Speech Emotion Recognition using Cross Correlational Database with Feature Fusion Methodology

G. Chandrika Sri Lakshmi, K. Sri Sundeep, G. Yaswanth, Niranjan S R Mellacheruvu, Swarna Kuchibhotla, Venkata Naresh Mandhala

Abstract*: Speech emotion recognition provides an interface for communication between the human and the machine. Classifying the emotion based on the speech signals is not that easy task since we need to take into account the conditions like noisy data, changes in voice due to cold/cough and so on because the voice of a person will not be the same when he/she is suffering from cold/cough, or when he/she consumed alcohol. In this paper we just extracted some of the features like Volume, Energy, MFCC, and Pitch in order to classify the emotion into happy/sad/anger/neutral. In this paper MFCC plays a major role for classifying the emotions into happy/anger/sad/neutral. The concept of Cross-Correlation is that we first make use of Berlin Database and train the model using Berlin database and then we will test the same model using Spanish Database. The main role is that to test whether the model taken produces the same output (emotion) for both the Spanish and Berlin Databases that is we need to prove that the model taken is independent of the language used. Accordingly, a function is developed in MATLAB for Identification of an Emotion for any Audio File given as an input [1].

Index Terms*: MFCC (Mel Frequency Cepstral coefficient), SVM (Support Vector Machine), RDA (Regularized Discriminant Analysis), LDA (Linear Discriminant Analysis), kNN (k Nearest Neighbor).

I. INTRODUCTION

Since, we are in a modern era the development of different types of gadgets is increasing at a higher rate. With this development of modern technology humans are becoming lazy to get things done by their own. Since, a lot of gadgets are being developed there need to be some technology which attracts humans to buy them. Speech emotion recognition is such a technology with which we can directly communicate with the gadgets in order to get our needs to be done without any physical effort. Speech emotion recognition made our life much easier and simpler. Human Speech Emotion is used in some of the applications like intelligent robots, smart household appliances like TV, Washing Machine, Air conditioning, Microwave Oven etc., Self-Driving cars and many more applications. A lot of work was carried out in this domain in order to perceive the emotions from the human speech but Speech emotion recognition is not that easy because it need to discriminate between the feelings happening naturally and the feelings happening artificially that is feelings produced via prepared on-screen actors. The degree of closeness/correlation is high between the emotions like happy, anger, sad and neutral. Identifying these emotions is not that easy because the features of such emotions don't have a high degree of uniformity. The survey of the existing work covers that a large portion of the present work depends on Mel Frequency Cepstral Coefficients (MFCC's) as an essential component for Speech Emotion Recognition. In [2], the feature that is used for classifying the speech data into various emotion categories is MFCC. The emotions are classified by employing Artificial Neural Networks. Another work in [3] uses utterance as the key parameter for lexical investigation of the expressed content. This methodology, however profoundly solid and exact, is language subordinate and can't be utilized to model a general Speech Emotion Recognition framework. In an altered methodology, just 2 highlights, to be specific MFCC and Pitch have been utilized with the end goal of grouping of feelings into different classifications [4]. This system figures out how to set up a decent harmony between computational volume and execution precision of the ongoing procedures, yet neglects to give great outcomes in uncommon circumstances, similar to when the individual is excessively irate or upbeat, prompting fast variances in Pitch and MFCC values. A moderately old work here processes measurable estimations of parameters, for example, mean, standard deviation, extend, most extreme esteem, least esteem and middle of Pitch and Energy to frame an 11 dimensional element vector [5]. This is a truly established methodology of utilizing fundamental insights into the zone, and has an exactness of 84% for the irate and unbiased states and 91% for the cheerful state. This exactness is great and appropriate for must purposes, however just when the bimodal framework classifier is utilized as the outward appearances and acoustic data are consolidated at the element level. Now, in the coming sections we mainly concentrate on the Speech Databases, Pre-Processing of the signal, Feature Extraction, Classification, Results and Conclusion.

Revised Manuscript Received on April 25, 2019.

G. Chandrika Sri Lakshmi, K. Sri Sundeep, G. Yaswanth, B.Tech , Dept.of CSE, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur, AP, India..

Niranjan S R Mellacheruvu, Asst Professor, Dept of ECE, Vikas College of Engineering and Technology, Vijayawada rural, Nunna, AP, India.

Swarna Kuchibhotla, Venkata Naresh Mandhala, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur, AP, India.

II. SPEECH DATABASES

We have considered two databases namely Berlin and Spanish emotional speech databases for emotion classification. The speech samples in these both databases are given by different people which means that the databases are multi-speaker databases, so it provides a possibility to perform speaker independent tests. These databases contain the following emotions like anger, boredom, disgust, fear, happiness, sadness and neutral. Berlin database contains 500 speech samples and are simulated by ten professional native German actors, five male and five female [13]. The number of speech files are shown in the Table 1.

Table. I Describing the number of files taken for each emotion in both Berlin and Spanish Database

Emotions	**No. of files in database**	
	Berlin	**Spanish**
Anger	124	183
Bored	83	183
Happy	70	183
Fear	64	183
Sad	61	183
Disgust	48	183
Neutral	78	183
Total	**528**	**1281**

For each emotion in Spanish database it contains 183 sentences which includes numbers, words, sentences etc. It takes recordings from two professional actors, that is one male and the other is female. Among these 183 files, 1-100 comes under Affirmative sentences, 101-134 comes under Interrogative sentences, 135-150 are paragraphs, 151-160 are Digits, 161-183 comes under Isolated words.

III. METHODOLOGY

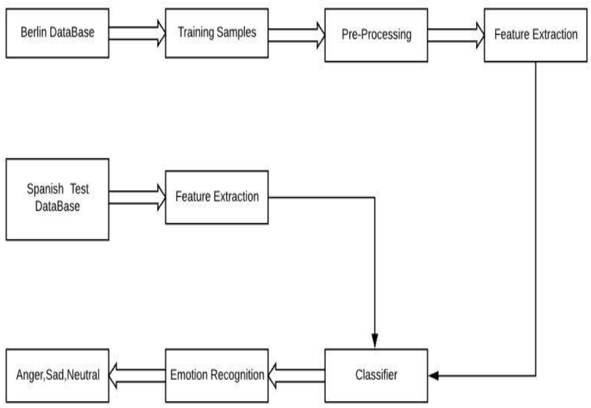

Fig. 1 Block diagram representing the process/ methodology for speech emotion recognition process.

The block diagram of the overall process of the speech emotion recognition is as shown in Fig1. The following is the procedure to consider. We have taken the Berlin Speech Database. We then taken the samples of the Berlin Database and used them for Training process. For each speech sample we need to perform Pre-Processing which includes the following methods like Silence Removal, Pre-Emphasis, Normalization, Windowing and then we can obtain the processed signal with which we need to perform further process. With this processed speech signal we need to extract features which is known as Feature Extraction. Some of the features are energy, pitch comes under prosody features and MFCC comes under spectral features. The extracted features are fed to a classifier model with which we can recognize the emotion which is known as Emotion Recognition. The output can be any of the emotions like Anger, Sad or Neutral. This ends the training process. Now, the testing process is as follows: First we need to take the Spanish Speech database for testing purpose. Now, we need to extract features form the Spanish Database which comes under Feature Extraction. The extracted features are fed to a Classifier from which we can get the outcome as any emotion.

A. *Pre-Processing:*

In order to get a pure signal which needs to be used in the next stage (Feature Extraction) the following pre-processing methods are used silence removal, pre-emphasis, normalization and windowing. A classic speech signal consists of two main parts: The first one carries the information about the speech and the second one consists of silent or noise sections. The verbal (informative) part of speech is again divided into three types:

(a) The voice speech
(b) unvoiced speech
(c) silence.

Voiced speech tend to be louder like the vowels /a/, /e/, /i/, /u/, /o/. Voiced speech comprises for the most part of vowel sound. It is created by driving air through the Glottis, proper alteration of the strain of the vocal lines brings about opening and shutting of the lines, and a generation of practically intermittent beats of air as shown in Fig 2. These heartbeats energize the vocal tract. Psychoacoustics tests demonstrate that this part holds the most important part of the data of the speech and along these lines holds the keys for describing a speaker. Another definition is if the sound is made by the vocal lines or via air being constrained through the mouth depression with the tongue and lips molded to highlight a tone range.

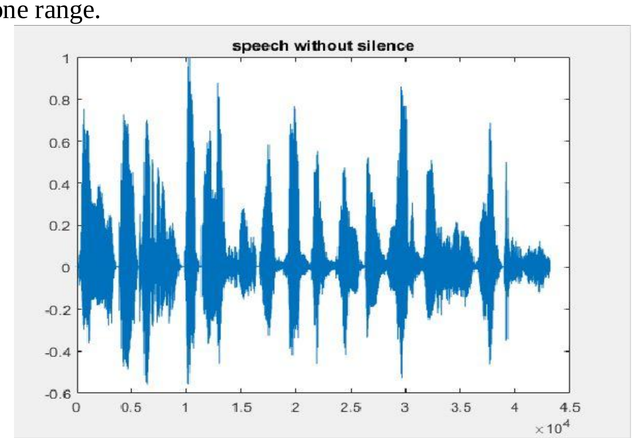

Fig. 2 Analyzing the speech signal without silence/ voiced speech signal.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

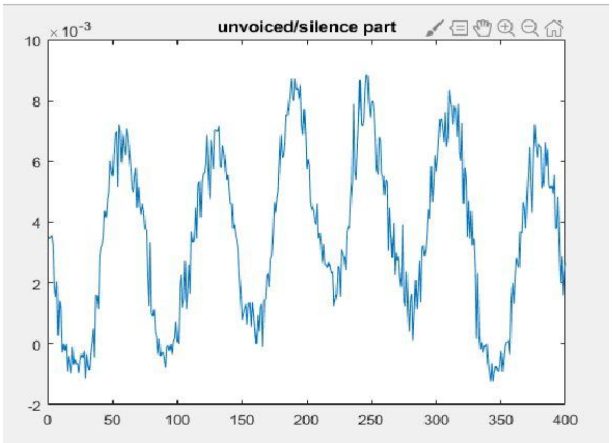

Fig. 3 Analyzing the unvoiced speech signal/ silence part of the speech signal

Unvoiced speech is non-intermittent, irregular like sounds, brought about via air going through a limited choking of the vocal tract as when consonants are spoken as shown in Fig 3. The last classification is Silence, when there is no vibration of the vocal ropes after the air is released from the lungs. Because of the encounters and perceptions experienced by man throughout the hundreds of years it turned out to be simple for him to recognize emotions, for example, when an individual is angry, his tone raises, and his appearance ends up stern. In the meantime when an individual is happy, he talks in a melodic tone, subsequently there is a look of merriment all over and the substance of his discourse is fairly wonderful. In view of these perceptions, an individual can rapidly recognize the condition of the speaker whether he is happy, sad, energy or others states. Preprocessing methods mainly includes.

Silence Removal: The speech signal generally incorporate numerous pieces of silence. The silence signal isn't vital in light of the fact that it doesn’t contain data. There are a few strategies to expel these parts, for example, Zero Crossing Rate (ZCR) and Short Time Energy (STE). Zero crossing rate is a proportion of number of times in a given time the amplitude of the speech signals goes through an estimation of zero. Short time energy is defined as a measure of energy of signal in time interval.

Pre-Emphasis: The pre-emphasis of the speech signal is the most essential strides of preprocessing at high frequency. It's utilized to get practically identical amplitude for all formant. To satisfy the task, the speech signal is gone through a high-pass filter (FIR).

Normalization: It is a procedure for adjusting the volume of sound to a standard dimension. Normalization uses the signal sequence divided by highest value of the signal to make sure that each sentence has a comparable volume level.

Windowing: The most prominent window is Hamming window. It has impact of smoothing the edges and lessening the impact of side projection. The window time is typically 25ms and covers each 10ms.

The taken speech signal will be first divided into frames, and then for each frame the features will be extracted as shown in the Fig 4.

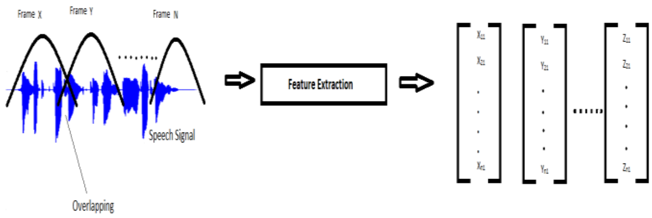

Fig. 4 Process of dividing the speech signal into frames and extracting features for each frame

B. *Feature Extraction*

Once, Pre-processing is done the next step is Feature Extraction where we need to extract some of the features like Pitch, Volume, MFCC (Mel Frequency Cepstral Coefficient) from the speech signal. Since, we are talking about Feature Extraction we need to know about the types of features that are present. There are two types of features namely Prosody features and Spectral features.

1. Prosody Features: Prosody features are also known as suprasegmental phonology. Suprasegmental means that meaning an element of an expression other than the consonantal and vocalic parts, for instance stress and intonation. This is a collective term used to describe variations in Pitch and volume levels. In this paper we are dealing with some of the prosodic features like Pitch and Energy.

1.1 (a)Pitch: The meaning can differ based on variations in pitch levels. The raising of pitch level and the falling of pitch level indicates some of the feelings like astonishment and boredom. When a person is astonished his/her pitch level may raise and when a person is bored his/her pitch levels will be low. It indicates that as the pitch level changes the meaning also changes.

1.2(b)Volume: Volume is the main indicator for indicating some of the emotions like anger, sad etc. If a person is angry there will be a raise in the volume of their voice. Unknowingly, person voice or volume will be raised when he/she is anger. If a person is sad his/her voice will be very low that is the volume of their speech will be low. Since, the person is sad he/she will be out of the mood and so the volume of their speech will be very low which indicates that he/she is sad.

2. Spectral Features: The spectral features are frequency based features which uses Fourier Transform to convert the signal which is in time domain to frequency domain. In this paper, we used MFCC (Mel Frequency Cepstral Coefficient) as spectral feature.

or the prosody features (Volume and Pitch) their first and second order differentiation are taken which provides useful information and that information is also considered (Luengo etal. 2005), [13]. Energy and pitch were estimated for each frame together with their first and second derivatives, providing six features per frame and applying statistics like Mean, Variance, Minimum, Range, Skewness, Kurtosis giving a total of 36 prosodic features [13] MFCC features represents the short term power spectrum of a speech signal, based on a linear cosine transform of a log power spectrum on a nonlinear melscale of frequency[13]. The procedure for implementing MFCC is shown in Sato and Obuchi(2007), Vankayalapati and SVKK Anne (Vankayalapati and SVKK Anne), [13]. Similar to prosody statistics, spectral statistics are calculated using the statistics as shown like Mean, Variance, Minimum, Range, Skewness, Kurtosis. The extracted MFCC are 18 and the number of filter banks used are 24 [13]. Eighteen MFCC Coefficients and their first and second derivatives are estimated for each frame giving a total of 54 spectral features [13]. The statistics as mentioned earlier are applied to these 54 values so totally 54×6 = 324 different features are calculated [13].

Using, these Prosody and Spectral features independently we cannot get more accurate results. So, here in this paper we are introducing the concept of Feature Fusion which means combining both the Prosody and Spectral features together. Using Feature Fusion we can get more accuracy when compared to using features independently.

In this project we made use of Berlin and Spanish databases for classifying the emotions of the wav files. If we use Prosody features or Spectral features independently the accuracy of classifying the emotion rate is less when compared with Feature Fusion method. Feature Extraction plays an important role because it is one of the method from where we can draw some conclusions about classifying emotions into happy, sad, anger or neutral. As said earlier based on Volume we can say whether the person is anger or sad. Based on Pitch we can say that whether the person is astonished or bored. So, Feature Extraction plays a major role in Speech emotion recognition

C. *Classification*

After the Feature Extraction stage the features are given as input to the classifier which then classifies the emotion of the speech signal. The main role of the classifier is to classify the emotions into happy, sad, neutral and anger. We have taken two databases namely Berlin and Spanish database out of which we used two-third of the samples for training the classifier and the remaining one-third of the speech samples for testing purpose. The training data samples are used for preparing the classifier and the test data samples are used for testing the classifier accuracy of correctly classifying the emotions into happy, sad, neutral, anger. Some of the classifiers are LDA (Linear Discriminant Analysis), SVM (Support Vector Machine), RDA (Regularized Discriminant Analysis) and kNN (k Nearest Neighbor). Out of these classifiers RDA produces more accuracy compared with others. In this paper we made use of RDA classifier for classifying the emotions into their respective classes.

LDA (Linear Discriminant Analysis):

LDA is used to separate two or more classes of objects. In this paper we have taken many wav files which belongs to different classes like happy, anger, sad and neutral. LDA task is to separate the wav files to their respective classes. The wav files of similar emotions comes under the same class. LDA suffers from singularity issue when dealing with high dimensional and low sample size speech data [13]. The singularity issue is the main drawback of LDA and so we need to go for another classifier

SVM (Support Vector Machine):

SVM is a classifier that deals with separation of classes with a suitable hyperplane. A hyperplane is used to separate two or more classes. A hyperplane is used to divide the classes that means that all the similar data belonging to one class falls one side of the hyperplane and the data that belongs to the another class falls on the other side of the hyperplane. Out of many possible hyperplanes SVM selects the optimistic hyperplane for separating the class

RDA (Regularized Discriminant Analysis):

RDA is same as LDA but the major difference is that LDA suffers from singularity problem whereas RDA doesn't suffer from singularity problem. RDA is one of the classifiers which produces accurate results when compared with the other classifiers. In this paper we made use of RDA classifier for classifying emotions.

KNN (K Nearest Neighbor):

KNN is a classification algorithm which determines what group the data point belongs to. It compares the data point with the data groups around it. It calculates the Euclidean distance from the present data point to all other data groups around it. The data point is then belongs to the group for which the calculated Euclidean distance is less when compared with the other group. Suppose that there are set of points belonging to one group and the set of points belonging to the another group. There is a new data point which belongs to one of the groups. KNN is then used to determine to which group the data point belongs to based on the Euclidean distance between the points in the group and to the data point. The data point is then belongs to the group for which the measured Euclidean distance is less.

IV. EXPERIMENTAL RESULTS

To evaluate the outcomes of different classifiers like SVM, LDA, RDA, KNN on different emotions like happy, neutral, anger and sad emotional classes, we considered two phases namely Baseline results and Feature fusion through which recognition tests were carried out.[13]

A. *Baseline Results:*

As shown in the Table 2 in the first phase using prosody and spectral features of both the Berlin and Spanish databases the classifiers have been trained and their accuracies are as shown in Table 2. Examining Table 2, it indicates that spectral features provides higher accuracy % when compared to prosody features. The accuracy % is in the scope of 41-69% for both Berlin and Spanish databases. Spectral features have got an accuracy of 51-70% for all the classifiers, except for RDA it got an accuracy of 79.75%. Feature Fusion method is used in the next stage for better accuracy purpose.

Table. II Percentage accuracy of emotion recognition of various classifiers (LDA, RDA, SVM, KNN) over both the Berlin and Spanish databases taken over Prosody and Spectral features.

Classifiers	Databases			
	Berlin		Spanish	
	Prosody (%)	Spectral (%)	Prosody (%)	Spectral(%)
LDA	43.2	52.50	41.3	51.75
RDA	68.50	79.75	45.0	62.50
SVM	55.50	69.75	62.25	64.75
KNN	56.50	60.75	59.75	68.50

Table. III Emotion Recognition accuracy of various classifiers (LDA, RDA, SVM, KNN) over both the databases Berlin and Spanish using Feature fusion method

Classifiers	Databases	
	Berlin (%)	Spanish (%)
LDA	62	60
RDA	80.7	74
SVM	75.5	73
KNN	72	71.5

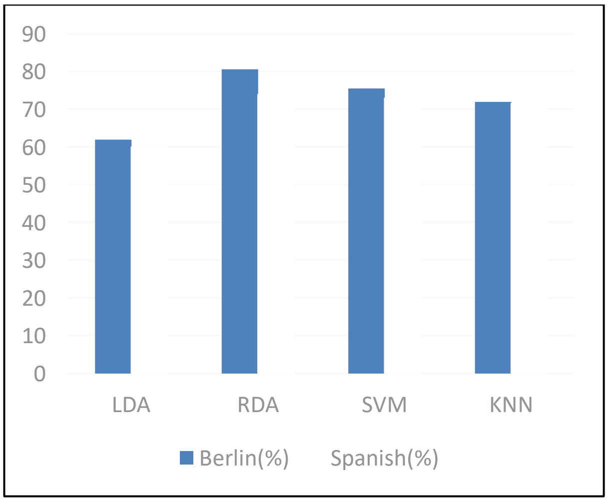

Fig. 5 Accuracy of various classifiers over both the databases using Feature Fusion method

B. *Feature Fusion Results*

Feature Fusion is done by combining both the prosody features and the spectral features. Feature Fusion is done to improve the accuracy of the classifiers when compared with baseline results as appeared in Table 3. When compared with the other classifiers the accuracy percentage is more for RDA and SVM. The overall performance of these classifiers for both the databases is shown in Fig 6. Horizontal axis deals with the name of the classifier whereas the vertical axis deals with the precision rate. The accuracy of each classifier is improved by 21% roughly for each classifier when compared with standard results.

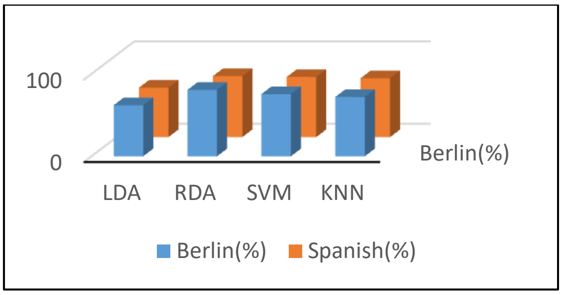

Fig. 6 Comparison of emotion recognition accuracy using different classifiers for both Berlin and Spanish Database

C. *Analysis of results for each emotion*

Table IV describes about the confusion matrix of RDA classifier. By analyzing the results we can conclude that the emotions happy is combined with anger and the emotion neutral is combined with sad. Prosodic features can segregate the feelings (happy, anger) from high arousal space to feelings (neutral, sad) from low arousal space. However, confusion matrix exists among the feelings in a similar arousal state.

Table. IV.i Percentage accuracy of emotion recognition using feature fusion method for both (a) Berlin database speech utterances

Berlin	Happy	Neutral	Anger	Sad
Happy	21	3	3	3
Neutral	4	19	0	7
Anger	6	0	23	1
Sad	0	3	0	27

Table IV.ii Percentage accuracy of emotion recognition using feature fusion method for both (b) Spanish Database speech utterances

Spanish	Happy	Neutral	Anger	Sad
Happy	19	3	6	2
Neutral	0	16	4	10
Anger	12	0	15	3
Sad	1	1	0	28

The accuracy of the recognition of every emotion are analyzed independently with all the four classifiers and are shown in Table 5. In the Table 2 the left section and the top column of the table talks about the classifiers and the emotions respectively. Each cell speaks about the recognition accuracy of the emotion by the comparing classifier. For every emotion with the Berlin database the emotion recognition rate of the classifiers RDA and SVM are both comparable with one another for every one of the emotions. The emotion anger is identified more with kNN classifier for the two databases. The efficiencies of these classifiers is shown in Fig 7.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The bars with different colors indicates the efficiencies of LDA, kNN, SVM and RDA separately. Examining the emotions of happy, neutral, anger and sad are spoken by the bars from left to right

Table V.i. Percentage of recognition accuracy for emotions (happy, neutral, anger, sad) with variety of classifiers using both the data bases (a) Berlin Database (b)Spanish DB Feature Fusion (Prosody + Spectral) Classifiers

(a) Berlin Database	**Feature Fusion (Prosody + Spectral)**			
Classifiers	Happy	Neutral	Anger	Sad
LDA	49	59	68	72
RDA	73	70	83	97
SVM	70	65	74	93
KNN	55	63	93	77

Table V.ii Percentage of recognition accuracy for emotions (happy, neutral, anger, sad) with variety of classifiers using both the data bases (b) Spanish Database

(b)Spanish DB	**Feature Fusion (Prosody + Spectral)**			
Classifiers	**Happy**	**Neutral**	**Anger**	**Sad**
LDA	49	56	65	70
RDA	71	60	67	97
SVM	67	77	65	83
KNN	63	70	88	65

Table. VI Confusion matrix of cross-correlational databases

Berlin/Spanish	**Happy**	**Neutral**	**Anger**	**Sad**
Happy	20	2	5	3
Neutral	7	22	0	1
Anger	4	6	18	2
Sad	0	9	6	15

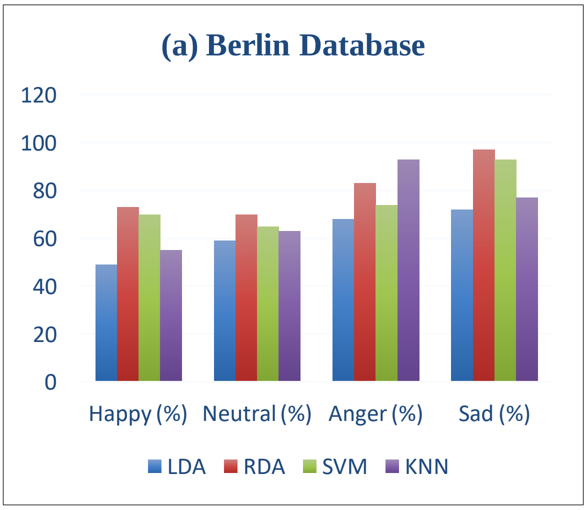

Fig. 7.1 Represents the efficiency of the classifiers using both the databases (a) Berlin

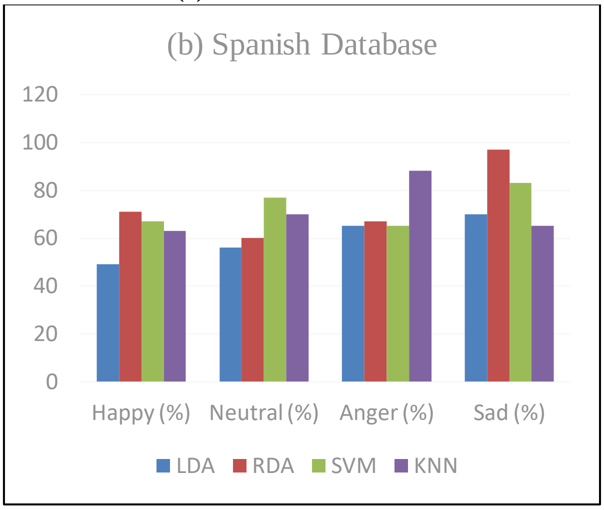

Fig. 7.2 Represents the efficiency of the classifiers using both the databases (b) Spansih

Till now, we have done training and testing using the same database. For example, using Berlin Database we have done both the training and testing part for our model constructed. But, our main motive is speech emotion recognition using cross correlational database and we aim to show that the model is independent of the database used. Below, the Table 6 represents the cross correlation databases (i.e.) we just trained the model using Berlin database and for testing the same model we have used Spanish database and the confusion matrix is shown below in Table 6.

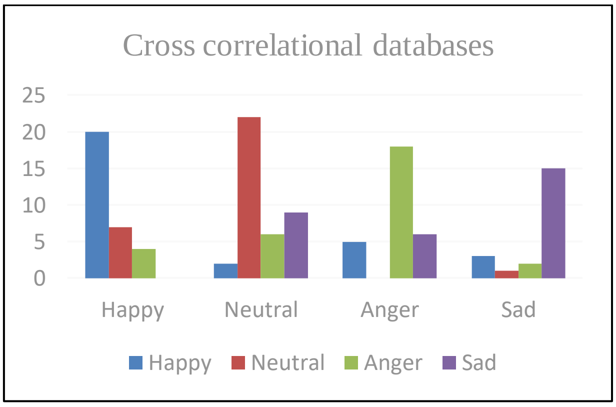

Fig. 8 Represents the comparison of the Confusion matrix of cross correlation databases

Table. VII Represents the results when tested with Spanish database of happy emotion.

Emotion	**Result**
Happy	20
Neutral	2
Anger	5
Sad	0
Fear	2

V. CONCLUSION

Since, our project is based on cross correlational databases we aim to prove that the model taken is independent of the language. It can be explained as below, we first used Berlin Database to train the model and after successful training we underwent through the process of testing in which we used Spanish Database to test the same model and we achieved the proof that the mode is independent of the language taken. A confusion matrix is obtained as shown in Table 6 and that table represents the fact that: We first taken Berlin Database consisting of 30 wav files and trained them and then we tested the same model using Spanish Dataset consisting of 30 wav files of happy emotion and we used MATLAB for implementation and the results are as shown in Fig 8. As shown in Fig 8, it is very clear that the model taken was independent of the language used. The proof is above in Fig 8 it indicates that when we used Berlin Database for training the model and then testing the same model with Spanish database of happy emotion we got the count for happy as 20 out of 30 wav files taken. It's the prove indicating that the model taken is independent of the language taken.

REFERENCES

1. Videla, L. S., Rao, M. R. N., Anand, D., Vankayalapati, H. D., & Razia, S. (2019). Deformable facial fitting using active appearance model for emotion recognition doi:10.1007/978-981-13-1921-1_13 Retrieved from www.scopus.com
2. Gurram, D., & Narasinga Rao, M. R. (2017). A comparative study of support vector machine and logistic regression for the diagnosis of thyroid dysfunction. International Journal of Engineering and Technology(UAE), 7(1.1), 326-328. Retrieved from www.scopus.com
3. Mane, S. U., & Narasinga Rao, M. R. (2017). Many-objective optimization: Problems and evolutionary algorithms - a short review. International Journal of Applied Engineering Research, 12(20), 9774-9793. Retrieved from www.scopus.com
4. Mane, S. U., & Narasinga Rao, M. R. (2017). Many-objective optimization: Problems and evolutionary algorithms - a short review. International Journal of Applied Engineering Research, 12(20), 9774-9793. Retrieved from www.scopus.com
5. Mane, S. U., & Narasinga Rao, M. R. (2017). Many-objective optimization: Problems and evolutionary algorithms - a short review. International Journal of Applied Engineering Research, 12(20), 9774-9793. Retrieved from www.scopus.com
6. Mane, S. U., & Narasinga Rao, M. R. (2017). Many-objective optimization: Problems and evolutionary algorithms - a short review. International Journal of Applied Engineering Research, 12(20), 9774-9793. Retrieved from www.scopus.com
7. Mane, S. U., & Narasinga Rao, M. R. (2017). Many-objective optimization: Problems and evolutionary algorithms - a short review. International Journal of Applied Engineering Research, 12(20), 9774-9793. Retrieved from www.scopus.com
8. Razia, S., Narasingarao, M. R., & Bojja, P. (2017). Development and analysis of support vector machine techniques for early prediction of breast cancer and thyroid. Journal of Advanced Research in Dynamical and Control Systems, 9(Special Issue 6), 869-878. Retrieved from www.scopus.com
9. Swetha, K., & Narasinga Rao, M. R. (2016). Dynamic searchable encryption over ditributed cloud storage. Asian Journal of Information Technology, 15(23), 4763-4769. doi:10.3923 / ajit.2016.476 3.4769
10. Kolla Bhanu Prakash, Dorai Rangaswamy M.A. and Ananthan T.V. (2014), "Feature extraction studies in a heterogeneous web world", International Journal of Applied Engineering Research, Research India Publications,Vol.9,No. 22, pp- 16571-79. 6
11. Kolla Bhanu Prakash, Dorai Rangaswamy M.A. and Ananthan T.V. (2014), "Feature extraction studies in a heterogeneous web world", International Journal of Applied Engineering Research, Research India Publications,Vol.9,No. 22, pp- 16571-79. 6
12. ElAyadi,M.,Kamel,M.S.,&Karray,F.(2011).Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition, 44(3), 572–587.
13. Nwe, T. L., Foo, S. W., & De Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. Speech Communication, 41(4), 603–623

AUTHORS PROFILE

G. CHANDRIKA SRI LAKSHMI. pursuing Bachelor of Computer Science at Koneru Lakshmiah Education Foundation,Guntur.Interested in Research on speech emotion recognition.

K. SRI SUNDEEP pursuing Bachelor of Computer Science at Koneru Lakshmiah Education Foundation,Guntur.Previously,Published an international journal in speech emotion recognition.

G. YASWANTH pursuing Bachelor of Computer Science at Koneru Lakshmiah Education Foundation,Guntur.

Niranjan S R Mellacheruvu, B.Tech from NGIT, affiliated to JNTU Hyderabad in 2002 with E.Cnt.E branch.M.Tech from VIT UNIVERSITY, Vellore. T.N. in 2005 with Power Electronics EEE, Ph.D pursuing from NIT, Trichy, Tamil Nadu. Dept. of EEE. Currently Working as an Asst professor in dept of ECE in Vikas College of Engineering and Technology.

Swarna Kuchibhotla, M.Sc in Physics from Andhra University , M.Tech in CSE from Andhra University and Ph.D in CSE from Acharya Nagarjuna University. Currently Working as an Associate Professor in Koneru Lakshmaiah Education Foundation. Interested in Research area of Speech Processing, Machine Learning, Image Processing.