THOLETI, G.S.R., GHANTA, D., CHILUKURI, N.V.S.G.S.S. and BANO, S. 2022. Vocal source builds divergence in gender recognition. In *Kumar, A., Senatore, S. and Gunjan, V.K. (eds.) Proceedings of the 2nd International conference on data science, machine learning and applications (ICDSMLA 2020)*, 21-22 November 2020, Pune, India. Lecture notes in electrical engineering, 783. Singapore: Springer [online], pages 171-183. Available from: https://doi.org/10.1007/978-981-16-3690-5_16

Vocal source builds divergence in gender recognition.

THOLETI, G.S.R., GHANTA, D., CHILUKURI, N.V.S.G.S.S. and BANO, S.

2022

Vocal Source builds divergence in Gender Recognition

Guru SreeRam Tholeti[1], Deepika Ghanta[2], N.V.S Guru Sai Sarma Chilukuri[3]
Shahana Bano[4]
Department of CSE[1], Department of CSE[2], Department of CSE[3], Department of CSE[4]
Koneru Lakshmaiah Education Foundation
Vaddeswaram, India
[1]usertgsr@gmail.cm
[2]deepikaghanta15@gmail.com
[3]gurusai21102000@gmail.com
[4]shahanabano@icloud.com

Abstract: Gender Classification is one of the crucial problems in fields of AI. Can a machine be able to classify Gender (Male/Female) ? Yes, the above problem is done using various image classification techniques which uses feature extraction from given set of images. As human being's we can know the gender of the person when we talk to them , can this be achieved by a machine? This is what we are going to work in this research. This research is going to examine the performances of different Machine Learning algorithms and Deep learning algorithms on Gender Classification based on voice. We have used Multi Layer Perceptron (MLP), Random Forest, Decision Tree and Logistic Regression models and compare their performance to find out the best classifier for our data set. We got 96.84% accuracy using MLP, 96.42% using Random Forest, 96.21% using Decision Tree and 89.37% using Logistic Regression. Multi Layer Perceptron stands high with a modest difference in accuracy.

Keywords: Multi Layered Perceptron (MLP), Random Forest, Decision Tree, Logistic Regression, Gender, Voice.

1. Introduction

Performing different tasks on speech is one of the valuable and interesting work now a days, mostly in "**Natural Language Processing"** (NLP) and other branches of AI. Gender identification is considered to be one of the major problems in the field of signal processing [6]. Gender Differences occur in Voice based vocal folds , length of vocal folds, pitch of vocal box and use age of articulators. A close scrutiny, some of the human vocal features, reveals that classifying gender goes way beyond just the frequency and the pitch of a person [6]. The Data Set "Voice.csv" is used in our work taken from Kaggle. This is a classification work in which we have 2 classes and 21 features in our dataset, Male represented as 1 and Female represented as 0. In crime detection, it will be helpful. People commit different types of crimes through phone calls or voice messages. We have used Multi Layer Perceptron (MLP), Random Forest, Decision Tree and Logistic Regression models. Among all the models Multi Layer Perceptron gave the highest accuracy and evaluation metrics such are AUC-ROC curve, confusion matrix, Recall score, accuracy. MLP solves the problem stochastically so it is helpful fitness approximation. Gender classification has applications like, it is able to improve the intelligence of a surveillance system analyse the customer's demands for store management, and allow the robots to perceive gender etc[7].

2. Related Works

Steve Jadav [6] in his work performed gender Classification using SVM and achieved an training accuracy of 89.9% and testing accuracy of 91% by using default parameters . Later he tuned his hyper-parameters and he raised his testing accuracy of 96.6% and testing accuracy of 97% ,So in this work he found the right hyper-parameters as a result both the testing and training accuracy raised

S.M. Badhon, Md. H. Rahaman, Farea Rehnuma Rupon [2] in their work they compared Gradient boosting model, Random Forest and Logistic Regression models on human bengali voices by pre-processing the voice ,Gradient boosting is a ensemble models , which are very efficient on a larger data set they actually divide the whole data set into smaller features and use individual models to achieve highest accuracy they achieved an accuracy of 99.13% for SVM classifier 98.25% for Random Forest classifier and 91.62% for Logistic regression model

P. Gupta, S. Goel, A. Purwa [7] int their work used stacked model consists of SVM , CART and Neural Network in their work all the models will produce some classification based on some features allocated to them and finally the decision is taken by using Random forest which uses majority vote technique for taking a decision which yields an accuracy of 96.73% in which CART model achieved 95.05% accuracy, Neural Network with an accuracy of 95.57% and SVM with an accuracy of 95.78% so all together stacked achieved an highest accuracy that is mentioned above.

In our work we used the same data set used by steve Jadev [6] and used Multilayer perceptron model ,Random forest model ,decision tree model and logistic regression model for classification and we achieved 96.84% accuracy for MLP and achieved 97.09% ROC score and 100% recall score which states that 0 persons are misclassified as females who are actually males.

3. Methodology

3.1 Multi Layered Perceptron (MLP)

The imported data is then split into k folds where k=2. An MLP model is built. The weights and biases are initialized with random values. In the initial iteration, an input pattern is presented and output values are calculated. This process is continued for until the loss value reach the global minima by updating the weights and biases. Once the above process is done, results will be displayed and the process is stopped.

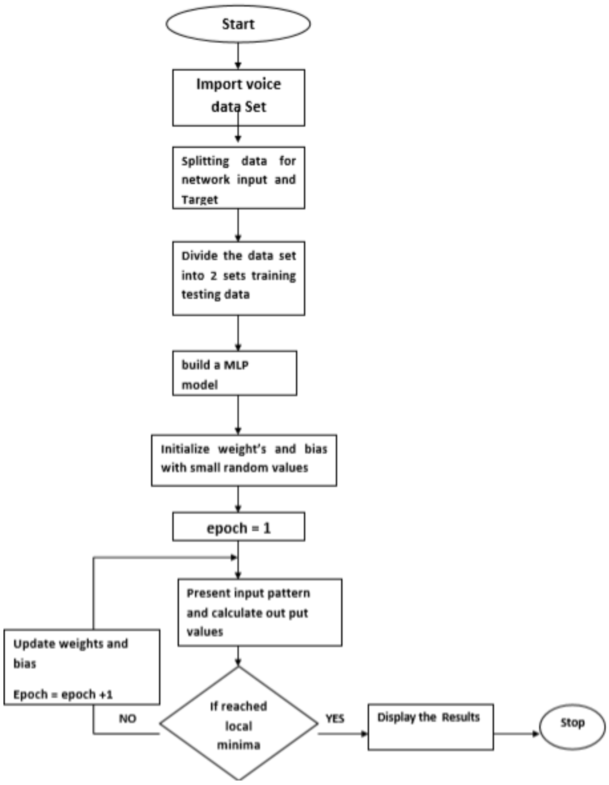

Fig 3.1: Multi Layered Perceptron Flow Chart

3.2 RandomForest

The imported data is splitted into training and testing data. It creates multiple number of decision trees. The outputs are collected from each decision tree. The decision of the model is obtained by decision given by maximum number of trees in the forest is taken as the final decision. Then the new data point is classified into the corresponding class.

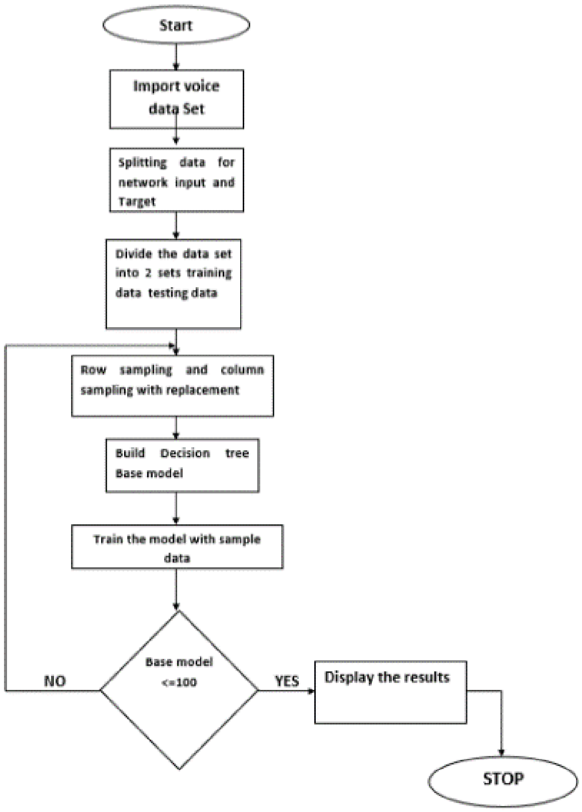

Fig 3.2: Random Forest Flow Chart

3.3 DecisionTree

The imported data is splitted into training and testing data with a split ratio. The system entropy is calculated in the beginning. The information gain of each input feature is calculated. The features with maximum information gain will be selected, this attribute will become a node in the decision tree. Continue to expand the decision tree with combination of the attributes from root to that level. Once the tree is completely derived with given features then display results and stop.

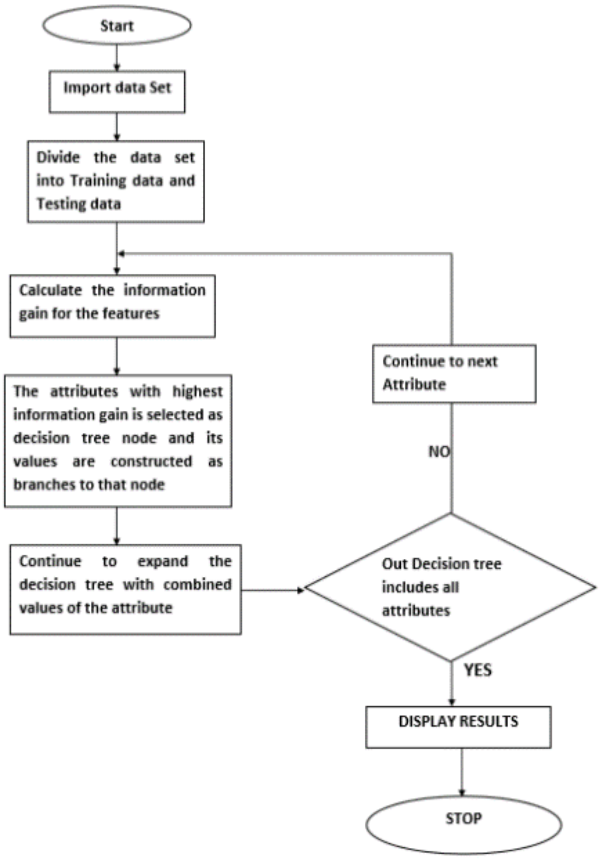

Fig 3.3: Decision Tree Flow Chart

3.4 LogisticRegression

Initially the Voice data is imported into our environment. The data is splitted into train and test set with a split ratio of 70:30. Then the equation of the linear model is sent to the sigmoid function. The sigmoid function generates a number which is between 0 to 1 which will help us in classifying the given point i.e the sigmoid will generate the probability of the person to be placed in male and female. The optimized parameters for theta are calculated by using gradient descent optimizer. Then the classified results are displayed.

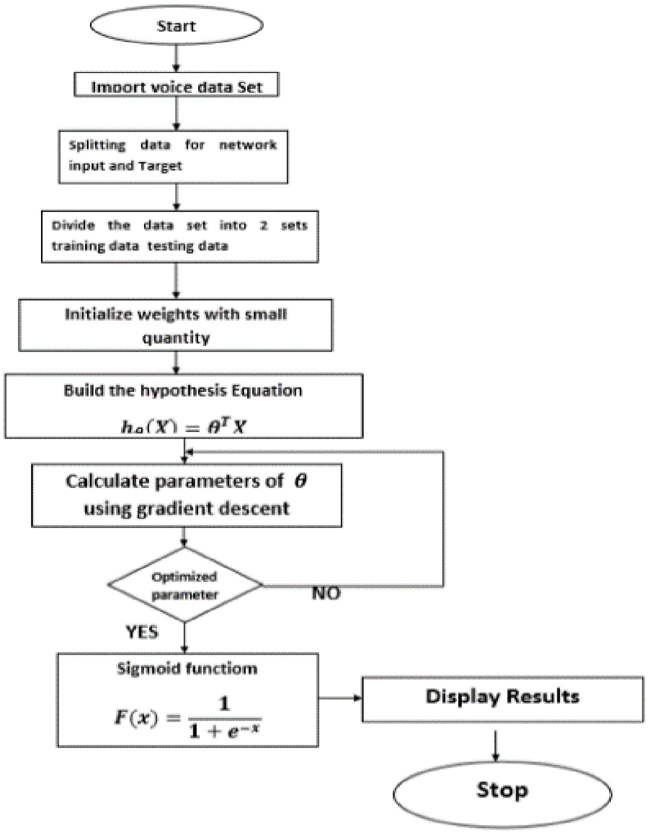

Fig 3.4: Logistic Regression Flow Chart

4. Procedure

Data Set

Voice Data set is taken from Kaggle website which is source for several datasets. There are 21 features and 3168 records present in dataset. Since data set that we are considering having 2 classes(Male , Female) our work is confined to binary classification

Some features in dataset:

- Meanfreq (mean frequency)
- Sd (standard deviation)
- Median

- Mode
- centroid
- Skew
- Gender

4.1 Multi Layered Perceptron (MLP)

Multi Layer Perceptron is used for data which is not linearly separable. MLP model used in my work consists of 1 input layer, 2 hidden layers and 1 output layers, we used “ReLU” Activation function (Eq.2) in hidden layers and “SIGMOID” Activation function(Eq.1) for output Layer.

$$F(s) = \frac{1}{1+e^{-s}} \ --- \ \text{Eq. 1}$$

Sigmoid Activation Function

$$\mathrm{F(s)} = \max(0, \mathrm{s}) --- \text{Eq. 2}$$

ReLU Activation Function

Input layer of our model consists of 20 neurons each one indicates individual features in our dataset the neurons present in input layers are linear i.e. the input and outputs of the input neurons are same. The outputs of input layer are connected to the hidden layers nodes which are computational neurons which will compute the output using below Equation (Eq.3).

$$output = \sum_{i=1}^{n}(w_i x_i) + bias \ --- \ \text{Eq. 3}$$

W → weight of connection; x → input value

Finally the output of output layer is compared with original output in training data, which is known as loss function ,here our main goal is to minimize this loss to global minima,loss is caliculated by Eq.4.

$$loss = \sum_{i=1}^{n}(y-\hat{y})^2 \ --- \ \text{Eq. 4}$$

In the above formula

Y → o/p of the record in training data

ŷ → o/p obtained by the network

initially we got a loss of : 0.0391

To reduce the loss function, we use gradient descent optimizer (Eq.5)

$$W(new) = W(old) - \eta * \frac{\partial loss}{\partial weight} \ --- \ \text{Eq. 5}$$

$\eta \rightarrow$ Learning rate of the network

Finally we got a loss of **:** 0**.**0167; Accuracy **:** 96.84

4.2 RandomForest

Random Forest classifier is an boosting technique of ensemble model, in random forest the base model used is decision tree where the and the final decision is based on majority vote technique, my model creates 100 decision trees.

The complete training data is random sample with replacement and given as input to all the decision trees , while testing the testing data is given to all the trees and the all the decision that is given by majority of the trees in the forest is taken as the final decision of the tree, the main advantage of Random forest is it overcomes the high variance problem faced by decision tree, my random forest model gives an accuracy of 96.42% and F1 score of 96.42%.

4.3 Decision Tree

A decision tree is a decision building or a decision support algorithm with the help of tree like structure in which the nodes indicates the attributes and the branches indicates the decision, decision tree is build on 2 things mainly, Information gain and Gini index, our decision tree model use information gain (Eq.6) as a criteria and for calculating information gain we required entropy given by Eq.7.

$$\mathrm{InformationGain(T, X)} = \mathrm{entropy(X)} - \mathrm{Entropy\ (X, T)} \ --- \ \mathrm{Eq.\,6}$$

T $\rightarrow$ Target variable; X $\rightarrow$ Individual attribute

$$\mathrm{Entropy(X)} = -\mathrm{plog_2p} - \mathrm{qlog_2q} \ --- \ \mathrm{Eq.\,7}$$

X $\rightarrow$ individual attribute; p $\rightarrow$ probability of class1; q$\rightarrow$probability of class2

By using above to formulas we got “meanfun” attribute as our main root with 0.5 information gain.
Accuracy of the model on the given data set is 96.21 and F1 Score of 96.27

4.4 Logistic Regression

Logistic regression is mainly used for data which is linearly non separable, Logistic regression hypothesis function (Eq.8) can be given by

$$H_\theta(x) = \theta^T * x \ --- \ \mathrm{Eq.\,8}$$

$$\theta^T \rightarrow weight\ vector;\ x \rightarrow attributes$$

and applies an activation function namely sigmoid function to it. The sigmoid function take the predicted value from linear model and converts into a value between 0 and 1.

The sigmoid function is given by Eq.9,

$$Sigmoid(\theta^T * x) = \frac{1}{1+e^{-\theta^T * x}} --- \text{Eq. 9}$$

According to probability value of sigmoid function, we classify our input data into classes present in our dataset. We use a cost function (Eq.10), which will be optimized to find the best parameters of theta.

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^i log h_\theta\left(x^i\right) + (1-y^i)\log(1-h_\theta(x^i))\right] --- \text{Eq. 10}$$

To minimize the cost function we used a gradient descent optimizer and it can be computed by using Eq .11

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)x_j{}^i --- \text{Eq. 11}$$

for logistic regression model over voice data set we got an accuracy of 88.53% and F1 score of 89.06% and Cost is 0.325.

5 Results

5.1 Multi Layered Perceptron(MLP)

In the result of MLP classifier we have Accuracy of testing data as 96.84% (From Fig 5.1) and an accuracy of 96.16% on training data(From Fig 5.2), from confusion metrics we have that 162 persons are classified correctly into female class and 145 persons are classified correctly into male class and there are 0 persons who are male and classified into female and 10 persons who are classified into male but actually female. The Roc Score of our model is 97.09% which indicates the trust level of the model.

```
print("\nAccuracy score: %f" %(accuracy_score(testY,predictions_MLP) * 100))
print("Recall score : %f" %(recall_score(testY,predictions_MLP) * 100))
print("ROC score : %f\n" %(roc_auc_score(testY,predictions_MLP) * 100))
print(confusion_matrix(testY,predictions_MLP))
print("F1 score : %f\n" %(f1_score(testY,predictions_MLP, zero_division=1)*100))
```

```
Accuracy score: 96.845426
Recall score : 100.000000
ROC score : 97.093023

[[162  10]
 [  0 145]]
F1 score : 96.666667
```

Fig 5.1: Result of MLP model

```
mlp = MLPClassifier(hidden_layer_sizes=(50,3),max_iter=301)

y_t = y_t.reshape(-1, 1)
print(x_t.shape,y_t.shape)
mlp.fit(x_t,y_t)
print("Training set score: %f" % mlp.score(x_t,y_t))
print("Test set score: %f" % mlp.score(x_tes,y_tes))
```

```
(2217, 20) (2217, 1)

E:\Anaconda\lib\site-packages\sklearn\neural_network\_multilayer_perceptron.py:934: DataConversionWarning: A column-
s passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

Training set score: 0.961660
Test set score: 0.968454
```

Fig 5.2: Training and Testing Set Scores of MLP model

5.2 RandomForest

In the result of Random Forest classifier we have Accuracy of testing data as 96.42% (From Fig 5.3), from confusion metrics we have that 459 persons are classified correctly into female class and 458 persons are classified correctly into male class and there are 16 persons who are male and classified into female and 18 persons who are classified into male but actually female. The Roc Score of our model is 96.42% which indicates the trust level of the model.

```
print("\nAccuracy score: %f" %(accuracy_score(y_test,RF_ypre) * 100))
print("Recall score : %f" %(recall_score(y_test, RF_ypre) * 100))
print("ROC score : %f\n" %(roc_auc_score(y_test, RF_ypre) * 100))
print(confusion_matrix(y_test, RF_ypre))
print("F1 score : %f\n" %(f1_score(y_test, RF_ypre, zero_division=1)*100))
```

```
Accuracy score: 96.424816
Recall score : 96.624473
ROC score : 96.425444

[[459  18]
 [ 16 458]]
F1 score : 96.421053
```

Fig 5.3: Result of Random Forest model

5.3 Decision Tree

In the result of Decision Tree classifier we have accuracy of testing data as 96.21% (From Fig 5.4), From confusion metrics we have that 450 persons are classified correctly into female class and 465 persons are classified correctly into male class and there are 19 persons who are male and classified into female and 17 persons who are classified into male but actually female. The Roc Score of our model is 96.42% which indicates the trust level of the model.

```
print("\nAccuracy score: %f" %(accuracy_score(y_test,dt_ypre) * 100))
print("Recall score : %f" %(recall_score(y_test, dt_ypre) * 100))
print("ROC score : %f\n" %(roc_auc_score(y_test, dt_ypre) * 100))
print(confusion_matrix(y_test, dt_ypre))
print("F1 score : %f\n" %(f1_score(y_test, dt_ypre, zero_division=1)*100))
```

```
Accuracy score: 96.214511
Recall score : 96.074380
ROC score : 96.217062

[[450  17]
 [ 19 465]]
F1 score : 96.273292
```

Fig 5.4: Result Of DecisionTree model

5.4 Logistic Regression

In The result of Logistic regression classifier we have Accuracy of testing data as 89.37% (From Fig 5.5), From confusion metrics we have that 381 persons are classified correctly into female class and 469 persons are classified correctly into male class and there are 25 persons who are male and classified into female and 76 persons who are classified into male but actually female. The Roc Score of our model is 89.15% which indicates the trust level of the model.

```
print("\nAccuracy score: %f" %(accuracy_score(y_test,lr_ypre) * 100))
print("Recall score : %f" %(recall_score(y_test, lr_ypre) * 100))
print("ROC score : %f\n" %(roc_auc_score(y_test, lr_ypre) * 100))
print(confusion_matrix(y_test, lr_ypre))
print("F1 score : %f\n" %(f1_score(y_test, lr_ypre, zero_division=1)*100))
print("Cost : %f\n" %np.sqrt(((lr_ypre - y_test) ** 2).mean()))
```

```
Accuracy score: 89.379600
Recall score : 94.939271
ROC score : 89.154537

[[381  76]
 [ 25 469]]
F1 score : 90.279115

Cost : 0.325890
```

Fig 5.5: Result Of LogisticRegression model

We perform classification on same dataset using four model their Accuracy and F1 Score is given by “Table 1”.

In “Table 2” we can find the different metrics in “Multi Layer Perceptron” model such as accuracy , Recall , Roc Score and F1 score.

S No.	Model name	Accuracy	F1 Score
1	MLP	96.84	96.66
2	Random forest	96.42	96.42
3	Decision Tree	96.21	96.27
4	Logistic Regression	89.37	90.27

Table 1: Accuracy and F1 Score Of All Models used

S No.	Model	Accuracy	Recall	ROC Score	F1 Score
1	MLP	96.84%	100%	97.093%	96.66%

Table 2: Accuracy, Recall, Roc-Score, F1-Score of MLP model

6 Conclusions

In this work we tried to classify gender based on voice, we performed this work only on 2 types of genders, the 3rd type of gender is not considered in our work. Now a days we can handle many apps based on voice in which gender classification is done first, in our work we took 1584 male and 1584 female records and achieved an accuracy of MLP is 96.84 and error of 0.03 and 0.94 True positive rate and 1.00 true negative rate.

7 References

[1] T. Roshini, P. V. Sireesha, D. Parasa and Shahana Bano, “Social Media Survey using Decision Tree and Naive Bayes Classification”, 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) doi:10.1109/icct46177.2019.8969058 .

[2] S.M. Badhon1, Md. H. Rahaman2, Farea Rehnuma Rupon3 , “A Machine Learning Approach to Automating Bengali Voice Based Gender Classification” , 2019, 8th International Conference on System Modeling & Advancement in Research Trends (SMART). doi:10.1109/smart46866.2019.9117385

[3] R. RamyaSri, S. IshaSanjida, D. Parasa and Shahana Bano, "Food Survey using Exploratory Data Analysis," 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) doi:10.1109/icct46177.2019.8969016

[4] B. Hassan, R. Ahmed, Bo Li, O. Hassan,T. Hassan, ”Autonomous Framework for Person Identification by Analyzing Vocal Sounds and Speech Patterns”, 2019 5th International Conference on Control, Automation and Robotics (ICCAR), Beijing, China, 2019, pp. 649-653, doi: 10.1109/ICCAR.2019.8813463.

[5] Yasmin G., Mullick O., Ghosal A., Das A.K. “Gender Recognition Inclusive with

Transgender from Speech Classification". 2019 12 December Emerging Technologies in Data Mining and Information Security. Information Security, 89–98. doi:10.1007/978-981-13-1951-8_9

[6] Steve Jadav , "Voice-Based Gender Identification Using Machine Learning" , 2018 4th International Conference on Computing Communication and Automation. (ICCCA), Greater Noida, India, 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777582.

[7] P. Gupta, S. Goel, A. Purwar," A Stacked Technique for Gender Recognition Through Voice", 2018 Eleventh International Conference on Contempo- rary Computing (IC3) Noida, 2018, pp. 1-3, doi: 10.1109/IC3.2018.8530520.

[8] S. Chaudhary and D. K. Sharma, "Gender Identification based on Voice Signal Characteristics," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 2018, pp. 869-874, doi: 10.1109/ICACCCN.2018.8748676.

[9] Raahul, A & Sapthagiri, R & Pankaj, K & Vijayarajan, Vijayan. (2017). Voice based gender classification using machine learning. IOP Conference Series: Materials Science and Engineering. 263. 042083. 10.1088/1757-899X/263/4/042083.

[10] Świebocka-Więk, Joanna. (2017). Gender Recognition Based on Speaker's Voice Analysis. 539. 80-85. 10.1007/978-3-319-48944-5_8.

[11] Dharamraj yadav, Shashwat Shukla, Bramah Hazela,"Gender classification using face image and voice",2015 , IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 5, Ver. V (Sep. – Oct. 2015), PP 20-25.

[12] M. Kumari and I. Ali, "An efficient algorithm for Gender Detection using voice samples," 2015 Communication, Control and Intelligent Systems (CCIS), Mathura, 2015, pp. 221-226, doi: 10.1109/CCIntelS.2015.7437912.

[13] M. Abdollahi, E. Valavi and H. A. Noubari, "Voice-based gender identification via multiresolution frame classification of spectro-temporal maps," 2009 International Joint Conference on Neural Networks, Atlanta, GA, 2009, pp. 1-4, doi:10.1109/IJCNN.2009.5178984.

[14] Yen-Liang Shue and M. Iseli, "The role of voice source measures on automatic gender classification ," 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, 2008, pp. 4493-4496, doi: 10.1109/ICASSP.2008.4518654.

[15] Rahimi Azghadi, Mostafa & Bonyadi, Mohammad reza & Shah-Hosseini, Hamed. (2007). Gender Classification Based on FeedForward Backpropagation Neural Network. 299-304. 10.1007/978-0-387-74161-1_32.