Middle East Journal of Applied Science & Technology (MEJAST)
Volume 8, Issue 1, Pages 40-55, January-March 2025

A Machine Learning Approach for Identifying Gender Based on Bengali Vocal Cues

Al Arman Ovi[1], Iffath Tanjim Moon[2] & Md. Julker Nayeem[3*]

[1,2]Department of Computer Science & Engineering, Pundra University of Science & Technology, Bogura-5800, Bangladesh. [3]Department of Computer Science & Engineering, International Islami University of Science and Technology Bangladesh, Dhaka-1349, Bangladesh.
Corresponding Author (Md. Julker Nayeem) Email: julkernayeemt72@gmail.com[*]

DOI: https://doi.org/10.46431/MEJAST.2025.8104

Article Received: 13 November 2024 *Article Accepted: 19 January 2025* *Article Published: 26 January 2025*

ABSTRACT

In this research, an advanced system getting-to-know model is being carried out to properly examine and analyze multiple Bangladeshi vocal cues to diagnose gender correctly. This study is implemented in quite a few domain names, including customer service, where determining the gender of a caller can provide extra-personalized interactions. Additionally, voice-activated assistants use it to customize responses and beautify the consumer experience. To beautify provider shipping and offer demographic insights and gender recognition the use of voice evaluation is also utilized in safety systems, transcription services and sociolinguistic research. Voice is stricken by environmental way of life elements such as smoking, acid reflux sickness, air pollutants, warm weather weight reduction, city air pollutants and the horrible effect of energy on Bangladeshi fitness. In this study, we obtained data on voices from male and female participants living in Bangladesh. To provide a consistent and convenient method of research, we first converted each recording to Waveform Audio File Format (WAV). We then extracted the most significant voices from these WAV recordings and converted them to statistical data. Then, we preprocessed this statistical information to put it together for in-depth analysis. After preprocessing, we used report visualization to understand the traits and patterns observed within the voice recordings. This holistic approach enables a complete assessment of voice data to gain the goals of our study. So, the idea of this venture is to train a machine learning model with updated information processing techniques which can be expected should be gender according to voice notes. We goal to establish a dependable gender identification set of rules based on modern-day findings and big-scale facts.

Keywords: Voice analysis; Gender classification; Acoustic features; Machine learning; Vocal biometrics; Machine learning classification; Bengali language dataset; Gender recognition system; Random forest; Logistic regression; Bangla audio data; Audio data.

1. Introduction

Voice analysis helps spot gender in many fields today. Customer service bots use it to tailor responses, making callers happier. Security systems add it as an extra check, working with other body scans. Doctors rely on it to spot voice issues and track treatment results, since some health problems change how men and women sound differently. The manner in which robots and virtual assistants speak is shaped by voice gender technology. These voices sound more human when imitating the sex of the person they are talking to. Moreover, this know how arranges audio documents by speech, thus making searching for information easier than ever before. With its wide applications, it enhances experiences as well as increases security and facilitates medical demands. From help desks to hospitals, this tool packs a wallop in how we deal with technology and each other. Voice gender ID packs a punch in customer service and healthcare. This technology has a huge impact on our digital conversations. It's not just neat; it's flipping the script on our online connections. Talk about a game-changer! This stuff blows my mind. We then extracted the hugest voices from these WAV recordings and converted them to statistical facts. Then, we preprocessed these statistical facts to place it together for in-intensity analysis. After preprocessing, we used document visualization to apprehend the tendencies and styles discovered in the voice recordings. This holistic approach enables an entire evaluation of voice information to benefit the goals of our look at. So, the idea of this undertaking is to educate a system gaining knowledge of model with up-to-date information processing strategies which could anticipate because it has to be gender in accordance voice notes. We intention to set up a reliable gender identification set of policies based at the contemporary-day findings and big-scale records. The research is sound on voice analysis used for the identification of gender in audio data transmitted by Bangladeshi speakers.

The entire mechanism is working with the system of data collection, what is called pre-processing that is converting data to the WAV numeric form, and finally, the CSV data storage. The next step is that the audio data undergoes normalization, elimination of noise and additional features, such as Mel-frequency cepstral coefficients (MFCCs) are extracted. The data is, in the following step, represented through plots and graphs that can help understand the distribution and patterns of the whole dataset. This step-by-step procedure guarantees the correctness of the identification of gender in the case of the Bangladeshi speaking population.

1.1. Study Objectives

(a) AI Model Development: Develop a state-of-the-art artificial intelligence model capable of classifying gender based on the features of the Bengali accent.

(b) Contribution to Phonetics and Linguistics: Deepen the understanding about the phonetics and linguistics of the Bengali language.

(c) Computational Linguistics Applications: Tool up and give valuable insights for computational linguistics, ASR, and HCI.

(d) Culture-Free and Accessible Technology: Provide relevant, off-the-shelf solutions crossing cultural barriers to further inclusiveness.

(e) Enhance Speech Recognition Efficiency: Address language-specific challenges to make speech recognition technologies more accurate and productive.

(f) Validate Machine Learning in Societal Contexts: Analyze the interplay between machine learning, needs of society, and linguistics.

(g) Support Diverse Applications: Enable applications in the areas of accessibility solutions, forensic analysis, personalized AI systems, and regional language preservation.

1.2. Problem statements

(a) Gender recognition is important in voice-based systems to make them more personalized and efficient, (b) Most systems today are made for English and don't work well for other languages, especially Bengali, (c) Bengali speech has its own unique sounds and patterns, which makes gender identification difficult, (d) There is no strong machine learning system that can identify gender from Bengali voices accurately, (e) Current systems don’t consider the linguistic and cultural differences, leading to lower accuracy for Bengali speakers, and (f) This research aims to create a machine learning system that can accurately identify gender using Bengali speech.

2. Literature Review

V.S.K. Reddy & R. Surendran [1] introduce experiments using random forests and a new decision tree scheme demonstrates effective methods for predicting human voice recognition irrespective of gender. Random Forest algorithm is an AI-based calculation that is a variant of Recursive Element Disposal. It was shown that the Novel Decision Tree algorithm iteratively using a random forest in the data set is very accurate in predicting male or female voice recognition with an accuracy of 97.79%. S. Jadav [2] Gender determination is an important signal processing problem, traditionally solved by image classification techniques. Recently, researchers have focused on sex classification using vocal features. This paper explores the efficiency and importance of machine learning algorithms in gender-based voice recognition, focusing on feature selection and shape reduction, as well as gender-related characteristics sexual decisions.

V.S. Kone et al. [3] A Machine Learning model for recognizing age from voice using Deep Learning techniques is discussed in this paper. Gender can also be detected using this model; it employs Robust Scalar, Principal Component Analysis (PCA) and Logistic Regression algorithms for that purpose. This model employs an approach of identification of a dataset's best predictive age algorithm by utilizing a pipeline based on a grid-search technique. Over 91% on the gender, over 59% of the tested set's age demography were predicted correctly though not all accurately which shows how much work remains undone and what it takes to do so.

M. Markitantov & O. Verkholyak [4] this piece discusses a new technique based on deep neural networks that can be employed to determine one's sex or the number of years they have been in existence. It has been used and proven successful in the analysis of Genders, which are German speeches where the experiment was carried ousted upon. In comparison to the conventional ways, the effectiveness of the program ranges from 48.41% (percentage correct) without regard to classes to 57.53% for gender and 88.80% for age separately. K. Chachadi & S.R. Nirmala [5] this paper devises a comparison between a neural network model that uses features such as MFCC and the Mel spectrogram in order to distinguish men from women through speech patterns. With experiments done on the Mozilla voice dataset, it was discovered that attaining 94% accuracy for gender recognition through speech processing required that both MFCCs and melt spectra be used together at one time point while keeping separate their features as shown in Table 2 below table 2 for illustration purposes only. The authors concluded that using a combination of both features led to an increased accuracy of 94.32% making this method preferable for speech problems including automatic gender identification or authentication.

W. Li et al. [6] this paper introduces a practical language and gender identification system for non-semantic voice data, which combines language and gender detection models. The system does not depend on the speaker or the text they are reading, obtaining high accuracy rates—85.25% and 93.2%, respectively—making this system viable for Human Computer Interfaces. R.R. Nair & B. Vijayan [7] The paper's aim is to produce a gender-identifying machine that is able to learn the given real-world input of humans at the time it speaks, and thereby make it possible for this recognized voice to be partially informed about personal identification information as compared it with others – even though we all know that so much has not yet been retrieved from those who made calls last time round we had an achievement at hand which was quite something! H. Harb & L. Chen [8] this paper examines acoustic and pitch elements along with several classification schemes to distinguish males from females in multimedia indexing. It is evident that the combination of the features and the classifiers is more effective than one classifier. A neural-network system with acoustic and pitch features for instance, attains 90% classification accuracy when using 1 second segmentations; however, its realizable accuracy rate goes up to 93% due to some practical factors. M.A. Uddin et al. [9] Speech recognition is an evolving area that has contributed greatly towards human-machine interaction, and has equally identified gender as crucial to the means of enhancing security, robotics as well as AI tools. Some essential frequency, spectral entropy, and MFCC among others are commonly used for this type of automated gender classification. Among numerous ML classifiers, KNN and SVC constantly outperform in TIMIT, RAVDESS and user-created dataset. This research presents a dual layer approach where extensive feature selection and data cleaning is employed to obtain a KNN accuracy of 96.8% on the TIMIT data set to showcase its efficiency in gender recognition. A. Majkowski et al. [10] discusses gender identification based on Polish speech signals using supervised machine learning techniques. A speech database was developed, and audio features were extracted using Python for signal processing and R for feature calculations. A neural network was trained using both CPU and GPU, with the GPU significantly accelerating the training process. The model achieved an accuracy of 92.4%, highlighting the effectiveness of the approach and the computational advantages of GPU utilization. M. Buyukyilmaz & A.O. Cibikdiken [11] suggests an MLP model for the automatic gender classification from voice that falls in deep learning models category. It uses a data set of 3168 voice samples male and females and performs acoustic analysis for feature extraction. The model achieved 96.74 percent test accuracy demonstrating impressive gender dimension identification capability of the model. Also, interactive web application was built to showcase real-world application of the model.

L. Rabiner et al. [12] considers seven pitch detection algorithms and presents a comparison of their performances for the same speech database, which contains eight utterances spoken by three males, three females, and one child. Recordings were made with telephone, close-talking microphones, and wideband set-ups. A "standard" pitch contour was generated semi-automatically and compared against the pitch contours generated by the seven algorithms, which are AUTOC, CEP, SIFT, PPROC, DARD, LPC, and AMDF. Errors were quantified in terms of pitch period deviations, gross pitch errors, and voiced-unvoiced classification errors. Ranking was done for each algorithm based on the performance across recording conditions and speaker pitch ranges to get an idea about their relative effectiveness. M.A. Yusnita et al. [13] Automatic Gender Recognition system takes inspiration from the human cognitive capability of classifying the gender of a speaker as male or female, and it uses algorithms trained for such classification. This study analyzed the speech data of 93 speakers by feature extraction using Linear Prediction Coefficients. The preparation of data using normalization, pre-emphasis, frame blocking, and windowing was used. The LPC coefficient orders were varied while optimizing the AGR system. The training was done by using a feed-forward neural network with an MLP. The proposed system has obtained an average recognition accuracy of 93.3%, showing that it is quite consistent in the detection of male and female genders. S. Tolmeijer et al. [14] Voice gender is usually defined by pitch and is a very common design feature in VAs; it often reinforces negative stereotypes. Despite its prevalence, only limited research has considered users' perceptions of voice gender in VAs. The current study examines gender stereotyping and trust formation in a voice assistant in an online experiment with 234 participants, while manipulating pitch and gender for comparison, it also included a gender-ambiguous voice. Results showed implicit stereotyping in VAs but found no significant difference in trust toward gender-ambiguous versus gendered voices, thus showing their possible usage for more general commercial applications.

3. Methodology

Recognizing gender through voice was our primary focus in this project. Firstly, we gathered an eclectic mix of sounds containing equal representation of men as well as women voices. Our material came from different internet sources and personal records aiming at consistent ages, accents, and speech. Afterward, each audio clip got shortened using personal applications so that all lasted for 15 seconds only making sure of consistency at all times.

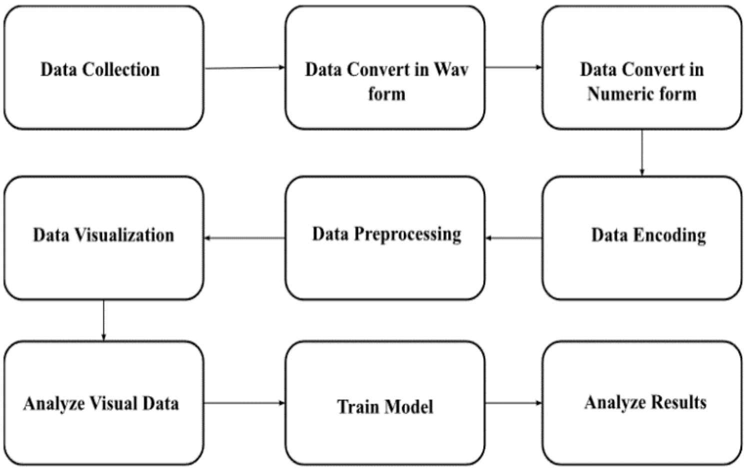


Figure 1. Methodology for gender Identification from Vocal Cues: Feature engineering and Machine Learning

3.1. Data collection

The dataset for this project was collected from Bangladeshi vlogs available on YouTube and Facebook. The selection of these platforms was based on their widespread usage and the availability of diverse vocal samples from various demographics in Bangladesh.

3.2. Data trimming

To ensure that it is appropriate for training with a gender recognition model, we have to trim the audio data. This means looking for clear speech while disregarding those that aren't as such – e.g., noises from other sources or irrelevant talks during silence periods. Clear voice including silent gaps has been kept while adding some kind of effect on them depending on what happens around it thereby maintaining their actual tone; nothing else should come out except what had been said originally where it is situated within these several seconds before proceeding abruptly into the next ventilating matter at once! In order to make things more understandable using this method has guaranteed success because it eliminates ambiguity caused by such factors as changes in volume levels among others which may mask otherwise obvious attributes such as pitch or tempo variations by reducing other types of disturbances.

3.2.1. Data trimming tools

There some tools are mentioned below:

Librosa:

✓**Functionality**: It is used for the analysis and processing of audio signals.

✓**Features**: Silence detection functions, feature extraction etc. – all this is possible with Librosa. There is also a possibility to remove silence in audio and analyze some audio characteristics necessary for any machine learning data preparation tasks.

Pydub:

✓**Functionality**: This is a simple audio manipulation library that is easy to use.

✓**Features**: Ability to cut, concatenate and apply effects on audio files.

3.3. Data conversion in wav

The gathered audios were then prepared and checked for compatibility with machine learning model. Hence, the audio files collected were converted into WAV format. WAV is an uncompressed audio format that is widely used allowing for a higher quality of sound preserving which is suitable for different tasks which involve processing audios.

3.3.1. Data Conversion Tools

There some tools are mentioned below:

✓**FFmpeg**: A comprehensive multimedia framework used for converting audio and video files.

✓**Pydub**: A Python library used for manipulating audio files.

3.4. Data conversion in numeric

To convert WAV audio files into numerical features which can be further used in machine learning models and data analysis processes, we must extract relevant audio characteristics from it, store them in a CSV file where we work with some important features of audio data. The important features:

✓**Zero Crossing Rate (ZCR):** Reflects the frequency of signal modifications from high-quality to bad or vice versa, providing insights into the presence of abrupt modifications or periodicity in audio signals.

✓**Energy (Root Mean Square Energy - RMSE):** Computes the average power of the signal over time, critical for obligations like speech popularity and sound category via discerning between quiet and loud segments.

✓**Spectral Centroid:** Indicates the "middle of mass" of the strength spectrum, presenting a measure of the dominant frequency content inside the sign and supporting in identifying the tumbrel traits of audio.

✓**Spectral Bandwidth:** Defines the range of frequencies occupied through the sign, presenting statistics approximately its spectral spread, useful for distinguishing between sounds with different spectral shapes.

✓**Spectral Contrast:** Measures the distinction in value between peaks and valleys inside the spectrum, assisting within the discrimination of tonal components from background noise or non-harmonic sounds.

✓**Spectral Roll-off:** Determines the frequency under which a positive percent of the full spectral power is living, assisting in identifying the dominant spectral traits and the presence of high-frequency components.

✓**Mel-Frequency Cepstral Coefficients (MFCCs):** Extracts the spectral capabilities of audio signals, especially effective in speech and tune evaluation through shooting traits like timbre, pitch, and depth versions.

✓**Chroma Feature:** Represents the pitch content of track, supplying a concise representation of musical concord and facilitating obligations which include chord popularity, melody extraction, and style class.

✓**Tonnetz (tonal centroid functions):** Represents the tonal content of music by using modeling harmonic relationships among musical notes in a geometric space, aiding in harmonic evaluation and chord transcription.

✓**Harmonics-to-Noise Ratio (HNR):** Measures the ratio of harmonic components to noise in a signal, supplying insights into the readability and richness of voiced sounds, crucial in voice analysis and synthesis obligations.

3.4.1. Data conversion tools

There some tools are mentioned below:

✓**Librosa:** A library for audio and music analysis.

✓**Pandas:** A data manipulation and analysis library.

3.5. Data visualization

Visualization of data proves crucial when try to comprehend and interpret audio file extracted characteristics, whereby it entails coming up with graphical representations for them to pinpoint patterns, trends, and insights necessary for the development of gender recognition model.

3.5.1. Data visualization tools

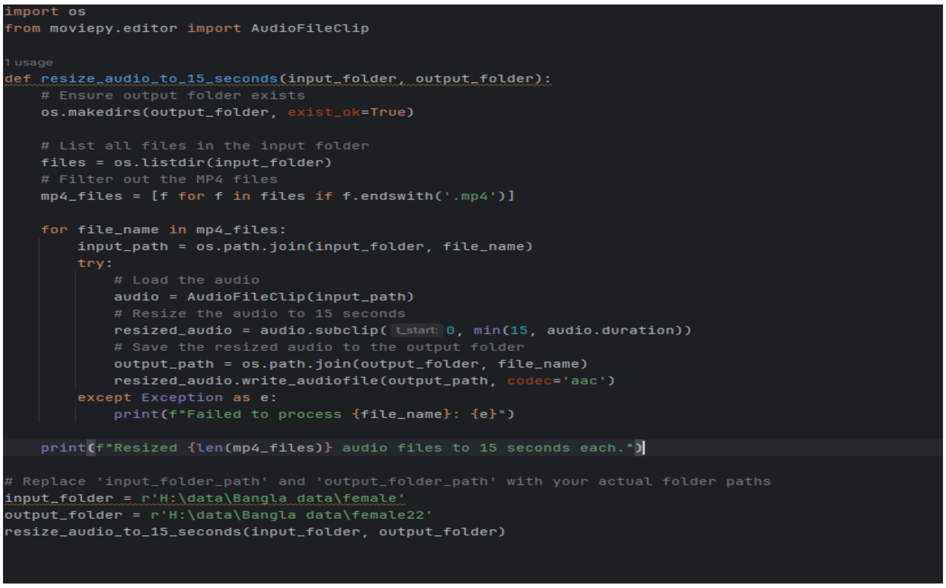
✓**Matplotlib:** A plotting library for creating static, animated, and interactive visualizations.

✓**Seaborn:** A library based on Matplotlib that provides a high-level interface for drawing attractive statistical graphics.

✓**Scikit-Learn:** A machine learning library that includes tools for dimensionality reduction.

4. Experimental Work

The experimenter collected an equal audio data set from females and males, converted it into WAV format, extracted Mel-frequency cepstral coefficients features, normalized it by scikit-learn, and visualized the data for analysis.



```
import os
from moviepy.editor import AudioFileClip

1 usage
def resize_audio_to_15_seconds(input_folder, output_folder):
    # Ensure output folder exists
    os.makedirs(output_folder, exist_ok=True)

    # List all files in the input folder
    files = os.listdir(input_folder)
    # Filter out the MP4 files
    mp4_files = [f for f in files if f.endswith('.mp4')]

    for file_name in mp4_files:
        input_path = os.path.join(input_folder, file_name)
        try:
            # Load the audio
            audio = AudioFileClip(input_path)
            # Resize the audio to 15 seconds
            resized_audio = audio.subclip( t_start: 0, min(15, audio.duration))
            # Save the resized audio to the output folder
            output_path = os.path.join(output_folder, file_name)
            resized_audio.write_audiofile(output_path, codec='aac')
        except Exception as e:
            print(f"Failed to process {file_name}: {e}")

    print(f"Resized {len(mp4_files)} audio files to 15 seconds each.")

# Replace 'input_folder_path' and 'output_folder_path' with your actual folder paths
input_folder = r'H:\data\Bangla data\female'
output_folder = r'H:\data\Bangla data\female22'
resize_audio_to_15_seconds(input_folder, output_folder)
```

Figure 2. Experimental work for Data Preparation

4.1. Data preparation

We collated an array of audio recordings by both male and female speakers at the first instance. To yield a representational data set, we got them from different online platforms and personal archives. Afterwards, they were reduced to 15 seconds in order to standardize the input data that would be used in next steps of processing. In this case, making use of tools such as ffmpeg and pydub would guarantee all these files are in WAV format for the purpose of consistency.

4.2. Feature extraction

To the end of obtaining features, the data was converted from audio form to numeric forms that could be used by machines. One of the methods that was utilized is known as Mel-Frequency Campestral Coefficients (MFCC). This method helps in taking important aspects of the sound signals. They are important in audio classification as they can easily represent phonetic characteristics in speech signals. We extracted them using librosa library on audio clips which were useful based on their characteristics of audio data. And we describe the process of audio data conversation from MP4 to WAV. And also describe the process of audio data conversation from audio to numeric form and save the audio data in Comma-Separated Values (CSV) file.

4.3. Data preprocessing

After extracting them, it was necessary to perform some operations on pre-processing steps they were taken before training. So that they fell within an appropriate range (between 0 and 1), but also so that they would be comparable to other data that we might have, if not general rescaling only but hard limits can stop us from having any further analysis done at all even if by mistake ∈ [0,1]. This scaling operation was simplified thanks to the scikit learn library.

4.4. Data visualization

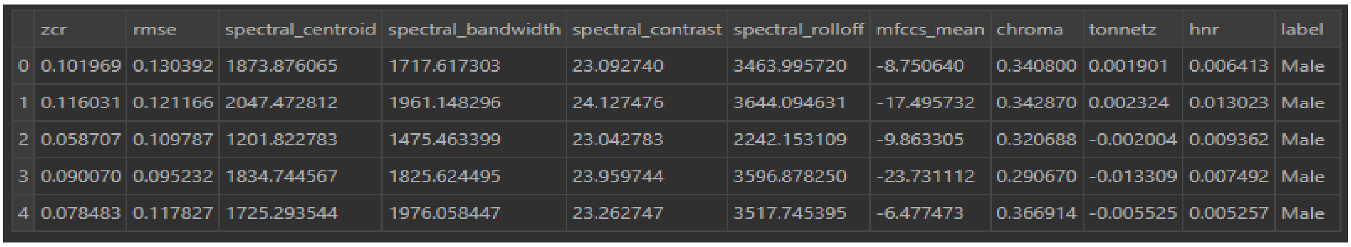
There are two methods of visualizing the data with a connection to AI, which include the transformation of spectrograms into waveform plots; creation of confusion matrices as well as performance graphs that will depict how our model performed in classifying genders. At the same time, the aim on this paper was to complement sounds with pictures so that gender identification through sound could become more perceptible.

5. Outcome and Discussion

We get the final outcome based on our prepared dataset. We prepare the dataset for Gender Identification from Vocal Cues.

5.1. Final outcome based on dataset discussion

We curated our dataset with care so that it contained an equal number of audio recordings from both men’s and women’s voices at different ages, with varying accents and speaking styles to ensure it was suitable for a wide range of purposes. By using Mel-Frequency Cepstral Coefficients (MFCCs) for feature extraction, as well as implementing normalization techniques, we made significant improvements to the quality of our dataset for use in machine learning.



	zcr	rmse	spectral_centroid	spectral_bandwidth	spectral_contrast	spectral_rolloff	mfccs_mean	chroma	tonnetz	hnr	label
0	0.101969	0.130392	1873.876065	1717.617303	23.092740	3463.995720	-8.750640	0.340800	0.001901	0.006413	Male
1	0.116031	0.121166	2047.472812	1961.148296	24.127476	3644.094631	-17.495732	0.342870	0.002324	0.013023	Male
2	0.058707	0.109787	1201.822783	1475.463399	23.042783	2242.153109	-9.863305	0.320688	-0.002004	0.009362	Male
3	0.090070	0.095232	1834.744567	1825.624495	23.959744	3596.878250	-23.731112	0.290670	-0.013309	0.007492	Male
4	0.078483	0.117827	1725.293544	1976.058447	23.262747	3517.745395	-6.477473	0.366914	-0.005525	0.005257	Male

Figure 3. Flatten representation of Datasets

5.2. Columns discretion

The basic statistical measures were used to describe the dataset, the distribution of the dataset was also described. The count refers to the number of records for the specific variable so that the number of data

Points used for the analysis is sufficient. The mean gives the central value as a result giving information on the average yield. A standard deviation (Std) depicts dispersion of data with reference to the mean it results in pointing out variability. The Min and Max are used to depict the extent of range and how spread the data is. The percentiles of 25% (Q1), 50% (Q2), and 75% (Q3) gives a better view on the data distribution where Q1 stands for lower quartile, Q2 is median, and Q3 is the upper quartile. These statistics give a picture of the whole set of data as a whole, which is very important for the further analysis of the data and formulation of models.

Table 1. Discretion table of Datasets

Type	ZCR	RMSC	Spectral Centroid	Spectral Bandwidth	Spectral Contrast	Spectral Rolloff	MFCCS Mean	Chroma Feature	Tonnetz	HNR
Count	512.0	512.00	512.00	512.00	512.00	512.00	512.00	512.00	512.00	512.00
Mean	0.09	0.11	1879.33	1909.55	23.75	3554.76	-12.16	0.32	0.00	0.01
Std	0.03	0.03	433.69	302.65	1.46	860.98	6.84	0.05	0.00	0.03
Min	0.02	0.01	680.26	1055.90	18.05	1183.52	-31.21	0.19	0.04	0.00
25%	0.07	0.09	1605.73	1727.04	22.80	2998.39	-16.29	0.28	0.00	0.00
50%	0.09	0.11	1821.37	1896.43	23.62	3465.54	-11.60	0.31	0.00	0.00
75%	0.10	0.13	2084.40	2090.58	24.53	3961.19	-7.76	0.35	0.00	0.01

5.2.1. Columns analysis

We used Matplotlib together with Seaborn libraries to understand data characteristics. Particularly, distribution, box as well as scatter plots assisted us in observing frequency parts on a dataset as well as examining how different the features were and what correlations existed between them. On this basis we could identify features necessary for further processing and building gender recognition model within our final year project.

5.2.2. Waveform of data

This waveform illustrates the characteristics of both male and female voice patterns.

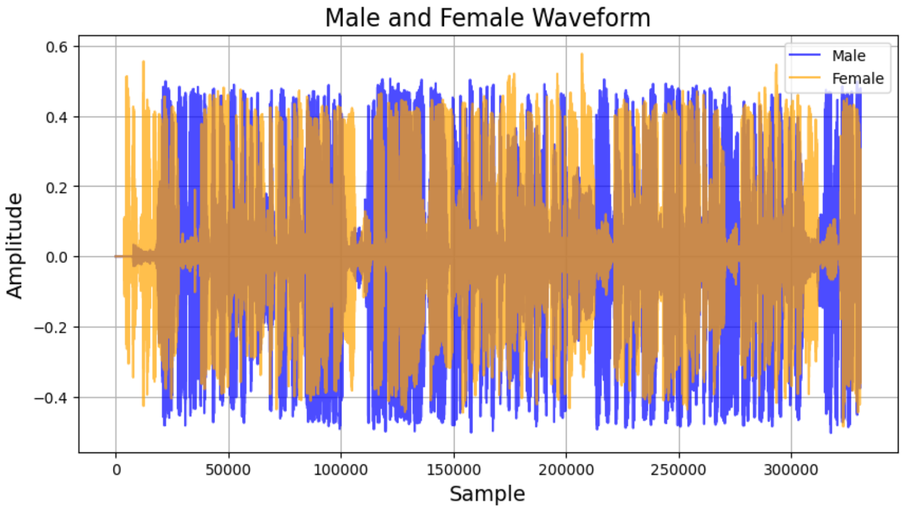
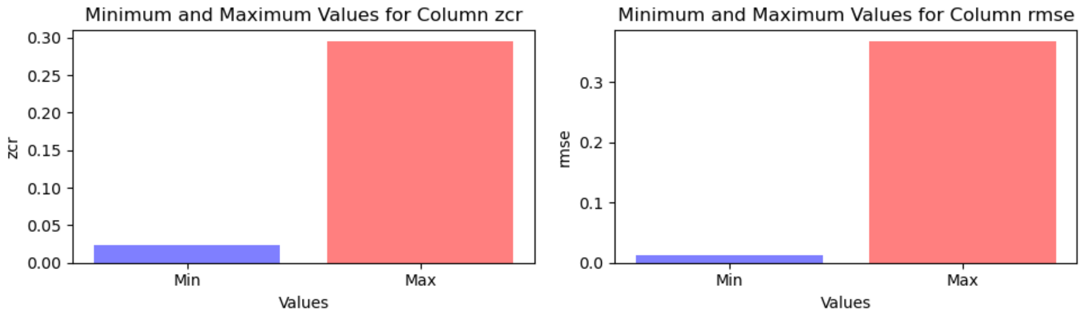


Figure 4. Waveform of Male and Female audio Data

5.2.3. Bar plot of columns data

We have created bar plots to visualize the maximum and minimum values of different features in our dataset. These plots show us what range of values each feature takes on, while also highlighting the gender differences in voice pitch.

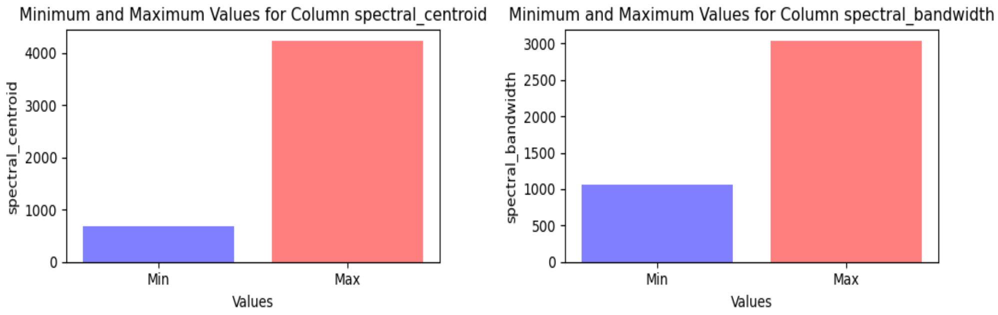


(a) (b)

Figure 5. Bar plot of Maximum and Minimum data of ZCR & RMSE

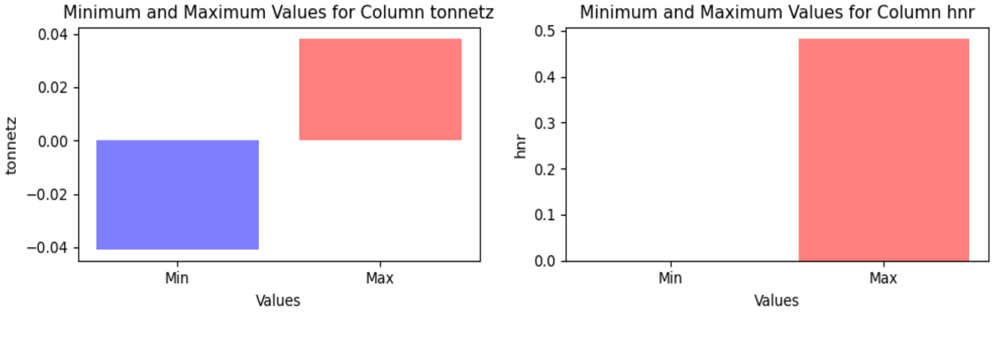
5.2.4. Scatter Plot of columns Data

Scatter plots were used to visualize the relationships between pairs of features in our dataset. When different combinations of MFCC features are plotted against one another, different clusters for male vs. female voices are observed.



(a) (b)

Figure 6. Bar plot of Maximum and Minimum data of Spectral centroid & Spectral bandwidth



(a) (b)

Figure 7. Bar plot of Maximum and Minimum data of Tonnetz & HNR

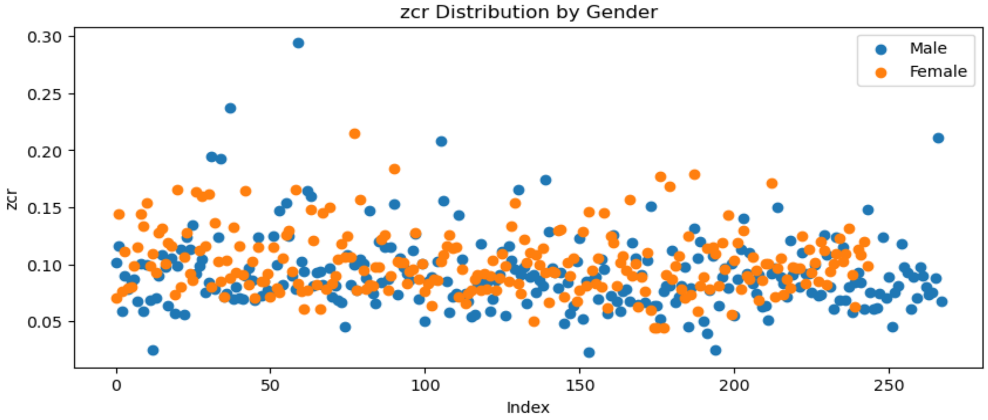


Figure 8. Scatter plot of ZCR for target columns

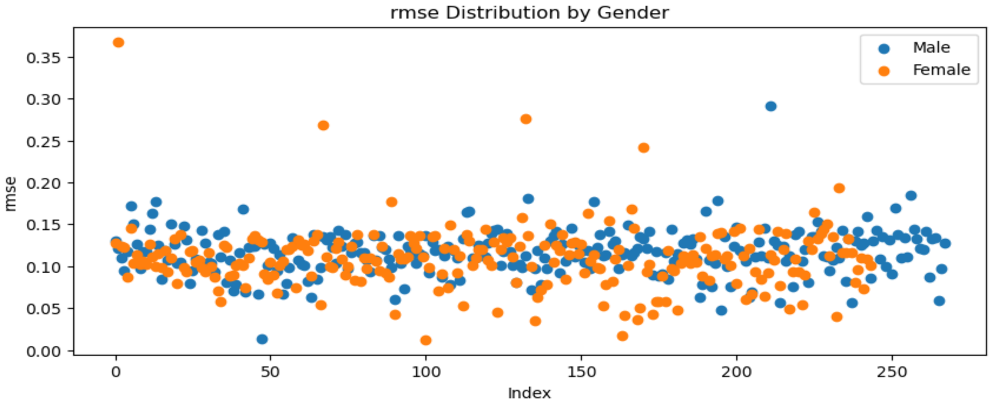


Figure 9. Scatter plot of EMSE

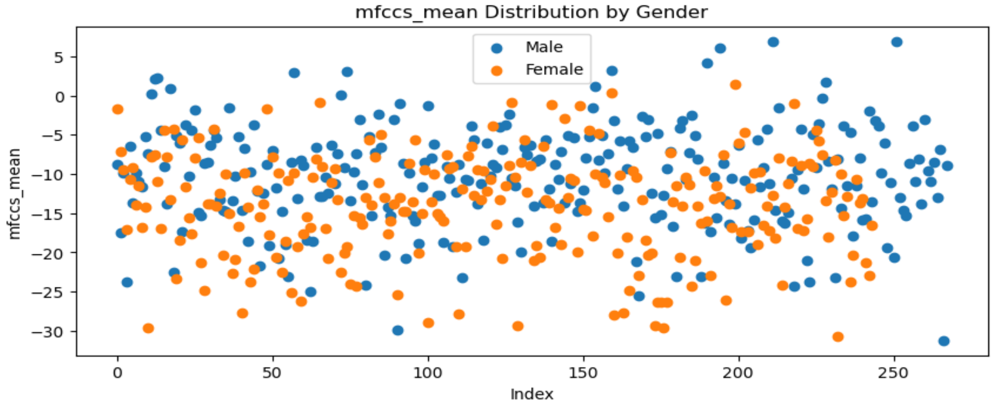


Figure 10. Scatter plot of MFCCS mean

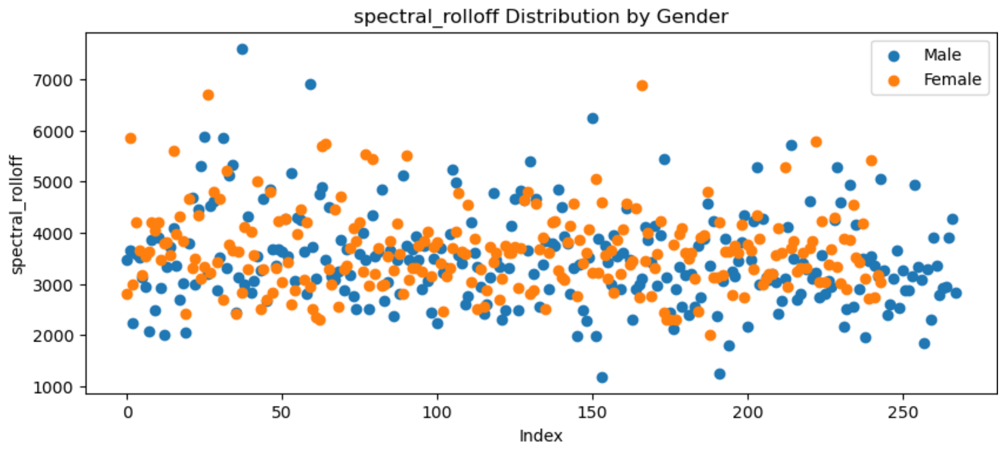


Figure 11. Scatter plot of Spectral Rolloff

5.2.5. Network diagram of correlated features

In order to visualize the relations between various attributes within our dataset, we developed a network diagram that indicated those connections. In this diagram, nodes represented attributes while edges showed how much one attribute correlates with another.

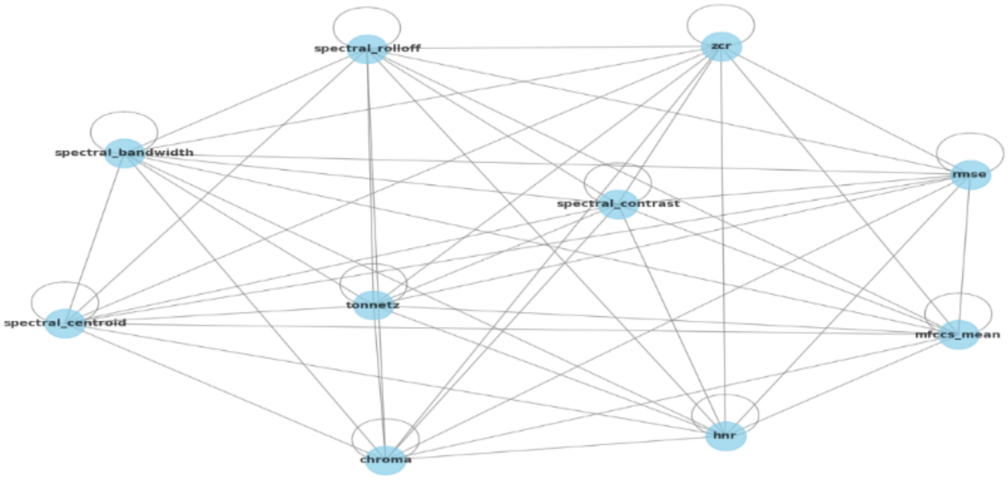


Figure 12. Network Diagram of all Columns

5.3. Results

The results of the study point out the effectiveness of the machine learning approaches in the gender class of the Bengali audio speech signals. The Random Forest model was able to record a training accuracy of 100% and a testing accuracy of 86%, which indicates its role in understanding and capturing even the most complex patterns in the data set. However, the difference in training and testing accuracy shows signs of overfitting in some cases, where the model has learnt too much about particular training samples and is not able to learn about new ones. Alternatively, the Logistic Regression model has produced results that show consistency around the training accuracy of 75% and a testing accuracy of 76% portraying some measure of consistent generalization. These findings relate the need for choosing a suitable model for tasks with numerous and complex vocal features in consideration of accuracy and generalization trade-offs. Further model optimization and more feature selection could improve the robustness of these models making them suitable for real world application.

5.3.1. Results validations

Table 2. Classification reports

Models	Label	Precision	Recall	f1-score	Support
Random Forest Classifier	Male	0.87	0.87	0.87	1284
	Female	0.87	0.87	0.87	1270
	Accuracy			0.87	2554
	Macro avg	0.87	0.87	0.87	2554
	Weighted avg	0.87	0.87	0.87	2554
Logistic Regression	Male	0.75	0.77	0.76	1284
	Female	0.76	0.76	0.75	1270
	Accuracy			0.76	2554
	Macro avg	0.76	0.74	0.76	2554
	Weighted avg	0.76	0.74	0.76	2554

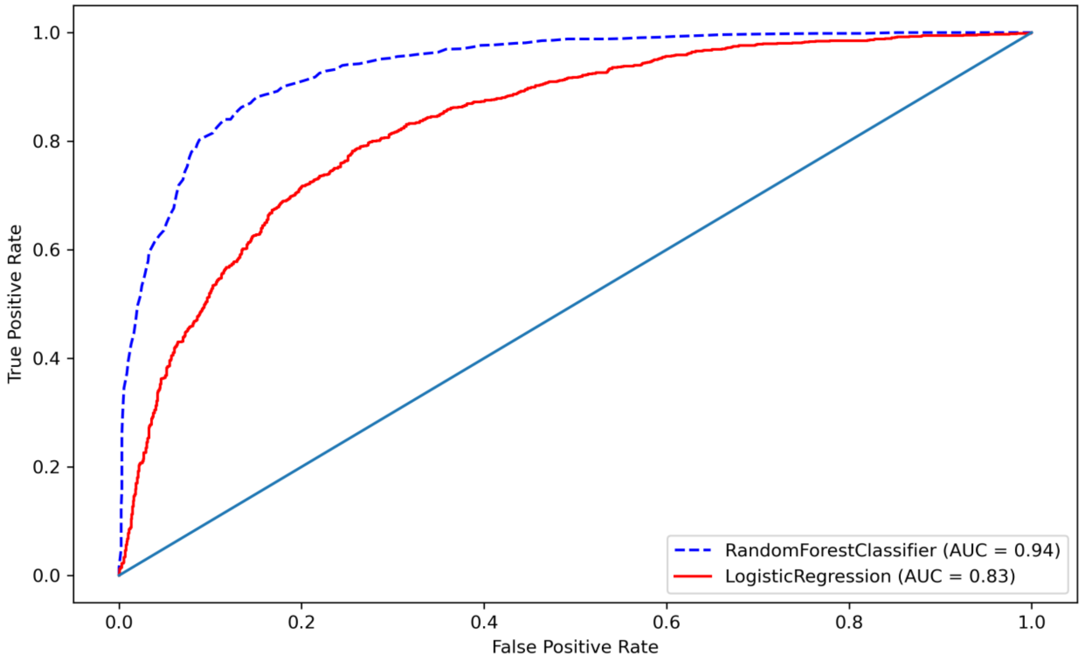


Figure 13. RoC Curve for models

The performance of Random Forest and Logistic Regression models with respect to the task of gender classification is presented in Table 2. It is found out that the Random Forest model is more efficient than the Logistic Regression model with the Random Forest model achieving an accuracy rate of 87% and a consistent precision, recall and F1-scores of 0.87 on the male and female classes. Logistic Regression model on the other hand records an accuracy rate of 76% with relatively lower precision and F1 scores showing that Random Forest model is more effective for this task. The ROC curves of the Random Forest and the Logistic Regression models used for gender classification are shown in Figure 14. The ROC curve of the random forest model has an AUC score of 0.94, showing that this model is highly discriminatory and performs well in terms of sensitivity and specificity. The AUC for the logistic regression model is much lower at 0.83, which suggests it moderate capability when it comes to class separation. In terms of the ROC, random forest performs better than logistic regression which allows the random forest to model the more complicated aspects of the Bengali acoustic signals and hence is preferred for this type of classification task.

The Random Forest and Logistic Regression models' confusion matrices pertaining to gender classification are displayed in Figure 15. With respect to the Random Forest model, males and females are classified correctly about 1112 and 1107 times respectively, with false negatives standing at 163 and 171 for males and females respectively. On the other hand, the performance of the Logistic regression is comparatively lower, where 988 males and 942 females were classified correctly, while false negatives surged to 328 males and 296 females. These outcomes also illustrate quite clearly the dominant of Random Forest model as it can be noted that it outperformed the rest with regards to the classification errors. The better results with this technique demonstrate more accurate identification of the target population's gender proneness based on the Bengali speech samples, as the sensitivity, and specificity ratios are more favorable.

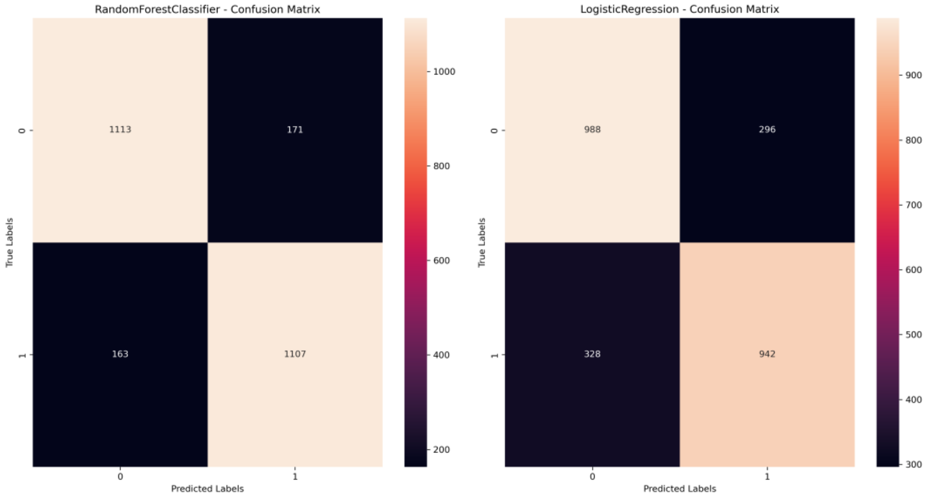


Figure 14. Confusion Matrix for models

5.4. Limitations

In this field, a major limitation is the lack of previous studies and the lack of open sources. To open some sources, it requires money. The literature review is an important part of any research because it helps to identify the scope of work that has been done so far in the research area. Very few datasets are available for this research. For a shortage of time, we can't prepare a large size of dataset. In our research, it would require a high-configuration computer, which we didn't have.

6. Conclusion

This research will create an ASLM aiming to identify the gender by analyzing several Bangladeshi voice samples. This study gets applied in fields that are both business related and pure research such as in customer service, safety systems, transcription services and sociolinguistics. In conclusion, voice evaluation offers demographic data in the study and enhances the appearance of the consumers' experiences. Smoking act, increase in acid reflux, air pollutants and energy impacts on Bangladeshi fitness are the aspects that influence voice. Voices given by the participants, male and female, were recorded in this study. Drawing from Hilbert, the study adopted a consistent and convenient method of research, by converting each recording to WAV, extracting the most significant voices and preprocessing the statistical data for in-depth analysis. Data analysis was conducted through report visualization to help grasp the characteristics and tendencies peculiar to the voice records. Using modern information processing approaches, it is aimed at training a machine learning model for determining gender based on voice notes with the development of a gender identification set of rules based on the findings of the current study and big data. The overall size of our dataset may be a focus of future research. A dataset's processing normally takes a long period. We are going to create a High Accuracy Machine Learning Model in the future. A supervised machine learning model is what we would attempt. Support Vector Machines, Naive Bayes, Decision Trees, Random Forests, and KNN are a few examples of supervised machine learning models. Subsequently, we discover that the model is trained using both our own dataset and a combination of our dataset and an existing standard dataset. Our goal is to create software.

Declarations

Source of Funding

This study did not receive any grant from funding agencies in the public, commercial, or not–for–profit sectors.

Competing Interests Statement

The authors declare no competing financial, professional, or personal interests.

Consent for publication

The authors declare that they consented to the publication of this study.

Authors' contributions

All the authors made an equal contribution in the Conception and design of the work, Data collection, Simulation analysis, Drafting the article, and Critical revision of the article. All the authors have read and approved the final copy of the manuscript.

Availability of data and material

Authors are willing to share data and material according to the relevant needs.

References

[1] Reddy, V.S.K., & Surendran, R. (2023). Human Voice Recognition System to Predict the Gender using Random Forest Algorithm. In IEEE 2nd International Conference on Edge Computing and Applications (ICEC AA), Pages 946–951. https://doi.org/10.1109/icecaa58104.2023.10212186.

[2] Jadav, S. (2018). Voice-based gender identification using machine learning. In IEEE 4th International Conference on Computing Communication and Automation (ICCCA), Pages 1–4. https://doi.org/10.1109/ccaa.2018.8777582.

[3] Kone, V.S., Anagal, A., Anegundi, S., Jadhav, P., Kulkarni, U., & Meena, S.M. (2023). Voice-based gender and age recognition system. In IEEE International conference on advancement in Computation & Computer Technologies (InCACCT), Pages 74–80. https://doi.org/10.1109/incacct57535.2023.10141801.

[4] Markitantov, M., & Verkholyak, O. (2019). Automatic recognition of speaker age and gender based on deep neural networks. In Speech and Computer: 21st International Conference, SPECOM, Istanbul, Turkey, Proceedings, Pages 327–336, Springer International Publishing. https://doi.org/10.1007/978-3-030-26061-3_34.

[5] Chachadi, K., & Nirmala, S.R. (2022). Voice-based gender recognition using neural network. In Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces, Pages 741–749, Springer Singapore. https://doi.org/10.1007/978-981-16-0739-4_70.

[6] Li, W., Kim, D.J., Kim, C.H., & Hong, K.S. (2010). Voice-based recognition system for non-semantics information by language and gender. In IEEE 3rd International Symposium on Electronic Commerce and Security, Pages 84–88. https://doi.org/10.1109/isecs.2010.27.

[7] Nair, R.R., & Vijayan, B. (2019). Voice based gender recognition. International Research Journal of Engineering and Technology, 6(5): 2109–2112.

[8] Harb, H., & Chen, L. (2005). Voice-based gender identification in multimedia applications. Journal of Intelligent Information Systems, 24: 179–198. https://doi.org/10.1007/s10844-005-0322-8.

[9] Uddin, M.A., Hossain, M.S., Pathan, R.K., & Biswas, M. (2020). Gender recognition from human voice using multi-layer architecture. In IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA), Pages 1–7. https://doi.org/10.1109/inista49547.2020.9194654.

[10] Majkowski, A., Kołodziej, M., Pyszczak, J., Tarnowski, P., & Rak, R.J. (2019). Identification of gender based on speech signal. In IEEE 20th International Conference on Computational Problems of Electrical Engineering (CPEE), Pages 1–4. https://doi.org/10.1109/cpee47179.2019.8949078.

[11] Buyukyilmaz, M., & Cibikdiken, A.O. (2016). Voice gender recognition using deep learning. International Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA), Pages 409–411. https://doi.org/10.2991/msota-16.2016.90.

[12] Rabiner, L., Cheng, M., Rosenberg, A., & McGonegal, C. (1976). A comparative performance study of several pitch detection algorithms. IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(5): 399–418. https://doi.org/10.1109/tassp.1976.1162846.

[13] Yusnita, M.A., Hafiz, A.M., Fadzilah, M.N., Zulhanip, A.Z., & Idris, M. (2017). Automatic gender recognition using linear prediction coefficients and artificial neural network on speech signal. In IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Pages 372–377. https://doi.org/10.1109/iccsce.2017.8284437.

[14] Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J.S., Leimeister, J.M.M., & Bernstein, A. (2021). Female by default?–exploring the effect of voice assistant gender and pitch on trait and trust attribution. Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Pages 1–7. https://dl.acm.org/doi/abs/10.1145/3411763.3451623.