ISSN No:-2456-2165

Voice Gender Recognition

Sakshi Malhotra[1]
Department of Information Technology Galgotias College of Engineering and Technology
Greater Noida, India

Kushagra Singh[2]
Department of Information Technology Galgotias College of Engineering and Technology
Greater Noida, India

Mukul Nag[3]
Department of Information Technology Galgotias College of Engineering and Technology
Greater Noida, India

Kuldeep Nirmal[4]
Department of Information Technology Galgotias College of Engineering and Technology
Greater Noida, India

Abstract:- Due to its widespread application in a variety of circumstances, the gender classification system has become more and more relevant including social media platforms and criminal investigations. Prior studies in this area have mostly focused on discrimination against both men and women. Nevertheless, since transgender persons have just received legal recognition, it has been vital to create techniques for accurately diagnosing gender from a specific voice, which can be a challenging undertaking. To extract pertinent characteristics from a training set that may be utilised to create a model for gender categorization, researchers have employed a number of techniques. Following then, a vocal signal's gender may be ascertained using this model. The study makes three significant contributions: first, it provides a thorough analysis of well-known voice signal features using a well-known dataset; second, it investigates a variety of machine learning models from a variety of theoretical families to classify voice gender; and third, it uses three well-known feature selection algorithms to select the features that have the greatest potential to improve classification models.

***Keywords*:-** *Python, Machine Learning, Transformer, Ten-Sorflow, Spectrogram, Matplotlib, Pandas, HTML, CSS, Django.*

I. INTRODUCTION

Gender can be determined by using speech and voice recognition technology. Based on the frequency and volume of a person's voice and speech, the ear has an excellent system for determining their gender. A method known as machine learning voice recognition employs machine learning algorithms to assist computers in comprehending and interpreting spoken language. The algorithm is fed a lot of labelled data during the training process, allowing it to learn from examples and get better over time. When the AI calculation has been prepared, it tends to be utilized to perceive and decipher human discourse continuously. Virtual assistants, speech-to-text transcription, language translation, and voice-controlled devices like smart speakers and home automation systems are just a few of the many uses for this technology.

II. LITRATURE REVIEW

After studying the existing publications that are relevant to the idea of our proposed system, we found out that a huge amount of voice gender recognition models are built on Support Vector Machine (SVM), CART, Random Forest, and deep learning techniques like Multilayer Perceptron (MLP) and GBM [1][2][3][7][8][9]. Several programs have helped with gender classification in the past, but only for men and females, not transgender people One of the findings was that it uses gender and language to identify the language of spoken utterances and identify the speaker's gender based on their voice. Some of them use orthographic transcription accomplished by recording and analyzing the speaker's speech using the Gaussian Mixture Model and MFCC feature ex- traction called automatic speech recognition (ASR) and semi- supervised learning [4][5][13] focus male and female only. The transformer input layer is used to implement ASR and Trans- former Keras voice recognition, Librosa, Mel Spectrogram which interprets audio waves to identify transgenders from male and female voice datasets. In order to create classifiers that are more accurate, current research focuses on fusing ensemble learning strategies with semi-supervised learning frameworks. Making emotional speech understandable in order to determine the speaker's gender makes the challenge even more intriguing [6][10]. Recently, a number of cutting-edge gender recognition methods based on several biometrics, in- cluding the face, body form, and voice, have been presented. The worst one of them is relying solely on voice. Voice verification, gender categorization from voice, and native (mother tongue) linguistic context were all explored. Four classifiers are used in a stacked ensemble for gender voice identification as the basis classifiers: LR, KNN, SVM, SGD, and LDA[11][12][14].

III. PROPOSED SYSTEM

The VOICE GENDER RECOGNITION proposes to solve the classification of genders through voice not only in males and females but also to include transgenders. System design of multiple technologies are included in voice gender recognition. The system uses Python and machine learning for the backend and HTML, CSS, Django, and JavaScript for the front-end development. Here, algorithms will help in the

ISSN No:-2456-2165

classification of the genders by matching the extracted data from the provided input with the stored dataset. In order to host the programme on the website, a user interface is also created. The Voice Gender Recognition GUI is designed in the first place, on a website. Users can offer speech data to be recognised as input on the website thanks to a voice fetch mechanism. The model is created using machine learning methods in the way that follows:

- Extraction of data Machine learning is a process in whichthe data is extracted from an original source, and then it is processed to obtain information. In order to extract the data from the .wav audio file, we use Spectrogram.
- Training the data Fitting a model to fresh data is the process of machine learning. The main aim of the training is to identify the model's ideal parameters. To discover these parameters in practice, we often employ an opti- misation approach (such as gradient descent). Machine learning training is the process of developing a model from unstructured data. We use Python, Spectrogram, Transformer, TensorFlow, matplotlib, PyCharm, NumPy,Pandas and Jupyter Notebook.
- All the acoustic features are extracted with the use of librosa library, it is then stored in .csv file for further processing. Pattern matching is the collection of tasks performed by various Machine Learning tools like pan- das, matplotlib etc., for identifying trends.

A. *Supervised Learning and Classification of Rules*

One can classify rules based on the information they contain with supervised learning. It can be very useful for finding patterns in our data that will help us decide which rule to apply to a new situation or case more effectively.

B. *Voice Gender Recognition*

The ML-based voice identification tool takes the .wav audio file as input from the user, and will first extract the features using R Programming, it is then stored in .csv file for further processing. Pattern matching is performed to identify the trends by using various ML tools like matplotlib, and pandas etc., Based on the pattern matchings and various trends the decision logic algorithms like SVM, XGBoost etc., come into play for identifying gender and put the final decision.

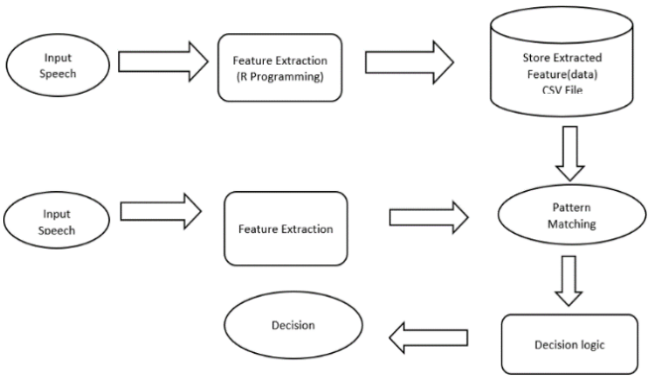

Fig 1 System Flow Diagram

C. *CSV Dataset*

Wav file is a sound format that can be played in Windows Media Player, QuickTime and even iTunes. The .wav exten- sion stands for Waveform Audio File and it was created by Microsoft.

➢ *Implementation*

This project's primary goal is to address the issue of trans- gender voice gender recognition. The implementation makes use of the following programmes, devices, or frameworks:

D. *Web Design*

Any system includes a large portion called the Graphical User Interface (GUI). An Audio file entry Page has been made with the aid of HTML, CSS, Django, and JavaScript utilising the UI of VOICE GENDER RECOGNITION. The website is utilised to get user voice data in.wav file format. The audio file is further processed as input speech to themachine learning model, which in the backend performs pattern-matching algorithms to give the most accurate result possible which is then displayed to the user.

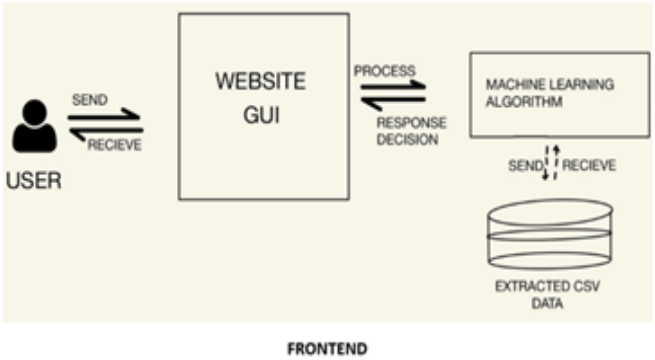

Fig 2 GUI Implementation

E. *Workflow Graph*

Figure 4, explains how the whole workflow is done in the whole project. The graph completes each step of the project very accurately. The training phase describes how the model

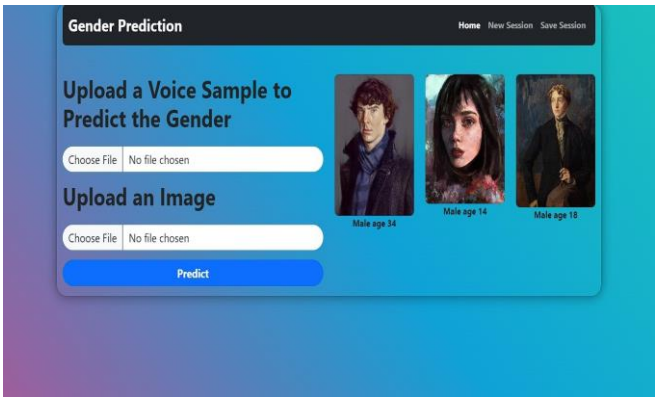

Fig 3 Web UI for input

Is being trained for male, female, and transgender voices and which is then tested with an untouched dataset in the testing phase.

ISSN No:-2456-2165

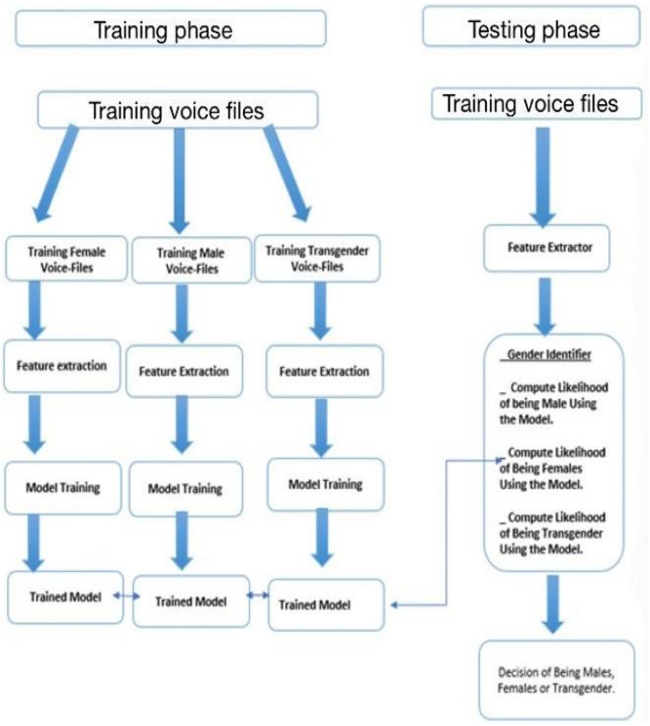

Fig 4 Web UI for Input

F. Workflow & Process Definition for Final Model

The process of analyzing a speech signal to extract relevant information in a form that is smaller than the speech signal it- self is known as speech analysis. Multiple application domainsmake extensive use of AI and machine learning. Vectors of features An ordered list of numerical properties of observed phenomena is called a feature vector. A prediction-making ma- chine learning model uses it as input features. Decisions can be made by humans by analyzing qualitative data. A conceptual framework that standardizes communication between diverse networks is provided by reference models.

G. Transformer Keras Voice recognition

A transformer is a machine-learning technique that utilizes self-attention mechanisms for sequential input data processing.It has become more widely used as a method for completing numerous natural language processing (NLP) tasks, including

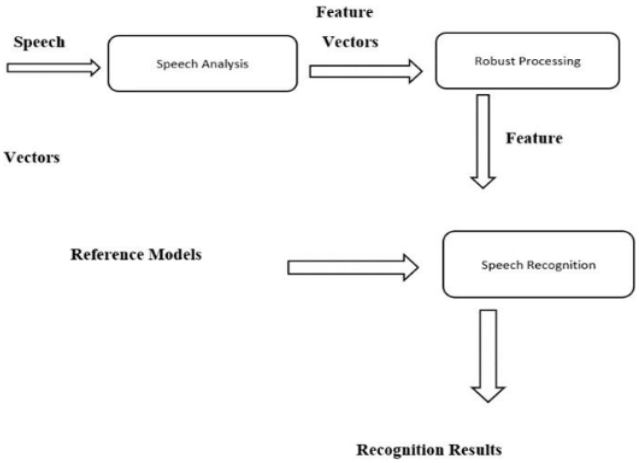

Fig 5 Workflow & Process Definition for Final Model

Text summarization, sentiment analysis, and language trans- lation. The self-attention feature enables the model to weigh different parts of the input in the representation of the input sequence. To do this, attention weights are calculated for each point in the input sequence depending on how similar the current position is to all other positions in the series. Thereare two types of transformers: an encoder and a decoder. The former processes the input sequence, generating hidden representations for each input position. The latter generates the output sequence, based on the encoder's hidden rep- resentations. Unlike RNNs that process inputs sequentially, transformers can parallelly process input sequences, making it one of their key advantages. Many computer machine learning systems, including those that comprehend human speech, do picture analysis, and employ voice recognition, frequently use a component of the apparatus called the spectrograph.

H. Spectrogram

Machine learning methods can be used to analyse spectral data and extract relevant features from the data. These algo- rithms can find patterns in spectral data that are challengingto notice with the naked eye. By examining the patterns and correlations in the spectrum data, machine learning algorithms are able to find individual spectroscopic fingerprints that are unique to particular materials or compounds. Spectrum data analysis and insight extraction are now made possible by machine learning, which was previously impractical orimpossible to do. As spectroscopic research develops, machine learning is anticipated to play a bigger role in the analysis and interpretation of spectrum data.

I. Dataset

Here we are Using CSV Files for our model. The fields meanfreq, sd, median, Q25, Q75, IQR, skew, Kurt, sp. ent,sfm, mode, centroid, meanfun, minfun, maxfun, meandom, mind,maxdom, range, modding, label correspond to the sam- ple's gender. Acoustic qualities are specified in the remaining fields. Along with the pre-processed dataset, the training data also includes the raw voice samples.WAV files kept in a different location.

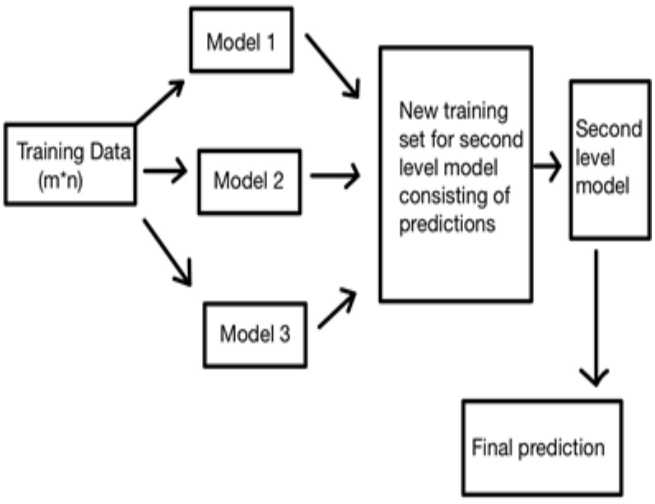

Fig 6 Architecture Diagram

J. Pearson Correlation of Features

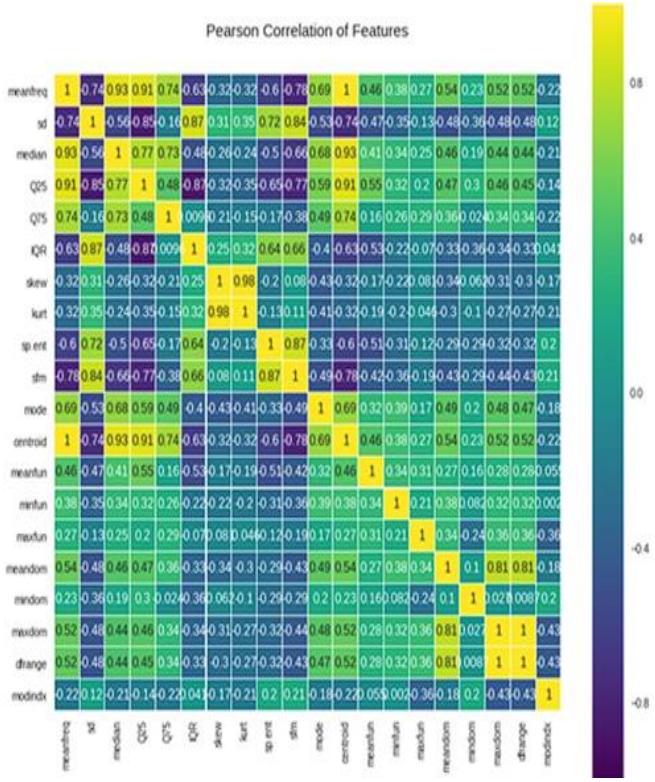

Fig 7 Pearson Correlation of Features

The Pearson correlation measures the degree to which two variables have a linear relationship. A value of -1 indicates a total negative linear correlation, a number of 0 means there is no association, while a value of +1 means there is a completelypositive correlation.

IV. RESULT

After giving the model a defined input under accurate bounds, the output is generated in terms of Male, Female and Transgender. The précised output is displayed to the user through GUI:

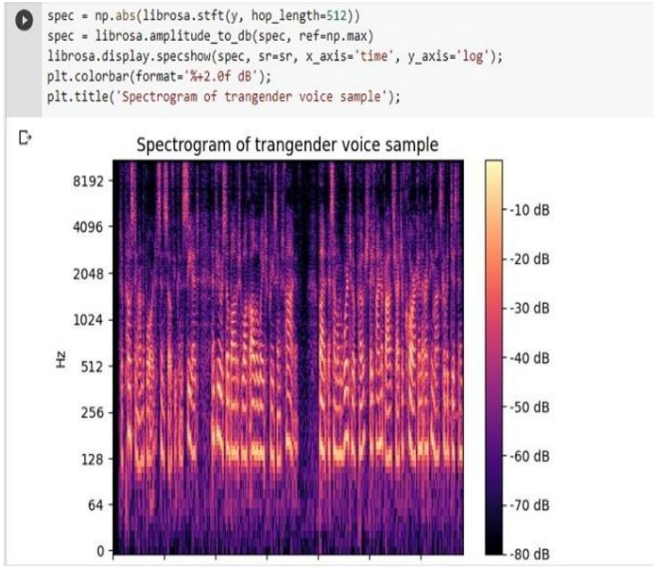

Fig 8 Result

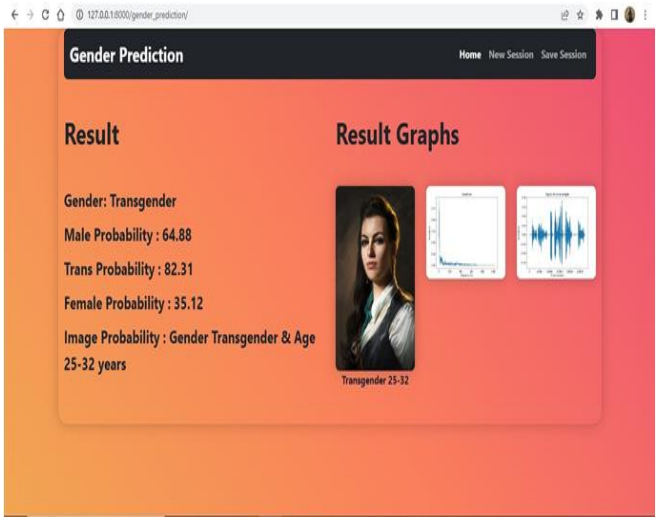

Fig 9 GUI Representing Result

V. CONCLUSION & FUTURE WORK

VOICE GENDER RECOGNITION is the tool we are building to identify the gender of a person based on vocal data and information. In this paper we are presenting the design, the proposed system in this paper and the implementation. Speech and sound specialists have found it challenging to deduce a person's gender based on their voice, even when employing a number of technologies, such as CRM systems' Effective advertising and marketing strategies. Investigating the voice of the culprit in crime scenes. improving conversation systems and other human-computer interaction (HCI) technologies. In the medical sector, diagnosing people with voice issuesmay be highly helpful. By taking over duties that don't require humans, it increases efficiency. The purpose of acousticcharacteristics is to react to sound waves. Gender recognition is a method for figuring out a speaker's gender category. The length, strength, frequency, and filtering of an acoustic signal may be learnt from the signals of a recorded voice. Adaptation of music for waiting rooms so that different kinds of music can be played depending on age and gender. It can help in various other fields like Robotics, finance and banking, computer vision, etc.

ACKNOWLEDGMENT

We appreciate Ms Shakshi Malhotra, our project mentor, forsupporting us and assisting us in resolving some of the most difficult difficulties in creating our project.

REFERENCES

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

[2] G. Sharma and S. Mala, "Framework for gender recognition using voice," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 32-

37, doi:10.1109/Confluence47617.2020.9058146.
[3] K. Zvarevashe and O. O. Olugbara, ”Gender Voice Recognition Using Random Forest Recursive Feature Elimination with Gradient Boost- ing Machines,” 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 2018, pp. 1-6, doi: 10.1109/ICABCD.2018.8465466.
[4] W. Li, D. -J. Kim, C. -H. Kim and K. -S. Hong, ”Voice-Based Recognition System for Non-Semantics Information by Language and Gender,” 2010 Third International Symposium on Electronic Commerce and Security, Nanchang, China, 2010, pp. 84-88, doi: 10.1109/ISECS 2010.27.
[5] N. M and A. S. Ponraj, ”Speech Recognition with Gender Identification and Speaker Diarization,”2020 IEEE International Conference for Inno- vation in Technology (INOCON), Bangluru, India, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298241.
[6] Livieris, Ioannis & Pintelas, Emmanuel & Pintelas, P.. (2019). Gen- der Recognition by Voice using an Improved Self-Labeled Algo- rithm. Machine Learning and Knowledge Extraction. 1. 492-503. 10.3390/make1010030.
[7] Büyükyılmaz, Mücahit & Ç ıbıkdiken, Ali. (2016). Voice Gender Recog- nition Using Deep Learning. 10.2991/msota-16.2016.90.
[8] S. Chaudhary and D. K. Sharma, ”Gender Identification based on Voice Signal Characteristics,” 2018 International Conference on Ad- vances in Computing, Communication Control and Networking (ICAC- CCN), Greater Noida, India, 2018, pp. 869-874, doi: 10.1109/ICAC-CCN.2018.8748676.
[9] L. Jasuja, A. Rasool and G. Hajela, ”Voice Gender Recog- nizer Recognition of Gender from Voice using Deep Neural Net- works,” 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 319-324, doi: 10.1109/ICOSEC49089.2020.9215254.
[10] M. Kotti and C. Kotropoulos, ”Gender classification in two Emotional Speech databases,”2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 2008, pp.1-4, doi: 10.1109/ICPR.2008.4761624.
[11] M. Wang, Y. Chen, Z. Tang and E. Zhang, ”I-vector based speaker gen- der recognition,”2015 IEEE Advanced Information Technology, Elec- tronic and Automation Control Conference (IAEAC), Chongqing, China, 2015, pp. 729-732, doi:10.1109/IAEAC.2015.7428651.
[12] O. Iloanusi et al., ”Voice Recognition and Gender Classification in the Context of Native Languages and Lingua Franca,” 2019 6th In- ternational Conference on Soft Computing & Machine Intelligence (IS- CMI), Johannesburg, South Africa, 2019, pp.175-179, doi: 10.1109/IS-CMI47871.2019.9004306.
[13] Al khammash, Eman & Hadjouni, Myriam & Elshewey, Ahmed. (2022). A Hybrid Ensemble Stacking Model for Gender Voice Recognition Approach Electronics.11.1750.10.3390/electronics11111750.
[14] Fahmeeda, Sayyada Ayan, Mohamed & Shamsuddin, Mohamed & Amreen, Aliya. (2022). Voice-Based Gender Recognition Using Deep Learning. International Journal of Innovative Research & Growth. 3.649-654.