www.ijcrt.org © 2023 IJCRT | Volume 11, Issue 7 July 2023 | ISSN: 2320-2882

IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A STUDY ON FEATURE SELECTION ALGORITHM AND FEATURE REDUCTION TECHNIQUE FOR SPEECH RECOGNITION

[1] Mampi Devi,[2] Dr..Manoj Kr Sarma

[1]Research Scholar,[2]Associate Professor

[1]Computer Science & Engineering Department,

[1]Assam Down Town University, Guwahati, India

Abstract: Feature selection is the basic part in the area of research. The outcomes and goal achievement of the research are mainly affected by the features chosen for the research in any field. Till date it has been observed that mostly used acoustic features for speech recognition are that Mel-Frequency Cepstral Coefficient and its derivatives. Diminishing the number of features unveils to further vigorous estimates of typical constraints and runs up the classification process more swiftly. It is very conclusive for concurrent speaker recognition while running such classification task on small source devices. Feature extraction can be implied as a phase to lessen the dimensionality of the contributed data, a lessening which obviously primes to some information malfunction. Mostly, during the speech recognition process, the speech signals are separated by using frames and then from each frame feature mining is done. In this paper, we are going to focus on some commonly used feature selection algorithms and feature dimension reduction techniques by comparing the effectiveness on the model functioning.

Index Terms - Speech features, LFCC, spectrogram, formant, SFS, RFE, CFS, LDA, PCA.

1 INTRODUCTION

The speedy expansion of the technology and growing social plea of computerization has been leading the speech recognition system as one of the highest preferred software in several plans. In some Speech recognition system, the skill renovates speech into text and provides digital control to users for the devices by dialogue instead of using conservative apparatuses such as key stroke, buttons, keyboards etc. In the present generation Some common examples of such software are Google voice, digital assistants, blue-tooth in car, Alexa etc. The speech signals mask huge totals of information including basic gender information and emotion status which can be notable relatively more uncomplicatedly by humans than the computers. Now a days, the research on vocal sound recognition primarily emphases on the credentials of single information, which is not sufficient to identify with the proper meaning of speech. The speech signals are converted into an arrangement of feature vectors during the procedure of feature mining. After that these vectors are shifted to the classification stage. Miscellaneous language speech recognition,
Manuscript summarization, news broadcasts reclamation and spoken discourse system, speaker credentials and appreciation etc. are some speech-related applications in our day-to-day life which have been rapidly growing in recent years. Moreover, it has been noted that the data objects is an important resource both in automatic speech recognition as well as in emotion recognition from speech. However, emotional speech recognition is frequently data-driven: by means of labeled speech samples a classifier is trained that can subsequently be used to categorize unidentified or hidden data [1].

Now a days, another very widely applied technology is the Orthogonal investigational design; which is basically functional in agriculture field of the most developed countries. It has been successfully applied in many areas of our country too. The development of orthogonal investigational strategy has made a upright use of the table— “Orthogonal” to make plans for test. An insignificant quantity of tentative conditions can be elected in many tests, but should be strong and supposition to discover the finest process environments through that numbers. The parameters are termed as feature that can mark the outcome of the experiment in a good track. The state of the diverse factors is called level. Main objective of the Orthogonal experiment is to novelty the finest grouping of the prevailing factors. In the progression of incisive the mandatory test, times are found to be less than the systematic technical steps. There are three rules followed in constructing an orthogonal table: a) Usually a hard mathematical theory is needed for Orthogonal table, but in case factor level is two, the table is very familiar to paradigm b) Factors: It means inside the orthogonal table, the amount of each level is identical to the average and between any two columns of different levels of the total measure of groupings is identical to the average. Therefore, while positioning orthogonal experiments, all kinds of aspects collection is unbiased. In the table, every row explains an experimentation scheme, which is a grouping of various factors in state; each column explains that the corresponding factors of the state. c) The experimental results analysis: Investigation of adjustment can make a distinction between experimental results and fault caused by the vacillation of deviations between the experimental results. Thus, the system of mathematics covers the absence of the underprivileged analysis method. We can grow the jurisdictive that the alteration of level triggered by the variance between the experimental results as per the theory of difference analysis. If the tentative outcomes fluctuations caused by the variations of factor levels within the fault range or have slight change with the error, the variation of this factor level can determine the root of a substantial variations in outcomes. On the contrary, if factor levels’ variation will alleviate in the experimental results than the error range, then we can certainly say that the factor has a considerable brunt on the investigational out comes. The main resolution of this examination is to determine the things that have a substantial influence factor concluding the data overall.

Abbreviations and Acronyms

Some of the abbreviations for the short forms of speech features used in this research are listed below.

MFCC	Mel Frequency Cepstral Coefficient
HFCC	Human Factor Cepstral Coefficient
LFCC	Loudness Factor Cepstral Coefficient

2 Back Ground of The Features regarding Speech

The basic idea for research in any area always depends upon the background knowledge in the same area. In our research paper, the background knowledge consists of structure of speech signals and speech features and these are defined in depth in the following sub sections.

2.1 Pre-Processing

This is the first step of speech processing or speech recognition, where the speech signal is passed through a preemphasis filter. It helps to up lift the liveliness of sophisticated frequencies included in the signal, which are reduced during the speech signal creation from vocal tract.

2.2 Features for Speech Recognition

Not a specific parameter is linked only to the speech content or the speaker who speaks, however there exists so many speech feature parameters. Appropriate parameters choosing is most important to do dealing out or investigation, apart from other significant features intrusion produced and focusing the feature invoice signals that can direct the content to supplementary identification of the features which is suitable to the content class. The most commonly used parameters as we have found till date is: Pitch, Fundamental Frequency, Mel-Frequency Cepstral Coefficients along with it sub bands and domain energy related to amplitude, time, MFCC, HFCC, LFCC and along with their different combinations.[2].

2.3 Syllable Structure

Structure of a syllable consists of mainly two parts: verbalization features and pronunciation that is hidden in the initial, intermediary or final position of a word. A syllable contains mainly three chunks or parts likely: onset, nucleus and coda. The onset and the coda together is termed as consonants and sometimes consonant clusters may be termed. And these are appeared at the origination and the termination of the syllable correspondingly.

Rule of syllabification is to observe the different patterns of syllables that may conduct experiment with monosyllabic, disyllabic, tri syllabic and polysyllabic. And in Assamese the same rule formulate an algorithm for dividing the particular language words into syllables.

Types of Assamese Syllable Structure : V,VV,CV,VC,CVV,CCV,CCVV,CVC,CVVC,CCVC,CCVVC, CCCV,CVCC,VVC.Among the syllable types mentioned above, very much restricted in use and have occasional occurrences in the language’s vocabulary. The Assamese language highly prefers syllables with an onset except an unusual favour on the structure VC and V.

The most canonical syllable types in Assamese are the simple CV i.e. consonant vowel (EXAMPLE:/Xi/means ‘he’) which is a sequence of core syllables consisting only two items in the skeletal tier) and a more complex one is (C)(C)V(V)(C) i.e. consonant consonant vowel vowel consonant, but it does not allow CCCVV(C) and(C)CCV(V)CC sequences.

2.4 Vowel Phoneme

There are mainly three types of vowel phonemes in Assamese and they are short vowels, long vowels and nasal vowels. Otherwise, in Assamese language there are total eleven vowels. In this project we are applying feature selection algorithm and feature reduction technique on all the eleven vowels to visualize the effects of communication and depending on the results we have choose any one of the techniques.

3 Feature Extraction : Features with diagram and Procedure

It has seen from previous researches that the “open SMILE” tool is a common that can be used to generate the features like pitch, HFCC, LFCC, MFCC, MELBS, HFBS, Zero Crossing Ratio (ZCR) etc. In this paper, some of these features have been used along with their first order delta coefficients. In Table 1 a short of the produced features have been given. However, the linked functional can be useful to each contour of the low-level-descriptors, such asmeans,5thmoments,6thmoments, segments, percentiles, durations, onsets, Coefficients of DCT , Regression of both linear and quadratic etc. A total of 55utterance-level audio features were extracted related to fundamental frequency(f0), energy, HFCC, LFCC and MELBS, LFBS, PLP and its sub bands etc. like total 29 features that were extracted by some previous researchers [4]. The audio feature mining has done using the latest version of the PRAAT software.

Table 1. produced features

1	2	3	4	5	6	7	8	9	10
Mfcc1	Mfcc2	HFCC	LFCC	MELBS	HFBS	LFBS	Harmonicity	Modulation Energy in joul	MSME in joul

11	12	13	14	15	16	17	18	19	20
fundamental frequency in hz	mean frequency	median frequency	standard deviation	formant frequency	formant bandwidth	Zero crossing ratio	plp1	plp2	plp3

21	22	23	24	25	26	27	28	29
plp4	temporal energy	Teager energy operator	Kurtosis	Skewness	Max	Min	5th moment	6th Moment

Two important features LPC(Fundamental frequency) and Formant Frequency are showninFigure.1.

In Figure 2, the Illustration of audio feature extraction is done such as: wave form, spectrogram, pitch track respectively.

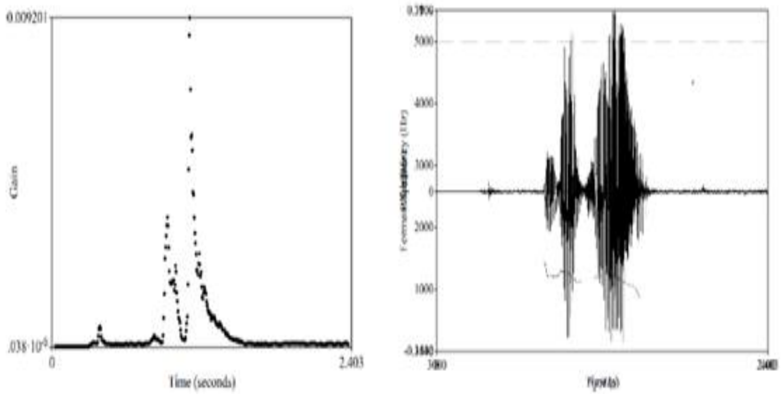

FIGURE 1 : lpc & formant frequency

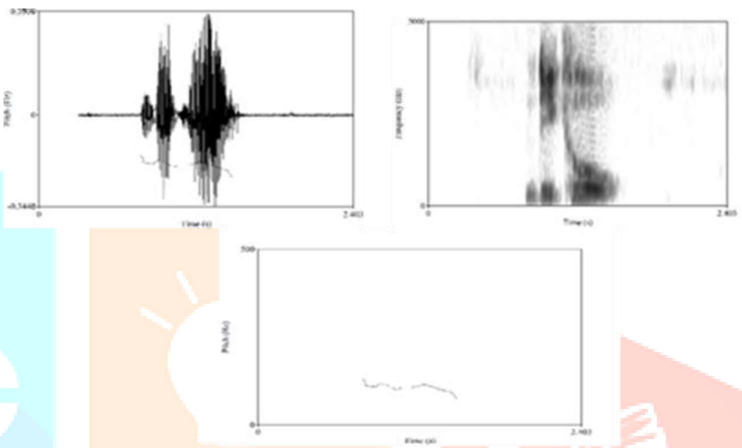

FIGURE 2: form, spectrogram, pitch track

The whole model of the System architecture can be erected from the fundamental frequency(f0) extraction was completed with Speechs having System software by RAPT algorithm. Some of the features extracted from f0 contour are: Mel frequency, maximum& minimum frequency, mean and median fundamental frequency and standard deviation of first and second Gaussian of Amplitude., Mel frequency band spectrum, standard deviation of Mel freq. first order difference, etc.[9],HFBS and LFBS.

To evaluate the Energy features ; some steps are followed;1ststep:-the using Butterworth filter has been used to filter the signal in bands and from that filtered signal energy has been calculated at frame level via the Hamming window of20ms with a step size of 5ms. Some of the energy features extracted are: MSME, standard deviation, minimum, maximum of fundamental frequency and range of normalized energies in the original speech signal and speech signal in the frequency bands 0-0.50kHz,1-1.5kHzand2-2.5kHz. Statistical features collected like; skewness, kurtosis, Teager energy operator in frequencies of the standardized energies in the novel speech signal and speech signal in the identical frequency bands. A number of descriptive spectral features like: Zero Crossing Rate, Spectral Roll of, Skewness and Kurtosis etc. were extracted and some these are suitable for musical instrument identification [6]. Duration of features collected via microphone and PRAAT software were used to excerpt duration features based on heeding supported by waveform and spectrogram forms. The features that were extracted of different durations like: voiced speech duration, unvoiced speech duration, voiced-to-unvoiced speech interval ratio, average voiced-to-unvoiced speech interval ratio, speech rate (phone/s), voiced-speech-to-sentence level speech interval ratio.

4 Feature Selection Algorithms

As the simple models yield a reduced amount of training time and advance the over simplification skill by tumbling over fitting, so in Feature choosing procedure only simple models are basically applied. The main objective of the feature selection algorithm is to remove the features that are superfluous and irrelevant. Moreover, a very important role played by the feature selection algorithm for better results in any field of research.

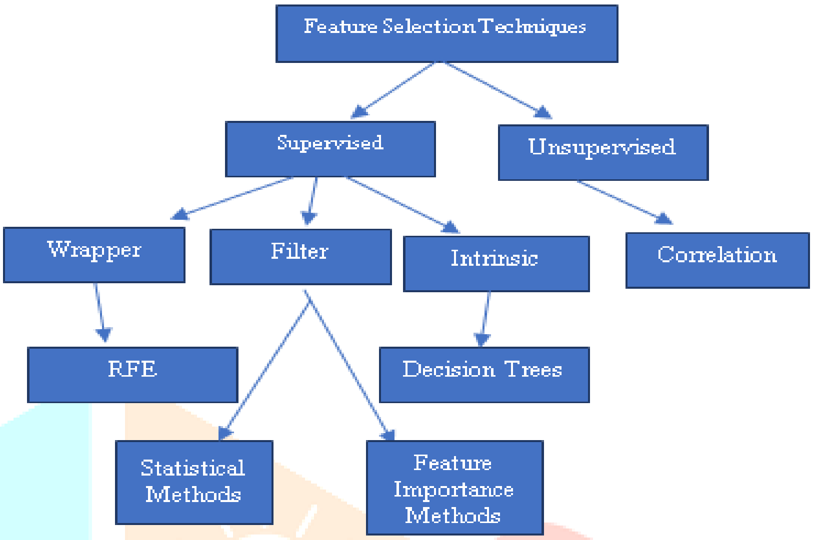

FIGURE3:the classification of feature selection techniques

4.1 SFS vs SBS vs SFB

Sequence Forward Selection (SFS), SBS (Sequence Backward Selection), SFB (Sequence Forward Backward) are some of the commonly used feature selection algorithms. In these techniques, the feature selection is accomplished by using an individual algorithm founded on a discriminative measuring role. This process supports to confiscate uncommunicative, superfluous or piercing features. The SFS algorithm is a lower most (bottom) up and about search method, where one feature is added at a time. At first, the finest feature is cautiously chosen and then the role is assessed for grouping with the remaining candidate features and then the greatest fresh feature is added [11], like that it continues. The delinquent with the SFS algorithm is that, as soon as a feature is added (which may be come impractical on far along with the growth of the feature set), it cannot be detached. However, the SBS is a top down process that begins with the whole feature set and at each step the foulest feature is cast off such that the compact set gives a extreme value of the standard function. The SBS gives improved outcomes but is computationally extra intricate. Sequential forward backward search bids reimbursement sof both SFS and SBS by the uses of Plus l-Take Away R algorithm. In every step, l features are supplied to the existing feature set and R features are detached. Until the required feature set size is attained, the process is being carrying on. Though we have not used this algorithm in our dataset for feature set selection, but the full feature sets(audio)with the distance as a criterion as stated by Bhattacharyya has been casted-off by us. The diffusion of classes was expected to be Gaussian. The searching of feature was accomplished with l=2 and r=1, i.e. one feature was supplemented at each step. SFB is such an algorithm, where the probability of an observation sequence is computed for a given model in a very convenient, recursive and efficient way. This algorithm acts as an interface algorithm for hidden Markov model. Moreover, this algorithm can be calculated in linear time, where as a brute force algorithm that checks all possible state sequences would be exceptional over the length of the sequence.

```
sfs1.k_feature_idx_
(0, 8,9,10,11,13,15,19,23,25,28)
new_dataframe.columns[1:][list(sfs1.k_feature_idx_)]
Index(['mfcc_1','ModulationEnergyinjoul','MSMEinjoul','fundamentalfrequencyinhz','meanfrequency','stand
arddeviation','formantbandwidth','plp3','kurtosis','Min','skewness'],dtype='object').
```

TABLE2 : accuracy table by sfs

Accuracy Types	Accuracy Obtained before SFS	Accuracy Obtained after SFS
Training Accuracy	68.18	79.54
Testing Accuracy	45.45	45.45

4.2 FS via RFE

The most popular Feature selection is Recursive Feature Elimination (RFE) as its effortless to organize and exploit along with its effectiveness at selecting those features within a training dataset that are added or mainly applicable in forecasting the object variable. There are two vital formation options when using RFE: the opposition of selecting the quantity of features and the preference of the algorithm used to support for choosing features. Thus both the formations can be explored, although the concert of the method is not robustly reliant on these hyper parameters being configured on form[8].
RFE in a word, is belongs to a wrapper-type feature selection algorithm as well as filter-based feature selection algorithm. For the first term wrapper, it has been used a changed machine learning algorithm and used in the core of the method that is wrapped by RFE; thus used to support features [11] selection. While for the second reference, it has been observed that instead of selecting each feature individually RFE picks only those features having the principal (or slightest) score. Therefore, in technical language it can be said that "RFE" as a wrapper-style feature selection algorithm where filter-based feature selection (on the inside) used too .
In RFE, the technique starts by probing for a subset of features including all the features within the training dataset and effectively eliminating the features until the favourite number left overs. This is attained by suitable the given machine learning algorithm used in the prime of the model, ranking features by prominence, removing the minimum significant features and the model re-fitting method. Thus, the process is recurred while waiting for a quantified number of features leftover.

```
rfecv.fit(X,y)
   rfecv.transform(X)
   print(); print(rfecv)
   print(); print('Optimal no of features:{}'.format(rfecv.n_feature))
   print(); print(np.where(rfecv.support_== False)[0])
```

Output :(55, 14)

4.3. FS via Embedded Technique

Random Forest belongs to a supervised feature selection technique which outfits both of the decision trees and bagging methods. The knowledge behind of it is to train dataset as resampled according to a technique called "bootstrap". Feature selection using Random Forest originates in the sort of Embedded approaches. Embedded approaches combine the potentials from both the filter and wrapper methods. They are executed via the algorithms having their individual inbuilt feature selection approaches. Some of the advantages of this approaches are :
They are exceedingly precise.
They take a broad view improved.
They are understandable
Respective sample comprises of a random subset of the original columns and is used to fit a decision tree. The quantity of models and the quantity of columns are hyperparameters to be adjusted. Lastly, the estimation of the trees are diverted organized for manipulate the mean value which ids for regression or by means of easy elective.
In bagging method the outputs of the single decision trees are averaged as well as the standard error decreases and ultimately the variance of the model according to bias-variance halts beyond. So, Random Forest has developed very well-known in the past years of Feature Selection Field.

4.3.1 Random Forest Features Selection Procedure

#Random Forest
from sklearn.ensemble import RandomForestRegressor

Random forests mainly comprise of 4 –12 hundred decision trees where every individual tree is built over a random extraction of the interpretation from the dataset and the features that are randomly extracted. All the trees never attempt all the features or all the interpretation, and thus the trees are de-correlated that leading to less inclination for over-fitting. Moreover, each tree is a categorization of yes-no questions too based on a single or combination of multiple features. The question arises for each node is on division of the dataset mad about two buckets where, each of them accommodating remarks that are more alike along with themselves and dissimilar that with other. Hence, the prominence of each feature is computed from the purity of every buckets.

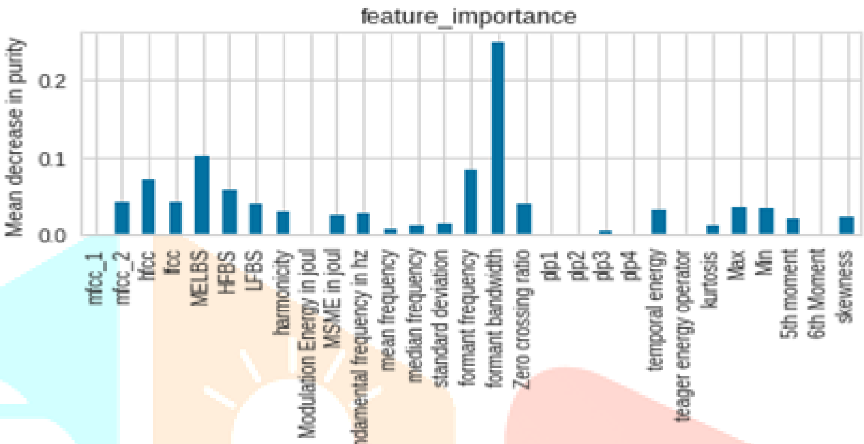

FIGURE 4 : random forest

4.4 FS via Ensemble Technique:

Ensemble learning technique includes Extremely Randomized Trees Classifier and Extra Trees Classifier (all together), which accumulate the results of the multiple de-correlated decision trees assembled in a "forest" to yield it's classification consequence. In perception, it is very analogous to a Random Forest Classifier and only diverges from it in the approach of creation of the decision trees in the forest.

In the second one i.e. (Extra Trees Forest), each 'Decision Tree' is built from the inventive training trial at first. After that, each tree is granted with a random sample of k features from the feature-set at each node. Further, each decision tree must choose the best feature to split the data grounded on some measured standard, such as "Gini Index" which is stereotypical. The method of informal sample of features straightens to the formation of numerous decision trees that are de-correlated. For accomplishing the feature selection via the mentioned forest structure during the erection of the forest, the standardized entire drop in the mathematical measures used in the decision of each feature along with the split (in case the Gini Index is utilized in the construction of the forest) is intended. Thus the calculated value is called the Gini Importance for the respected feature. To accomplish the feature selection, each feature is organized in downward order rendering to the Gini Importance of each feature and the manipulator picks the top 'k' features as per choice.

#Filter_method::Feature Importance
from sklearn.ensemble import ExtraTreesClassifier

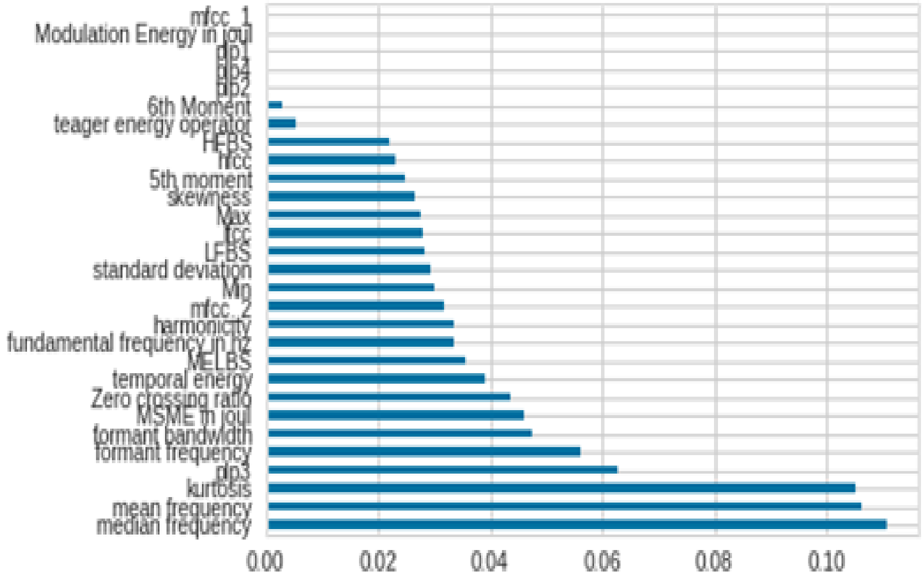

FIGURE 5: extra tree classified features

4.5 Feature Selection via Correlation or CFS Method :

We have also applied a CFS method on the features for this research paper, which has been compared later on. This technique estimates the subgroup of features and only discriminate features are picked, which are having a high correlation with a class occurrence [12]. CFS grades the features using a empirical estimation function based on correlation. It is used to measure the resemblance among the features. On the other hand, CFS rejects unrelated features that have fewer connection with the class label. The CFS standard function is as follows:

$$CFS = MAX\left[\frac{rcf1 + rcf2 + \cdots \dots \dots . +rcfk}{k + \sqrt{(rf1f2 + ..rfifj \dots . +rfkfk - 1)}}\right] \qquad Eq(1)$$

Here, rcfi denotes the feature classification correlation, k as the number of features and rfifj as the correlation between the features. The selected features are only if the classifiers of the SER is achieved.
The correlated features are depicted in figure 6 below.

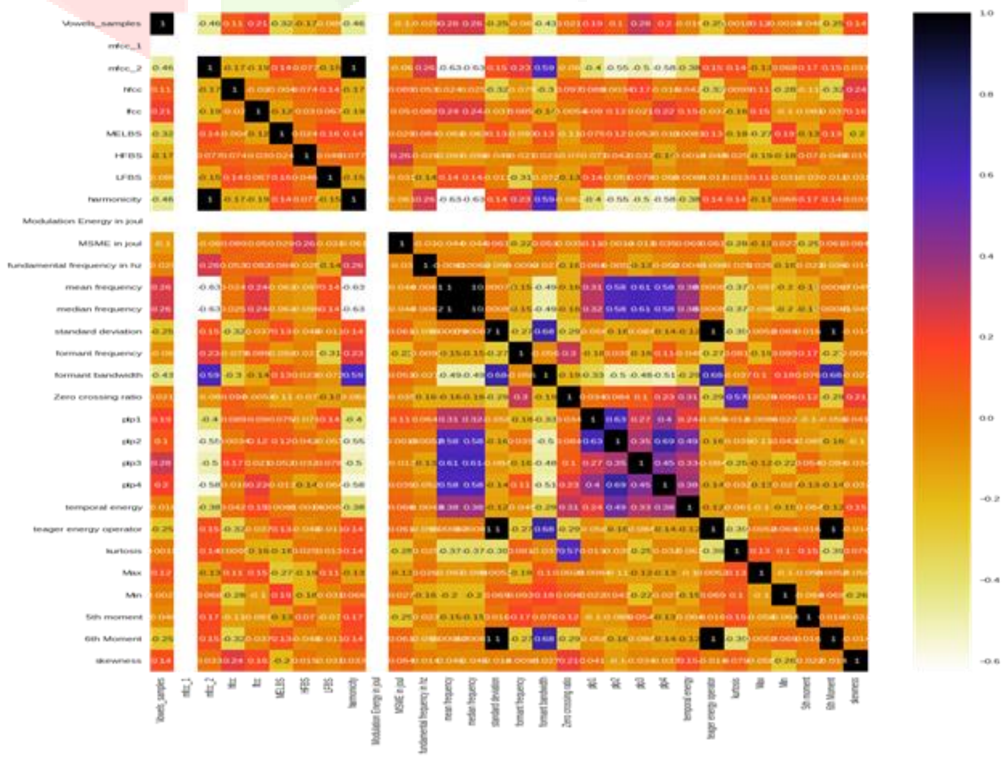

FIGURE6: cfs

5 Feature Reduction Techniques

The Dimensionality reduction can be referred in simple words as the technique of reducing the feature set dimension. Generally, in machine learning the features' dataset contains hundreds of columns (means the features) or an array of points, making an immense range in a three-dimensional space. By applying dimensionality reduction, the number of columns to measurable counts can be diminished or bring down, thereby converting the three-dimensional sphere into a two-dimensional circle.

The vulgarity of dimensionality reduction is a phenomenon that arises while working with the analyze and envisage the data in high-dimensional spaces that do not occur in low-dimensional spaces.

moreover, the dimensionality lessening has some other advantages, such as:

- It removes clamor and superfluous features.
- It supports improving the model's accurateness and concert.
- It enables the practice of algorithms that are incapable for more significant magnitudes.
- It diminishes the extent of loading interstellar required (less data needs lesser loading interstellar).
- It bandages the data, that lessens the calculation time and simplifies faster exercise of the data.

5.1 PCA vs LDA

A feature set's dimensionality can be abridged by means of some statistical means to make the most of the appropriate information well-maintained by putting on a linear conversion. And the method is applied on x; like x = Wz In the above transformation, x represents a feature vector in the compact feature space; while z represents the novel .

feature vector [9]; and W signifies the transformation matrix. PCA is broadly used for extracting crucial features. And those are mined subsequently removing noise [5] of a high dimensional data set. On the other hand, for

enhancing the separability amid the classes, LDA exploits the ratio and of course it's between-class variance to within-class variance. The PCA and LDA methods contain feature focusing and bleaching, covariance estimation

and eigen disintegration. We have used PCA instead of LDA as a linear transformation technique for feature lessening, thereby applied only as linear transformation technique.

Another term a simplification of Fisher's linear discriminant method denotes the linear discriminant analysis, which is broadly pragmatic in statistics, pattern recognition, and machine learning [7]. The main objective of the LDA technique is to develop a linear combination of features. So that it can further depict or extricate between two or more objects classes. Maximization of class separability is signified via LDA indicator data[15]. On the other hand, objects belong to the similar class are contrasted via projection and then different classes' objects are set away from each other.

```
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel("no of components")
plt.xticks(np.arange(1, 30, step=1))
plt.ylabel("cumulative explained variables")
plt.title('The no of components needed to explain variance')
plt.axhline(y=0.99, color='r', linestyle='-')
plt.text(0.5,0.85, '99% cut of threshold', color= 'red', fontsize=16)
```

In figure7 we have obtained that 10 components are needed for 99% of threshold value , while in figure8 [19] the researcher obtained 10-12 components for achieving 95% of threshold value.

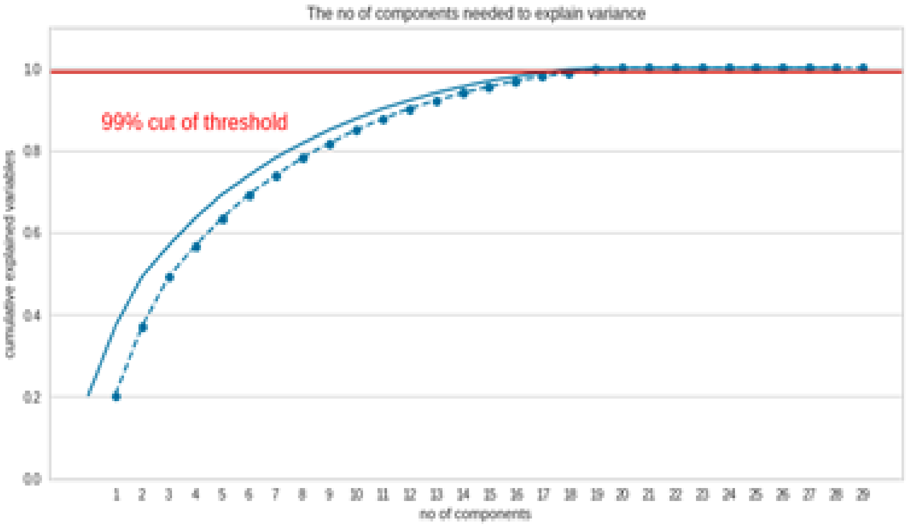

FIGURE7: pca threshold

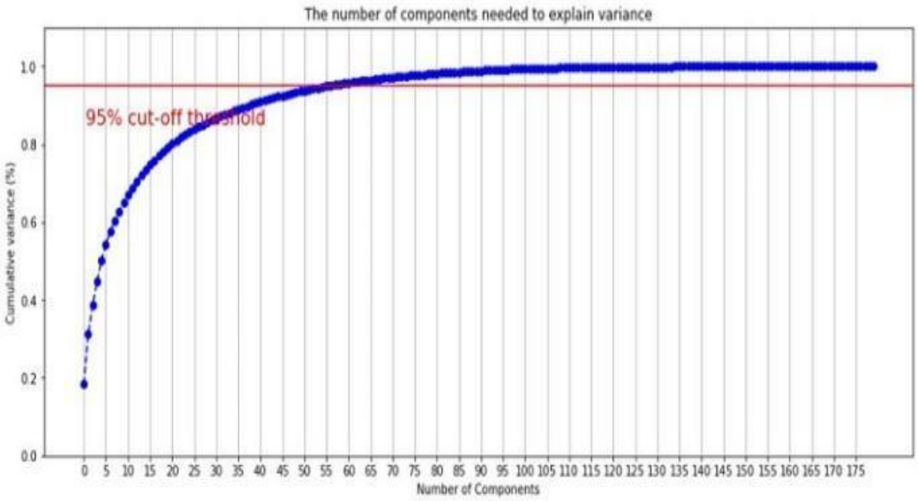

Figure 8:plot indicating the number of components needed to illustrate variance

plt.scatter(features_principal['P1'], features_principal['P2'])#, c = cvec)
plt.legend((b, y), ('Label 0', 'Label 1'))
We have achieved different scatter diagrams for label 1 and two labels (label 0 & label1) with feature set using PCA as depicted below in figure9 and figure10 below.

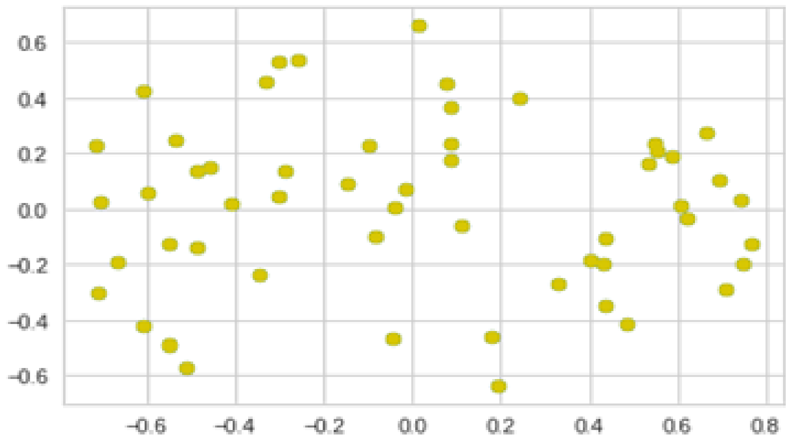

FIGURE 9: pca label1

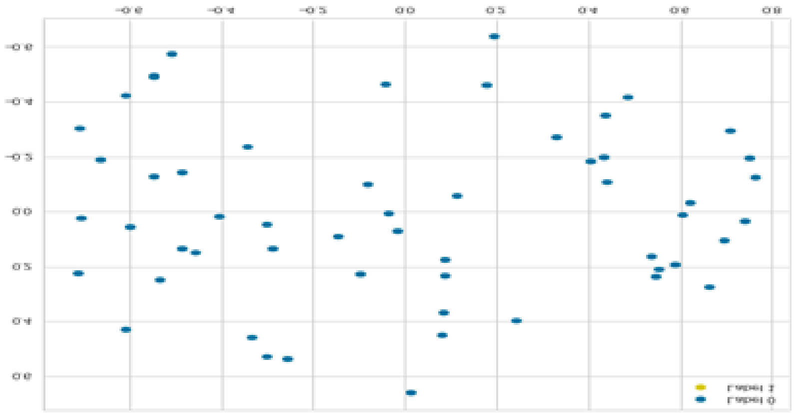

FIGURE 10: pca label 0 & label 1

5.2 Unsupervised FS via Non Negative Matrix Factorization (NMF)

A non-negative matrixis perturbed into the product of two non-negative ones by the NMF. Thus, the NMF method becomes an appreciated means in the areas that are mostly disturbed with non-negative signals in some fields like instance, illustration, astronomy, space science, etc. The Lee & Seung amended the NMF technique of the multiplicative apprise rule through inclusion of doubts, in view of lost data, equivalent calculation along with consecutive erection.

Such insertions subsidized for constructing the NMF method firm and linear. Moreover, NMF never removes the mean of the matrices and thus constructs unphysical zero-negative fluxes, for which it can store extra info than that by the PCA method.

Sequential NMF having characterization of a steady factor base throughout building a linear modeling process and thus makes it the seamless resources in astronomy. Sequential NMF can also reserve the flux in the straight tomography of circumstellar structures in astronomy for noticing exoplanets and guiding pictures of circumstellar disks.

5.3 Unsupervised FS via Generalized Discriminant Analysis or shortly (GDA)

This analysis is a type of nonlinear discriminant analysis, where the kernel function operator controls the analysis. Its vital theory challenges very meticulously to that of support vector machines (SVM), as the GDA technique supports high-dimensional feature space [13] via mapping of the input vectors. However, LDA search for finding a projection of variables into a lower-dimensional space. But, GDA do the same by making the most of the ratio of between-class scatters to within-class scatter.

5.4 Missing Values Ratio

When we reconnoiter a given dataset, its very important to find the absent values of the dataset. The first and foremost phase to arrangement of the missing values identification with the reason of that. Consequently, we can change the missing values or bead them overall by by means of some suitable methods. This method is seamless for the states once there a few missing values are seen.

But still there is question on what to do if there are too many missing values like 50% or above. In these circumstances, a threshold value can be set and use the absent values ratio method. When the threshold value increases the destructiveness of the dimensionality reduction will be more. When the percentage of the absent values in a variable beats the threshold then the variable can be eliminated.

Usually, data columns containing abundant missing values hardly contain useful information. Therefore, all the columns containing data having higher absent values than the set threshold values can be eliminated.

6. Results & Discussion

We have found some results for the three clustering techniques with different set of features. Some of the results obtained by other researchers with closely related works along with our obtained results are mentioned below.

6.1 Experimental Results and Some Survey Results

The dataset in our system is not a inbuilt one as the Assamese language have not any inbuilt speech dataset. We have constructed the dataset with the eleven vowels with the utterances of five speaker in a studio type environment. And we have extracted total 30 features for each utterance. However, we have choosed different feature scale from different feature selection and reduction technique at different times with different feature selection and feature reduction technique. The Spectral Clustering is a graph based traditional clustering technique[10]. But in this research, to finalize any one of the feature selection or reduction technique; we have first applied three clustering techniques namely KMeans, PAM and CLARA on the different feature scales and for then compared the accuracies for each of the clustering technique with different feature scales found from different feature selection and reduction techniques. From the accuracies found, we have decided our feature combination[14]. The results are discussed in the following tables.

Table2: clustering accuracies with randomly selected features

Clustering Algorithm	No. of Features					
	5	10	15	20	25	30
KMEANS	.60	.56	.51	.49	.48	.39
PAM	.67	.65	.61	.58	.56	.50
CLARA	.65	.63	.60	.55	.53	.48

TABLE3: clustering accuracies with features selected by using correlation

Clustering Algorithm	No. of Features				
	Top 5	Top 10	Top 15	Top 20	Top 25
KMEANS	.65	.60	.56	.53	.50
PAM	.76	.73	.60	.61	.58
CLARA	.70	.65	.59	.55	.52

In table 8[17], the researcher has obtained spectral ratio of 3.1(90%:30%) with 10% filtered features for PCA with KMeans. However, in our research we have found 76% for top 5 features with PCA using K-Means without filtering as mentioned in table9.

TABLE 4 : overall success rate of phoneme recognition

Phonemes	Phoneme Success Rate
Vowel	95.25%
Initial Consonant	94% 3
Last Consonant	93.2%

TABLE5 : clustering time based on mushroom dataset

No. of cluster	K-Mean	K-Medoid	Clara
2	12558	6708	2278
3	13369	7005	3354
4	14430	7845	4399
5	15491	6646	5492
6	15487	7012	5412
7	15008	7023	5504
8	14587	6987	5624

TABLE6 : clustering accuracies with respect to word documents

Algorithm	No. of Clusters	No. of documents	Clustering Efficiency
K-Means	0	18	27.78%
	1	3	66.67%
	2	39	25.64%
	3	25	24%
	4	15	26.67%
K-Medoids	0	43	18.60%
	1	11	54.15%
	2	7	42.85%
	3	35	14.2%
	4	4	50%

In table 6[16] the researcher has obtained 66.67% highest accuracy for 1 cluster with 3 documents using K-Means and 54.15% for 1 cluster with 11 documents using K-Medoids. Wherever, in our research we have achieved 71% for top5 features with K-Means and 75% with top 5 features for PAM(K-Medoids) in table7.

TABL7: clustering accuracies with features selected by using feature importance

Clustering Algorithm	No. of Features				
	Top 5	Top 10	Top 15	Top 20	Top 25
KMeans	.71	.67	.60	.57	53
PAM	.75	.69	.62	.59	.57
CLARA	.72	.68	.61	.58	.54

TABLE 8: ser after filtering with the spectral ratio confidence comparing of the pca-based diarization.

Filtered(%)	0	10	20	30	40	50
PCA	4.1	3.1	2.6	2.6	2.4	2.1

TABLE 9: clustering accuracies with features selected using pca

Clustering Algorithm	No. of Features				
	Top 5	Top 10	Top 15	Top 20	Top 25
KMeans	.76	.69	.63	.59	55
PAM	.87	.84	.70	.65	.60
CLARA	.72	.69	.61	.59	.57

In table 8[17], the researcher has obtained spectral ratio of 3.1(90%:30%) with 10% filtered features for PCA with KMeans. However, in our research we have found 76% for top 5 features with PCA using K-Means without filtering as mentioned in table9.

6.2 Discussion on the Results

Results mentioned in table 4 are obtained by Talukdar Pallabi et.al., by using Hybrid neural framework with K-Means clustering for Assamese spoken words, where they achieved better accuracies for the Assamese vowels with two specific features .However, in our research we have done the KMeans only for Assamese vowels without Hybrid neural framework using various feature dataset and achieved good accuracies with less no. of features. In the table 5 obtained by Kelde Dharmendra et.al. for an IRIS dataset of mushroom, it has been observed that with the increasing number of clusters the time based dataset of KMeans, Kmedoids and Clara is also increasing, but maximum value is different for the KMeans, KMedoids and Clara.
However, in the table6 obtained by Balabantaray Rakesh Chandra et. al., KMeans efficiency is good for only 3 documents, while Kmedoid efficiency is good for 11 no. of documents, where each cluster number defines documents of a particular domain. From the above result for hundred documents it can be determine that the clusters obtained using K-means algorithm is more efficient than the clusters obtained from K-Medoids algorithm.
In our experiment we have found in table7 some similar results of better efficiency for KMeans and PAM than that of Clara though it is for only vowels with various features.

In table 8, results obtained by Orith Toledo-Ronen and Hagai Aronowitz, the confidence ratio is found good for less filtering with PCA using KMeans. In this paper, a confidence measure for speaker diarization based on the spectral ratio of the eigen values of the Principal Component Analysis transformation was computed on the pre-segmented audio before diarization performing on the conversation. In our experimental table9 we have observed that clustering accuracy of KMeans 76% and PAM 87% is better than Clara 72% and that is for the topmost features after applying PCA for the whole dataset.

7 Conclusion& Future Work

After doing survey of some latest papers as mentioned in the reference part, some reasonable conclusions can be drawn for the features selection as well as feature dimension reduction techniques. The most important conclusion is that the Feature selection techniques are more helpful in both supervised and unsupervised techniques of machine learning algorithms. However, feature reduction algorithms are more helpful for unsupervised machine learning techniques only depends upon the size of the dataset. From the results obtained in the tables we have drawn another conclusion that a dataset having less(smaller size) data, feature dimension reduction is not so useful. And speech feature data set are mainly unlabelled, however labelling is also possible in some of iris datasets.
Application of feature reduction technique in large dataset for unlabelled data having word boundary, noise removal and tone identification will be our future work in this field.

Acknowledgement:

This paper entitled “A Study on Feature Selection Algorithm and Feature Reduction Technique for Speech Recognition” would not have been possible without the support and encouragement of my parents at first place as they directly offer their help and cooperation towards the completion of the work. I am very fortunate for having such a good parents.
Secondly, I would like to express my sincere gratitude and heartfelt thanks to my guide Dr. Manoj Kr. Sarma who is an Associate Professor of the Assam Down town University.
Last but not the least, I offer my sincere gratefulness to Dr. Jyotismita Talukdar, an associate professor of Tezpur University, whose support and inspiration have been benefited me enormously. This work would not have been possible without
the respected Talukdar madam.

References

[1]Xu-Kui,Yang.; Liang He, Dan Qu.; Wei-Qiang, Zhang.; and Michael T., Johnson(2016).Journal on Audio, Speech, and Music Processing,DOI 10.1186/s13636-016-0086-9, P2-10.

[2]C.Poonkuzhali.;R.Karthiprakash1.;Dr.S.Valarmathy.;andM.Kalamani.(November 2013).International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 11, ISSN (Print) : 2320 – 3765,P-5671-5678.

[3]M., Sidorov.; Ch, Brester,; W, Minker and E.Semenkin(2014).Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).P-3481–3485.

[4]H., Muthusamy.; K., Polat. and S., Yaacob.(in PLOS ONE) open access research article. March 23, 2015, DOI:10.1371/journal.pone.0120344,P1-20.

[5]M., Sidorov.; Ch..Brester.; and A. ,Schmitt.(September 2015).Proceedings of 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), Vols 1-5. Dresden, Germany – Baixas,P2699-2703.

[6]R., Loughran.; A., Agapitos.; A.,Kattan.; A.,Brabazon.; and M., O’Neill.(in press). July 2017, Evolutionary Intelligence 10(1),DOI:10.1007/s12065-016-0150-5,P1-32

[7]P., Chauhan.; G. Date.; and N.,Nagarkoti.(2018). St. Francis Institute of Technology, Mumbai, a Project Report, SFIT Library.

[8]M.Z., Ansari.; T., Ahmad.; and A.,Fatima.(November 2019).International Journal of Recent Technology and Engineering, Volume-8, Issue-4,P-1-9.

[9]S.,Kolita.;and P.,Acharjee.(May2019) International Journal of Engineering Research & Technology,Volume-8,Issue-5,P-207-210.
[10]M., Kumar; S., Narayanan; K. J., Han; and T.Jin.,Park.(March 2020).IEEE Signal Processing Letters,Volume- 27, P. 381-385.
[11]S., Reddy.; BANDELA;andT.K.,KUMAR.(June2020).Radioengineering. Jun2020, Vol. 29 Issue 2, P.353-364.
[12]Talukdar Pallavi et.al.,"Recognition of Assamese Spoken Words using a Hybrid Neural Framework and Clustering Aided Apriori Knowledge"
[13]S.,F., Ali.; Zaidi.; and L., Xu(2021).Journal of Physics: Conference Series, Volume 2224,2nd International Symposium on Automation, Information and Computing (ISAIC 2021), doi:10.1088/1742-6596/2224/1/012119,IOP publishing.
[14]Z., Hameed; W.Ur., Rehman. ;W., Khan.;N.,Ullah.;and F.,R., Albogamy.(in Mathematics December2021), 9, 3172. https://doi.org/10.3390/math9243172. https://www.mdpi.com/journal/mathema,P.1-19.
[15]Kelde Dharmendra et.al., "A Framework for Comparative Study of K-Mean, K-Medoid and Clara" in International Journal of Engineering Science and Computing, July 2016, Volume 6 Issue No. 7,p-1694-1698.
[16]Balabantaray Rakesh Chandra et. al., "Document Clustering using K-Means and K-Medoids" in International Journal of Knowledge Based Computer System, Volume 1 Issue 1 June 2013, http://www.publishingindia.com
[17]OrithToledo-Ronen andHagaiAronowitz," Confidencefor Speaker DiarizationusingPCASpectralRatio" in INTERSPEECH2012 at ISCA's13th Annual Conference, September 9-13,2012.
[18]Ayantika Nath,ShikhaNema, "Clustering Visualization and Class Predictionusing Flask of Benchmark Dataset forUnsupervisedTechniquesinMachinelearning" in InternationalJournalofInnovativeTechnologyandExploring,November2020, ISSN:2278-3075,Volume9, issue7,P-285-298.
[19]Misbah Farooq et.al., "Impact of Feature Selection Algorithm on Speech Emotion Recognition Using Deep Convolutional Neural Network" in Sensors October 2020, 20, 6008; doi:10.3390/s20216008.
[20]W., Jia.; M., Sun.; J., Lian.; and S., Hou.(Jan 2022)open access article in Springer ,Complex & Intelligent Systems volume 8, P.2663–2693.