International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Classifying Emotional traits from Speech file using Machine Learning

Pooja Nayak S, S G Hiremath, Arun Biradar

Abstract**:** ***With the proliferated usage of speech as a voice command over various devices in existing times, it offers hands-free as well as comfortable experience for operating the current trends of devices. However, the devices are operated just on the basis of recognition of the speech commands and not its context. It is now well known that context factor plays a contributory role in mechanism an artificial intelligence in such devices where a comprehensive analytical modeling can be carried out. In the direction of extraction contextual information from speech, it is first necessary to recognize the speech in it logical form which can be disrupted because of various external causes. Apart from this, even if the recognition is carried out well than the next challenging part will be to extract the emotional factor within the speech. The existing review shows that research-based approaches are not sufficient enough to extract meaning full information which causes a greater degree of impediment in performing classification. Therefore, the proposed system introduces a novel and cost effective framework where machine learning is utilized in a very unique manner for performing this task. The paper discusses about an analytica approach which perform feature extraction followed by applying machine learning to show that proposed system offers faster response time with higher matching found during the testing operation phase.***

Keywords**:** ***Speech, Classification, emotion, Machine Learning, SVM***

I. INTRODUCTION

In the era of the human and machine nitration and artificial intelligence it becomes quite imperative to have an efficient and accurate system where the emotion can be classified from the speech signal inputs. This problem is a problem of type classification where the system is non-linear to takes the variable length input and output is one label out of multiclass labels. The problem of speech classification is quite challenging due to the fact that the amplitude and nodes of expression is different for the different people while expressive in vocal expressions along with the mismatch and inadequacy of manual annotations and lack of information from the linguistic aspects.

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
Pooja Nayak S*, Research Scholar, EWIT, Bangalore, India. Email: pooja.7610@gmail.com
S G Hiremath, Professor and Head, Department of ECE, EWIT, Bangalore, India.
Arun Birader, Professor, CMR University, Bangalore, India.

A typical system of the emotion-speech classification takes the input speech signal from which the feature extractor module extraction the feature the most critical features are selected which are potentially important then the feature based classification takes place by the designed classifier and finally the as an output the label of the type or class of the speech is exhibited. A typical process flow diagram of the speech signal classifier for emotion labeling is shown in the fig.1.

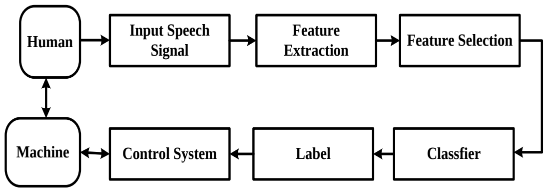


Fig.1 Process flow of speech signal classifier for emotion labelling

What, who and how a speech is said bring a model possibility for the perfect intelligent human-machine interaction-based systems. What gives what has been said and who provides who is the speaker and how explains that what is the emotion of the speaker. Generally, the emotions are classified into categories like:1) happy, 2) sad, 3) surprise, 4) anger and the last of is 5) neutral means normal. The task of human-emotion recognition by the machine directly is not possible so in the era of HMI the classification of the human recognition through the speaker speech-signal input indirectly provides a machine capability to the machine to work on the basis of human emotions. The accuracy and effectiveness of the system depends upon the robustness and strength of the algorithms for the feature extraction, feature selection and model training and testing mechanism. Another dimension of the research challenge is that the usual features consider for the speech recognition is not suitable for the speech-signal emotion classification, therefore new kind of feature and feature extraction methods is the need to take up this problem. Another approach of combining both the image and voice can reduce the computation overhead and many of the literature advocates the use of pre-trained image network along with the SVM. The popular machine learning models includes :1) SVM, 2)ANN, 3)HMM, 4)ELM, 5)GMM, 6)ISLRS, 7)VQ and 8)KNN as described in table II in the section 2.0 and the highest accuracy is found in the cases, where SVM is used. In this paper a forward and backward mel-frequency wrapping and SVM based speech-signal emotion classifier is presented to achieve the higher accuracy in less computational complexities on the basis of energy and sample mapping.

The scope of the research problem solution has wide prospects as the algorithms shall be suitable for the different languages spoken across the globe. Section II discusses about the background of the proposed study followed up by discussion of research problem in Section III. Section IV briefs of the research methodology while the algorithm is discussed in Section V. Section VI discusses about result analysis while section VII briefs of conclusion of study.

II. REVIEW OF LITERATURE

In order to perform an effective classification of emotional factors from the speech signal, it is essential to perform extraction of significant features of a speech. Fig.2 highlights the existing classification of the speech features which are basically of 4 types viz. continuous, qualitative, spectral, and TEO based.

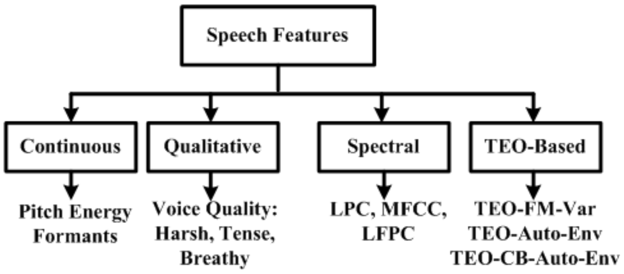


Fig. 2. Categories of speech features

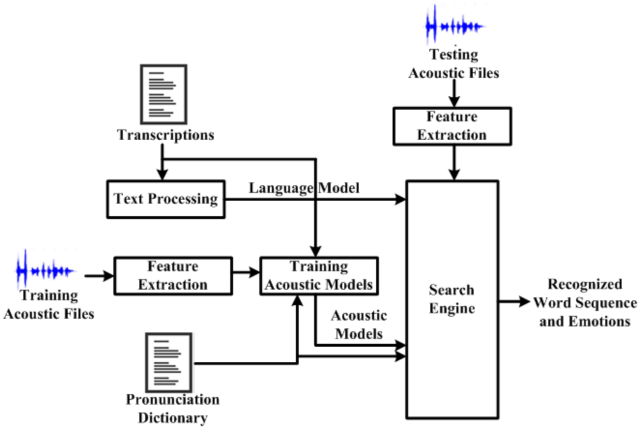


Fig.3. Standard Architecture of emotion recognition

Fig.3 highlights the standard architectures that are frequently used by the existing researchers where the speech files are initially digitized to obtain the transcript file followed by textual processing of it using specific language model. The training acoustic files are further used in parallel for extraction of features and play a significant role in training acoustic model along with the pronunciation dictionary. Finally the test files are used over a search engine to check the effectiveness of recognition system. The effectiveness of the system of the spoken dialogues solely depend s upon the accurate classification of the speech signals. In the work of (Z. Wang and I. Tashev, 2017) proposes a mechanism by adopting the concept of the deep learning for the purpose of encoding the notes of the speech in a vector which maintain an equal size using a mechanism of activation at neural layer at the hidden layer which goes through the classification process. The model was validated on the Mandarin dataset to observe its effectiveness [1]. The uses of machine learning is gaining popularities as the models adopts the non-linearity of the system, in the work of (Yi-Lin Lin and Gang Wei, 2005) the emotional classification problem is modeled using two different apaches of learning namely a) support vector machine and b) hidden Markov model. In these paper methods like sequential forward selection and Mel frequency cepstrum coefficient is suggested [2]. In this continuation, (J. Umamaheswari and A. Akila, 2019), explores the hybrid approach of using pattern recognition network along with KNN is exploited with the use of MFCC and the matrix that exhibits the co-occurrence of the gray pixels for the better classification accuracy [3]. Generally, most of the traditional methods of speech-signal emotional classification problem takes those features that are used for the simple speech recognition, whereas it Is not necessary that it maps correctly in the system. The approach by (L. Cen, W. Ser and Z. L. Yu), explores on the specific features as canonical to the emotion aspects and classify it using neural network and found that only at one third of the feature set provides the same accuracy that ensures that reduction in computational overheads [4]. The exploitation of the both image and speech signal is used to reduce the time complexities and getting higher accuracy and implement into the real-time systems. This kind of approach is being proposed by (M. N. Stolar, M. Lech, R. S. Bolia and M. Skinner, 2017), where they exploited the potential of pre-trained network along with SVM [5]. The survey by (Y. B. Singh and S. Goel, 2018), discusses about the popular databases, and machine learning models to be applicable for the system, the machine leaning models specifies by them are typically listed in the table 1.

Table.1. Machine Leaning Models used for SSECP [6]

ISLRS	ELM	GMM	SVM
Incomplete Sparse Least Square Regression	Extreme learning machine	Gaussian Mixture Models	Support Vector Machine
ANN	**VQ**	**KNN**	**HMM**
Artificial Neural Networks	Vector Quantization	k-Nearest Neighbors	Hidden Markov Model

The research problem of the SSECP is gaining popularity of various language across the globe, The work by (P. J. Manamela, M. J. Manamela, T. I. Modipa, T. J. Sefara and T. B. Mokgonyane, 2018) uses database of one of the African language and uses a tool for the feature extraction and the analysis taken place using a datamining tool WEKA and validate its performance with KNN and SVM [7]. Few very archival journals in this domain are studies and the inference of those papers is listed in the table 2.

III. RESEARCH PROBLEM

The significant research problems associated with the emotion recognition from the speech file are as follows:

- Extraction of accurate feature which are also lightweight in its charecteristics is quite challenging to find. Inappropriate selection of feature will downgrade the performance of recognition.

- Frequently used machine learning approaches are quite iterative in its operation and have dependencies of multiple numbers of files in increasing order creating overhead.
- Existing database are quite challenging to determine the emotion and hence lead to tradeoff with the computational model.

Therefore, all the above research problems act as a motivating factor to carry out further investigation for robust model. The statement of problem is "*Developing a framework with effective feature extraction for facilitating accurate emotional classification system.*"

Table.2. Summary of Existing Approaches

Cite No.	Feature Extraction	Classifier	Remarks
R. Lotfian et. al, 2019 [8]	low-level descriptors (LLDs), Mel-frequency cepstral coefficients (MFCCs	deep neural networks (DNNs)	Datbase:MSP-Podcast corpus
S. Zhang et.al, 2018 [9]	high-level feature, segment-level features, speech feature extraction	Deep Convolutional Neural Networks (DCNN)	Image net dataset, public datasets, speech datasets
A. D. Dileep et. al, 2014 [10]	Gaussian mixture model (CIGMM)	support vector machine (SVM), GMM-based classifiers, DK SVM-based classifiers, LMGMM-based Bayesianclassifier	2002 and 2003 NIST SRE corpora.
R. A. Khalil et. al, 2019 [11]	Energybased features such as Linear Predictor Coefficients (LPC), Mel Energy-spectrum Dynamic Coefficients (MEDC), Mel-Frequency Cepstrum Coefficients (MFCC)	Gaussian Mixture Model (GMM), Hidden Markov Model (HMM)	Emo-DB and SAVEE datasets
T. Sobol-Shikler et.al, 2010 [12]	Vocal feature	Rule-based and decision-tree classifiers.	Mind Reading database
L. Guo et.al, 2019 [13]	spectrogram-based statistical features, empirical features	kernel extreme learning machine (KELM)	Emo-DB and IEMOCAP databases
R. Xia et.al, 2017 [14]	Deep Belief Network (DBN), Supra-segmental features	support vector machine (SVM) classifier	IEMOCAP and SEMAINE
E. Väyrynen et. al, 2013[15]	Prosodic and acoustic features	low-dimensional manifold learning	MediaTeam emotional speech corpus
S. Zhang et al. 2018 [16]	audio-visual segment features	SVM classifier	public audio-visual emotional databases,BAUM-1s database, RML database,
K. Audhkhasi et.ai 2013 [17]	Globally-variant locally-constant Modelgmm	Logistic regression, Naive Bayes, and J48 decision tree	emotional speech classification datasets
K. P. Seng et.al 2018 [18]	prosodic and spectral features	(OKL-RBF) neural classifier	Cohn-Kanade dataset (CK+),
H. Meng et.al, 2014 [19]	video and audio features	Multi classifier k-NN and HMM Classifier	AVEC 2011 video, AVEC 2011 audio, PAINFUL dataset
Y. Zong et.al, 2018 [20]	micro-expression feature extraction	SVM classifier	SMIC (HS),SMIC (VIS),SMIC (NIR),
L. Mion et.al, 2008 [21]	PCA	minimum distance classifier	Database of recordings.

IV. RESEARCH METHODOLOGY

The implementation of the proposed study has been carried out using analytical research methodology (Fig.4). For this purpose, the proposed system captured the speech from subjects with different state of emotion. The data were captured from the different subjects where the subject performs expression of phrases that are significant of different forms of mood. All the subjects are basically English speaking while they utter the terms and phrases using different states of their emotion.

The proposed system make use of power spectrum coefficient that are signified by very short term and characterized by the cosine transformation of linear form performed over the spectrum of log power over a frequency of speech with non-linear attributes. This approach leads the generation of the coefficient from the speech sample that further permits for effective representation of the speech. Following steps were performed for this purpose viz. the input signal is subjected to the Fast Fourier Transform followed by mapping of the spectrum power where the overlapping windows can be used in triangulated form. The next option is to obtain the logarithmic value of the power of the signal for the given frequencies. All the unique cosine transform are obtained from the set of the logarithmic power considering it to be a signal. Finally the coefficients of the speech are obtained from the yielding spectrums. This principle is simpler version of extracting features from the speech files and thereby it assists in performing classification of the speech file effectively.

This method also performs extraction of the speech file and performs representation of the features of the speech signals. The method considers the spectral with shorter time period that has essential information about the voice quality. For the purpose of computing the coefficient, it is required that cosine transform should be calculated with respect to the real value of the logarithmic followed by performing the signal over the respective scale of frequency. The approach then performs pre-emphasis of the segments of the speech file which are further subjected to the windowing process. The approach also uses hamming window that can further address the problems of leakage effect. The next part of the implementation is about smearing the energy from the correct frequency of the signal into the adjacent one. The technique also assists in resisting any form of break over the speech signal considering the temporal domain and this possibility is always there if fast Fourier transformation is applied. The process of the feature extraction has dependencies on various essential attributes e.g. shift of the frame, duration of frame, coefficient of emphasis, number of the channels in the filter bank, limits of the lower and upper frequency, size of sample frames, etc. The next apart of the implementation is about applying supervised learning approach towards the processed signal obtained in the prior steps. However, prior to performing machine learning approach, it is necessary to perform labeling of the training data. Adoption of this approach assists in significant better form as there is a discrete mechanism of differentiating various classes of moods. After extraction, the number of the speech fractional components increases and creates more high dimensional spaces and therefore, this machine learning approach is one of the suitable approach for performing classification of emotions for a given speech.

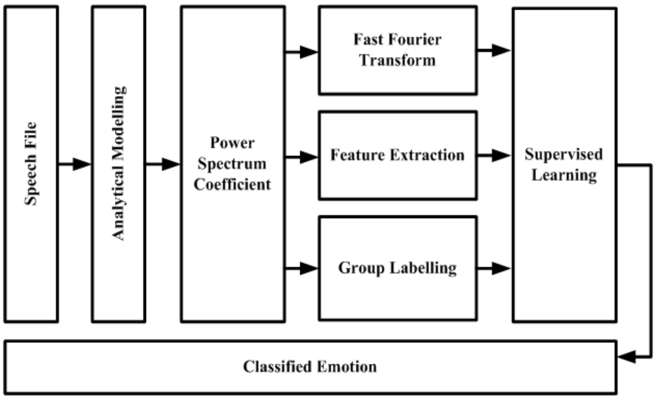


Fig.4. Proposed Research Methodology

V. ALGORITHM IMPLEMENTATION

The prime purpose of the proposed algorithm is to carry out the classification of the genre of the digitized speech signal using machine learning language. The algorithm takes the input of n (number of speech files) which after processing yields an outcome of CV (classification vector). The significant steps of the algorithm implementation are as follows:

Algorithm for speech classification
Input: n (number of speech files)
Output: cv (classification vector)

Start
1. **For** i=1:n
2. $[y_1\ F_s] \rightarrow f_1(n)$
3. frac $\rightarrow (L-w_{len})/s$
4. **For** i=1:frac
5. w $\rightarrow$ H(s) //53
6. y $\rightarrow f_2(w)$
7. FC $\rightarrow$ y
8. $\alpha \rightarrow f_3(\phi)$
9. end
10. cv $\rightarrow$ apply $\theta(\alpha)$
11. **End**

End

The discussions of the algorithmic steps are as follows: After taking the input of the speech file, the algorithm reads the complete speech file and extracts a test sample from it (Line-2). The function $f_1(x)$ extracts a definitive number of sample over its given length. The algorithm also initializes number of windows as well as samples of window. The algorithm then normalizes the signal by dividing it by the maximum score of absolute value of signal obtained followed by extracting the length of the signal. The next part of the implementation is about computing the fractional component *frac* which is computed by difference of length of signal *L* and length of window w_{len} and the difference is divided by steps, where steps represents number of window samples. This operation is followed applying hamming operation H towards the length of the window. Considering all the fractional components of the signal frac (Line-4), the algorithm then computes the empirical value of the processed window (Line-5) where hamming window H is used along with signal *s* (Line-5). The next part of the implementation is about applying Fast Fourier transform operation over the window using an explicit function $f_2(x)$ (Line-6) in order to obtain the processed signal *y*. The algorithm than proceeds towards extracting both first and second order derivative of processed signal y (Line-7) which is further followed up by extracting statistical parameter of mean, standard deviation, median, minimum, and maximum over the signal processed signal *y*. The algorithm make use of the power spectrum factor connected with the speech file that is of short term, while the idea is completely based upon cosine transform of linear form of the frequency. A function $f_3(x)$ is constructed for this purpose (Line-8) that extracts the coefficients extracted from the speech signal with an aid of FIR filter of first order. This is basically a form of speech signal of reemphasized form is now proceeded for using Fourier transform of short time considering duration of frame and the shift of the frame as well as window function for performing analysis. In the algorithm step-8, the variable ϕ represents a set of signal, sampling rate, duration of frame, pre-emphasized coefficient, lower and higher limits of frequency, quantity of filter bank channels, and sine lifter attribute. All the obtained features have been summed up and absolute value is obtained. This operation gives rise to generation of the dataset feature. This completes the above operation while the next phase of the implementation is about performing classification considering the test dataset.

For this purpose, a training set data is constructed considering the data feature followed up by indexing the groups. All the unique group indexes are obtained as well as length of number of classes are obtained. For all the number of classes, the proposed system checks of the group of the trained signals are equivalent to the unique groups. Only the information matching with this condition is selected to be subjected to the new function θ (Line-10). This new function is responsible for applying supervised learning algorithm in order to generate a training model. Fig.5 highlights the complete process flow diagram to exhibit that the proposed system is high progressive inspite of using machine learning approach and there is no complex iterative forms in the entire process.

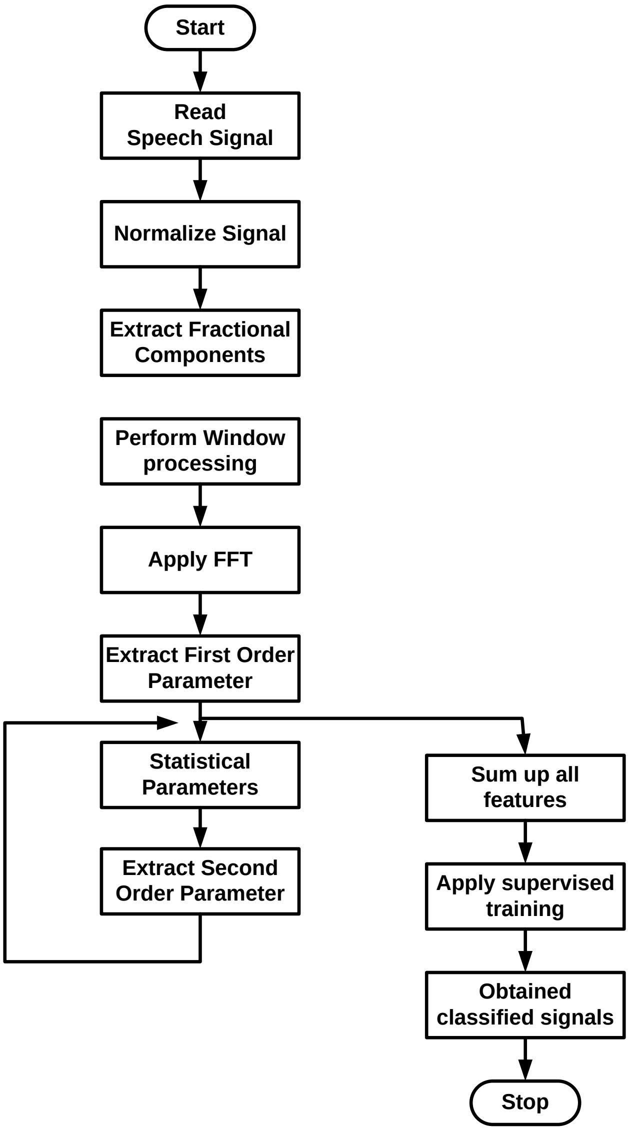
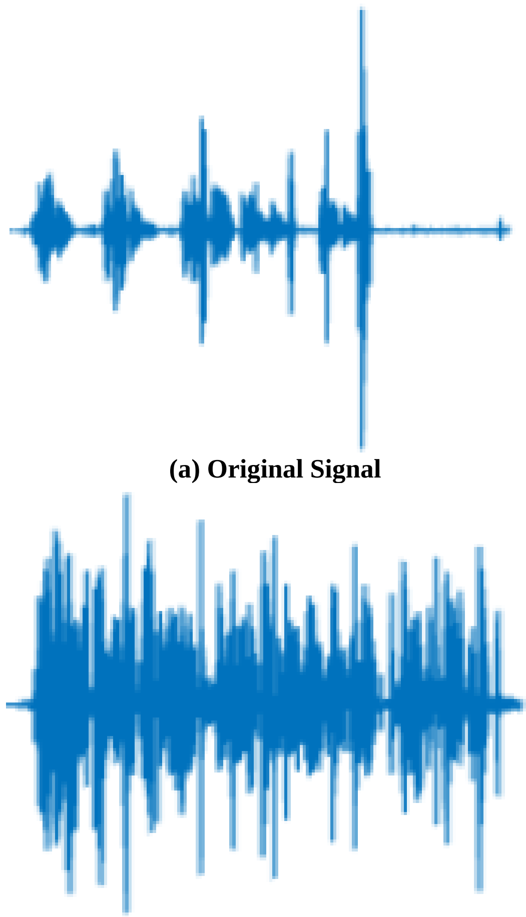


Fig.5. Steps of Algorithm Implementation

VI. RESULTS AND ANALYSIS

The proposed system is scripted in MATLAB where the speech samples are taken from analog signal that are later digitized. The samples are then subjected to normalization process to ensure that all the signals are of short duration and have similar range of volume. All the samples range approximately to length of 1s and it is distinguished from another speech sample using 20-30s. Normally, the properties of the speech signals alter very slowly with respect to time and this property basically assists in investigating the window of shorter time for the speech signal. Normally, the sample of the original speech file is enough for extracting the features of speech from emotional viewpoint over shorter window of time. The proposed system considers supervised machine learning algorithm i.e. Support Vector Machine to perform training. During training, the proposed system uses 80% of dataset for training while 10% of the dataset is used for performing validation while the next 10% of the dataset is used for performing testing. The system model for the speech-signal based emotion classifier is tested with the x different types of the speech-signal notes, each corresponding to a class of the speech-signal emotion on the basis of the score computation. Table 3 shows the different speech-signal pattern



(a) Original Signal

(b) Processed Signal

Fig.6. Analysis of Classified Boring Emotion

(a) Original Signal

(b) Processed Signal
Fig.7. Analysis of Classified Cool Emotion

Fig.8. Analysis of Classified Happy Emotion

The proposed system performs classification of 6 types of emotional factor from the speech viz. i) bore, ii) calm, iii) cool, iv) sad, v) hot, and vi) happy.

Table.3. Algorithm Processing Time for Classification

Emotion	Processing Time
Bore	0.455
Calm	0.487
Cool	0.548
Sad	0.478
Hot	0.496
happy	0.501

The algorithm processing time is highlighted in Table 3.The ambiguity is very large in the one pattern itself, where one speaker can express their emotion in one way , whereas another speaker can express in the different ways. This can be noticed in the both the cases of boring and cool, where the pattern is different in two different cases. A closer look into the numerical outcomes of classification time shows that there is no significant deviation or fluctuation of higher order in the algorithm processing time

VII. CONCLUSION

This paper has presented a discussion of the framework that is capable of performing emotional classification for a given set of speech files. The contribution of the proposed system are as following: i) the proposed system is found to offer 98% of approximated accuracy when a comprehensive testing and validation is carried out, ii) the response time for algorithm to perform classification is almost instantaneous, iii) the complete process of performing extraction of speech feature is highly progressive and less iterative which can offer better evidence towards cost effectiveness.

REFERENCES

1. Z. Wang and I. Tashev, "Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 5150-5154.
doi: 10.1109/ICASSP.2017.7953138
2. Yi-Lin Lin and Gang Wei, "Speech emotion recognition based on HMM and SVM," *2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 2005, pp. 4898-4901 Vol. 8.
doi: 10.1109/ICMLC.2005.1527805
3. J. Umamaheswari and A. Akila, "An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN," *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, 2019, pp. 177-183.
doi: 10.1109/COMITCon.2019.8862221
4. L. Cen, W. Ser and Z. L. Yu, "Speech Emotion Recognition Using Canonical Correlation Analysis and Probabilistic Neural Network," *2008 Seventh International Conference on Machine Learning and Applications*, San Diego, CA, 2008, pp. 859-862. doi: 10.1109/ICMLA.2008.85
5. M. N. Stolar, M. Lech, R. S. Bolia and M. Skinner, "Real time speech emotion recognition using RGB image classification and transfer learning," *2017 11th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Gold Coast, QLD, 2017, pp. 1-8.
doi: 10.1109/ICSPCS.2017.8270472
6. Y. B. Singh and S. Goel, "Survey on Human Emotion Recognition: Speech Database, Features and Classification," *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida (UP), India, 2018, pp. 298-301.
doi: 10.1109/ICACCCN.2018.8748379
7. P. J. Manamela, M. J. Manamela, T. I. Modipa, T. J. Sefara and T. 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)*, Durban, 2018, pp. 1-7. doi: 10.1109/ICABCD.2018.8465403
8. R. Lotfian and C. Busso, "Curriculum Learning for Speech Emotion Recognition From Crowdsourced Labels," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 815-826, April 2019.
doi: 10.1109/TASLP.2019.2898816
9. S. Zhang, S. Zhang, T. Huang and W. Gao, "Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching," in *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1576-1590, June 2018.
10. A. D. Dileep and C. C. Sekhar, "GMM-Based Intermediate Matching Kernel for Classification of Varying Length Patterns of Long Duration Speech Using Support Vector Machines," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 8, pp. 1421-1432, Aug. 2014.
doi: 10.1109/TNNLS.2013.2293512
11. R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," in *IEEE Access*, vol. 7, pp. 117327-117345, 2019.
doi: 10.1109/ACCESS.2019.2936124
12. T. Sobol-Shikler and P. Robinson, "Classification of Complex Information: Inference of Co-Occurring Affective States from Their Expressions in Speech," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1284-1297, July 2010. doi: 10.1109/TPAMI.2009.107
13. L. Guo, L. Wang, J. Dang, Z. Liu and H. Guan, "Exploration of Complementary Features for Speech Emotion Recognition Based on Kernel Extreme Learning Machine," in *IEEE Access*, vol. 7, pp. 75798-75809, 2019.
doi: 10.1109/ACCESS.2019.2921390
14. R. Xia and Y. Liu, "A Multi-Task Learning Framework for Emotion Recognition Using 2D Continuous Space," in *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 3-14, 1 Jan.-March 2017.

15. E. Väyrynen, J. Kortelainen and T. Seppänen, "Classifier-based learning of nonlinear feature manifold for visualization of emotional speech prosody," in *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 47-56, Jan.-March 2013.
16. S. Zhang, S. Zhang, T. Huang, W. Gao and Q. Tian, "Learning Affective Features With a Hybrid Deep Model for Audio–Visual Emotion Recognition," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 3030-3043, Oct. 2018.
17. K. Audhkhasi and S. Narayanan, "A Globally-Variant Locally-Constant Model for Fusion of Labels from Multiple Diverse Experts without Using Reference Labels," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 769-783, April 2013.
18. K. P. Seng, L. Ang and C. S. Ooi, "A Combined Rule-Based & Machine Learning Audio-Visual Emotion Recognition Approach," in *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 3-13, 1 Jan.-March 2018.
19. H. Meng and N. Bianchi-Berthouze, "Affective State Level Recognition in Naturalistic Facial and Vocal Expressions," in *IEEE Transactions on Cybernetics*, vol. 44, no. 3, pp. 315-328, March 2014.
20. Y. Zong, W. Zheng, X. Huang, J. Shi, Z. Cui and G. Zhao, "Domain Regeneration for Cross-Database Micro-Expression Recognition," in *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2484-2498, May 2018.
21. L. Mion and G. De Poli, "Score-Independent Audio Features for Description of Music Expression," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 458-466, Feb. 2008.

AUTHORS PROFILE

Pooja Nayak S, She has completed her BE in computer science and engineering in the year 2009 , MTech in Computer networks engineering in the year 2011, worked as assistant professor in SSIT, Tumkur from 2012-14, assistant professor at EWIT, Bangalore from 2014 to 2017, currently research scholar under VTU

Dr. S G Hiremath, is a highly skillful professor with over 20 years of experience in the field of engineering education. He has been awarded Ph D from Anna University in the faculty of ECE. His interest in research areas of Instrumentation, Digital Signal Processing, Neural Networks, Control Engineering, Hardware Design, Electronics Circuit Design, Instrumentation Sensors, Process Controls, PLC s and SCADA, Modeling & Simulation, Neural Networks, Telecom and related Teaching and promotional areas. He is awarded doctorate for his Research in the areas of Bio Medical Signal processing, Electronic circuit design, Chemical sensors, and Neural Network modeling/Simulation. He has more than 70 research publications at peer reviewed international journals and conferences.Five research scholars are doing research under his guidance and four of his scholars were awarded with Ph D degree from visveswaraya technological university Belagavi for the contributed works in the field of signal processing and Neural Network modeling. He is one of the main resource persons in the fields of Electronics, Especially in domains like Medical Signal Processing, Embedded System Design, Mathematical modeling and simulation, etc. He has given keynote address in various international and national conferences all over India. He has organized various workshops, Seminars and Conferences in the field of Signal processing and System Designing. At present he is working as Professor in the department of Electronics and Communication Engineering in East West Institute of Technology, Bengaluru, India. He is also consultant for companies in Bengaluru, Pune, Hyderabad, Chennai etc. His project consultancies includes in the field of Signal Processing, Neural Networks and Embedded Systems Designing.

Dr. Arun Biradar is an academician having more than 22 years of experience in teaching, development, administration and leadership. He completed B.E. in Computer science and Engineering from Gulbarga University and M.Tech in Software Engineering from Mysore University and Ph.D. in Computer Science Engineering from Swami Ramanand Teerth Marathwada University, Nanded. His research interests are Wireless Ad-hoc Networks, Computer Networks, Software Engineering, Genetic Algorithms, Machine Learning, IoT and Cloud Computing. He is a recognised guide under Visvesvaraya Technological University, Belagavi, and has many students pursuing their Ph.D. under his guidance. He is a member of various professional bodies and has been a part of the NAAC and NBA Advisory Committee to many Colleges. He is an active member of IEEE