognition
R. Arefi Shirvan and E. Tahami, (2011) [48]	GA	Voice Analysis

Mehdi Hosseinzadeh Aghdam, (2012) [39]	ACO	Speaker Verification
Hassanzadeh, K Faez and G Seyfi (2012) [21]	FA	Speech Recognition
Mansour Sheikhan (2013) [40]	ACO	Speech Synthesis
Poonkuzhali et al. (2013) [10]	ACO	Speech Recognition
A.V. Ermilov (2014) [3]	SA	Modeling Speech Feature
Xing Wei and Xiaojin Yang (2014) [56]	ACO	Speech Recognition
A Shahzadi et al. (2015) [2]	GA	Speech Emotion Recognition
Lihui DU and Yueguang Li (2015) [30]	ACO	Speech Emotion Recognition
J.Sirisha Devi and Srinivas Yarramalle (2015) [25]	ACO	Speaker Verification
Fooad Jalili and Milad Jafari Barani (2016) [16]	ACO	Speech Recognition
Fabíola Araújo, José Filho, Aldebaro Klautau (2016) [15]	GA	Speech Syntheses
Ahmed Al-Hmouz et.al. (2017) [5]	PSO	Speaker Identification

D. Classification and Recognition

In general, acoustic speech recognition is categorized into three different methods namely, acoustic-phonetic approach, pattern recognition, and artificial intelligence method meanwhile the classification of the accurate speech data was categorized by using machine learning algorithms and computational intelligence. The parameter of the classification system was assessed only through the huge amount of training classes. The acquired feature sets are tested with an exact match of trained speech features in each class.

From the literature study, it's evident that many of the researchers proposed typical classification and recognition models namely Neural Networks (NN), Artificial Neural Networks (ANN), Deep Neural Network (DNN), Fuzzy model, Multilayer Perceptron (MLP), Naïve Bayes Classifier, ELMkernel classifier, Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM) - Universal Background Model (UBM) which are significant to build the definite classifier and recognition system.

The optimal SVM achieves 97.43 % and the conventional method reaches 96.48 for fifty words (Xueying Zhang and Yueling Guo, 2009) [57]. The GMM chooses the optimal subset to avoid the load on GMM for recognition and also reduces the complexity incurred by mRMR combined with PSO (S Rajarajeswari, Shree Devi B N and Sushma G, et al., 2013) [50]. The speech verification system using PSO-SVM reduction rate of time achieves 23.5% (Thaer M. Taha, Hazem M. El-Bakry et al., 2016) [53]. Quadratic crossover PSO (Q-PSO) increases the recognition accuracy in the Hindi speech database (Mohit Dua, Rajesh Kumar Aggarwal, and Mantosh Biswas, 2018) [34]. The speech recognition technique using NN attains

88.58%, HMM attains 78.33%, PSO attains 84.34% and IPSO based HMM achieves 97.14% of accuracy (Lokesh Selvaraj and Balakrishnan Ganesan, 2014) [31]. The APSO achieves 97.8% accuracy compared to the existing technique SVM linear kernel function (Bright Kanisha and Ganesan Balarishnanan, 2016) [9]. Among the overall study, PSO optimization technique is well suited for recognition and classification which enhances the performance of recognition and classification systems significantly because of the fast convergence performance.

VI. Conclusion and Future Work

One of the most prominent and difficult problems in artificial intelligence is the interaction between machines and humans as like human to human. Thus it is necessary to enhance the performance of the recognition system to achieve higher productivity. Hence, the optimal features are opted from high dimensional space by introducing any one of the feature selection techniques based on a metaheuristic algorithm.

In this paper, various applications of speech processing converged with nature-inspired computing paradigms are reviewed to solve the wide range of challenges existing in the real world. In order to solve the diverse sort of detriments in feature extraction and classification, the substantial fitness function and parameter computed in the evaluation of the metaheuristic algorithm are examined. The overall literature identified that the preliminary experimental analysis on the existing work reveals that the nature-inspired problem-solving optimization algorithm has eventually surpassed the performance of conventional work and ultimately affords a significant solution to abstruse problems. Likewise, the global optimum is explored by many authors who predominantly rely on choosing the relevant fitness function and its performance with respect to the nature of the dataset. The optimistic results originated from the in-depth study show that the metaheuristic algorithms are the efficient method that improves the parameters, performance of the developed speech recognition model in resolving the numerous speech and language processing applications. In the future, this study could be extended to focus on the nature-inspired algorithm to overcome the issues in different speech processing applications such as voice analysis, speaker-dependent, and speaker-independent speech recognition models.

References

[1] A. Dev and P. Bansal, Robust Features for Noisy Speech Recognition using MFCC Computation from Magnitude Spectrum of Higher Order Autocorrelation Coefficients, Journal of Computer Applications,10(8) (2010) 36-38.

[2] A. Shahzadi, A. Ahmadyfard, and A. Harim, Speech emotion recognition using nonlinear dynamics features, Turkish Journal of Electrical Engineering & Computer Sciences, 23 (2015) 2056-2073. DOI:10.3906/elk-1302-90

[3] A. V. Ermilov, Modeling Speech Features Via Simulated Annealing Algorithm, Discrete and continuous models and applied computational science, 2 (2014), 354-358.

[4] A. Zabidi, Lee Yoot Khuan, W. Mansor, I. M. Yassin, R. Sahak, Optimization of MFCC parameters using Particle Swarm Optimization for diagnosis of

infant hypothyroidism using Multi- Layer Perceptron, in Proc Annual International Conference of the IEEE Engineering in Medicine and Biology, (2010)1417-1420, DOI: 10.1109/IEMBS.2010.5626712.

[5] Ahmed Al-Hmouz, Khaled Daqrouq, Rami Al-Hmouz and Jaafar Alghazo, Feature Reduction Method for Speaker Identification Systems Using Particle Swarm Optimization, International Journal of Engineering and Technology (IJET), 9 (3) (2017) 1714-1723. DOI:10.21817/ijet/2017/v9i3/170903045.

[6] Asl, Laleh Badri, Nezhad, Vahid Ma, Speech Enhancement Using Particle Swarm Optimization Techniques, in Proc. International Conference on Measuring Technology and Mechatronics Automation, (2010) 441-444. DOI:10.1109/icmtma.2010.510

[7] B. Xue, M. Zhang, W. N. Browne and X. Yao, A Survey on Evolutionary Computation Approaches to Feature Selection, IEEE Transactions on Evolutionary Computation, 20(4) (2016)606-626, DOI: 10.1109/TEVC.2015.2504420.

[8] Bolun Chen, Ling Chen and Yixin Chen, Efficient ant colony optimization for image feature selection, Signal Processing, 93(6) (2013)1566-1576, DOI:10.1016/j.sigpro.2012.10.022.

[9] Bright Kanisha, Ganesan Balarishnanan, Speech Recognition with Advanced Feature Extraction Methods Using Adaptive Particle Swarm Optimization, International Journal of Intelligent Engineering and Systems, 9(4) (2016) 21-30, DOI:10.22266/ijies2016.1231.03.

[10] C.Poonkuzhali, R.Karthiprakash, Dr.S.Valarmathy, and M.Kalamani, An Approach to Feature Selection Algorithm Based on Ant Colony Optimization For Automatic Speech Recognition, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,2(11)(2013).

[11] Czyzewski.A, Kaczmarek. A, Kostek.B, Intelligent processing of stuttered speech, Journal of Intelligent Information Systems, 21(2003)143–171, https://doi.org/10.1023/A:1024710532716.

[12] E. Avci, Z. H. Akpolat, Speech recognition using a wavelet packet adaptive network based fuzzy inference system, In Expert Systems with Applications, 31(3) (2006)495-503. DOI:10.1016/j.eswa.2005.09.058.

[13] E. Saraç and S. A. Özel, An Ant Colony Optimization Based Feature Selection for Web Page Classification, The Scientific World Journal, 2014 (2014)1-16, DOI:10.1155/2014/649260.

[14] F. L. Huang, An Effective Approach for Chinese Speech Recognition on Small size of Vocabulary, Journal of signal and image processing, 2(2) (2011) 48-60. DOI:10.5121/sipij.2011.2205.

[15] Fabíola Araújo, José Filho and Aldebaro Klautau, Genetic algorithm to estimate the input parameters of Klatt and HLSyn formant-based speech synthesizers, Biosystems, 150 (2016)190–193, DOI:10.1016/j.biosystems.2016.10.002.

[16] Fooad Jalili and Milad Jafari Barani, Speech Recognition Using Combined Fuzzy and Ant Colony Algorithm, International Journal of Electrical and Computer Engineering (IJECE), 6(5) (2016) 2205-2210.

[17] G. C. Batista, W. L. Santos Silva and A. G. Menezes, Automatic speech recognition using Support Vector Machine and Particle Swarm Optimization, in Proc. *IEEE Symposium Series on Computational Intelligence (SSCI),* Athens, (2016)1-6. DOI: 10.1109/SSCI.2016.7850125.

[18] Gambardella L. and M. Dorigo, Ant-Q: A Reinforcement Learning approach to the traveling salesman problem, in Proc. of ML-95, Twelfth International Conference on Machine Learning, Tahoe City, CA, A. Prieditis and S. Russell (Eds.), Morgan Kaufmann,1733(1995), 252–260, DOI:10.1016/b978-1-55860-377-6.50039-6.
[19] Ganapathiraju A, Hamaker J. E, and Picone J, Applications of support vector machines to speech recognition, IEEE Transactions on Signal Processing, 52(8) (2004) 2348-2355, DOI:10.1109/tsp.2004.831018.
[20] Guowen Wang, Shixin Luo, Li He, Gang Yin, Application BP neural network in the speaker recognition Based on Chaos Particle Swarm Optimization Algorithm, Advanced Materials Research, 765-767 (2013) 2805-2808.
[21] H. Kanan, K. Faez, S. M. Taheri, Feature Selection Using Ant Colony Optimization (ACO): A New Method and Comparative Study in the Application of Face Recognition System, P. Perner (Ed.): ICDM 2007, LNAI 4597, (2007) 63–76.
[22] Hariharan Muthusamy, Kemal Polat, Sazali Yaacob, Particle Swarm Optimization Based Feature Enhancement and Feature Selection for Improved Emotion Recognition in Speech and Glottal Signals, PLoS ONE, 10(3) (2015) 1-20, DOI:10.1371/journal.pone.0120344. [29]
[23] Ibrahim M. El-Henawy, Walid I. Khedr, Osama M. ELkomy, Al-Zahraa M.I. Abdalla, Recognition of phonetic Arabic figures via wavelet based Mel Frequency Cepstrum using HMMs, Journal of Housing and Building National Research Center, 10(2014) 49-54.
[24] J. Jiang, Z. Wu, M. Xu, J. Jia and L. Cai, Comparing feature dimension reduction algorithms for GMM-SVM based speech emotion recognition, Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific, (2013) 1-4, doi: 10.1109/APSIPA.2013.6694336.
[25] J. Sirisha Devi and Srinivas Yarramalle, Multi-Objective Optimization Problem resolution based on Hybrid Ant-Bee Colony for Text Independent Speaker Verification, I.J. Modern Education and Computer Science, 1(2015) 55-63.
[26] J.S. Lee and C.H. Park, Hybrid simulated annealing and its application to optimization of hidden Markov models for visual speech recognition, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 40(4) (2010) 1188-1196.
[27] K.M.Ravikumar, Balakrishna Reddy, R.Rajagopal, H.C.Nagaraj, Automatic detection of syllable repetition in read speech for objective assessment of stuttered disfluencies, in Proc. of World Academy Science, Engineering and Technology (2008) 270–273.
[28] Kalamani, M., Valarmathy, S., Poonkuzhali, C., & Catherine J N. (2014), Feature selection algorithms for automatic speech recognition. 2014 International Conference on Computer Communication and Informatics, 1-7, DOI:10.1109/iccci.2014.6921797.
[29] Korvel, G., O. Kurasova, and B. Kostek, Comparative analysis of spectral and cepstral feature extraction techniques for phoneme modelling, in Proc. 11th International conference MISSI (2018) 480-489.DOI:10.1007/978-3-319-98678-4_48.
[30] Lihui DU and Yueguang Li, Recognition of practical English speech emotion using improved Quantum Ant Colony Algorithm, International Symposium on Computers & Informatics (ISCI 2015).

[31] Lokesh Selvaraj and Balakrishnan Ganesan, Enhancing Speech Recognition Using Improved Particle Swarm Optimization Based Hidden Markov Model, Hindawi Publishing Corporation, The Scientific World Journal, Volume 2014, Article ID 270576, 1-10.
[32] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, C. Wellekens, Automatic speech recognition and speech variability: A review, Speech Communication, 49(10-11) (2007)763-786.
[33] M. Dorigo, M. Birattari and T.Stutzle, Ant Colony Optimization: Artificial Ants as Computational Intelligent Technique, IEEE Computational Intelligent Magazine, (2006).
[34] M. Dua, R. Aggarwal and M. Biswas, Optimizing Integrated Features for Hindi Automatic Speech Recognition System, Journal Intelligent System, 29(1) 2020, 959–976, DOI:10.1515/jisys-2018-0057.
[35] M. M Kabir, M. Shahjahan and K Murase, A new hybrid ant colony optimization algorithm for feature selection, Expert Systems with Applications, 39 (2012) 3747–3763.
[36] M. Manjutha, Dr. P. Subashini, Particle Swarm Optimization based Voice Activity Detection for Stuttered Tamil Speech, International Journal of Computer Engineering and Applications, 11(9) (2017) 1-14.
[37] M. Manjutha, P. Subashini, M. Krishnaveni and V. Narmadha, An Optimized Cepstral Feature Selection method for Dysfluencies Classification using Tamil Speech Dataset, in Proc. IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, (2019), 671-677.
[38] M.Dorigo and K.Socha, An introduction to ant colony optimization, Handbook of Metaheuristic, Brussels: IRIDIA, 26(1) (2006).
[39] M.H. Aghdam. An Improved Ant Colony Optimization Algorithm and its Application to Text-Independent Speaker Verification System, JAISCR,2 (4) (2012)301-315.
[40] M.Sheikhan, Synthesizing Supra segmental Speech Information Using Hybrid of GA-ACO and Dynamic Neural Network, in Proc. 5th Conference on Information and Knowledge Technology (IKT) (2013) 175-180, doi: 10.1109/IKT.2013.6620060
[41] Mahesha. P, Vinod.D.S, Support vector machine-based stuttering dysfluency classification using GMM supervectors, Int. J. Grid and Utility Computing, 6(3/4) (2015)143–149.
[42] Mahesha.P, Vinod. D. S, Feature based classification of dysfluent and normal speech, in Proc. Second International Conference on Computational Science, Engineering and Information Technology - CCSEIT 12, (2012) 594-597.
[43] Manikandan J, Venkataramani B, Girish K, Karthic H and Siddharth V, Hardware implementation of real-time speech recognition system using TMS320C6713 DSP, in Proceedings of IEEE International Conference on VLSI Design, 250-255, 2011. 45
[44] Manjula.G, Shivakumar.M, Geetha.Y. V, Adaptive optimization based neural network for classification of stuttered speech, in Proc. 3rd International Conference on Cryptography, Security and Privacy - ICCSP ’19 (2019) 93-98.
[45] N. Najkar, F. Razzazi and H. Sameti, A novel approach to HMM-based speech recognition systems using particle swarm optimization, Mathematical and Computer Modelling 52 (2010), 1910-1920, DOI:10.1016/j.mcm.2010.03.041.

[46] R. Alhutaish and N.Omar, Feature Selection for Multi-Label Document Based on Wrapper Approach through Class Association Rules, International Journal on Advanced Science Engineering and Information Technology, 7(2) (2017),DOI: 10.18517/ijaseit.7.2.1040.

[47] R. K. Aggarwal; M. Dave, Filterbank optimization for robust ASR using GA and PSO, International Journal of Speech Technology, 15(2), (2012),191–201, DOI:10.1007/s10772-012-9133-9

[48] R.A. Shirvan and E. Tahami, Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method, Biomedical Engineering (ICBME), 2011 18th Iranian Conference of, 14-16 (2011), 278-283, DOI: 10.1109/ICBME.2011.6168572.

[49] R.Mehmood, W.Shahzad and E. Ahmed, Maximum Relavancy Minimum Redundancy Based Feature Subset Selection using Ant Colony Optimization, Journal of Applied Environmental and Biological Sciences, 7(4), (2017)118-130.

[50] S Rajarajeswari, Shree Devi B N, Sushma G, Optimal Feature Selection of Speech using Particle Swarm Optimization Integrated with mRMR for Determining Human Emotion State, International Journal of Computer Applications, 74(10) (2013) 48-52.

[51] S. Nemati, M.E. Basiri, Particle Swarm Optimization for Feature Selection in Speaker Verification, in Proc. International Conference on the Applications of Evolutionary Computation, 6024, (2010) 371–380, DOI:10.1007/978-3-642-12239-2_39

[52] T. A. Souza, M. A. Souza, W. C. d. A. Costa, S. C. Costa, S. E. N. Correia and V. J. D. Vieira, Feature Selection based on Binary Particle Swarm Optimization and Neural Networks for Pathological Voice Detection, 2015 Latin America Congress on Computational Intelligence (LA-CCI),(2015) 1-6, DOI: 10.1109/LA-CCI.2015.7435962.

[53] Widana, I.K., Dewi, G.A.O.C., Suryasa, W. (2020). Ergonomics approach to improve student concentration on learning process of professional ethics. *Journal of Advanced Research in Dynamical and Control Systems*, *12*(7), 429-445.

[54] Thaer M. Taha, Hazem M. El-Bakry, Amal Ibrahim, Samir Abd-Elrazik, Magdi Z. Rashad, Fast Sound Verification Using Support Vector Machine and Particle Swarm Optimization Algorithms, International Journal of Advanced Research in Computer Science & Technology (IJARCST 2016), 4(1), (2016)78-83.

[55] W. Zha and G. K. Venayagamoorthy, Comparison of Non-uniform Optimal Quantizes Designs for Speech Coding With Adaptive Critics and Particle Swarm, IEEE Transactions on Industry Applications, 43(1) (2005)674-679

[56] Wisniewski.M, Kuniszyk-Jozkowiak.W, Smolka.E, Suszynski.W, Automatic detection of prolonged fricative phonemes with the hidden markov models approach, Journal of Medical Informatics and Technologies, 11(2007) 294-298.

[57] X. Wei and X. Yang, Speech dynamic time warping based on ant colony optimization algorithm, 3rd International Conference on Consumer Electronics, Communications and Networks, (2013) 602-604, DOI: 10.1109/CECNet.2013.6703403.

[58] X. Zhang and Y. Guo, Optimization of SVM Parameters Based on PSO Algorithm", Fifth International Conference on Natural Computation, (2009) 536-539, DOI: 10.1109/ICNC.2009.257.