ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FEATURE SELECTION AND CLASSIFICATION OF SPEECH DATASET FOR GENDER IDENTIFICATION: A MACHINE LEARNING APPROACH

RIZWAN REHMAN[1], KAUSTUVMONI BORDOLOI[2], KANKANA DUTTA[3], NOMI BORAH[4], PRIYAKHI MAHANTA[5]

[1,2,3,5]assistant Professor, Dibrugarh University, Centre For Computer Science And Applications, India

[4]assistant Professor, Dibrugarh University, Duiet, India

E-mail: [1]rizwan@dibru.ac.in, [2]kaustuvmoni@dibru.ac.in, [3]kankanadutta@dibru.ac.in, [4]baruahnomi@gmail.com, [5]kaustuvmoni@dibru.ac.in

ABSTRACT

In speech analysis, gender identification is one of the most complex tasks. Gender can be traced from the acoustic parameters like formants (F1, F2, F3, F4) or the pitch (F0). Therefore it is very important to identify which feature or features can classify the dataset efficiently in terms of a male and female speakers. This paper is an attempt to classify the dataset more accurately using fewer features i.e. among F0, F1, F2, F3, and F4. For the feature selection, the Fisher score algorithm is used to find out the most discriminative feature that can be used for the classification of the gender from the speech data set. Then to cross-validate the result obtained using the Fisher score algorithm we have applied the Tree-based algorithm. The results of both the algorithm comply with each other as F0 or pitch is the most distinctive feature among all with both the algorithms. Since the result of both the algorithm comply with each other we have then performed the classification by applying logistic regression, KNN classifier, SVM, and Decision tree algorithms. We have then evaluated and compared the accuracy of each of the features using these classification techniques. The finding of this study will provide the statistical means to identify the best feature for gender identification from the acoustic characteristic.

Keywords: *Feature Selection, Classification, Gender Identification, Statistical Methods*

1. INTRODUCTION

Gender identification is one of the prominent parts of a good speech or speaker identification system. When a speaker speaks a sentence the listener apart from decoding the linguistic information also detects the paralinguistic information such as the gender of the speech, his or her age, and other characteristics of the speaker. This information is also known as the voiced information.

Accordingly, the voiced information consists of many acoustic parameters. Two of the important acoustic parameters are the spectral formant frequencies and the fundamental frequency. The fundamental frequency also denoted as F0 is determined by the glottis only. This fundamental frequency is also referred to as the pitch [1]. On the other hand, the spectral formants are the resonances determined only by the position of the vocal tract.

The formants are also referred to as F1, F2, F3, F4 i.e the first, second, third, and the fourth formant. These acoustic features plays an important role in identifying the information from the speech signal.

Most of the previous studies for gender identification using acoustic properties is based primarily on the fundamental frequency estimation which in itself is a challenging task to estimate [2]. Therefore the proper selection of the acoustic features based on some statistical techniques is needed to identify the relevant features for proper gender identification.

Thus the objective of the current study is to apply different statistical approach to identify the appropriate acoustic features for the correct gender identification.

In order to identify the gender of the speaker using different statistical approaches effectively, the data preprocessing is the first important step. In the preprocessing step the most relevant feature(s) among all or the feature(s) that can distinctively identify the data set elements in terms of the target variable are found [3]. This increase the accuracy of the prediction and the comprehensibility as well.

Journal of Theoretical and Applied Information Technology
30th November 2020. Vol.98. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Another advantage is that the classifiers can avoid overfitting with proper feature selection hence they have an improved ability to generalize.

The two models for the feature selection can be broadly categorized as the wrapper and the filter [4]. Among the two, the filter model in its preprocessing step selects the subgroup of the features which is classifier independent. This makes the filter method fast and less complex computationally and hence making it a popular choice of use. In addition to this, after the task of feature selection is over any of the classifiers can be added.

We have first used the Fisher Score Algorithm to find the best feature among all [5]. Then for the validation of the result obtained we have again applied a tree-based classifier using the ExtraTree Classifier function of the Sklearn module of python [6]. Once we have figured out the best features using these algorithms we have then tested the performance using Logistic Regression, kNN, SVM, and Decision tree classifier using the python programming. Then a comparison is made to highlight the feature that can produce a better gender identification among all.

In the present study, python programming is used for the statistical analysis. Python uses the Maximum likelihood estimation (MLE) algorithm [7]. The estimator that is extracted using MLE algorithm are also identical to the ordinary least square method (OLE) which is a very popular and widely used method [8].

2. METHODOLOGY

2.1 Feature Selection

At first, to select the best feature(s), Fisher Score Algorithm is used. Fisher Score Algorithm finds the discrimination of two sets of real numbers. If we have a set of two real numbers then the Fisher Score algorithm measures the discrimination between the two sets.

Consider a training vector a_i, i=1,.....p, where Z_+ and Z_- are the number of positive and negative instance the Fisher Score of the k^{th} feature is defined as :

$$F_{(k)} \equiv \frac{\left(\bar{a}_k^{(+)} - \bar{a}_k\right)^2 + \left(\bar{a}_k^{(-)} - \bar{a}_k\right)^2}{\frac{1}{z_+ - 1}\sum_{i=1}^{z_+}\left(a_{i,k}^{(+)} - \bar{a}_k^{(+)}\right)^2 + \frac{1}{z_- - 1}\sum_{i=1}^{z_+}\left(a_{i,k}^{(-)} - \bar{a}_k^{(-)}\right)^2} \quad (1)$$

where,

$\bar{a}_k$, $\bar{a}_k^{(+)}$, $\bar{a}_k^{(-)}$ are the average of the k^{th} feature of the whole positive and negative data set.

$a_{i,k}^{(+)}$ is the k^{th} feature of the i^{th} positive instance and $a_{i,k}^{(-)}$ is the k^{th} feature of the i^{th} negative instance.

The discrimination among the positive and negative sets is given by the numerator and denominator and it shows the discrimination within the sets. The larger Fisher Score indicates that the feature is more discriminative hence the score can be considered for the purpose of feature selection.

The ExtraTree Classifier function of the sklearn module of python is then used to cross- check the result of the Fisher Score results. ExtraTree Classifier uses a meta estimator that fits a number of randomized decision tree or ExtraTree on the sub-samples of the dataset. It improves the accuracy of the prediction and handles overfitting using averaging.

2.2 Classification

Classification falls under the class of supervised learning [9]. The class of the data element is predicted with the help of classification. For modeling, a training dataset consisting of several examples of the input and the outputs are required by the classifier to learn.

The model hence derived will predict the class labels of the input data from the training dataset. Therefore the training dataset must include data points of class labels.

In the present study, the target or the dependent variable i.e. male and female are categorical in nature therefore we have used logistic regression first to fit the model. The regression model in general consists of an output variable that is defined by the input variable(s). There can be one or more input variables.

In the present work the gender is considered as the dependent variable while the other variables in the set i.e. F0, F1, F2, F3, F4 (formants) are considered as the explanatory variables to best fit the regression line.

Next, we applied the kNN classification which is based on the nearest neighbor. This method does not build any internal model. It just keep track of all the instances that are available with it and then on the basis of the similarity measure it classifies a new case.

Euclidean distance as distance metrics is used. The main advantage of using kNN are that it is simple in implementation and its robustness to the noisy data. But the fact that this algorithm has to find the distance of each occurrence to all the training samples makes it slow. Then the Decision Tree classification is used. This algorithm can classify data with the help of the input data and its classes. It is done by generating a sequence of rules that can classify the data. Given a sample of data this classifier will find a predictor from the input variables to divide the sample into different categories. The predictor which is chosen is the one with minimum cost or highest accuracy. The Decision tree is good with both numerical and categorical data. Next, the Support Vector Machine classifier is applied. This classifier depicts the training set as points in space that have a distinct separation based on the category. If a new point is encountered it is added by predicting the category as well as space they fall in. This algorithm is used mainly because of its efficiency in the high dimensional space and also because of its efficiency in terms of memory utilization.

3. EXPERIMENTAL SETUP

For the current study the first four formant frequencies of the fourteen vowels i.e. {**i, i:, e, e:, a, a:, o, o:, u, u:, í, í:, é, é:**}, of Mising language (spoken in the North-Eastern part of India) as uttered by the male and female speakers are considered [10]. The parameters and their specifications for the current work are shown in *Table 1.*

Table 1: Parameters and Specifications

Sl.No.	Parameters	Specifications
1.	Speakers	Male: 10 Female: 10
2.	Database	20 Speakers X 14 vowels X 3 sessions = 840 total vowels sample.
3.	Recording Environment	Laboratory
4.	Device	PHILIPS overhead microphone with noise cancellation
5.	Channel	Mono
6.	Sampling Frequency	22050 Hz
7.	Re-Sampling Frequency	Male: 10000 Hz Female: 11000 Hz
8.	Speaker Background	Graduate and Post Graduate Students.
9.	Software	PRAAT

We have used python programming and a jupyter notebook to implement the feature selection and classification algorithms. We have used 10-fold cross-validation for the SVM classification [11]. After the feature selection, random set of features are selected for the accuracy calculation of the classification. The next set of features are evaluated and used only if their result is better than the previous results.

The model is cross-validated using the pipe-svc estimator [12]. The learning as well as the validation curve are checked to track the accuracy of the test and the training data. In the present work, the dataset has been divided into 80% training and 20% testing set.

4. RESULTS AND DISCUSSIONS

It is important to find the predictive power of each of the features to find its effectiveness in predicting the gender of the speaker. To understand the discriminative power of each of the feature the pair plot function of the Seaborn package is used [13]. This will help to graphically judge the effectiveness of each of the features with respect to the gender of the speaker. Figure 1 represents the pair plot of each of the features where label 1 represents male while label 0 represents female speakers. It is clear from Figure 1 that the feature F0 can clearly make a distinction between the male and female speakers, as evident from the blue and orange dots, since we can see a clear separation between the blue and orange dots.

Journal of Theoretical and Applied Information Technology
30th November 2020. Vol.98. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

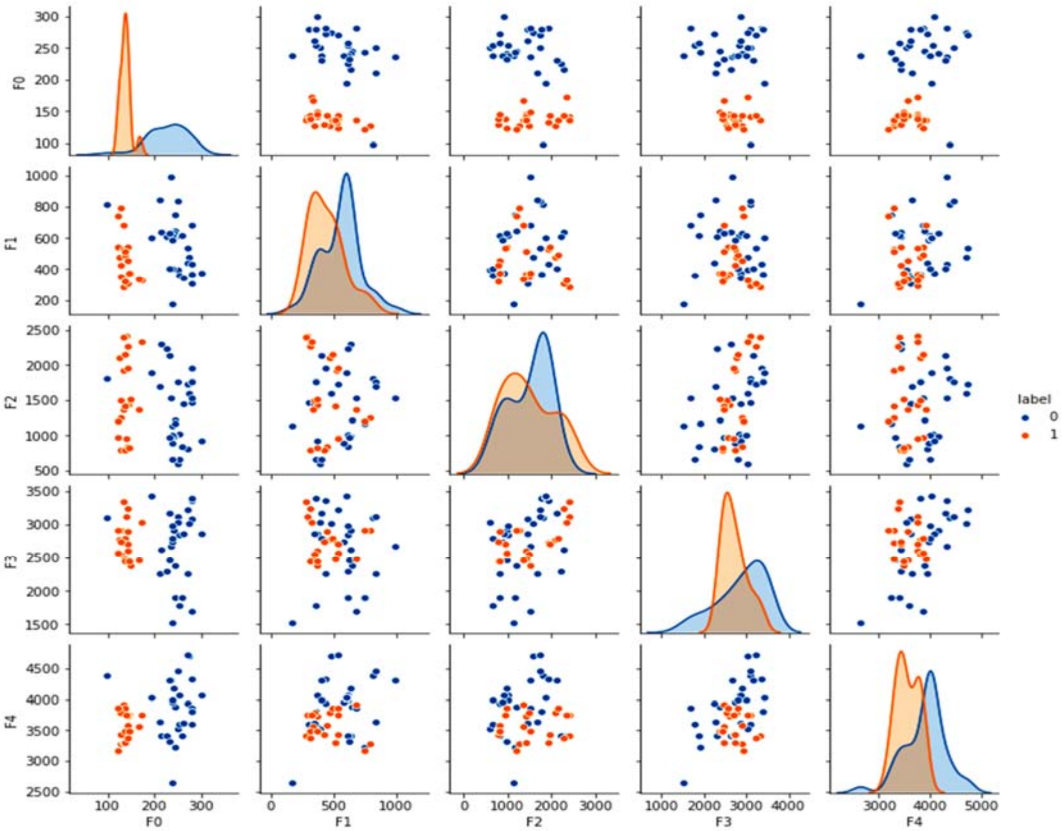

Figure 1: Pair plot of each of the features.

It is also evident that F0 with respect to F1,F2,F3, and F4 shows a clear distinction between the speakers while for others the dots are mixed up and do not have a clear distinction which makes them a poor discriminator. This suggests that the feature F0 has got more predictive power than the other features. In order to verify the same with the help of a proper algorithm, the Fisher Score algorithm is applied first and the same is crossed checked with a tree-based algorithm. Therefore these two steps can also be considered as preprocessing steps before performing the classification.

Table 2, shows the F-Score value of each of the attributes in the dataset. To find the F-Score the model was fit using Logistic Regression as well as kNN model. The accuracy score using Logistic Regression is 0.86 while that of kNN is 0.93, as shown in Figure 2. The result suggests that kNN method gives better accuracy. Therefore kNN model is used to fit the model to evaluate the F-Score.

F-Score has ranked the attributes as F0, F4, F1,F3,F2. With the highest F-Score value of 127.15, F0 is the attribute with the highest predictive power followed by F4 with an F-Score of 16.42 which far less than the score of F0. Figure 3 represents the bar plot of all the attributes.

Table 2: F-Score values

	F Score	P Value	Support	Attribute
0	127.151427	3.210108e-17	True	F0
4	16.419306	1.327865e-04	True	F4
1	10.124941	2.205449e-03	True	F1
3	2.093722	1.524997e-01	False	F3
2	0.049695	8.242626e-01	False	F2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

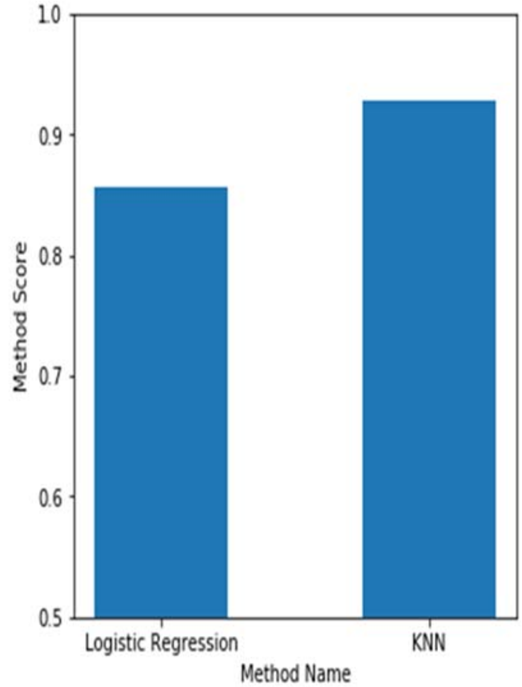

Figure 2: Accuracy score of the models

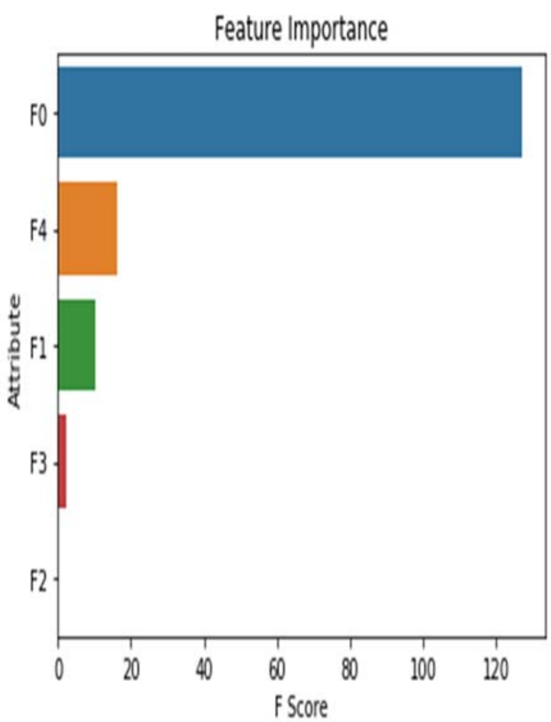

Figure 3: Bar plot using F-Score

The bar plot as shown in Figure 3, shows the comparative feature importance of all the attributes.

The tree-based algorithm namely Extra tree classifier is then applied to cross-check the result. Extra tree classifier outputs the classification result by aggregating the results from all the decision trees in a forest. The method is same as in the random forest classifier, only the technique of constructing the decision tree is a forest is different. In extra tree classifier each of the decision tree is constructed from the original sample of the training set. After that at each node of the test node a random sample of *k* features is provided. These *k* features are from the set of features from which the best feature has to be selected by each of the decision tree. The bar plot of the result is shown in Figure 4.

This complies with the result of F-Score with F0 having the highest score value of 0.49. Hence the tree-based algorithm also suggests that F0 has the highest predicting power.

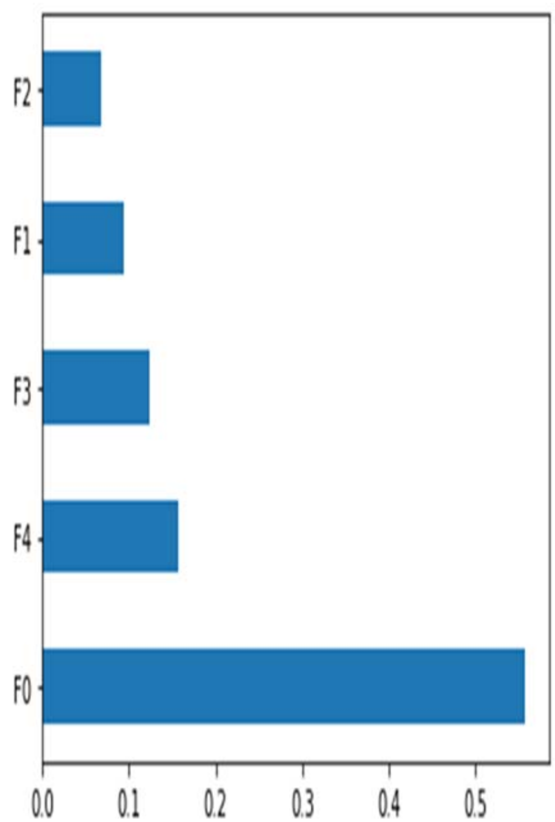

Figure 4: Bar plot using Extra Tree Classifier

Figure 4 also suggest a slight change in the prediction of the ranking of the attributes F3 and F1. With tree-based algorithm F3 is placed above F1 unlike in the F-Score algorithm. But the difference in the value is very less with F1 having a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

score of 0.12 and F3 having a score of 0.14. Still, we have a common attribute F0 with the highest value with both the algorithms.

Once we have ranked the features the next objective is to verify and test the same with different classification models. The objective is to check how well the best feature is able to classify the target i.e. male or female speakers with respect to other features in the list.

The classification models selected for the purpose are Logistic Regression (L.R.), Decision Tree (D.T.), KNN and SVM. Each of the models were tested for the accuracy with respect to all the parameters as well as individual parameters using jupyter notebook. The result from each model is shown in *Table 3*.

Table 3: Classification Score (Accuracy) of the models

Meth-od Name	All Param-eters	F0	F1	F2	F3	F4
L.R.	0.85	0.96	0.66	0.60	0.58	0.73
D.T.	0.92	0.98	0.57	0.71	0.78	0.85
KNN	0.82	0.92	0.66	0.78	0.57	0.84
SVM	0.71	0.92	0.71	0.64	0.50	0.86

In *Table 3* , the accuracy score using the parameter F0 is higher for all the classification models. This is also supported by the fact that pitch plays an important role in gender identification. When all the parameters were considered together then it is found that the decision tree classifier has higher accuracy in comparison to other classification models.

Thus it is clear from all the classification models that the attribute F0 provides a better accuracy in gender identification. *Table 3*, also suggests that the attribute F4 also has better accuracy in comparison to the rest of other attributes. Thus apart from F0, F4 is also a good predictor for all the models. Thus for building a gender identification model the formants F0 as well as combination of the formants (F0, F4) can play a significant role.

Next it also very important to validate the model. *Table 3* shows that the decision tree model gained an accuracy of 92% with all parameters and 98% with respect to F0 but then this model needs to be evaluated on metrics like cross-validation and learning curve validation for better acceptance.

First cross validation is performed so as to reduce the variation of the model. The effectiveness of the machine learning model can be easily tested using the cross validation technique. Specifically if the data sample size is limited then cross validation which is basically a re-sampling procedure can be used to evaluate the model more precisely. A train/test split method is used where the dataset is divided into 80% for training the model and 20% to test the model. k-fold cross-validation is performed with 10 folds.

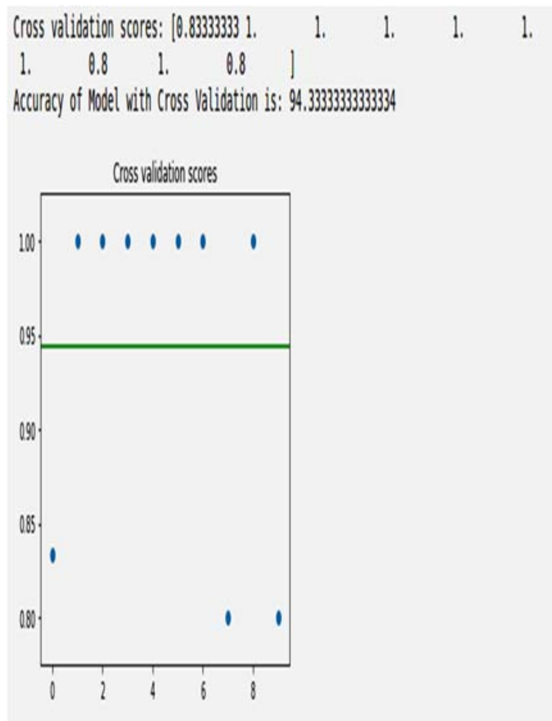

Figure 5: 10-fold cross validation

Figure 5 shows the result of 10 fold cross-validation with an acceptable accuracy of 94.33%. To cross check the over fitting problem learning curve method is applied. This is used to check if the model is overfitting to a given training set.

The goal is to generalize perfectly from the training data set. This will allow the prediction from new data sets. To check how well the machine learning model is able to learn and generalize from the new set of data, it is necessary to check the overfitting issue. if the overfitting is not checked properly it can lead to the poor performance of the model. The result is shown in the Figure 6.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

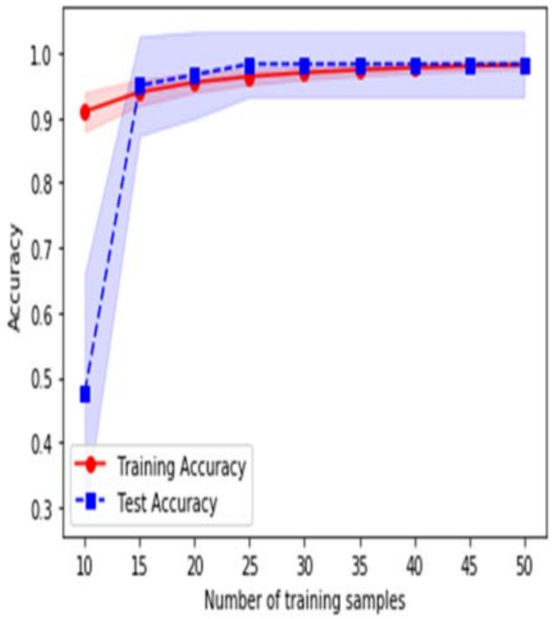

Figure 6: Learning curve

Figure 6 shows that the variation of the score with respect to test data is more but there is no trend of overfitting as it can be seen the plot that graph of the test score is improving along with the training sample.

Then a validation curve is plotted using regularization parameter C.

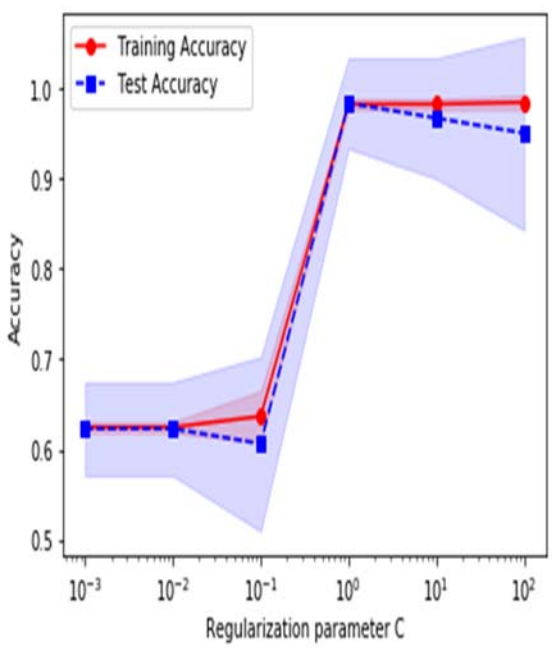

Figure 7: Validation curve

It can be seen in Figure 7 that the best value of C is 1 as over that the test accuracy is showing a decreasing pattern that can cause overfitting.

Once we have validated the model and checked the measures so that there is no overfitting the accuracy of the decision tree model can be cross-validate so as to cross check which parameter has better-predicting accuracy. *Table 4* shows the result of the cross-validation.

Table 4: Accuracy of Decision Tree Model with cross validation

Parameter	Accuracy	MAE	MSE	RMSE
F0	95.71	0.24	0.06	0.26
F1	68.57	0.49	0.27	0.52
F2	52.86	0.48	0.25	0.49
F3	72.86	0.48	0.24	0.50
F4	75.71	0.34	0.14	0.38

In *Table 4* it can be seen that the accuracy of the parameter F0 is highest with 95.71% and also the parameter like Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) values are also supporting F0 to be the most predictive parameter. Similarly, after F0, F4 with the accuracy of 75.71% is the second-best feature for predicting the gender of the speaker in the dataset. This result is in conformation with our earlier results obtained using different classification techniques.

The researchers has been using the formants specifically F0 and the Mel-frequency cepstral coefficients as the most common feature for the gender identification [14][15][16][17][18]. The reason behind using the fundamental frequency i.e. F0 as the most common feature is the fact that the average F0 for men is in the range of 100-146 Hz while that for female it is 188-221 Hz [19]. But the dependence on such mean value range for the gender identification sometimes also produces false result. Since there may be cases where the F0 of male and female voices might fall in vice-versa ranges. Therefore the feature selection using machine learning techniques can play a significant role in gender identification as it justifies the selection of the acoustic features based on the statistical method.

Table 5: Formant Frequencies and Speaker Type for Fourteen Mising Vowels

V. No.	Time and Pitch		Formant Frequencies				Gender
	Time	F0	F1	F2	F3	F4	SType
1	0.345669	235.22	993.25	1533.7	2673.19	4311.04	Female
2	0.282823	237.93	175.81	1141.23	1523.02	2646.61	Female
3	0.329955	244.25	648.34	1213.02	2374.64	3876.29	Female
4	0.3928	243.52	744.77	1173.47	1907.42	3228.28	Female
5	0.34568	97.99	817.42	1810.99	3106.41	4385.57	Female
6	0.209974	250.07	836.12	1754.96	3110.39	4465.11	Female
7	0.370915	280.34	305.03	1463.17	3089.26	3608.39	Female
8	0.298526	274.98	476.09	1587.20	3018.38	4711.17	Female
9	0.298526	250.44	396.31	598.56	3025.39	3523.31	Female
10	0.282823	270.12	532.27	1729.58	3227.51	4729.23	Female
11	0.3928	254.75	613.59	1023.28	2842.46	4063.04	Female
12	0.23568	240.60	633.00	984.39	2912.78	4172.27	Female
13	0.361372	271.81	439.59	807.94	2252.97	3938.34	Female
14	0.267109	254.08	358.67	664.20	1770.86	3595.68	Female
1	0.465136	281.4534	681.3576	1524.85	1691.76	3860.04	Female
2	0.465136	210.2277	841.7061	1693.78	2267.57	3639.37	Female
3	0.427925	215.7868	634.7584	2293.09	2624.32	3421.663	Female
4	0.372109	225.1178	610.3167	2223.39	2298.44	3417.81	Female
5	0.390726	193.36	601.04	1885.54	3432.01	4027.95	Female
6	0.409331	258.01	616.92	838.66	1892.83	3407.65	Female
7	0.446531	280.77	435.89	1945.78	3365.90	4325.87	Female
8	0.408209	136.46	514.97	1411.2	2486.48	3299.47	Female
9	0.446531	279.62	360.52	1765.93	3397.42	3805.67	Female
10	0.372109	279.90	364.22	1764.28	3400.40	3810.92	Female
11	0.744218	238.42	615.42	1012.40	2984.77	3987.79	Female
12	0.427925	237.40	584.04	898.33	2737.55	3962.14	Female
13	0.40932	300.13	372.88	916.40	2868.55	4064.08	Female
14	0.40932	237.25	401.82	655.04	2798.59	3999.39	Female
1	0.390714	193.36	601.04	1885.54	3432.05	4027.94	Female
2	0.390714	193.36	601.04	1885.54	3432.05	4027.95	Female
3	0.390714	193.36	601.04	1885.54	3432.05	4027.95	Female
4	0.390714	193.36	601.04	1885.54	3432.05	4027.95	Female
5	0.399501	231.64	627.65	968.12	2462.93	3325.52	Female
6	0.390714	193.36	601.04	1885.54	3432.05	4027.93	Female
7	0.354093	231.27	406.73	2126.03	3164.93	4304.65	Female
8	0.39949	262.26	344.33	1451.01	2861.70	3618.83	Female
9	0.390714	193.36	601.04	1885.5	3432.05	4027.95	Female
10	0.390714	193.36	601.04	1885.54	3432.05	4027.93	Female

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

11	0.390714	193.36	601.04	1885.54	3432.05	4027.94	Female
12	0.390714	193.36	601.04	1885.54	3432.05	4027.94	Female
13	0.390714	193.36	601.04	1885.5	3432.05	4027.93	Female
14	0.390714	193.36	601.04	1885.5	3432.05	4027.94	Female
1	0.277347	123.13	535.12	972.46	2558.85	3858.55	male
2	0.298481	135.33	685.32	1362.1	2477.90	3915.73	male
3	0.275215	145.73	368.79	1423.29	2602.40	3752.65	male
4	0.351814	141.68	366.6	1428.37	2608.30	3737.89	male
5	0.213946	126.60	475.21	2096.70	2774.96	3795.88	male
6	0.303753	136.28	491.42	2154.23	2798.28	3857.46	male
7	0.258844	141.41	294.77	2416.3	3111.6	3740.84	male
8	0.261497	127.83	348.59	1491.10	2543.46	3795.65	male
9	0.198095	135.47	283.06	2400.71	3339.67	3404.01	male
10	0.198095	135.47	283.069	2400.71	3339.67	3404.01	male
11	0.294229	122.82	539.62	972.42	2562.05	3862.16	male
12	0.302676	138.65	531.85	954.80	2698.3	3757.19	male
13	0.429694	138.03	319.96	800.001	2452.22	3486.33	male
14	0.365238	145.75	372.81	823.05	2443.47	3483.67	male
1	0.408209	127.53	792.49	1256.33	2913.07	3278.68	male
2	0.343753	122.17	743.72	1209.34	2919.86	3169.92	male
3	0.365238	133.05	524.76	1925.03	2730.64	3299.66	male
4	0.408209	143.02	542.01	1961.55	2710.22	3430.47	male
5	0.408209	136.46	514.97	1411.25	2486.48	3299.46	male
6	0.408209	136.46	514.97	1411.2	2486.48	3299.46	male
7	0.451179	143.23	311.22	2267.53	3237.61	3373.77	male
8	0.386723	172.18	325.98	2336.62	3034.22	3744.88	male
9	0.429694	148.15	372.21	1519.50	2373.95	3485.64	male
10	0.451179	166.54	337.59	1364.09	2463.52	3564.91	male
11	0.429694	143.76	452.01	836.50	2893.86	3573.32	male
12	0.386723	128.88	428.04	800.25	2748.11	3435.99	male
13	0.429694	138.03	319.96	800.00	2452.22	3486.33	male
14	0.3652	145.75	372.81	823.05	2443.47	3483.67	male

Table 5 represents the snapshot of the dataset. To study the relationship between the formant F0 i.e. pitch and the gender of the speaker a scatter plot is drawn from the values in the dataset as represented in Table 5. The plot is shown in the Figure 8.

For uniformity we have converted the gender (SType) values as 1 for female and 0 for male. Since there are only two variables i.e. male and female, the relationship between both is easily depicted using the scatter plot. In the Figure 8, the dotted line with respect to the ordinate 1 represents the female speaker while the dotted line with respect to the ordinate 0 represents the male speaker.

From the plot is evident that there is a significant difference in the pitch of the male and the female speaker. Hence the result of scatter plot is also in confirmation with the previous findings. This also validates the authenticity of the dataset used.

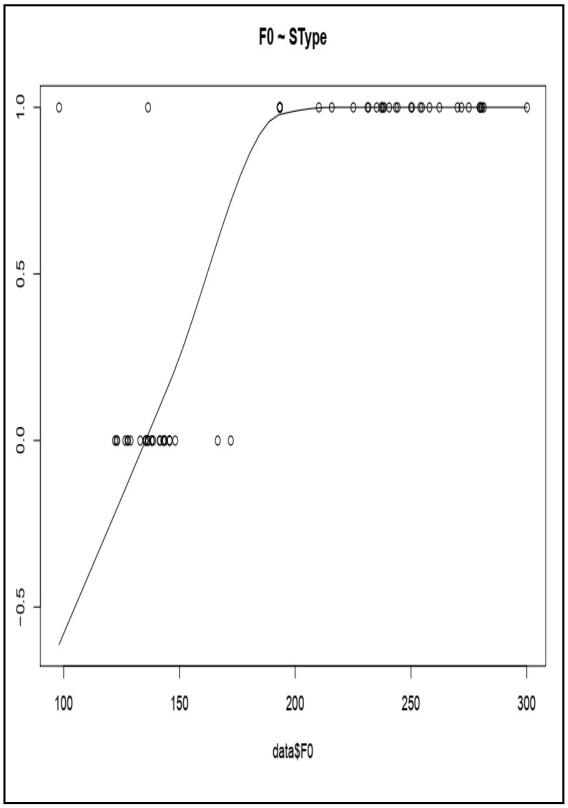

Figure 8: Scatter plot F0 and Gender

5. CONCLUSION

In this paper, we applied the machine learning approach to find the most appropriate feature for the gender identification. Selecting the best feature plays an important role in proper identification of the gender from the acoustic dataset. We first performed the feature selection using the Fisher score algorithm and Extra Tree Classifier model. The result obtained appears to be promising with respect to the acoustic dataset. We found that decision tree classification model provided better accuracy in the gender identification using both single parameter i.e. F0 (pitch) and also using all the parameters together. It is also observed that apart from F0, F4 can also be a good predictor. The cross-validation result of the model based on the decision tree method also confirms the fact that F0 (pitch) is the best parameter among all. Thus application of machine learning based approach provide better and validated result for the feature selection.

For the present work only four algorithms i.e. Logistic Regression, Decision Tree, kNN, and SVM were used but the application of some more algorithm and validation methods like grid search can provide more information and insights to the feature that can best classify the gender of the speaker.

REFRENCES:

[1] Taylor, P. "Text-to-Speech Synthesis" *Cambridge: Cambridge Univeristy Press,* 2009.

[2] Pondhu, Laxmi, Kummari, Govardhani. "Performance Analysis of Machine Learning Algorithms for Gender Classification" , *1109/ICICCT.2018.8473192*, 2018, pp. 1626-1628.

[3] Sharma A., Imoto S., Miyano S., "A top-r feature selection algorithm for microarray gene expression data", *IEEE Transaction on computational biology and bioinformatics.* Vol. 9 (3), 2012, pp. 754-64.

[4] Kohavi R., John G. H., "Multiple comparison in induction algorithms", *Machine Learning,* vol. 38 (3), 2000, pp. 309-338.

[5] Yang, Jian & Yang, Jingyu & Zhang, David, "Median Fisher Discriminator: a robust feature extraction method with applications to biometrics", *Frontiers of Computer Science in China*, vol.2. ,2008, pp. 295-305.

[6] Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python", *JMLR 12* ,2011, pp. 2825-2830.

[7] D. C. de Souza, C. C. Cavalcante and R. F. Vigelis, "Modified maximum likelihood estimator," *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janerio, 2016, pp. 1-5, doi: 10.1109/SAM.2016.7569748.

[8] V. G. Drozdova and R. V. Akhpashev, "Ordinary least squares in COST 231 Hata key parameters optimization base on experimental data," *International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, Novosibirsk, 2017, pp. 236-238, doi: 10.1109/SIBIRCON.2017.8109878.

[9] C. A. Ul Hassan, M. S. Khan and M. A. Shah, "Comparison of Machine Learning Algorithms in Data classification," *24th International Conference on Automation and Computing (ICAC)*, Newcastle upon Tyne, United Kingdom, 2018, pp. 1-6, doi: 10.23919/IConAC.2018.8748995.

[10] Taid, T. R., " An Introduction to Mising Phonology and Grammar". *Guwahati: ABILAC,* 2016.

[11] Z. Wen, R. Zhang, K. Ramamohanarao, J. Qi and K. Taylor, "MASCOT: Fast and Highly Scalable SVM Cross-Validation Using GPUs and SSDs," 20 *IEEE International Conference on Data Mining, Shenzhen,* 2014, pp. 580-589, doi: 10.1109/ICDM.2014.35.

[12] Andreas Muller, Sarah Guido; "Introduction to Machine Learning with Python", *First Edition; Riley O' publication* ,2016, pp-305–320.

[13] Waskom, M. et al., *mwaskom/seaborn: v0.8.1 (September 2017)*, Zenodo. Available at: https://doi.org/10.5281/zenodo.883859.

[14] S. Ergunay, E. Khoury, A. Lazaridis, and S. Marcel, "On the vul- ¨nerability of speaker verification to realistic voice spoofing," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems* (BTAS), 2015, pp. 1–6.

[15] T. N. Sainath, R. J. Weiss, A. Senior, K. W. Wilson,and O. Vinyals, "Learning the speech front-end with raw waveform CLDNNs," in *Proc. of Interspeech*, 2015.

[16] R. Zazo, T. N. Sainath, G. Simko, and C. Parada, "Feature learning with raw-waveform CLDNNs for voice activity detection," in *Proc. of Interspeech*, 2016.

[17] Jinhyeok Jang, Hyunjoong Cho, Jaehong Kim, Jaeyeon Lee, andSeungjoon Yang, Facial Attribute Recognition by RecurrentLearning With Visual Fixation, *IEEE TRANSACTIONS ON CYBERNETICS*, FEBRUARY 2019, VOL. 49, NO. 2.

[18] Lin Yuan, Fanglin Chen, Ling-Li Zeng, Lubin Wang, and Dewen Hu,Gender Identification of Human Brain Image with A Novel 3D Descriptor, *IEEE/ACM Transactions On Computational Biology And Bioinformatics*, MARCH/APRIL 2018, VOL. 15, NO. 2.

[19] G. Chen, X. Feng, Y. Shue, and A. Alwan, "On using voice source measures in automatic gender classification of children's speech," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH '10)*, 2010, pp. 673–676, Chiba, Japan.