REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

DECISION MAKING IN HEALTHCARE THROUGH MACHINE LEARNING ENHACEMENT WITH FUSION FEATURE

Devi Mampi[1*], Sarma Manoj Kumar[2]

[1*]Research Scholar, Faculty of Engineering, Assam down town University, Guwahati, Assam, 781026
[2]Associate Professor, Faculty of Engineering, Assam down town University, Guwahati, Assam, 781026

Abstract: Machine learning is becoming a vital tool for automating decision making processes in today's world. Machine learning algorithms evaluate information, spot trends and forecast outcomes to assist businesses in making wise decisions and also can help with disease diagnosis, prognostication, treatment plan personalization, resource allocation optimization by training machine learning algorithms on historical data too. In order to better understand how machine learning might be used on healthcare decision making, we did this research work.
We have tried to present a fine approach using clustering visualization to enhance health status based on some sentence and word analysis using some unsupervised machine learning algorithms like KMeans and Spectral clustering techniques that are most common to find hidden structures, correlations and trends in healthcare data based on speech signals that are not labeled. Fusion feature is an added advantage that has been created by combining several different distinct features. To facilitate pattern recognition and interpretation, we did cluster visualization using PCA as a feature reduction method on the features to verify the effectiveness of the suggested method on two of our primary datasets and lastly we have applied the same method on an online health dataset for comparison.
In the observation stage, the clustering visualizations have helped with health status results by revealing distinct cluster alterations linked to particular medical disorders. The overall research has impacted on significant potential applications on speech recognition and in future it may impact on speech therapy, real time disease detection and remote health monitoring systems.

Keywords*:* Assamese speech, MFCLBS, fusion, cluster, MFCFB.

1. Introduction

The Traditional methods generally involve invasive procedures for diagnostic tests. The basic purpose of this research is to discern changes in an individual's health by non-invasive means which has been a goal of medical research. The recent advancements in speech recognition, speech analysis and computational linguistics have opened up new avenues for health conditions that are both non-intrusive and cost-effective.
The human voice (Juang & Chen, 1998), which has been a reflection of various physiological, psychological factors, which has garnered significant role as a potential indicator of individual's health status (Delić et al., 2019; Kim et al., 1999)Some earlier models used for speech recognition are HMM,GMM (Ozerov et al., 2011) and some modern models like fuzzy logic & neural networks (Eljawad et al., 2019) used in previous research works. Also, many Assamese researchers have developed models for natural language processing (Chadha et al., 2015) or Assamese numeral speech recognition (Sarma & Sarma, 2011).
Speech analysis encompasses various linguistic levels, from the articulation of phonemes to the construction of complex sentences or words. Which levels are intricately linked to the physiological and psychological characteristics of the vocal tract, to make speech an ideal candidate for health status detection?
The subtle variations in speech recognition (Anusuya & Katti, 2010)patterns have been observed in individuals with various health conditions.
The most challenging and promising aspects of the research lies in deciphering the underlying patterns and creating a relationship tie linguistic content and vocal characteristics.
Developing a more accurate and efficient speech recognition system using clustering techniques to overcome the limitations of traditional recognition methods.
In the way to get the objective of the research, we had to create a system that can accurately group and categorize speech patterns or phrases into distinct clusters, enabling better decision-making processes across various applications, such as virtual assistants, transcription services, healthcare etc.
By implementing clustering algorithms, the aim is to enhance the recognition accuracy, reduce error rates, and improve the overall decision-making capabilities of voice signal identification systems in real time scenarios in unruly environments (Kim et al., 1999) or for robust speech (Afify &Siohan, 2004).
The context diagram of the general speech recognition system using clustering of speech features is depicted below.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

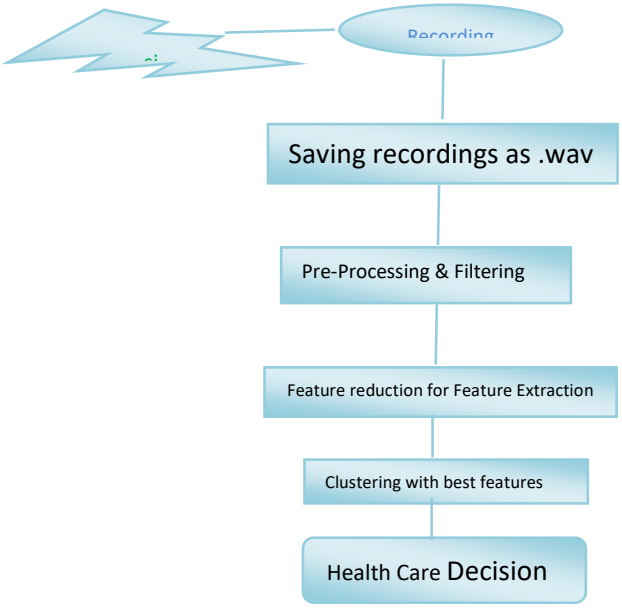

Figure 1: Speech Model Architecture

During this research, relating to the objective a comprehensive framework has been designed for developing some new fusion speech features and to apply an unsupervised approach (Marcu &Echihabi, 2002b)using clustering technique with these fusion features so that health status can be analyzed by visualizing the clusters. Our aim to uncover the hidden patterns in speech data with the application of fusion feature technique and implementing on health anomalies through the clustering visualization techniques which provide insights into early detection and treatment evaluation of health conditions. The main advantages of this research are accessibility, cost-effectiveness, data analytics and empowerment.

2. Description of Dataset

The speech recognition of Assamese language (Bora & Gupta, 2014; Talukdar et al., 2013)has been done by many previous researchers to achieve various objectives. Our research is to obtain healthcare status which is based on a primary dataset and a secondary dataset as well. As we know that a primary dataset is completely unique unlike the secondary dataset and the Assamese language does not have any built-in dataset, we have constructed our primary dataset. Some researchers have used the syllabication rules(Deng et al., 2005) for recognizing the Assamese words instead of constructed datasets and many researches have been done for the development of Assamese corpus(Sarma et al., 2013) which is a great achievement. In our research, between the two data sets, the first dataset construction consists of two pre-steps. After implementing these steps, the final construction of the dataset is done for the primary dataset. The second dataset is an online dataset named “overview_of_recordings” available in www.kaggle.com\Covid19datasets , where several speech features have been taken including both prosodic and statistical. The primary dataset with some feature values in .csv format is given below:

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

mfcc	hfcc	lfcc	MELBS	HFBS	LFBS	harmonicity	Modulation	MSME in c	fundament	mean freq	median fre	standard d	formant fre	formant ba	Zero cross	plp1	plp2
(20, 152)	-5.36E-06	1.71E-06	(0.3057666	(0.0225256	(0.0078373	(77760,)	0.070785	-0.99101	12	4.07E-05	0.000175	3182.725	0.00754	2650.482	0.160975	(1.0234154	(2.7885858
(20, 94)	-4.95E-06	-1.94E-05	(17.312849	(0.0051544	(0.0248802	(48096,)	0.070785	0.167269	24	0.005424	0.023112	3182.776	0.001724	2695.949	0.159554	(-3.959278	(-3.246114
(20, 123)	-4.34E-06	-2.34E-06	(-28.81371	(0.0079352	(0.0109643	(62496,)	0.070785	0.989802	6	0.005673	0.024363	6365.389	0.000554	3336.982	0.123117	(-3.820122	(-1.040900
(20, 134)	9.82E-06	-2.97E-05	(-32.49451	(0.0093920	(0.0041164	(68544,)	0.070785	0.334904	10	0.001638	0.007048	3182.736	0.003241	2714.681	0.152588	(-7.463684	(-2.356823
(20, 68)	-2.05E-05	-6.12E-07	(-27.09494	(0.0101609	(0.0345627	(34560,)	0.070785	-0.91642	8	0.023288	0.099388	3182.828	0.001363	2479.145	0.155711	(-3.289549	(-3.011672
(20, 91)	1.86E-05	-3.77E-06	(-20.05962	(0.0714859	(0.0016955	(46080,)	0.070785	0.961717	7	0.007192	0.03087	3182.781	0.006036	2530.526	0.129513	(-2.713878	(-1.129377
(20, 102)	1.17E-05	-3.44E-06	(-37.70540	(0.0023637	(0.0067626	(51840,)	0.070785	-0.99101	710	0.000127	0.000545	3182.766	0.003375	2561.052	0.157763	(-1.014277	(-2.760699
(20, 76)	1.40E-05	8.05E-07	(27.15369	(0.0601941	(0.0471239	(38880,)	0.070785	0.140609	6	0.00824	0.035323	3182.807	0.001186	2562.081	0.162398	(-5.263279	(-9.815822
(20, 57)	-6.82E-06	-4.16E-06	(28.844168	(0.0526392	(0.0398717	(29088,)	0.070785	0.231635	5	0.00141	0.006048	3182.862	0.001222	2477.319	0.132213	(5.3495894	(5.9637085
(20, 105)	1.82E-05	-1.20E-05	(-12.11574	(0.0021572	(0.0033552	(53280,)	0.070785	0.91403	4840	0.011911	0.050961	3182.763	0.001122	2594.435	0.17088	(1.4936333	(4.5582075
(20, 91)	-4.06E-05	-5.93E-06	(20.062255	(0.0145623	(0.0138692	(46368,)	0.070785	-0.89396	3	0.014777	0.063364	6365.424	0.000412	3449.675	0.119567	(5.0262684	(-4.208859
(20, 192)	-1.70E-05	5.50E-06	(38.890200	(0.0159878	(0.0226275	(97920,)	0.070785	-0.99101	8886	7.48E-05	0.00032	3182.708	0.007624	2850.428	0.156171	(1.7415263	(-1.240100
(20, 70)	-1.72E-05	-4.78E-06	(17.414231	(0.0035215	(0.0066349	(35424,)	0.070785	-0.93541	5	0.009663	0.041539	3182.823	0.006267	2560.231	0.185093	(4.3726941	(-7.590394
(20, 135)	-1.53E-05	9.38E-06	(18.973527	(0.0535567	(0.0050882	(68832,)	0.070785	0.898536	11	0.003138	0.013421	3182.736	0.004579	2668.696	0.164066	(-5.780798	(-2.520228
(20, 68)	2.02E-05	-3.96E-06	(-19.60482	(0.0034913	(0.1166143	(34560,)	0.070785	-0.12811	30	0.016948	0.072976	3182.828	0.002225	2343.027	0.151963	(1.7270135	(8.7840453
(20, 55)	-1.33E-05	1.30E-05	(26.307378	(0.0057068	(0.0006771	(27648,)	0.070785	0.804248	2130	0.013682	0.058743	3182.874	0.006695	2278.215	0.198775	(-8.223874	(-3.043044
(20, 137)	-1.23E-05	-9.52E-06	(-22.93999	(0.0075241	(0.0068852	(69696,)	0.070785	-0.99101	9	7.46E-05	0.000319	3182.735	0.013671	2832.81	0.163688	(-1.875858	(-3.297906
(20, 67)	-4.94E-06	-2.14E-06	(-25.90031	(3.7273280	(0.0001635	(33984,)	0.070785	0.735987	500	0.005223	0.022444	3182.831	0.005181	2439.818	0.209626	(8.3632619	(1.9907662
(20, 75)	-5.30E-06	2.03E-05	(-9.067599	(0.0355872	(0.0165682	(38304,)	0.070785	0.931149	6	0.003397	0.014558	3182.81	0.005874	2458.037	0.165254	(-6.084430	(-1.924753
(20, 56)	-1.69E-07	-3.75E-06	(9.7595550	(0.0287895	(0.0047295	(28224,)	0.070785	-0.84837	19	0.018476	0.079197	3182.869	0.009247	2239.994	0.200274	(4.9846748	(3.9028819
(20, 49)	3.07E-06	5.50E-06	(10.575914	(0.0145271	(0.0168286	(25056,)	0.070785	0.446988	2272	0.020291	0.086748	3182.897	0.005835	2543.518	0.103844	(2.2686545	(-2.423184
(20, 124)	9.51E-07	-1.68E-05	(36.519317	(0.0307380	(0.0237575	(63072,)	0.070785	-0.99101	8	0.000119	0.000509	3182.744	0.005943	2679.723	0.163228	(7.6605951	(1.1001544
(20, 114)	-1.75E-05	2.04E-06	(-6.196221	(0.0006345	(0.0014932	(58176,)	0.070785	-0.88167	113	0.00596	0.02557	3182.753	0.004189	2785.649	0.156139	(-6.546529	(-8.945562
(20, 92)	-1.53E-05	-1.12E-05	(-4.412526	(0.0082117	(0.0127434	(46656,)	0.070785	-0.69517	7	0.003676	0.015734	3182.78	0.002191	2603.264	0.150359	(1.1330671	(1.3187687
(20, 83)	-5.26E-06	-5.46E-06	(-22.03516	(0.0001693	(0.0047880	(42048,)	0.070785	-0.8397	3814	0.01501	0.064311	3182.795	0.001986	2365.316	0.144549	(-1.725323	(-2.423777
(20, 83)	-8.97E-06	-6.12E-06	(-23.15922	(0.0068715	(0.0127384	(42048,)	0.070785	-0.07144	597	0.027465	0.118047	3182.795	0.006687	2194.218	0.151973	(-1.081495	(-2.145301
(20, 89)	1.26E-05	-5.32E-06	(-24.53294	(8.9341722	(4.0362510	(45504,)	0.070785	-0.99101	17	0.000108	0.000462	3182.783	0.012453	2648.849	0.206773	(3.4665191	(7.5291817
(20, 81)	2.40E-06	8.53E-06	(13.287921	(0.0341435	(0.0050747	(41184,)	0.070785	-0.99904	5	0.007389	0.031693	3182.798	0.007623	2618.966	0.195771	(9.6616145	(7.3712268
(20, 150)	-1.70E-05	-2.95E-06	(-0.952259	(0.0569157	(0.0002221	(76608,)	0.070785	0.720032	140	0.001989	0.008522	3182.726	0.006779	2722.014	0.178451	(1.1501057	(1.5859068
(20, 48)	-7.65E-06	9.63E-06	(4.8980258	(0.0057216	(0.0043281	(24480,)	0.070785	-0.52135	16	0.020153	0.086332	3182.903	0.009257	2066.841	0.178752	(5.5728842	(2.8345154
(20, 75)	-6.16E-07	-6.46E-06	(-24.91250	(0.0051160	(4.5638730	(38304,)	0.070785	-0.99646	12	0.018988	0.081185	3182.81	0.005796	2042.911	0.135931	(8.6072426	(-1.811532
(20, 84)	2.60E-05	-6.36E-07	(4.8516508	(0.0440165	(0.0042517	(42624,)	0.070785	-0.9968	3867	0.0176	0.075516	3182.793	0.004156	2547.048	0.138434	(7.4681668	(1.6482803
(20, 87)	-3.70E-06	-1.54E-05	(17.762796	(0.0003251	(0.0014852	(44352,)	0.070785	-0.47429	552	0.011359	0.04855	3182.787	0.003855	2416.733	0.177235	(-6.408213	(-1.662286
(20, 104)	6.89E-06	-1.32E-05	(18.351353	(0.0014595	(0.0030428	(52992,)	0.070785	0.366268	8	0.002926	0.01258	3182.763	0.001968	2556.365	0.142883	(-3.754377	(-2.414248
(20, 59)	7.42E-06	-1.37E-05	(-24.47612	(0.1065615	(0.0004136	(29952,)	0.070785	0.283284	2667	0.016552	0.071189	3182.856	0.002226	2138.89	0.161489	(9.1095222	(4.6333426
(20, 77)	5.67E-06	-4.67E-06	(-19.97806	(0.0015512	(0.0065733	(39168,)	0.070785	0.045945	3022	0.010656	0.045635	3182.806	0.004648	2411.329	0.138811	(4.2086885	(3.3216976
(20, 101)	3.37E-06	-1.09E-07	(4.3677100	(0.0008704	(0.0094495	(51552,)	0.070785	-0.99101	9	7.68E-05	0.000329	3182.767	0.005081	2734.913	0.151517	(9.9237812	(1.8507485

2.1 Data Pre-processing

In this research, we have used an Audio Recorder App with two channels and 44100 HZ frequency range and selected for the 1st dataset and the 2nd dataset we have collected from a recording studio which was a previous conversation between a reputed doctor and some patients. For preprocessing, we have used LIBROSA as a library to do the feature extraction.

2.2 Data Filtering

Data filtering (Cheng et al., 2019)is very important for reducing the noise of the signal .In data filtering, we have used low pass filter with alpha value of 0.57 to make high-frequency noise-free audio signals. Wiener filter (Almajai& Milner, 2010; Thakuria& Talukdar, 2014) was a traditional filter among the various filters used by previous researchers

Alpha Value Influence: The alpha value, ranging between 0 and 1, controls the smoothing effect. 1 (e.g., 0.97) gives more weight to recent values, emphasizing the high-frequency components and closer to 0 gives more weight to past values, emphasizing the low-frequency components and reducing noise.

2.3. Feature Reduction

We have chosen feature reduction instead of feature selection (Anaraki et al., 2020; ArrutiIllarramendi et al., 2014; Zhang et al., 2020) technique as the feature selection have been already used by many researchers of the old days. The extracted features are reduced with PCA (Principal Component Analysis) which is a feature normalization(Huang et al., 2011) technique along with a feature dimension reduction technique (Dauda & Bhoi, 2014)so that no any missing value can affect the methodology. First, we stored the recorded audio file and collected audio files in .wav format and then we have used PCA(Dauda & Bhoi, 2014; Jia et al., 2022) with 28 speech features for preprocess the .wav files that have used. After this, the analysis identified several features for the subsequent steps. The 6th moment, 5th moment, PLP4, PLP3, and formant frequency stood out as promising options for further investigation.In some previous research works of speech technology, feature selection(Cohen &Zigel, 2002)techniques like regressions were used either for speaker segmentation(Delacourt & Wellekens, 2000)or in text dependent speaker(Zigel & Cohen, 2004) verification techniques.

In our research, the next step was to find out some best feature observations and to focus on a different set of features from subsequent analysis and clustering efforts as some of the features are statistical features. We have used "Mel-

frequency Cepstral Coefficients (MFCC), HFCC, Filter Bank of Mel Spectrogram, Zero Crossings Ratio and Root Mean Square" as the primary features. Excluding our observed feature values, we used a literature survey on the fusion of speech features and found more fusion features. With the help of best selected feature set and also the fusion feature set in place, then proceeded to perform clustering using Spectral clustering. These clustering techniques are commonly used in data analysis to group data points with similar characteristics or patterns into clusters or categories.

The research is to gain insights into the relationships and groupings within the audio data based on the chosen features, to extract meaningful information or patterns that could be valuable for various applications such as speech recognition, audio classification or signal processing. We have applied both the early (concatenation) and late (Addition & Averaging) fusion techniques for the fusion of features. The final Fusion Features developed for the project for the implementation of clustering technique are listed below:

Fusion_Feature Name	**Feature Combination**
mfcc_fb	mfcc+fb
mfccf_fb	mfcc+f0+fb
melbs_zcr	melbs+zcr
Plps	Avg(plp+plp2+plp3+plp4)
Fstft	Fb+stft

Table1: Fusion Features

2.4. Construction of Datasets

We have considered the "Assamese Language" for this research and collected three sentences from 350 people and all the people are from local areas of Assam. This technique we applied for the first dataset. Thus, the dataset consists of almost 1000 sentences namely "feeling good", "feeling not good" and "feeling bad" and 28 columns (i.e. 28 features).The sentences recorded in Assamese language are pronounced like "VAAL", "VAAL NOHOI", "BEYA". And for the second dataset, we have taken an online dataset collected from a Google. The second dataset is an online speech dataset used during Covid19 that consists of 1000 rows and 13 columns (13features). In the last phase, we have applied the concept of fusion technique by which we have developed total 5 fusion feature out of 10 from our existing extracted features. The different datasets are used for two types of recorded sounds with the extracted features without any speech segmentation. The recorded sounds are stored in the form of a .wav file and then extracted the speech features from the two types of sounds individually.

3. Methodology

The methodology involves the collection of speech from a diverse group of individuals and each with varying health issues. Through this we extracted linguistic features from the speech data samples, encompassing phonetic information, prosodic cues, and syntactic structures. Here we have captured the vocal features such as pitch, amplitude, and resonance. Through advanced clustering algorithms, we have found out multi-dimensional feature sets into clusters that reveal shared speech and vocal patterns among individuals with similar health status. First, we have tested on vowels and consonants. Then on the reduced feature set of both the datasets for getting some perfect clusters. In the first dataset clustering is done for making three clusters for the three individual sentences and in the dataset clustering is done for two types of people one cluster having Covid19 symptoms and the other not having any Covid19 symptoms.

3.1. Implementation Details:

For this research work, mainly KMeans and Spectral clustering (Marcu &Echihabi, 2002a; Wessel & Ney, 2004; Yadav & Singh, 2016) algorithms are used to implement on each of the two datasets individually with different sets of features at different times. Algorithm 1 shows the first clustering technique and algorithm 2 shows the second clustering technique that we have implemented. We have first implemented algorithm1 and after visualizing the clusters we have implemented algorithm2 for better visualization**.**

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

Algorithm 1 KMeans Clustering(*Nptel Clustering Algorithms*)

Input :
E={ e1,e2, ……..En} //Set of elements
C //No. of desired clusters
Output :
C //Set of clusters
KMeans Algorithm:
Assign initial values for M1,M……Mk
repeat
assign each item Ei to the clusters which has calculated new mean for each cluster;
until convergence criteria is achieved;

Algorithm 2 Spectral Clustering(*Nptel Clustering Algorithms*)

Input :
1.Let X^t be the set of points i.e. graph vertices) to be clustered at time t:
$X=(x_1,\ldots\ldots\ldots x_n, x_i \in R^d$ and C the no. Of clusters
2. Compare the affinity matrix W of x^t
3. Transform W into Laplacian matrix
4.Compute the first c smallest eigenvalues and the corresponding eigenvectors of L
5. Form the matrix $V=[v_1\ldots..v_k]$ containing the corresponding eigenvectors whose no. of rows is n(the size of the data)
6. Apply a clustering algorithm to cluster V
7.Assign the vertices x_i(i=1……..n data points) to cluster j if and only if the rows of the vectors v_1 were assigned to cluster j in step 6

3.2. Comparison of Our Methodology with IOT: Both IOT-based and machine learning-based projects can be valuable for health checkup applications. They serve different purposes and have distinct advantages. K-Means and Spectral clustering individually, gives the quality of clusters and any insights gained.

- **Real-time Monitoring**: IoT-based health checkup plans are able to give concurrent information on imperative signs and health parameters, for immediate detection of anomalies. On the other hand, Machine learning models often require a batch of data for analysis and might not offer real-time monitoring.
- **Reduced Data Transfer**: IoT devices preprocess data locally and give relevant information and thus it can reduce the amount of data to be transmitted and processed in the cloud, which can save the cost of bandwidth and cloud processing.
- **Reliability**: IoT devices are designed for specific purposes and are less prone to model degradation or performance variations over time.
- **Lower Cost**: IoT-based solutions can often be more cost-effective to implement and maintain than machine learning-based solutions, which may require more computational resources.
- **Customization**: IoT devices are used for specific health monitoring needs and can integrate a variety of sensors and actuators, for flexibility in designing solutions for different healthcare scenarios.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

3.3. Feature Fusion and Optimization

Fusion of Feature or the integration of characteristics from distinct feature or attribute, is an essential component of contemporary network design. It is commonly done using basic operations like summation or concatenation; however this may not be the best option. There are two current feature fusion approaches. One method is to merge two or more sets of feature vectors into a single amalgamation vector and then mine features in the upper dimensional true vector space. Another option is to join two sets of feature vectors using an intricate vector and during this study, we used two significant approaches to fuse features: early fusion transformed add, concatenate method, and late fusion averaging method. Then, mine features from the complicated vector. The two fusion techniques applied in this research are briefed below.

3.3.1. Addition Technique for Fusion Features:

It signifies that we will merge two vectors into one. For example, \ (A + B = C\). If \ (A\) and \ (B\) have the same shape, matrix multiplication can be used to project them as the same.

3.3.2. Concatenation Technique for Fusion Features

This signifies that we will concatenate the features into a vector. Concatenating \ (A\) and \ (b\) creates a 1*(m+n) vector. The context d iagram of the proposed speech recognition system using clustering of speech fusion features is depicted below.

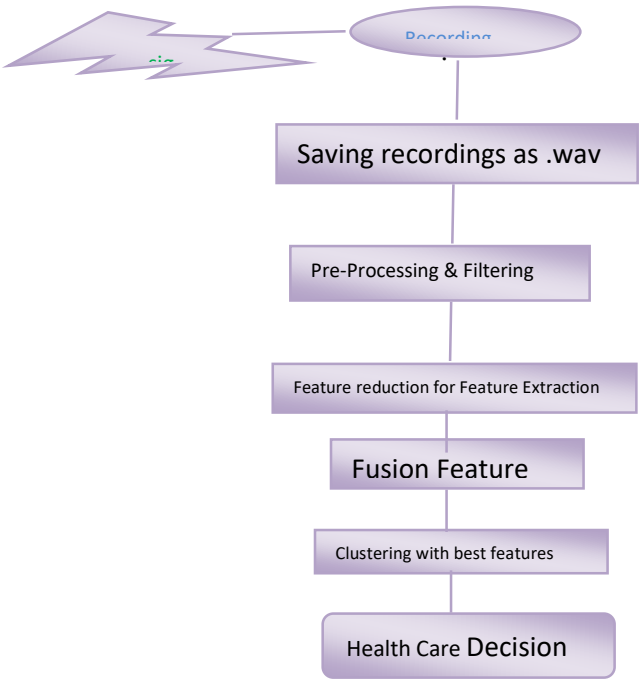

Figure 2: Proposed Speech Model Architecture

3.4. Significance

The research focuses on the potentiality of the development of fusion features and their impact on health monitoring system with the help of a non-invasive, real-time assessment tool. Some earlier model of speech emotion recognition (Chen et al., 2022)for speaker classification(Suresh et al., 2017) have also been analyzed. Our proposed system offers a holistic perspective, bridging the gap between linguistics(Metze et al., 2009), para linguistics(Cai et al., 2017), signal processing (Dragomiretskiy&Zosso, 2013), feature development and medical diagnostics(Allan &Arroll, 2014; Eccles, 2005; Huckvale& Beke, 2017). Moreover, the visual nature of the clustering results facilitates easy interpretation, allowing healthcare practitioners and researchers to quickly identify and understand the speech and vocal features indicative of particular health conditions.

4. Results

For quality of clustering, some common measuring indexes like the Silhouette score and Purity have been used. After clustering of the recorded speech and speech features, it is evaluated the results of the respective clustering techniques by considering three factors to compare the accuracies. The Silhouette score is a metrics which is used to assess the quality of clusters. Purity is another measuring index that measures the ability of clustering (*Nptel Clustering Algorithms*; Sazonov, 2010)and is applicable even when the no. of the cluster is different from the number of known classes. When it

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

comes to clustering, "purity" refers to how distinct and uniform the clusters are. It gauges how closely related a cluster's elements are to one another in terms of class or category. When a cluster has a high purity, it means that the majority of its items are from the same class; when it has a low purity, it means that some things are intermingled from different classes inside the cluster.

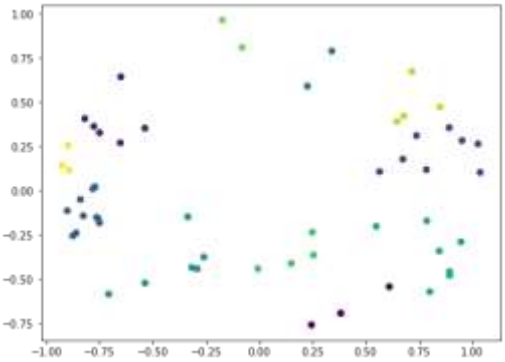

Figure3:Vowels_KMeans

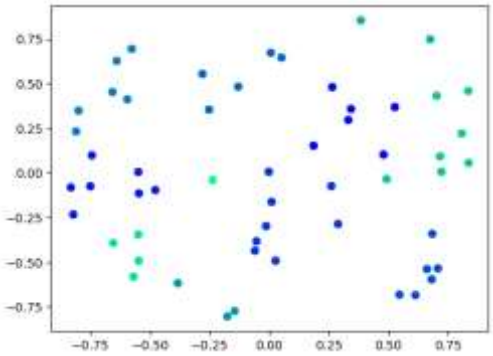

Figure4:Vowels_Spectral

In Figure1 and Figure2 we achieved the cluster results by testing with the 11 vowels of Assamese language. In other figures, we obtained the cluster visualization for three clusters with the first dataset in figure number 3,4,5,6 respectively.

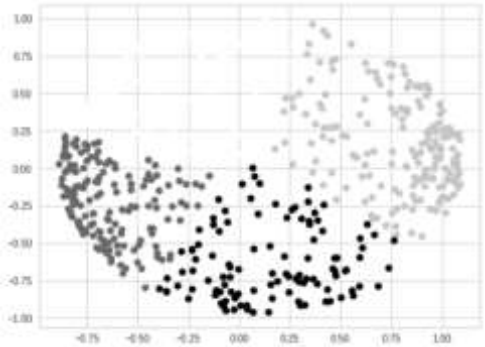

Figure5:1st_Dataset_KMeans

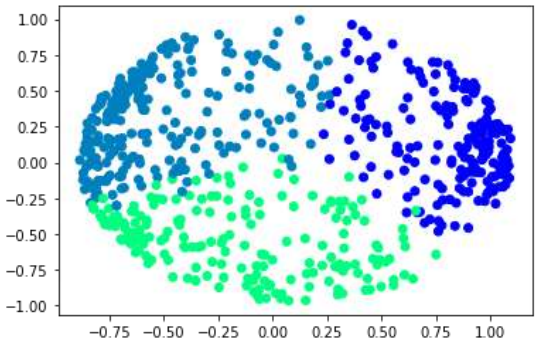

Figure6: 1st_Dataset _Spectral

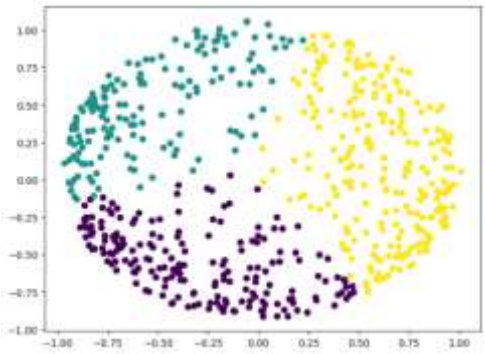

Figure7: 1st_Dataset_Fusion_KMeans

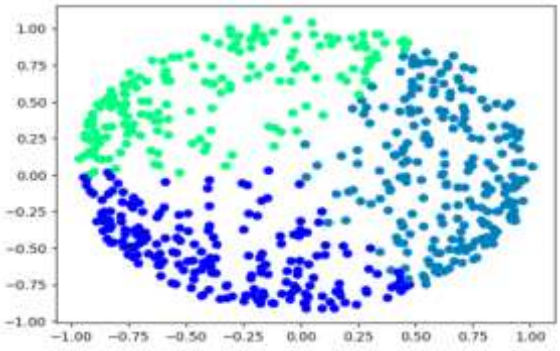

Figure8: 1st_Dataset _Fusion_Spectral

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

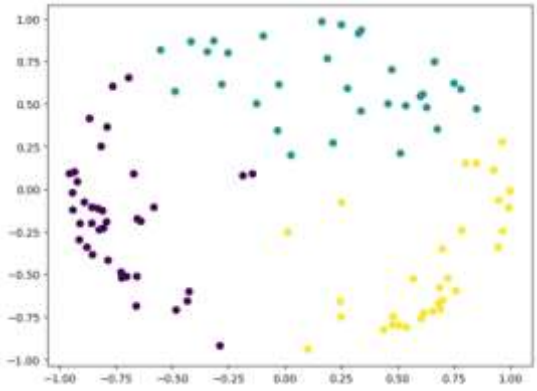

Figure9:1st_Dataset_Best_Fusion_KMeans

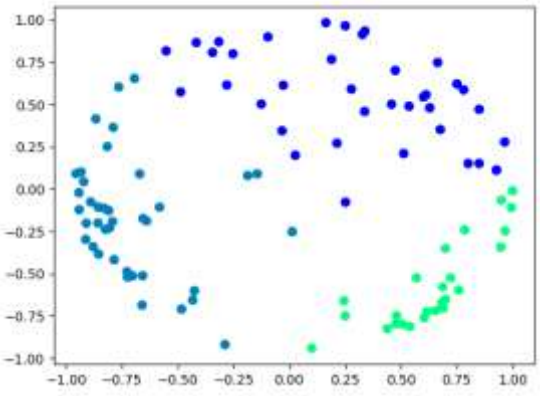

Figure10: 1st_Dataset_Best_Fusion_Spectral

The findings indicate that feature fusion, especially the first, third, and fifth fusion features(Bharali & Kalita, 2018; Metze et al., 2009) yields better results. Future research can delve deeper into feature selection, extraction, and fusion techniques to enhance the performance of speech analysis algorithms. This could involve exploring various combinations of features along with machine learning models to optimize accuracy.
The total purity of the clustering will be lower if Clusters 1 and 2 have high purities (almost 1) and Cluster 3 has a poor purity after mixing multiple classes. To sum up, purity quantifies the degree to which clusters are distinct and uniform in terms of the classes they comprise.

Clustering	Silhouette Coefficient	Purity
KMeans	**.55**	**.85**
Spectral	**.58**	**.91**

Table 2: Dataset1 1st feature and 5th fusion feature

Clustering	Silhouette Coefficient	Purity
KMeans	**.58**	**.90**
Spectral	**.63**	**.94**

Table 3: Dataset1 Top5 features and 5th fusion feature

	Silhouette Coefficient	Purity
KMeans	**.65**	**.87**
Spectral	**.68**	**.92**

Table 4: Dataset1 Top5 features and 3rd fusion feature

	Silhouette Coefficient	Purity
KMeans	**.69**	**.91**
Spectral	**.72**	**.97**

Table 5: Dataset1 Top5 features and 3rd & 5th fusion feature

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

	Top5 features	**Top5 & 1st fusion feature**	**Top5 & 2nd fusion feature**	**Top5 & 3rd fusion feature**	**Top5 & 4th fusion feature**	**Top5 & 5th fusion feature**
KMeans Clustering	**.765**	**.68**	**.71**	**.87**	**.79**	**.90**
Spectral Clustering	**.70**	**.72**	**.75**	**.92**	**.84**	**.94**

Table 6: Purity of Clustering for Database **1**

From the above tables, we have found that both the clustering indexes Silhouette score and purity increases when the number of features increased excluding some of the fusion features. The cluster visualization using PCA of the 2nd dataset for both the KMeans and Spectral clusters are mentioned below where the clustering is done or mainly two cluster, one cluster indicates patient having covid19 symptoms and another one not having any covid19 symptoms. Moreover, cluster measure indexes for both the datasets have been listed in table format below in tables 7 and 8 respectively.

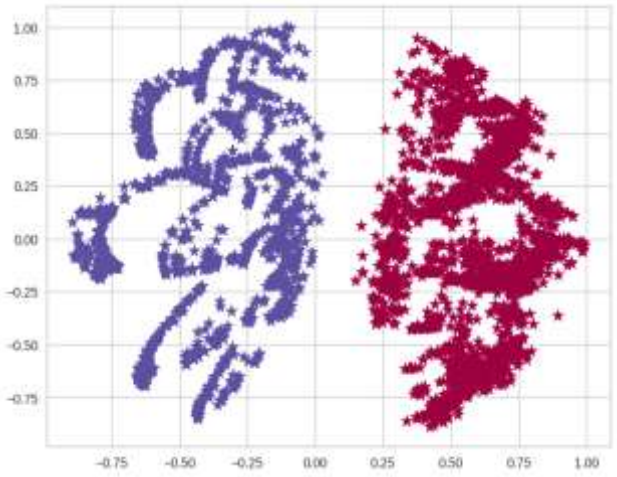

Figure 11: 2nd_Dataset_KMeans

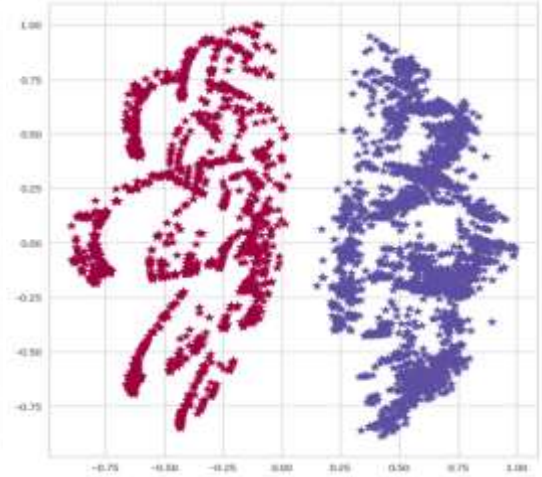

Figure 12: 2nd_Dataset_Spectral

Datasets	**KMeans Clustering (Silhouttee Coefficient)**	**Spectral Clustering (Silhouttee Coefficient)**
Dataset1(Three Assamese Sentences	**.75**	**.84**
Online Dataset	**.52**	**.66**

Table 7: Clustering Results for the all fusion features (mfcc_fb, mfccf_fb, melbs_zcr, plps, fstft for the 1st dataset)

Datasets	**KMeans Clustering (Purity)**	**Spectral Clustering (Purity)**
Dataset1(Three Assamese Sentences)	**.94**	**.98**
Online Dataset	**.68**	**.77**

Table 8: Clustering Results for the three fusion features (mfcc_fb, melbs_zcr, fstft)

The metrics we referring to, Silhouette coefficients and purity are commonly used to evaluate the machine learning results using clustering. Here's a breakdown of our observations:
1. Clustering of the top 5 features and fusion of 5 features: Silhouette coefficient and purity are good.
2. Clustering of the top 5 features and three fusion features: This combination gives even better results in terms of purity.
3. Clustering of 1 and 2 top features or 1 and 3 best fusion features: This combination is giving better results in terms of purity.
Surprisingly, in some cases, using only a few features (1 or 2) along with 3 or 5 best fusion features results in the best cohesion and separation, even if the Silhouette coefficient is not very high. The results obtained have pointed the number and choice of features which can significantly impact the quality of clustering. But in every observation, it is found that our feature clustering results are comparatively better than that of the clustering of the online (from Kaggle website)

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 05 April 2024, Revised: 15 May 2024, Accepted: 25 May 2024

dataset of speech features. From Tables 7,8 and 9, it has been observed that the clusters are well separated with tight coupling of the samples within each cluster which determines the good quality of clustering that has been done for all the clusters.

Spectral Cluster Measures	Dataset1	Online Dataset of Speech Features
Silhouette coefficient	.75(Avg)	.66(Avg)
Purity	.94(Avg)	.79(Avg)

Table 9: Final Comparison of Cluster Indexes

From the above table it has been observed that all the clustering measures with fusion features in all the datasets used during the research are better than that of the online dataset where most of the features are similar to our speech features contained in the datasets.

5. Conclusion

The clustering visualization of some particular words in the second dataset and some sentences in the first dataset holds promise as an innovative method for health status assessment. The research paper sets the stage for a deeper understanding of the intricate relationships between speech patterns and health conditions by merging linguistic and vocal analyses. We anticipate that the insights gained from this research will play a significant part towards the advancement of early detection systems, personalized healthcare, and improved wellness management. From the tables, we have concluded that the fusion of features is giving better results than that of without fusion. The first, third and fifth fusion features are giving better results than that of the other fusions. Integrating voice-based technologies into healthcare systems can contribute to sustainable development by improving healthcare access, efficiency, and effectiveness.

6. Future Scope

The clustering visualization in case of words and sentence level based on linguistic and vocal analyses be further refined and expanded for different health issues. Future research can also explore to more sophisticated algorithms for extracting and interpreting patterns in speech that are indicative of various health conditions which lead to more accurate and sensitive health assessment. In Future, it can focus on designing real-time or near-real-time monitoring systems that use speech patterns to identify potential health in the early stage. Future studies could investigate how tailored interventions and treatments can be developed by analyzing an individual's speech patterns.

In advanced stage of machine learning and artificial intelligence continue to incorporate these cutting-edge methods into speech analysis for health assessment which lead to more robust and automated systems for health monitoring. With the increasing use of speech data in health assessment, ethical considerations and privacy protection become crucial.

Acknowledgement

This paper entitled “Fusion Feature: Enhancing the Custer Visualization” would not have been possible without the support and encouragement of my Ph.D. guide Dr. Manoj Kr. Sarma in the first place as he directly offered his help and cooperation towards the completion of the work. Secondly I am very fortunate to have so good and supportive parent which always encourage me the research work .I would also like to convey my heartfelt appreciation and cordial thanks to two very great persons Dr. Aniruda Deka sir, an Associate Professor in the Department of Computer Science & Engineering Department of Assam Down town University, Dr. Jyotismia Talukdar madam, an Associate Professor in the Department of Computer Science & Engineering Department of Tezpur Univerisity and Dr. Laba r. Thakuria, the deputy register of the Rabindra Nath Tagore University, Assam.

References:

1. Afify, M., &Siohan, O. (2004). Sequential estimation with optimal forgetting for robust speech recognition. *IEEE Transactions on speech and audio processing*, *12*(1), 19-26. https://doi.org/10.1109/TSA.2003.819954.
2. Allan, G. M., &Arroll, B. (2014). Prevention and treatment of the common cold: making sense of the evidence. *Cmaj*, *186*(3), 190-199. https://doi.org/https://doi.org/10.1503/cmaj.121442
3. Almajai, I., & Milner, B. (2010). Visually derived wiener filters for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(6), 1642-1651. https://doi.org/10.1109/TASL.2010.2096212.
4. Anaraki, J. R., Moon, J., & Chau, T. (2020). Revisiting the Application of Feature Selection Methods to Speech Imagery BCI Datasets. *arXiv preprint arXiv:2008.07660*. https://doi.org/10.48550/arXiv.2008.07660
5. Anusuya, M., & Katti, S. K. (2010). Speech recognition by machine, a review. *arXiv preprint arXiv:1001.2267*. https://doi.org/https://doi.org/10.48550/arXiv.1001.2267
6. ArrutiIllarramendi, A., Cearreta Urbieta, I., Álvarez, A., Lazkano Ortega, E., & Sierra Araujo, B. (2014). Feature Selection for Speech Emotion Recognition in Spanish and Basque: On the Use of Machine Learning to Improve Human-Computer Interaction. https://doi.org/https://doi.org/10.1371/journal.pone.0108975

7. Bharali, S. S., & Kalita, S. K. (2018). Speaker Identification with reference to Assamese language using fusion technique. 2018 Second International Conference on Advances in Computing, Control and Communication Technology (IAC3T) (pp. 81-86). IEEE DOI: 10.1109/IAC3T.2018.8674020,
8. Bora, D. J., & Gupta, D. A. K. (2014). A comparative study between fuzzy clustering algorithm and hard clustering algorithm. *arXiv preprint arXiv:1404.6059*. https://doi.org/10.48550/arXiv.1404.6059
9. Cai, D., Ni, Z., Liu, W., Cai, W., Li, G., Li, M., Cai, D., Ni, Z., Liu, W., & Cai, W. (2017). End-to-End Deep Learning Framework for Speech Paralinguistics Detection Based on Perception Aware Spectrum. INTERSPEECH (pp. 3452-3456). http://dx.doi.org/10.21437/Interspeech.2017-1445(pp. ,
10. Chadha, N., Gangwar, R., & Bedi, R. (2015). Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey. *Int. J. Comput. Appl*, *131*(11), 28-31. https://doi.org/ 10.5120/ijca2015907471
11. Chen, L., Ren, J., Mao, X., & Zhao, Q. (2022). Electroglottograph-based speech emotion recognition via cross-modal distillation. *Applied Sciences*, *12*(9), 4338. https://doi.org/https://doi.org/10.3390/app12094338
12. Cheng, K., Zhang, C., Yu, H., Yang, X., Zou, H., & Gao, S. (2019). Grouped SMOTE with noise filtering mechanism for classifying imbalanced data. *IEEE Access*, *7*, 170668-170681. https://doi.org/doi: 10.1109/ACCESS.2019.2955086.
13. Cohen, A., &Zigel, Y. (2002). On feature selection for speaker verification. Proc. COST 275 workshop on The Advent of Biometrics on the Internet (pp. 89-92). https://www.semanticscholar.org/paper/ON-FEATURE-SELECTION-FOR-SPEAKER-VERIFICATION-Cohen-Zigel/9a831fc37dc5a1be473b68675730ddf6d561406c
14. Dauda, A., & Bhoi, N. (2014). Facial expression recognition using PCA & distance classifier. *Int J Sci Eng Res*, *5*. http://www.ijser.org/
15. Delacourt, P., &Wellekens, C. J. (2000). DISTBIC: A speaker-based segmentation for audio data indexing. *Speech communication*, *32*(1-2), 111-126. https://doi.org/https://doi.org/10.1016/S0167-6393(00)00027-3
16. Delić, V., Perić, Z., Sečujski, M., Jakovljević, N., Nikolić, J., Mišković, D., Simić, N., Suzić, S., &Delić, T. (2019). Speech technology progress based on new machine learning paradigm. *Computational intelligence and neuroscience*, *2019*. https://doi.org/https://doi.org/10.1155/2019/4368036
17. Deng, L., Droppo, J., & Acero, A. (2005). Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion. *IEEE Transactions on speech and audio processing*, *13*(3), 412-421. https://doi.org/doi: 10.1109/TSA.2005.845814
18. Dragomiretskiy, K., &Zosso, D. (2013). Variational mode decomposition. *IEEE transactions on signal processing*, *62*(3), 531-544. https://doi.org/doi: 10.1109/TSP.2013.2288675.
19. Eccles, R. (2005). Understanding the symptoms of the common cold and influenza. *The Lancet infectious diseases*, *5*(11), 718-725. https://doi.org/DOI:https://doi.org/10.1016/S1473-3099(05)70270-X
20. Eljawad, L., Aljamaeen, R., Alsmadi, M., Almarashdeh, I., Abouelmagd, H., Alsmadi, S., Haddad, F., Alkhasawneh, R., & Alazzam, M. (2019). Arabic voice recognition using fuzzy logic and neural network. *Eljawad, L., Aljamaeen, R., Alsmadi, MK, Al-Marashdeh, I., Abouelmagd, H., Alsmadi, S., Haddad, F., Alkhasawneh, RA, Alzughoul, M. & Alazzam, MB*, 651-662. https://ssrn.com/abstract=3422784
21. Huang, C.-L., Tsao, Y., Hori, C., &Kashioka, H. (2011). Feature normalization and selection for robust speaker state recognition. 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA) (pp. 102-105). IEEE,
22. Huckvale, M. A., & Beke, A. (2017). It sounds like you have a cold! Testing voice features for the Interspeech 2017 Computational Paralinguistics Cold Challenge. International Speech Communication Association (ISCA).
23. Jia, W., Sun, M., Lian, J., & Hou, S. (2022). Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, *8*(3), 2663-2693. https://doi.org/https://doi.org/10.1007/s40747-021-00637-x
24. Juang, B. H., & Chen, T. (1998). The past, present, and future of speech processing. *IEEE signal processing magazine*, *15*(3), 24-48. https://doi.org/doi: 10.1109/79.671130.
25. Kim, D.-S., Lee, S.-Y., & Kil, R. M. (1999). Auditory processing of speech signals for robust speech recognition in real-world noisy environments. *IEEE Transactions on speech and audio processing*, *7*(1), 55-69. https://doi.org/doi: 10.1109/89.736331
26. Marcu, D., &Echihabi, A. (2002a). An unsupervised approach to recognizing discourse relations. Proceedings of the 40th annual meeting of the association for computational linguistics (pp. 368-375).https://doi.org/10.3115/1073083.1073145
27. Metze, F., Polzehl, T., & Wagner, M. (2009). Fusion of acoustic and linguistic features for emotion detection. 2009 IEEE International Conference on Semantic Computing.(pp. 153-160). IEEE. DOI:10.1109/ICSC.2009.32
28. *Nptel Clustering Algorithms* http://nptel.ac.in/courses/106108057/module14/lecture34.pdf.
29. Ozerov, A., Lagrange, M., & Vincent, E. (2011). GMM-based classification from noisy features. International Workshop on Machine Listening in Multisource Environments (CHiME 2011),
30. Sarma, H., Saharia, N., Sharma, U., Sinha, S. K., & Malakar, M. J. (2013). Development and transcription of Assamese speech corpus. *arXiv preprint arXiv:1309.7312*. https://doi.org/10.48550/arXiv.1309.7312

31. Sarma, M. P., & Sarma, K. K. (2011). Assamese numeral speech recognition using multiple features and cooperative LVQ-architectures. *International Journal of Electronics and Communication Engineering*, *5*(9), 1263-1273. https://doi.org/http://scholar.waset.org/1307-6892/12746
32. Sazonov, E. (2010). Clustering (xu, r. and wunsch, dc; 2008)[book review]. *IEEE Pulse*, *1*(1), 74-76. https://doi.org/doi: 10.1109/MPUL.2010.937237.
33. Suresh, A. K., KM, S. R., & Ghosh, P. K. (2017). Phoneme State Posteriorgram Features for Speech Based Automatic Classification of Speakers in Cold and Healthy Condition. INTERSPEECH (pp. 3462-3466). DOI:10.21437/Interspeech.2017-1550
34. Talukdar, P., Sarma, M., & Sarma, K. K. (2013). Recognition of Assamese SpokenWords using a Hybrid Neural Framework and Clustering Aided Apriori Knowledge. *WSEAS Transactions on Systems*(7), 360-370. https://www.wseas.org/wseas/cms.action?id=6952
35. Thakuria, L. K., & Talukdar, P. (2014). Automatic Syllabification Rules for ASSAMESE Language. *International Journal of Engineering Research and Applications*, *4*(2), 446-450. www.ijera.com
36. Wessel, F., & Ney, H. (2004). Unsupervised training of acoustic models for large vocabulary continuous speech recognition. *IEEE Transactions on speech and audio processing*, *13*(1), 23-31. https://doi.org/doi: 10.1109/TSA.2004.838537.
37. Yadav, A., & Singh, S. K. (2016). An improved K-means clustering algorithm. *International Journal of Computing*, *5*(2), 88-103. https://doi.org/http://www.meacse.org/ijcar
38. Zhang, B., Titov, I., Haddow, B., &Sennrich, R. (2020). Adaptive feature selection for end-to-end speech translation. *arXiv preprint arXiv:2010.08518*. https://doi.org/10.48550/arXiv.2010.08518
39. Zigel, Y., & Cohen, A. (2004). Text-dependent speaker verification using feature selection with recognition related criterion. ODYSSEY04-The Speaker and Language Recognition Workshop 12(1), 19-26.,www.semanticscholar.org › paper › Text-dependent