International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

A Robust Isolated Automatic Speech Recognition System using Machine Learning Techniques

Sunanda Mendiratta, Neelam Turk, Dipali Bansal

Abstract: ***In order to make fast communication between human and machine, speech recognition system are used. Number of speech recognition systems have been developed by various researchers. For example speech recognition, speaker verification and speaker recognition. The basic stages of speech recognition system are pre-processing, feature extraction and feature selection and classification. Numerous works have been done for improvement of all these stages to get accurate and better results. In this paper the main focus is given to addition of machine learning in speech recognition system. This paper covers architecture of ASR that helps in getting idea about basic stages of speech recognition system. Then focus is given to the use of machine learning in ASR. The work done by various researchers using Support vector machine and artificial neural network is also covered in a section of the paper. Along with this review is presented on work done using SVM, ELM, ANN, Naive Bayes and kNN classifier. The simulation results show that the best accuracy is achieved using ELM classifier. The last section of paper covers the results obtained by using proposed approaches in which SVM, ANN with Cuckoo search algorithm and ANN with back propagation classifier is used. The focus is also on the improvement of pre-processing and feature extraction processes.***

Keywords: Speech recognition system, SVM, kNN, ANN, Cuckoo search optimization, ELM

I. INTRODUCTION

Ability to communicate is one of the most fundamental aspects of human behaviour. Through natural languages human communicate with each other verbally and in written form. Human communication written format is represented by vocalized form of human communication i.e., speech [1]. A high quality human computer interactive system has been developed by advancement in language and speech technologies. It has broad applications in education, entertainment and business and to make man-machine communication more user friendly human-computer interfaces are designed in which natural languages are used for interaction between users and machines [2]. As in case of human-human communication a loop of interaction is defined by flow of information between computer and human. The vocalized form of natural language speech or text make possible to communicate and vocalized form of human speech or communication is a most convenient way for human communication. It will lead to speech recognition system development and the machine understands the meaning of human speech. This is a difficult problem and relatively active area of research. The translation of spoken works into respective written scripts is done by speech recognition and language of speech is identified using Automated speech recognition (ASR) system and then in a respective natural language the segments of input speech is converted into respective units of text. By this an interaction between human and computer has become easier and systems have become user friendly [3]. And long term goal of HCI is minimizing the barrier between humans mental model. This model is on what they want to accomplish and computers support of the user's task. Preparation of structured documents, aircraft, data entry, speech to text processing and voice dialling like voice user interfaces are possible speech recognition applications in HCI. Helping persons to develop fluency with their speaking skills and listening to the proper pronunciation are used for learning different languages in ASR technology [4]. By use of speech to text programs physically disabled students can who suffer from strain injury to upper extremities be relieved to worry about handwriting. Without physically operating a keyboard or mouse, a computer can be use at home to search on internet by utilizing the speech recognition technology. Without the concern of spelling and other writing mechanics a students with learning disabilities can write better by the concept of speech recognition [2].

To facilitate the communication between machines and humans ASR can be used and in various applications a man-machine interaction and speed based applications are demonstrated. Communication interfaces for people with special abilities, translation devices, hands-free machine operations, dictation systems and voice-mail systems in telephony are its applications. On other hand noise free environment, vocabulary and language, low talking rates and speaker dependency are some of its limitations. So, to improve the results work has been done in this field by various researchers [5].

In the context of isolate word recognition (IWR) basic idea behind ASR can be explored. Independent of environment, speaker and device a conversion of speech signal into its equivalent text message is the goal of ASR [6]. It is a problem of pattern recognition in which features are extracted and a model is used for training and testing.

This paper is divided into various sections in which second section gives brief introduction of ASR architecture. The third section contains the brief details about machine learning and its use in ASR. This section also contains the review on use of SVM and ANN for speech recognition system.

Manuscript published on 30 August 2019.
*****Correspondence Author(s)
Sunanda Mendiratta*, Department of Electronics Engineering, J. C. Bose UST, Faridabad, India. *E-mail: sunanda.mendiratta@gmail.com*
Neelam Turk, Department of Electronics Engineering, J. C. Bose UST, Faridabad, India.
Dipali Bansal, ECE Department, FET, Manav Rachna International Institute of Research and Studies, Faridabad, India.

The fourth section of paper contains the methodology used in proposed work and comparison results obtained by various researchers using different classifiers. The fifth section of paper contains the simulation results obtained using proposed work. The paper is then ended with conclusion in the last section.

II. ARCHITECTURE OF ASR

The pre-processing, feature extraction and classification are three phases of basic Automatic Speech Recognition system.

A. Pre-processing:

A/D converter, Pre emphasis, windowing, background noise filtering and blocking are main tasks. Analog electrical signals are converted into a discrete valued or discrete time signal or are digitized is the first task of pre-processing in ASR. Quantization and sampling are two steps of analog to digital conversion process. The amplitude is measured at particular time in sampling and number of samples taken per second is the sampling rate. For application of speech recognition an 8 to 20 kHz sampling rate is used and about 10 kHz of frequency occur in speech as indicated by perceptual studies but within this considerable narrower range, speech remains intelligible.

Quantization factor is a second important factor that determines the type of scale used to represent intensity of signal. Sufficient information is capture by 11-bit numbers although only 8 bit numbers are achieved using log scale. The one of 256 distinct categories are used to classify each segment of signal. This is specifying by most of the current speech recognition systems. So, a stream of 8-bit numbers speech signal representation at the rate of large number of data or 10,000 numbers per second is a challenging task. The reduction of data into some manageable representation is a challenge for speech recognition system. Also to keep high SNR, background noise is filtered after completion of signal conversion task.

The presence of speech in the background of silence and noise from recording speech is an important problem in speech processing as a background noise or silence is also quantized with speech data while capturing speech. This is an end point detection problem as amount of processing can be kept at minimum by accurately detection of beginning and end. Pre-emphasis is the next step whose task is to emphasize important frequency components in a signal by spectrally flattening the signal.

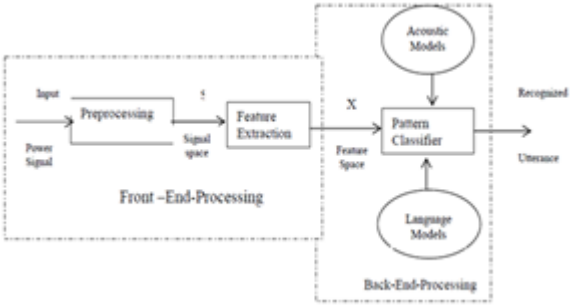


Figure 1: Architecture of ASR [7]

B. Feature Extraction/Parametric Transform:

Finding a set of properties of an utterance that have acoustic correlations in the speech signal is the goal of feature extraction. Or calculating the feature parameters that can be computed or estimated through processing of signal waveform. Parameterization of speech signal is feature extraction that includes the process of measuring frequency or energy response like important characteristics of the signal. Mel Frequency Cepstral Coefficients (MFCC), Missing feature approach, Linear Predictive Cepstral Coefficients (LPCC) and Wavelet as feature extraction processes are some of the ways of extracting features from speech signals [8].

C. Acoustic Modelling:

In this there is establishment of connection between acoustic information and phonetics. Speech is a temporal signal, so temporal models are used to map a speech unit into its acoustic counterpart Dynamic Bayesian Network (DBN), Artificial Neural Network (ANN) and Hidden Markov Model (HMM) models are used for this task. ANN is a general pattern recognition model that is used in ASR. In early years to achieve improvement in substantial performance a HMM approach was used [9].

D. Language Modelling:

Pr (w) or production of accurate value of probability of a word W is the goal of language modelling that generates probabilities and contains structural constraints available in language. After determination of word sequence by language modelling an occurrence of word probability take place. Each language has its own constraints for validity and with variation in speech application there is variation in complexity and method of modelling language. A grammar based approach is used to model a phone dialling like small vocabulary constrained on other hand stochastic approach is required in case of broadcast news transcription like large applications.

III. MACHINE LEARNING FOR SPEECH RECOGNITION SYSTEM

To recognize handwriting, speech and facial features patterns, sophisticated skills have been developed. Machine learning has taken birth from use of computer programs that make computers learn from past experience of above skills. There are four basic methods of learning on the basis of machine gaining knowledge to respond correctly are:

- Supervised learning
- Un-supervised learning
- Semi-supervised learning
- Active learning

In many commercial applications ASR systems are deployed but its problems are still unsolved. For acoustic modelling in ASR number of Machine learning techniques have been employed and in prediction problems, Markov models are used. The outcome can't be tied to particular state of Markov model in case of realistic problems although probability distribution associated with the states based HMMs can be used to estimate. ML is considered as most dominant techniques for ASR and considered as most significant paradigm shift in speech recognition [10].

It is considered as a starting point of speech recognizer. But training, decoding and evaluation are three classical problems associated with HMMs. A large amount of training data is required in HMMs to prevent the loss of performance by mismatch between conditions of training and testing. Its output densities are estimated by GMMs so GMM/HMM systems are considered as most prominent generative learning approach used in ASR [11]. Then ANN based approaches have developed and work done by various researchers justify the use of ANNs in ASR. In ASR multilayer perceptrons and ANNs use is justify for estimation of probability [12].

Support Vector Machine (SVM) is another alternate probability estimation technique and HMMs are generative models. In this likelihood is used to make decisions that generative model has produced on current pattern [13]. The robustness of ASR gets improved by the use of SVMs excellent generalization capabilities.

A. Use of SVM in speech recognition system:

Now a days a lot of research has been going on products and services based speech recognition. Machine learning paradigms are at the centre stage of speech recognition methodology. In man-machine assistive technology, mobile computing, and natural user interface are emerging applications of ASR technology. Osman Eray, et.al, (2018), have given a new design in which they have used SVM and ANN two important machine learning paradigms in ASR for Chhattisgarhi geographically important Indian dialect [15]. Maximum 50 isolated words of 15 speakers dataset is applied on conventional feed-forward SVM and ANN and compared their performance with Hidden Markov Model (HMM). With speaker variation experiments an independent and dependent speaker ASR tendency is investigated extensively. The numerical validation confirmed the stability and reliability of ASR. In the presence of noise good results have been achieved in various applications by the use of SVM. The advantages of SVM over ANNs have attracted its use in speech processing community. Its use is restricted in ASR due to high computational requirements where use of ANNs proves to be successful. The huge speech training databases with highly overlapped classes and millions of samples is a cause of high complexity of SVMs. The complexity of SVMs can be reduced by the use of weighted least squares (WLS) training procedure suggested by Rubén Solera-Ureña, et.al, (2012) that facilitate the compact semi-parametric model possibilities [16]. So, for real-time speech decoding on a connected-digit recognition task, use of proposed hybrid WLS-SVC/HMM system is allowed by reduction in complexity by two to three orders of magnitude with respect to conventional SVMs. Both in noisy and clean condition a good performance is achieved using proposed system but still there is need of more improvement to reach the maturity level of current context dependent HMM based recognizers.

In overall performance of ASR system a vital role is played in which parametric illustration of an input speech signal is acquired using optimal speech feature extraction. All important features of the signal are captured by the use of good feature selection algorithm and features extraction technique and reject irrelevant features. The pattern classification and recognition system performance can be influenced by critical task of feature selection. So, Sunanda Mendiratta, et.al, (2016) have proposed a hybrid of PSO and ABC for classification. It is used for optimal selection and extraction of features [17]. In this pre-processing, feature extraction and selection and classification are three stages in which wiener filter is used for pre-processing that reduces the level of noise. Then in feature extraction stage, 8 types of statistical and acoustic features are extracted and then selection of optimal set of extracted features is done using proposed ABC-PSO algorithm. After that SVM classifier is trained using optimized features and signal to corresponding text is displayed as the output using these optimized features. The MATLAB software is used to implement the proposed ASR that shows a high system performance for speech recognition.

In ML an important role is played by kernel function and Balwant A. Sonkamble, et.al, (2012), have considered the speech recognition problem as ML problem. For speech recognition generalization performance improvement a SVM is designed using Linear polynomial kernel (LP) a new kernel function by Balwant A. Sonkamble, et.al, (2012) [18]. The result shows an improvement in performance using LP kernel as compared to polynomial and linear kernel and also results in improvement of generalized performance ability of speech recognition system. A simple voice control and human-computer interaction can be achieved using speech recognition that is widely used in consumer electronics, industrial control and many other fields. For isolated words and specific people a higher performance speech recognition system has been presented by Xiu-Qing Zhang, et.al, (2010), in which they have combined characteristics of human physiology [19]. As core algorithm a dynamic programming algorithm and as a feature vector a Linear Prediction Coefficient and Mel-Frequency Cepstral Coefficient (LPMCC) is used to realize on a digital signal processor (DSP) system. Then further efficiency of real time processing is modified by using SVM a popular tool for ML task.

B. Use of ANN in speech recognition system:

How ASR system is efficiently designed for ubiquitous control is given by Anurag Bajpai, et.al, (2018). In which an algorithm based design is used that extract isolated words from a continuous speech signal and Mel Frequency Cepstral Coefficients (MFCC) is used to extract the features from voiced part of speech signal [20]. The training and pattern part is done using ANN and decision based Euclidean distance between tested and trained voiced commands are used to obtain an increased rejection of unauthorized speech commands. At pre-processing stage a signal SNR an improvement is achieved predicted from experimental results. As a basic appealing area of computerized sign process there is rise in speech processing. Discourse acknowledgment is the working of naturally perceiving the talked expressions of individual in view of information substance in discourse sign. Basic and regular managing amongst machine and human is empowered by discourse programmed discovery. In speech to text change, call directing, voice dialling, local apparatuses control, phone correspondence, robotization frameworks, text to speech transformation and lip synchronization are applications of communication.

On automatic Gujarati speech recognition a hybrid HMM/ANN approach is implemented by Sanjay Valaki, et.al, (2017) along with the combination of classification method and different feature extraction techniques [21]. Demonstration of indistinguishable word variety from different speakers is the fundamental difficulty of discourse acknowledgement. On the basis of social vernaculars, talking styles, sex and vocabulary estimate separate speakers can't be distinguished.

As compared to other medical techniques an acoustic analysis can be used to get definite results in identification of voice disorders. The first and second derivatives of MFCC are presented by Nawel Souissi, et.al, (2016) [22]. To demonstrate usefulness of the short term cepstral parameters and to conclude voice impairments detection efficient system a comparative study is established. The proposed system discriminatory is improved by the LDA based on projection and then ANN is used to classify the combination of every feature. The specificity, accuracy, area under curve (AUC), sensitivity and precision parameters are used to evaluate the performance of proposed system. The final results show that original MFCC features optimized combination with their 1st and 2nd derivatives gives the best performances. 87.82% was the achieved accuracy rate and 87.96% is AUC. For speech recognition use of Human auditory system is highly strong as compared to state of the art of ASR systems. In addition of robustness to speech recognition system a use of highly robust and compressed features are proven to be more efficient. A novel approach for feature extraction has been presented by Santosh Gupta, et.al, (2016), that proposed simple and efficient noise robust ASR system [23]. K-means is another popular approach for feature extraction but it is complex and time consuming. The different 65 words with low SNR (-5 dB) is used to perform the experiment on proposed noise-robust ASR system. For speech recognition a MFCC were used as features and ASR system is designed by using back propagation ANN. The speech recognition system is tested using feed forward neural networks and training algorithms for each type. The experimental results show that the proposed noise robust ASR system is highly accurate in recognition even in case of low SNR.

Mostly people use speech to interact with each other and as it is a most common method of communication, people also need to use speech to interact with machines. It can be predicted from the review that a lot of momentum has been gained by ASR. Sunanda Mendiratta, et.al, (2015), have used optimization technique based on Cuckoo Search Optimization (CSO) for ANN to make efficient ASR system [24]. In this work efficiency of neural network classification performance is improved using CSO. As mentioned in Sunanda Mendiratta, et.al, (2016), this work also consists of pre-processing, feature extraction and classification stages [17]. Where background noise present in the speech signal is removed in pre-processing stage then Linear Predictive Coding Coefficients (LPCC) and Mel Frequency Cepstrum Coefficients (MFCC) acoustic features are extracted from pre-processed signal. Then neural network is trained using these features and recognized the text correspond to the given speech signal using extracted features. MATLAB working platform is used to implement the proposed method and its results indicate that proposed approach gives better results as compared to ordinary ANN.

C. Phases of speech recognition system

There are two phases in which speech recognition system operate are:

- Training
- Testing

The stored database of the system speakers voice characteristics are learned by system in training phase. The extracted features vectors from voice signal are applied in the formation of reference model by using neural network training module. On other hand same feature vectors are extracted from test utterance with same process in testing phase. Testing is an actual recognition task and some matching techniques are used to test the degree of their match with the reference. The final decision level of match is achieved that determines that for further processing activities the test utterance is acceptable or rejected.

Gaussian Mixture Model (GMM), HMMs like different statistical models are used for pattern matching that consider temporal changes and underlying variations of acoustic pattern. Or the similarity between two sequences is measure using Dynamic Time Warping (DTW) that fluctuates in time or speed. It is used even in the case of non-linear variation or when there is variation is speaking speed during the sequence. Fuzzy Vector Quantization (FVQ), GMM, Crisp Vector Quantization (CVQ), Learning Vector Quantization (LVQ) and Self-Organizing Map (SOM) modelling techniques are used for speaker recognition systems. The principle used for clustering defines the success of each of the modelling techniques. Then different modelling techniques are used to model the speakers by extracted features from the utterances [31, 35]. There is comparison of each feature vector of test speech data with all the speaker models while testing and speaker model with minimum Euclidean distance or maximum posterior probability is considered as tentative speaker of speech frame. The final speaker of the test speech data will be the one that have assignment of recognizing a maximum number of frames. The amount of training data, testing data and number of Gaussian mixtures or codebook size tells the recognition rate of speaker.

IV. PROPOSED WORK

The ability of listening the spoken words is the speech recognition and it also included the identification of sounds present in it along with recognition of some known language. In this work, a new proposed speech recognition system results are compared. The proposed system comprises of three phases:

- Pre-processing
- Feature extraction and selection
- Classification

For pre-processing, Discrete Fourier Transform (DFT) has been used, that measures and fragment the noise level of selected recorded audio signal. Then most required features of audio signal are extracted for better classification. The important features considered in this work are pitch measure, word length, sampling point and an innovative phenome intelligibility feature, which is also extracted for the ASR purpose.

But there is also need to consider some more features to reduce the mannerism, dialects and accent level of recognized signal as variations in these features occur due to improper prediction of emotional speech signal. Due to this emotional feature extraction is also considered in the proposed system. In the second phase classifier training is performed on selected features from extracted features. Then these extracted features are used to classify the audio signal in last phase. The review of various classifiers shows that there is need to improve the existing traditional classifier.

The speech database is considered for processing purpose. Speech may be in the form of isolated or connected words, phrases or even sentences.

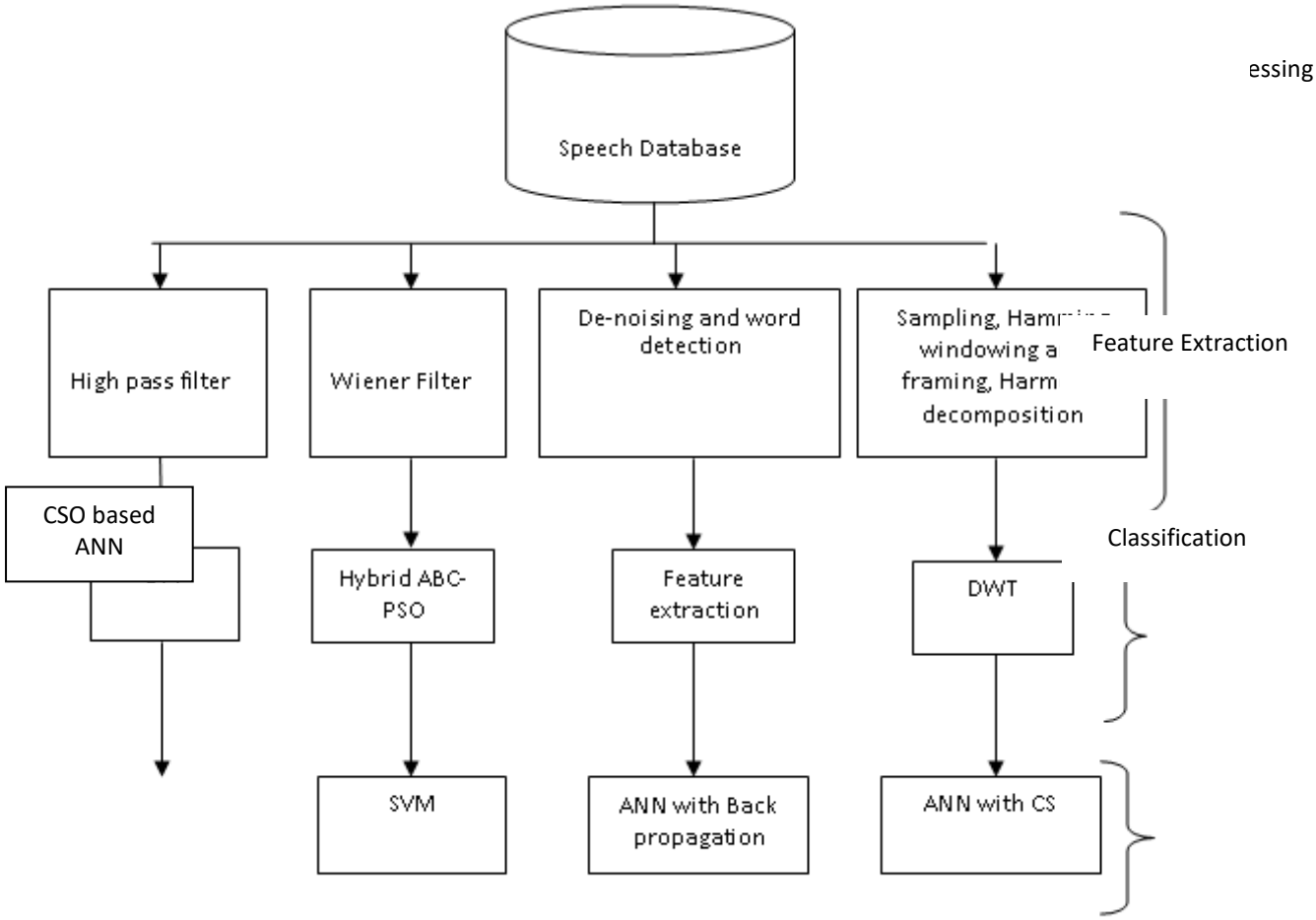


Figure 2: Proposed Techniques for Pre-Processing, Feature Extraction and Classification

In pre-processing phase we have used moving average, high pass filter for removal of background noise then it is framed & passed through window. For second phase of feature extraction Discrete Fourier Transform (DFT) is used to extract two kinds of acoustic features from the speech signal. They are:

- Mel Frequency Cepstrum Coefficients (MFCC); and
- Linear Predictive Coding Coefficients (LPCC)

Then these two extracted features are given to classification phase consisting of Cuckoo search optimization based ANN to recognize the corresponding text. Then a new approach has been proposed in which wiener filter is used for pre-processing purpose and eight types of features are extracted that is divided into two types are:

- Statistical features: Like variance, energy, mean, standard deviation, skewness
- Acoustic features: Like LPCC, MFCC and Pitch

After extraction, selection of the features is done using combination of Ant Bee Colony (ABC) with Particle Swarm Optimization algorithm (PSO) that selects the most optimal set of features. Then in the last SVM is used as classifier for classifying this most optimal set of features for recognition.

In new work de-noising and word detection is used for pre-processing purpose then features are extracted such as word length, sampling point and pitch - common features and Kurtosis, variance, mean and entropy - statistical features from it. For classification purpose they have used ANN with back propagation.

Then to improve speech recognition system Sampling, Hamming windowing and framing is performed. Harmonic decomposition is used to pre-process the speech dataset. The speech data set is further given for Discrete Wavelet Transform (DWT) as a feature extraction technique and then CS-ANN is used for classification.

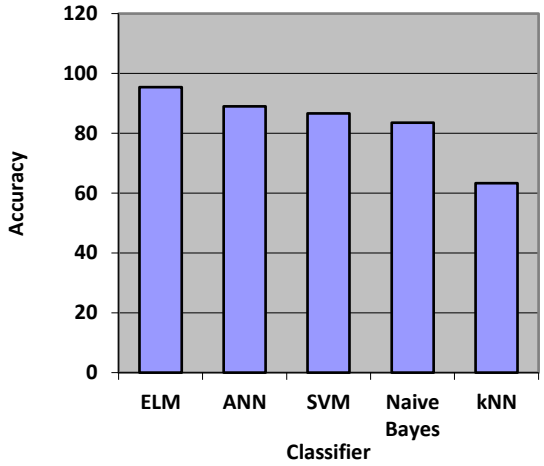
The Table 1 and Figure 3 gives the comparison results of work done by various researchers using different classifiers for Malayalam and Tamil language.

Sonia Sunny, et.al, (2013), have used ANN, SVM and Naive Bayes as classifier for speech recognition system that is tested for Malayalam database.

Table I: Comparison table of work done by various researchers in terms of accuracy using various classifiers

Classifier	Accuracy	Language
ELM [28]	95.45	Tamil
ANN [27]	89	Malayalam
SVM [27]	86.6	Malayalam
Naive Bayes [27]	83.5	Malayalam
kNN [28]	63.33	Tamil

ELM stands for Extreme Language Machine that is used by M. Kalamani, et.al, (2015), for automatic speech recognition and speech emotion recognition. On the basis of



continuous probability density function, weights between hidden and input neurons are assigned randomly. From results it has been proved that use of ELM in speech recognition system gives better results. Same authors have also used k-nearest network (kNN) for same work and tested it in terms of accuracy for Tamil language. The comparison shows that the accuracy of ELM is highest as compared to other classifiers.

Figure 3: Comparison of work done by various researchers in terms of accuracy using various classifiers.

V. RESULTS AND DISCUSSION

Table 2 shows the results obtained from various proposed classifiers in terms of accuracy. It is obvious from the table that results of accuracy obtained using ANN with cuckoo search optimization is better as compared to that using ANN with back propagation and SVM classifier. The hybridization of classifier with optimization algorithms results in improvement in accuracy of speech recognition system. Figure 4 shows the comparison of the hybrid classifiers (CSO-ANN) and the individual classifiers (SVM). CSO based ANN gives the highest accuracy.

TableII: Comparison table

Classifier	Accuracy (%)
CSO based ANN	89.65
SVM	83.89
ANN with Back propagation	62.00
ANN with Cuckoo Search	89.53

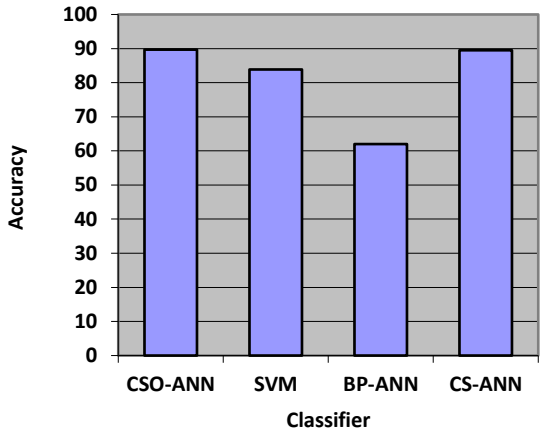


Figure 4: Comparison of proposed classifiers in terms of accuracy

VI. CONCLUSION

The translation of spoken works into respective written scripts is done by speech recognition and language of speech is identified using Automated speech recognition (ASR) system and then in a respective natural language the segments of input speech is converted into respective units of text. This makes human computer interaction easier and systems to be more user friendly. Three basic phases of ARS speech recognition system are pre-processing, feature extraction and classification. By use of machine in speech recognition have some problems like variation in accent, speed, volume, pronunciation and pitch. Speech may be distorted by a background noise and echoes, electrical characteristics. So, various researchers have worked on improving the speech recognition system to get better results. In this paper we have covered the review on work done using various machine learning approaches in speech recognition system. The work is done on improving the basic three phases of ASR. In the last section of this paper we have compared results obtained by researchers using various classifiers and improving the pre-processing and feature extraction. The work shows that traditional classifier results can be further improved by doing hybridization of it with other optimization algorithms. Hybridization of an algorithm with an optimization technique has given better results. More hybrid algorithms can be explored in the area of speech recognition systems to improve the accuracy of the system.

REFERENCES:

1. P. Sanderson, “Cognitive work analysis and the analysis, design, and evaluation of human-computer interactive systems, in proc: Computer Human Interaction, pp.220-22, 1998.
2. A.A.M. Abushariah, T.S. Gunawan, O.O. Khalifa, M.A.M. Abushariah, “English digits speech recognition system based on Hidden Markov Models”, in proc: Computer and Communication Engineering, pp. 1- 5, 2010.
3. M. Gavrilescu, “Improved automatic speech recognition system using sparse decomposition by basis pursuit with deep rectifier neural networks and compressed sensing recomposition of speech signals”, in proc: 10th International Conference on Communications, pp.1-6 2014.

4. T. Siddique, U. S. Tiwary, "Natural Language Processing and Information Retrieval", Oxford, 2008.
5. M. Ji, R. Srinivasan, D. Crookes, A. Jafari, "CLOSE-A Data-Driven Approach to Speech Separation", Audio, Speech and Language Processing, IEEE, Vol.21 No.7, pp.1355–1368, 2013.
6. X. Cai, W. Li, "Ranking Through Clustering: An Integrated Approach to Multi-Document Summarization", Audio, Speech and Language Processing, IEEE, Vol.21 No.7, pp. 1424-1433, 2013.
7. R. K. Aggarwal, Mayank Dave, "Implementing a Speech Recognition System Interface for Indian Languages", Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages, pp. 105-112, 2008.
8. B. Raj and Stern, "Missing feature approach in speech recognition", Signal Processing magazine, IEEE, pp. 101-116, 2005.
9. K.S. Rao, B. Yegnanarayana, "Modelling Syllable Duration in Indian Language Using Neural Networks. In proceeding of ICASSP Montreal, Qubic, Canada, pp 313-316, 2004.
10. Baker, L. Deng, J. Glass, S. Khudanpur, C.-H. Lee, N. Morgan, D. O. Shaughnessy, "Research developments and directions in speech recognition and understanding. Part I," IEEE Signal Process. Mag., Vol. 26, no. 3, pp. 75_80, May, 2009.
11. Bilmes, "What HMMs can do," IEICE Trans. Inf. Syst., pp. 869_91, Mar. 2006.
12. J. Stadermann, and G. Rigoll, "A hybrid SVM/HMM acoustic modelling approach to automatic speech recognition," in Proceedings of the Interspeech, Jeju island, Korea, pp. 661-664, 2004.
13. S. Zhang, A. Ragni, and M. Gales, "Structured log linear models for noise robust speech recognition," IEEE Signal Process. Lett, pp. 945-948, 2010.
14. Phuti J Manamela, Madimetja J Manamela, Thipe I Modipa, Tshepisho J Sefara, Tumisho B Mokgonyane, "The Automatic Recognition of Sepedi Speech Emotions based on Machine Learning Algorithms", 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (ICABCD), pp. 1-7, 2018.
15. Osman Eray, Sezai Tokat, Serdar Iplikci, "An Application of Speech Recognition with Support Vector Machines", 2018 6th International Symposium on Digital Forensic and Security (ISDFS), pp. 1-6, 2018.
16. Rubén Solera-Ureña, Ana Isabel García-Moral, Carmen Peláez-Moreno, Manel Martínez-Ramón, Senior Member, Fernando Díaz-de-María, "Real-Time Robust Automatic Speech Recognition Using Compact Support Vector Machines", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, pp. 1347-1361, 2012.
17. Sunanda Mendiratta, Neelam Turk, Dipali Bansal, "Automatic Speech Recognition Using Optimal Selection of Features Based On Hybrid ABC-PSO", 2016 International Conference on Inventive Computation Technologies (ICICT), pp. 1-7, 2016.
18. Balwant A. Sonkamble, D.D. Doye, "Use of Support Vector Machines Through Linear Polynomial (LP) Kernel for Speech Recognition", 2012 International Conference on Advances in Mobile Network, Communication and Its Applications, pp. 46-49, 2012.
19. Xiu-Qing Zhang, Shu-Wang Chen, "Speech recognition system based on DSP and SVM", Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, pp. 2313-2316, 2010.
20. Anurag Bajpai, Umang Varshney, Deepam Dubey, "Performance Enhancement of Automatic Speech Recognition System using Euclidean Distance Comparison and Artificial Neural Network", 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU), pp. 1-5, 2018.
21. Sanjay Valaki, Harikrishna Jethva, "A Hybrid HMM/ANN Approach for Automatic Gujarati Speech Recognition", 2017 International Conference on Innovations in information Embedded and Communication Systems (ICIIECS), pp. 1-5, 2017.
22. Nawel Souissi, Adnane Cherif, "Speech Recognition System Based on Short-term Cepstral Parameters, Feature Reduction Method and Artificial Neural Networks", 2nd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2016, pp. 667-671, 2016.
23. Santosh Gupta, Kishor M. Bhurchandi, Avinash G. Keskar, "An Efficient Noise-Robust Automatic Speech Recognition System using Artificial Neural Networks", International Conference on Communication and Signal Processing, pp. 1873, 1877, 2016.
24. Sunanda Mendiratta, Neelam Turk, Dipali Bansal, "Automatic speech recognition by cuckoo search optimization based artificial neural network classifier", 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI), pp. 29-34, 2015.
25. Sunanda Mendiratta, Neelam Turk, Dipali Bansal, "Isolated word recognition system for speech to text conversion using ANN", IIOAB, pp. 78-91, 2016.
26. Sunanda Mendiratta, Neelam Turk, Dipali Bansal, "Fuzzy based selection of dwt features for automatic speech recognition system for man machine interaction with CS-ANN classifier", IIOAB, pp. 222-240, 2016.
27. Sonia Sunny, David Peter S, K. Poulose Jacob, "Performance of different classifiers in speech recognition", IJRET, pp. 590-597, 2013.
28. M. Kalamani, Dr. S. Valarmathy, S. Anitha, "Automatic Speech Recognition using ELM and KNN Classifiers", International Journal of Innovative Research in Computer and Communication Engineering, pp. 3145-3152, 2015.

AUTHORS PROFILE

Sunanda Mendiratta received the B.Tech degree in Electronics and Communication Engineering from Kurukshetra University, Kurukshetra, Haryana, India in 1999. And her M.Tech degree in the same discipline in 2011 from Manav Rachna International University, Faridabad, NCR, India. She has an experience of 8 years of teaching in various prestigious engineering colleges in Faridabad, NCR, India. She is currently working towards Ph.D. degree in J. C. Bose University of Science and Technology, Faridabad, NCR, India. Her research interests include digital filtering techniques, artificial neural networks and speech processing.

Neelam Turk received B.E degree from North Maharasthtra University Jalgoan, India in 1998. She did her M.Tech. in Electronics and Communication Engg. from National Institute of Technology, Kurukshetra (India) in 2002. She received Ph.D. degree in Electrical Engg. From National Institute of Technology, Kurukshetra (India) in 2011. Currently she is working as Associate Professor in Electronics Engineering Department with the J. C. Bose University of Science and Technology, Faridabad, Haryana, India. Her research interest include MCSA, signal processing, Speech Processing, wireless communication.

Dipali Bansal is a doctorate in Biomedical Instrumentation and Bio signal processing from Jamia Milia University, New Delhi and an upcoming and young scientist. She is a professor in MRIIRS. Her research interests lie primarily in the areas of analysing human physiological signals and developing easy acquisition systems for these bio-signals using PC based systems. Her parallel areas of interest are development of algorithm in MATLAB for digital filtering, deriving HRV and implementing them on Digital Signal Controllers to achieve a compact solution to home health care. She is keen on latest microelectronics technology in developing implanted biomedical devices and other medical products using Smart Sensors and Integrated Microsystems.