International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

Text-Independent Automatic Dialect Recognition of Marathi Language using Spectro-Temporal Characteristics of Voice

[1]Rahesha Mulla, [2]Dr. B. Suresh Kumar
[1]School of Computer Science & Engineering, Sanjay Ghodawat University, Kolhapur, Maharashtra, India
rahesha@sanjayghodawatuniversity.ac.in
[2]Dean School of Computer Science & Engineering, Sanjay Ghodawat University, Kolhapur, Maharashtra, India
b.sureshkumar@sanjayghodawatuniversity.ac.in

Abstract— Text-independent dialect recognition system is proposed in this paper for Marathi language. India is rich in language varieties. Each language in turn has its unique dialect variations. Maharashtra has Marathi as official language and for Goa it is a co-official language . In literature there are very few studies available for Indian language recognition and then their respective dialect recognition. So research work available for regional languages such as Marathi is extremely limited. As a part of research work, an attempt is made to generate a case study of a low resourced Marathi language dialect recognition system. The study was carried out using Marathi speech data corpus provided by Linguistic Data Consortium for Indian Language (LDC- IL). This corpus includes four major dialects of Marathi speakers. The efficiency and performance evaluation of the explored spectral (rhythmic) and temporal features are carried out to perform classification tasks. We investigated the performance of six different classifiers; K-nearest neighbor (KNN), Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT) classifier , Stochastic Gradient Descent (SGD) classifier and Ridge Classifier (RC). Experimental results have demonstrated that the RC classifier worked well with 84.24% of accuracy for fifteen spectral and temporal features. With twelve MFCCs it has been observed that SGD has outperformed among all classifiers with accuracy of 80.63%. For further study, a prominent feature subset as a part of dimensionality reduction has been identified using chi square, mutual information and ANOVA-f test. In this chi-square based feature extraction method has proven to be the best over over mutual information and ANOVA f-test.

Keywords- Marathi dialect recognition, Spectral & temporal features, Dimension reduction,Chi-square, Mutual information , ANOVA f-test, Machine learning models, Prediction accuracy .

I. INTRODUCTION

Major challenge for existing research in Automatic Dialect Recognition (ADR) is to recognize dialectal variation in spokenform of language. An individual has own speaking styles due to dialect and accent variations because of social and regional background [5]. Person recognition by his/her voice is a form of biometric verification that identifies a person by unique voice characteristics [13].The spoken language identification system has three main parts, data collection, feature extraction, and language classification[26]. Problem of automatic dialect recognition can be treated as a problem under spoken language identification of computational linguistics (CL). Basically, computational linguistics is a field that includes study of linguistics and computer science. Computational linguistics aims to solve linguistic issues with the help of computers. Computational linguistics is an automatic processing of natural languages using computer programs. The form of language can be written(text) form or it may be spoken(audio) form . Natural Language Processing (NLP) is an alternate term for computational linguistics. There is much research scope in NLP to support wide variety of applications such as text based and dialogue based [17]. However, there are limited ASR and dialect recognition tasks are available for almost all the Indian languages [6]. As a part of future research in speech recognition , dialect recognition of low resourced language will get potential attension [18]. Its also important to understand that existing studies are not providing much comparative analysis of different feature extraction techniques with multi-classifier approach. Our research objective is to evaluate machine learning model performance using three feature selection techniques namely chi-square, mutual information and ANOVA f-test for six single classifiers.

II. SPEECH CORPUS

Collection of audio files and their transcriptions makes speech corpus .Speech corpus is the prime requirement of acoustic-feature based dialect recognition systems. Speech

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

corpus named "Marathi Raw speech Corpus "was requested and received from Linguistic Data Consortium-Indian Languages(LDC-IL) [20].This corpus has recordings(.wav) of type sentence level (S: SENTENCE). Detailed dataset description of speech corpus used for research work is given below in Table 1 .

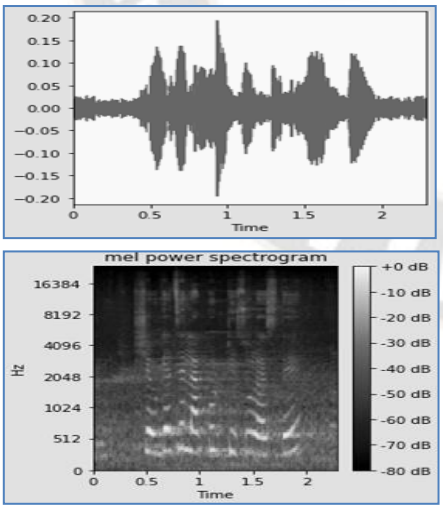
Table 1 Description of LDCIL-Marathi Raw speech corpus

Sr. No		Dataset description
1	Input type	Speech /Audio (.wav file with sampling rate 48.0 KHz)
2	Gender	Male and Female
3	Age	Group 1: 16-20, Group 2:21-50 and Group 3: 51 above
4	No of recordings	7555

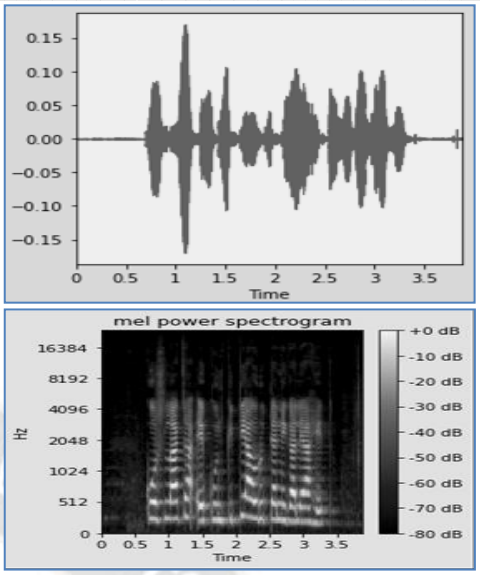
As mentioned above in Table 1 , the speech corpus used for study includes 7555 speech samples having a sampling rate of 48.0 KHz.

III. SPEECH SIGNAL AND FEATURE EXTRACTION

Speech signal processing is a prominent reaserch are for reserachers involved in recognition studies [8]. In general, non-stationary, complex human speech is composed of many frequency components; A spectrogram/sonogram is used to visualize the speech signal. Two geometric dimensions of speech signals are time and frequency. An example spectrogram for Marathi speech signal is shown in figure. 1. below.



(a) Marathwada speech signal



(a) Puneri speech signal

Figure1 Marathi speech variation and their respective spectrograms

For Machine Learning (ML) algorithms its important to understand rules of classification when raw input is provided in the form of image, text or signal. To accomplish this there is a need of numerical feature values [1]. However, feature extraction from a speech signal is a challenging research area with great significance to the speaker and speech recognition [9]. When speech signals are decoded by evaluating the frequency characteristics it becomes more useful. So, features selected for study are spectral features. Human speech is featured by monotonies in the occurrence of its basic elements.This can be temporal rhythmic pattern in speech [2]. Table 2 below gives spectral and temporal parameters and label/categories used for Marathi language recognition research work. To make a feature set, we used librosa speech processing library [31].

Table 2: Acoustic (Spectro-temporal) features and categories used for proposed study

Feature description (Spectral & Rhythmic)		
No. of features: 15		
Feature ID	**Feature Name**	**Description**
1	ch_stft	Chromagram
2	spec_ cent	Spectral Centroid
3	spec_bw	Spectral Bandwidth
4	spec _rolloff	Spectral Rolloff
5	Rmse	Root Mean Square Error
6	Zcr	Zero Crossing Rate
7	Mfcc	Mel Frequency Cepstral Coefficient
8	ch_cqt	Constant Q –Chromagram

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

9	ch_cens	Chroma Energy Normalized
10	mel_spec	Mel-scale spectrogram
11	spec_contrast	Spectral Contrast
12	spec_flatness	Spectral Flatness
13	Poly	Coefficient of nth order polynomial
14	Tone	Tonal Centroid
15	Tempo	Tempogram

Categories		
No. of categories:04		
Class ID	**Labels**	**Description**
L1	Marathwada	Marathi dialect being spoken in Aurangabad, Beed, Hingoli, Jalna, Latur, Nanded, Osmanabad and Parbhani
L2	Puneri	Marathi dialect being spoken in Pune
L3	Goa	Marathi dialect being spoken in Goa
L4	Vidharbha	Marathi dialect being spoken in Amravati, Yavatmal, Akola, Buldhana, Wardha, Washim, Chandrapur, Gadchiroli, Gondiya

IV. MARATHI DIALECT RECOGNITION

The purpose of Language Identification (LID) is to determine the language of an utterance. More challenging to this LID is to recognize close dialects within the language [22]. India is a center of prodigious heritage; it is a combination of different cultures and linguistic varieties [11]. Languages pronounced in India are from different language groups, the major ones being the Indo-Aryan languages. Marathi is one among an Indo-Aryan language category which is written in Devanagari script. Marathi is an official language of Maharashtra and co-official language of Goa state of India. Marathi is massively being spoken in Maharashtra. Maharashtra covers a larger geographical area having 34 different districts. According to different regions of Maharashtra, districts are divided into Marathwada, Paschim Maharashtra, Vidarbha, Konkan, and Khandesh [4]. Dialects are variations of the same language and are specific to geographical regions or social groups [7]. The dialect recognition is a special case of the language Identification. Automatic dialect recognition recognizes dialects of language from different speakers [28]. The dialect specific cues are available in speech at different levels. At the segmental level, the dialect specific information can be identified as a unique sequence of the shapes of the vocal tract for producing the sound units [21]. The challenge of the dialect recognition system is to differentiate the dialects of standard language because there exists a lot of similarities between dialects of language [25]. Basically, experiences have proven that speaking in native accent with ASR systems typically ends up with not much success, so such ASR systems can still be improved [24].

Text-independent automatic dialect recognition system for Marathi language is presented in this section. It is based on spectral/ rhythmic and temporal cues of speech signal. One of the spectral features are represented by Mel Frequency Cepstral Coefficients (MFCCs). As a part of study performance of six different classifiers namey, K-nearest neighbor (KNN), Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Stochastic Gradient Descent (SGD) and Ridge Classifier (RC) is recorded and analyzed. The study is carried out using speech corpus provided by LDC-IL. The Marathi speech corpus was collected from the LDC-IL for four Marathi dialects of Puneri, Marathwada, Vidharbha and Goa-based Marathi. A general recognition/identification system is a two-phase system of training and testing phase [15]. System consists of various steps such as data preprocessing, feature extraction, model building, model testing and model evaluation. To perform model training and testing, data splitting ratio is mentioned below in Table 3.

Table 3: Training and testing data (data splitting)

Sr. No.	**Data Splitting**	**Samples**
1	Model Training (70%)	5288
2	Model Testing (30%)	2267
3	Total	7555

For proposed system , 70 % of data (5288 samples) is used for training purpose and 30% of data (2267) is used for testing purpose. Several studies evidenced that no single machine learning algorithm performed better on all types of inputs. Therefore, its very much essential to present the comparison of various machine learning algorithms in order to get the “best” model for the particular dataset. In addition to this, research studies has shown that performance of automatic speech recognition (ASR) degrades when evaluated on a dialectal variation of the same language [7]. So, we performed dialect recognition using six traditional machine learning classifiers to check the performance variations and to find the most prominent classifier for this Marathi language case study.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

K-Nearest Neighbor: KNN classifies query instance based on samples (k-neighbors) that are most close to it [19].

ALGORITHM 1: K-nearest neighbor

Input: *The data set D= [x1, x2, xn], Sample query instance S_q& K closest training examples/neighbors in the feature space.*

Procedure:

1. *Select a value v as number of neighbors; K=v*
2. *Determine which distance function is to be used (Euclidean distance/cosine similarity measure/Minkowsky correlation /Chi square)*
3. *Choose a sample Sq to be classified and compute the distance d to its n training samples in data set D.*
4. *Sort the distances obtained and take K-nearest data samples*
5. *Assign the class 'y' to the S_q based on the majority vote of its K neighbors.*

Output: *The estimated class membership 'y' for sample query S_q where y= [y1, y2…, yn]*

- *Naive Bayes:*The Bayes' Theorem is based on conditional probability .Bayes' theorem, finds the probability of an event occurring, given the probability of another event that has already occurred. Bayes' theorem stated as*: P (x| y) = P (y| x) P (x)/P (y)* Here, P denotes probability, x and y are events [27].

ALGORITHM 2: Naïve Bayes

Input: *The data set, D= [x1, x2, xn], for training and sample query instance Sq for testing*

Procedure:

1. *Calculate the prior probability for a label of data set D*
2. *Find likelihood probability with each attribute xi for each class*
3. *Appy Bays formula and calculate posterior probability*
4. *Find class with higher probability, given sample query instance Sq belongs to the higher probability class*

Output: *The estimated class membership 'y' for sample query S_q where y= [y1, y2…, yn]*

- *Support Vector Machine:*The SVM algorithmic paradigm has a challenge by searching for "large margin" separators [23]. The objective of the SVM is tofind a hyper plane that clearly classifies the data points. Data points falling on either side of the hyper plane can be attributed to different classes. Also, the dimension of the hyper plane depends upon the number of features.

ALGORITHM 3: Support Vector Machine

Input: *The data set, D= [x1, x2, xn], for training and sample query instance Sq for testing*

Procedure:

1. *Input the feature dataset D*
2. *Apply SVM with kernel function (such as linear, polynomial, sigmoid, radial based function)*
3. *Specify line/hyperplane to perform classification*
4. *If obtained accuracy and validity is not satisfactory, then go to step 3*

Output: *A line or hyperplane separating data based on classes y= [y1, y2…, yn]*

- *Decision Trees :*A decision tree is a predictor that predicts the label associated with an instance x by traveling from a root node of a tree to a leaf [23]. Decision Tree models are created using 2 steps: Induction and Pruning. Induction is to build the tree Pruning is the process of removing the unnecessary structure from a decision tree, effectively reducing the complexity to combat over fitting with the added bonus of making it even easier to interpret.

ALGORITHM 4: Decision Tree classification

Input: *The set D of classified instances*

Procedure:

1. *Begin with original dataset D(root)*
2. *On each iteration, calculate entropy(H)and information gain (IG)of parameter xi€X of D*
3. *Select the parameter with smallest entropy(H) or largest information gain (IG)*
4. *Split the set D by selected attribute to generate subset*
5. *Continue on each subset for features those are yet unselected*

Output: *Decision tree*

- *Stochastic Gradient Descent :*It is an iterative algorithm that begins from a arbitrary point so known as stochastic. SGD works by making small, random updates to the parameters of a model to find the values that minimize a cost function.

ALGORITHM 6: Stochastic Gradient Descent

Input: *The set D of classified instances*

Procedure:

1. *Compute the gradient of the function.*
2. *Begin with random initial value for the parameters*
3. *Update the gradient function*
4. *Calculate the step sizes via gradient and learning rate.*
5. *Find new parameter*
6. *Repeat step 3 to 5 until gradient is nearly ZERO*

Output: *Minimum cost*

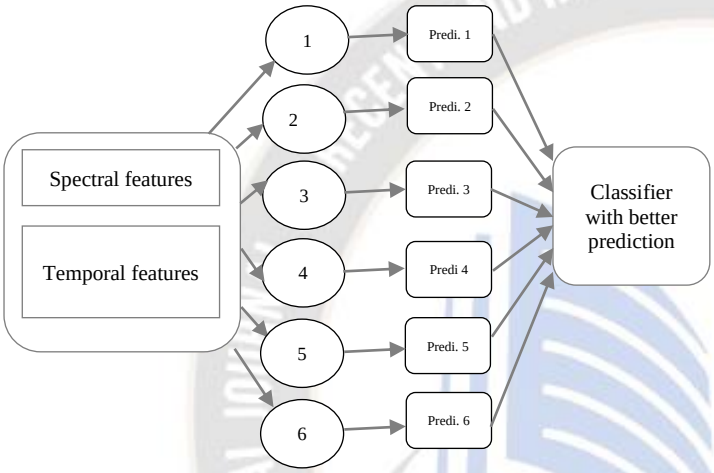
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

- *Ridge Classifier* Ridge regression is a classical data modeling method used in classification [10]. The RC based on ridge regression maps the labels into -1 and 1to solve the problem using regression technique.

ALGORITHM 6: Ridge classification

Input: *The set D of classified instances*
Procedure:
1. *Map the category/label*
2. *Use regression*
3. *Check for highest prediction*

Output: *Class membership based on highest-prediction value*



1:KNN, 2:NB, 3:SVM, 4: DT, 5:SGD, 6:RC
Figure 2. System Architecture

Study is carried out using acoustic features. Normally, acoustic features for Automatic Speech Recognition (ASR) are performed by considering Mel scale integrators [30]. In other words, Mel-frequency cepstrum coefficients (MFCCs) are essential acoustic cues to represent sound/speech signal [14]. Our research approach has generated MFCCs with 12-coefficients. In our research of Marathi dialect recognition, we made a comparative analysis on observed prediction accuracy of model. Recorded results are mentioned in Table 4 below. It provides performance statistics of parameter system accuracy, mean absolute error (MAE) and mean squared error (MSE) for an individual classifier.

Table 4: Prediction results of different classifiers: KNN, NB, SVM, DT, SGD and RC based on 15 spectral features and twelve MFCCs .

Performance Metric	**Classification Model**					
No. of features :15	**KNN**	**NB**	**SVM**	**DT**	**SGD**	**RC**
Prediction Accuracy	74.45 %	78.38 %	83.19 %	79.53 %	70.53 %	84.25 %
Mean Absolute Error MAE)	1.02	1.11	1.01	0.21	1.00	1.00
Mean Squared Error (MSE)	1.31	1.45	1.22	0.22	1.31	1.19
No. of MFCCs :12 coefficients	**KNN**	**NB**	**SVM**	**DT**	**SGD**	**RC**
Prediction Accuracy	79.04 %	77.59 %	80.32 %	79.70 %	80.63 %	80.19 %
Mean Absolute Error (MAE)	1.06	0.22	1.01	0.21	1.04	1.01
Mean Squared Error (MSE)	1.34	0.23	1.24	0.23	1.29	1.24

Prediction accuracy of six different classifiers based on 15 Spectro-temporal features and 12 MFCCs are visualized in fig 3 as a comparitive analysis.

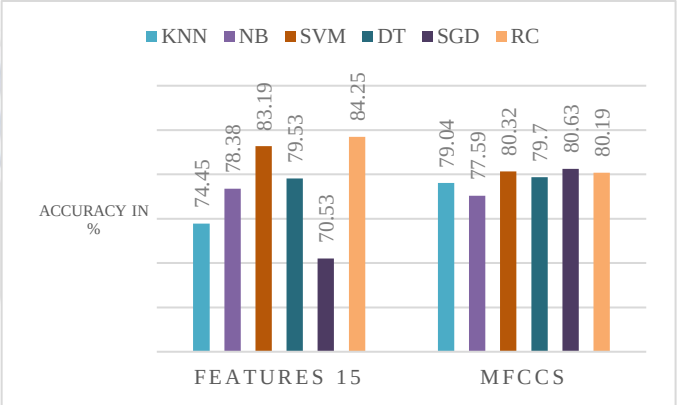


Figure 3. Prediction accuracy of classifiers based on 15 Spectro-temporal features and 12 MFCCs

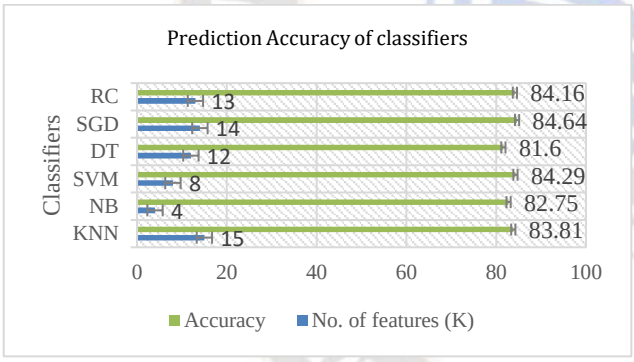
When dealing with huge number of features, we need to have a feature selection mechanism such as a chi-squared feature set [16]. It is well proven that feature extraction and dimension reduction plays a very important role in achieving better performance in the classification problem. The purpose of the feature extraction technique is to get the most compacted and essential set of distinct patterns to enhance the efficiency of the classifier. Dimension reduction can be combined with a feature extraction algorithm to select the most relevant and reduced feature set for classification [12] [29]. Additionally, feature reduction decreases the computational time taken by the learning algorithm. The feature extraction algorithms used here are chi-square, mutual information and ANOVA f-test.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

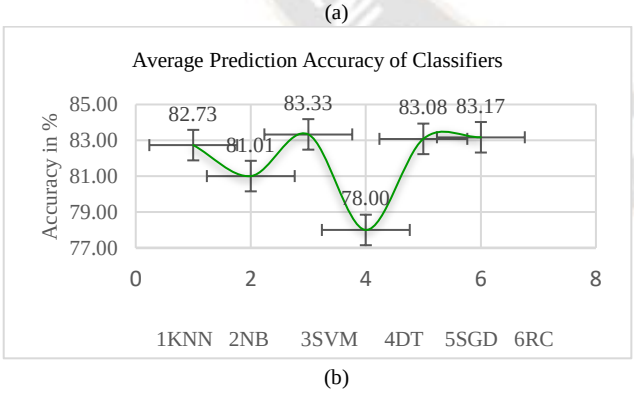
Chi-Square: Chi square test is best suitable when working with categorical feature. It gives degree of association between two categorical variables. Table 5 below presents classifier accuracy based on chi square test. . Prediction accuracy of classifiers based on chi-square feature selection technique is shown in figure. 4 .

Table 5: Chi square based feature selection

Chi square	KNN	NB	SVM	DT	SGD	RC	Feature (ID) selected
K=1	78.91	79.00	79.22	67.62	78.65	79.00	15
K=2	81.07	81.07	80.89	72.25	80.98	80.63	5
K=3	81.25	81.60	80.98	75.60	80.01	80.85	11
K=4	83.10	**82.75**	82.75	76.53	83.01	82.70	3
K=5	83.41	82.66	83.98	78.03	83.41	83.91	10
K=6	83.19	82.66	84.07	78.42	81.91	83.85	4
K=7	83.28	82.44	84.07	77.98	84.25	84.07	8
K=8	82.84	81.95	**84.29**	80.59	84.47	84.07	1
K=9	82.92	81.64	84.29	79.75	84.38	84.07	2
K=10	82.97	81.78	84.25	80.10	84.34	84.03	9
K=11	83.32	80.45	84.25	80.85	84.51	83.98	6
K=12	83.59	78.95	84.25	**81.60**	83.89	84.07	13
K=13	83.67	79.22	84.29	80.32	83.85	**84.16**	7
K=14	83.67	79.26	84.20	80.14	**84.64**	84.07	12
K=15	**83.81**	79.66	84.20	80.23	83.94	84.11	14
Avg	**82.73**	**81.01**	**83.33**	**78.00**	**83.08**	**83.17**	



(a)



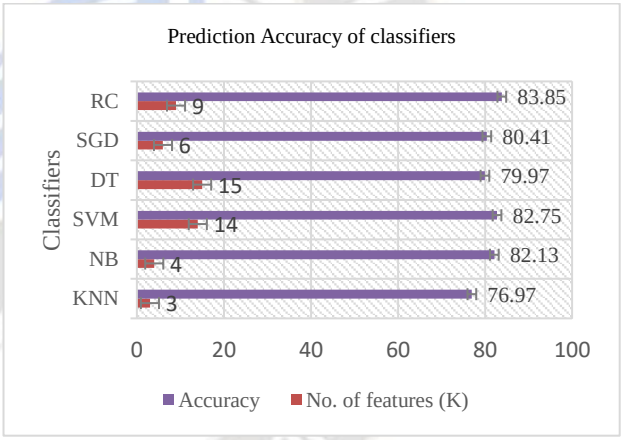
(b)

Figure 4 Chi square feature selection (a) prediction results of KNN, NB, SVM, DT, SGD and RC (Processing time: 18 sec) (b)Average accuracy of classifier

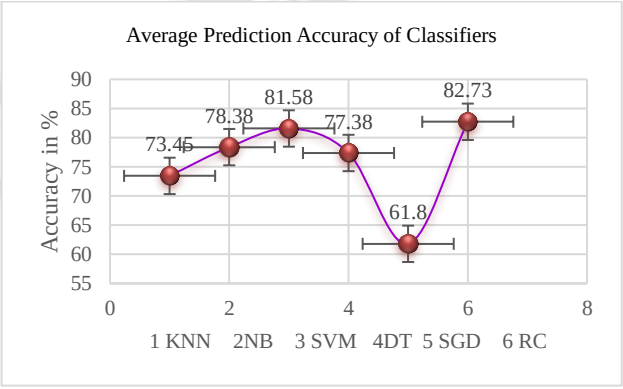
Mutual Information: Mutual information test is a measure between two random variables , say X and Y that enumerates the amount of information obtained about one random variable (X), through the other random variable(Y). Table 6 below presents classifier accuracy based on mutual information technique and results are presented in fig. 5.

Table 6: Mutual information-based feature selection

MF	KNN	NB	SVM	DT	SGD	RC	Feature (ID) selected
K=1	75.12	74.23	74.68	67.79	73.88	75.20	5
K=2	75.56	80.80	80.81	72.78	66.12	80.76	15
K=3	**76.97**	81.25	81.07	72.69	74.72	81.07	11
K=4	64.57	**82.13**	81.69	77.67	62.32	82.97	3
K=5	64.57	80.50	82.22	78.25	53.99	83.54	10
K=6	73.75	80.85	82.39	78.07	**80.41**	83.50	2
K=7	73.75	80.45	82.22	79.04	51.03	83.50	8
K=8	74.68	78.47	82.31	78.91	60.52	83.81	4
K=9	74.68	76.62	82.09	79.44	50.77	**83.85**	13
K=10	74.68	76.62	82.48	79.44	65.28	83.81	9
K=11	74.68	76.79	82.31	78.60	62.28	83.81	7
K=12	74.68	76.75	82.00	79.44	54.56	83.81	6
K=13	74.68	76.75	82.35	78.69	58.53	83.81	14
K=14	74.68	76.75	**82.75**	79.88	50.28	83.81	12
K=15	74.68	76.70	82.35	**79.97**	62.32	83.76	1
Avg	**73.45**	**78.38**	**81.58**	**77.38**	**61.80**	**82.73**	



(a)



(b)

Figure 5Mutual Info based feature selection (a) prediction results of KNN, NB, SVM, DT, SGD and RC (Processing Time: 24 min 31 sec) (b)Average accuracy of classifier

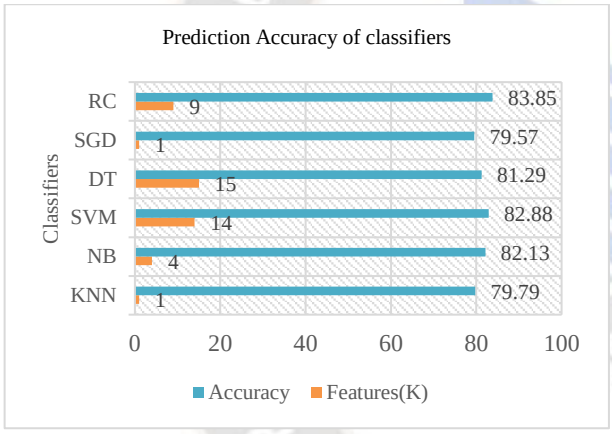
*ANOVA f-test:*ANOVA is more effective and efficient for categorical features. Table 7 below presents classifier

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

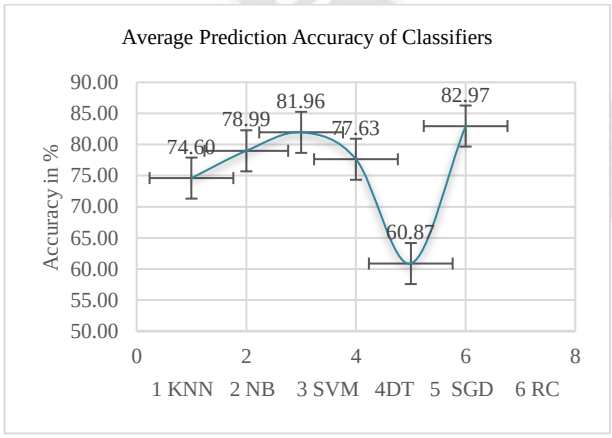
accuracy based on ANOVA f-test. Fig. 6 below shows prediction accuracy of different classifiers.

Table 7:ANOVA f-test feature selection & prediction results of KNN, NB, SVM, DT, SGD and RC

FT	KNN	NB	SVM	DT	SGD	RC	Feature (ID) selected
x=1	**79.79**	79.88	79.88	67.31	**79.57**	79.88	15
x=2	75.56	80.85	80.81	72.60	65.28	80.76	5
x=3	76.97	81.25	81.07	73.22	74.85	81.07	11
x=4	64.57	**82.13**	81.69	77.50	54.60	82.97	3
x=5	74.85	81.42	81.73	77.45	52.62	82.92	4
x=6	74.85	81.25	81.60	77.28	54.12	82.92	8
x=7	74.85	78.12	82.09	77.76	54.74	83.63	10
x=8	74.85	77.98	82.35	80.01	50.28	83.76	1
x=9	74.68	78.47	82.48	80.10	55.13	**83.85**	2
x=10	74.68	78.47	82.39	80.01	64.57	83.81	9
x=11	74.68	78.42	82.79	79.79	59.94	83.76	6
x=12	74.68	76.57	82.70	80.19	70.18	83.81	13
x=13	74.68	76.62	82.53	79.92	68.85	83.81	7
x=14	74.68	76.70	**82.88**	80.06	58.00	83.81	14
x=15	74.68	76.70	82.35	**81.29**	50.28	83.76	12
Avg	**74.60**	**78.99**	**81.96**	**77.63**	**60.87**	**82.97**	



(a)



(b)

Figure 6. ANOVA f-test based feature selection(a) prediction results of KNN, NB, SVM, DT, SGD and RC (Processing time: 29 min 1 sec) (b)Average accuracy of classifiers

This research work presents selection of features using chi-square, mutual information and ANOVA f-test. These techniques are then analyzed to identify effectiveness of feature and the performance of corresponding classifier.

As a part of summary, a prominent feature set and classifier with maximum prediction accuracy is presented in Table 8 below.

Table 8: Prominent feature set and model with maximum performance accuracy

	KNN	NB	SVM	DT	SGD	RC
Chi Square	83.81	82.75	84.29	81.60	84.64	84.16
(x)	15	4	8	12	14	13
Mutual Info	76.97	82.13	82.75	79.97	80.41	83.85
(x)	3	4	14	15	6	9
ANOVA f-test	79.79	82.13	82.88	81.29	79.57	83.85
(x)	1	4	14	15	1	9
x is the number of features						
Chi -square based prominent features set						
KNN	15,5,11,3,10,4,8,1,2,9,6,13,7,12,14					
NB	15,5,11,3					
SVM	15,5,11,3,10,4,8,1					
DT	15,5,11,3,10,4,8,1,2,9,6,13					
SDG	15,5,11,3,10,4,8,1,2,9,6,13,7,12					
RC	15,5,11,3,10,4,8,1,2,9,6,13,7					

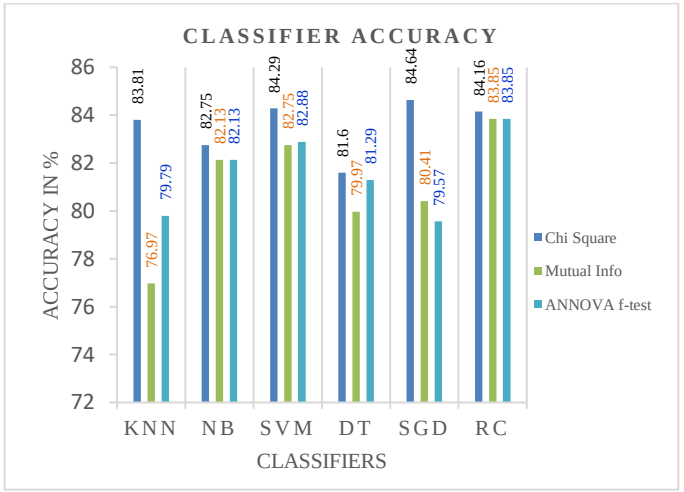


Figure 7 Comparison of dialect recognition performance accuracy by different classifiers based on feature extraction techniques

Here as shown in fig. 7 above, it is observed that chi-square-based feature engineering has proven to be the best for better performance of classifiers. With chi-squared technique-based feature extraction maximum accuracy is given by stochastic gradient descent (SGD) which is 84.64% and minimum is 81.16% with decision tree(DT).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5949
***Article Received**: 12 October 2022 **Revised**: 22 November 2022 **Accepted**: 26 December 2022*

V. CONCLUSION

In this paper, examination of the significance of spectral and temporal behaviors of speech signal has been carried out for dialect recognition of Marathi language. Experiments have been carried out on individuals and combinations of feature subset. Firstly dialectal spectro-temporal features are extracted by examining speech signal. Then, a prominent feature set is identified for individual classifiers based on maximum prediction accuracy. We have demonstrated a dialect recognition with maximum accuracy of 84.64% when used with ridge classifier for a set of 15 spectral and temporal features and 80.63% accuracy with 12 coefficients of MFCCs when applied on Stochastic Gradient Descent (SGD). Chi square-based feature extraction technique was proven to be the most efficient against mutual information and ANOVA f-test. For future work, we propose to use a reduced but prominent feature set to observe performance of deep neural networks.

DECLARATION

The authors have taken the prior permission from the LDC-IL to use the "Raw Marathi speech" dataset in our research study.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Rahesha Mulla: Data curation, Methodology, Validation, Investigation, Formal analysis, Software, Writing – original draft. Dr. B. Suresh kumar: Supervision, Writing – review & editing.

DECLARATION OF COMPETING INTEREST

The authors of this paper declare that they have noany competing interest.

ACKNOWLEDGMENT

The authors of this paper would like to express sincere gratitude towards LDC-IL for providing Marathi speech corpus for the research work.

REFERENCES

[1] Abro, S., Sarang Shaikh, Z. A., Khan, S., Mujtaba, G., & Khand, Z. H. (2020). Automatic hate speech detection using machine learning: A comparative study. Machine Learning, 10(6).

[2] Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2016). A multimodal spectral approach to characterize rhythm in natural speech. The Journal of the Acoustical Society of America, 139(1), 215-226.

[3] Ali, A., Dehak, N., Cardinal, P., Khurana, S., Yella, S. H., Glass, J., ... & Renals, S. (2015). Automatic dialect detection in arabic broadcast speech. arXiv preprint arXiv:1509.06928.

[4] Bansod, N. S., Dadhade, S. B., Kawathekar, S. S., & Kale, K. V. (2014, March). Speaker Recognition using Marathi (Varhadi) Language. In 2014 International Conference on Intelligent Computing Applications (pp. 421-425). IEEE.

[5] Biadsy, F. (2011). Automatic dialect and accent recognition and its application to speech recognition. Columbia University.

[6] Chittaragi, N. B., Limaye, A., Chandana, N. T., Annappa, B., & Koolagudi, S. G. (2019). Automatic text-independent Kannada dialect identification system. In Information Systems Design and Intelligent Applications (pp. 79-87). Springer, Singapore.

[7] Elfeky, M. G., Moreno, P., & Soto, V. (2018). Multi-dialectical languages effect on speech recognition: Too much choice can hurt. Procedia Computer Science, 128, 1-8.

[8] Etman, A., & Beex, A. L. (2015, November). Language and dialect identification: A survey. In 2015 SAI intelligent systems conference (IntelliSys) (pp. 220-231). IEEE.

[9] Gurbuz, S., Gowdy, J. N., & Tufekci, Z. (2000, April). Speech spectrogram-based model adaptation for speaker identification. In Proceedings of the IEEE SoutheastCon 2000.'Preparing for The New Millennium'(Cat. No. 00CH37105) (pp. 110-115). IEEE.

[10] He, J., Ding, L., Jiang, L., & Ma, L. (2014, July). Kernel ridge regression classification. In 2014 International Joint Conference on Neural Networks (IJCNN) (pp. 2263-2267). IEEE.

[11] Kale, S., & Prasad, R. (2018). Author identification on imbalanced class dataset of Indian literature in Marathi. International Journal of Computer Sciences and Engineering, 6, 542-547.

[12] Khalid, S., Khalil, T., & Nasreen, S. (2014, August). A survey of feature selection and feature extraction techniques in machine learning. In 2014 science and information conference (pp. 372-378). IEEE.

[13] Mamyrbayev, O., Mekebayev, N., Turdalyuly, M., Oshanova, N., Medeni, T. I., & Yessentay, A. (2019). Voice identification using classification algorithms. Intelligent System and Computing.

[14] Masood, S., Nayal, J. S., Jain, R. K., Doja, M. N., & Ahmad, M. (2017). MFCC, Spectral and Temporal Feature based Emotion Identification in Songs. International Journal of Hybrid Information Technology, 10(5), 29-40.

[15] Mohammed, T. S., Aljebory, K. M., Rasheed, M. A. A., Al-Ani, M. S., & Sagheer, A. M. (2021). Analysis of Methods and Techniques Used for Speaker Identification, Recognition, and Verification: A Study on Quarter-Century Research Outcomes. Iraqi Journal of Science, 3256-3281.

[16] Moh'd A Mesleh, A. (2007). Chi square feature extraction based svms Arabic language text categorization system. Journal of Computer Science, 3(6), 430-435.

[17] Muslim, E. M. (2007). An Introduction to Computational Linguistics Advantages & Disadvantages. journal of the college of basic education, 10(51).

[18] Nisar, S., & Tariq, M. (2018). Dialect recognition for low resource language using an adaptive filter

bank. International Journal of Wavelets, Multiresolution and Information Processing, 16(04), 1850031.

[19] Prasad, J. R., & Kulkarni, U. (2015). Gujrati character recognition using weighted k-NN and mean χ 2 distance measure. International Journal of Machine Learning and Cybernetics, 6(1), 69-82.

[20] Ramamoorthy, L., Narayan Choudhary, Gajanan R Apine & Apurva P Betkekar. 2019. Marathi Raw Speech Corpus. Central Institute of Indian Languages, Mysore.

[21] Rao, K. S., Nandy, S., & Koolagudi, S. G. (2010). Identification of Hindi dialects using speech. WMSCI-2010.

[22] Ren, Z., Yang, G., & Xu, S. (2019). Two-stage training for Chinese dialect recognition. arXiv preprint arXiv:1908.02284.

[23] Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.

[24] Sheng, L. M. A., & Edmund, M. W. X. (2017). Deep learning approach to accent classification. CS229.

[25] Shivaprasad, S., & Sadanandam, M. (2021). Dialect recognition from Telugu speech utterances using spectral and prosodic features. International Journal of Speech Technology, 1-10.

[26] Singh, G., Sharma, S., Kumar, V., Kaur, M., Baz, M., & Masud, M. (2021). Spoken Language Identification Using Deep Learning. Computational Intelligence and Neuroscience, 2021.

[27] Smola, A., & Vishwanathan, S. V. N. (2008). Introduction to machine learning. Cambridge University, UK, 32(34), 2008.

[28] Sreeraj, V. V., & Rajan, R. (2017, May). Automatic dialect recognition using feature fusion. In 2017 International Conference on Trends in Electronics and Informatics (ICEI) (pp. 435-439). IEEE.

[29] Subasi, A. (2019). Feature Extraction and Dimension Reduction, Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques, Academic Press.193-275

[30] Thomas, S., Ganapathy, S., & Hermansky, H. (2008, August). Spectro-temporal features for automatic speech recognition using linear prediction in spectral domain. In 2008 16th European Signal Processing Conference (pp. 1-4). IEEE.

[31] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference (Vol. 8).