

**AUDIO SPEECH RECOGNITION IN BENGALI LANGUAGE MALE FEMALE
AND THIRD GENDER BASED ON SUPERVISED LEARNING**

BY

**SHOEB SHEIKH
ID: 191-15-12610**

**TARANGA SAHA SEJUTI
ID: 191-15-12412**

AND

**ISNAT JAHAN ERA
ID: 191-15-12638**

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms. Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Ms. Zerín Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

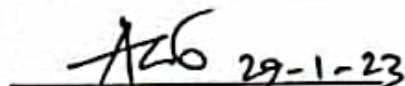
This Project titled **A machine-learning approach based on voice-activated gender classification for Bengali language**, submitted by SHOEB SHEIKH and TARANGA SAHA SEJUTI and ISNAT JAHAN ERA to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 29th January 2023

BOARD OF EXAMINERS



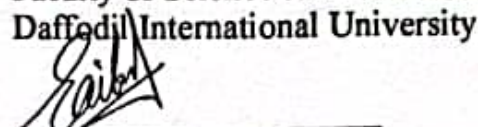
Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



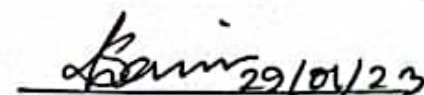
Arif Mahmud
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Saiful Islam
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



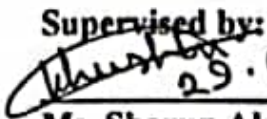
Dr. Shamim H Ripon
Professor
Department of Computer Science and Engineering
East West University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of Ms. Sharun Akter Khushbu, Lecturer, Department of CSE Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:


29.01.23

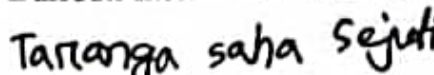
Ms. Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

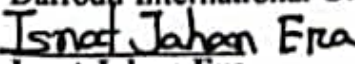
Co-Supervised by:


Ms. Zerine Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Submitted by:


Shoeb Sheikh
ID: -191-15-12610
Department of CSE
Daffodil International University


Taranga Saha Sejuti
ID: -191-15-12412
Department of CSE
Daffodil International University


Isnat Jahan Era
ID: -191-15-12638
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

Foremost, we would like to express our heartiest thanks and gratefulness to almighty God for His spiritual blessing makes it possible for us to complete the final year Research based Project successfully.

We are really grateful and wish our profound indebtedness to Ms. Sharun Akter Khushbu, Lecturer, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine Learning*” to carry out this project. Her endless patience, scholarly guidance, continual encouragement , constant and energetic supervision, constructive criticism , valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to Professor Dr. Touhid Bhuiyan, Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

–Is human voice frequency an integral part of vocal production, in which the vocal cords are the main source of sound? Using our hearing ability we can easily identify the classification of male, female and third gender voices. But using a machine learning approach, it's possible to find out the difference in voice. Identifying a voice from a natural voice without any kind of noise is a very hard task. Basically, we use MFCCs to find features in voice signals. Calculating discrete Fourier Transforms, Mel-spaced filter bank energies, and log filter bank energies for voice signals is what makes this possible. Recent work has shown that, to some extent, it's possible to identify gender from natural voice. It's one of the most important aspects of voice recognition. The voice's gender is irrelevant to the voice-to-text conversion. Nevertheless, identification of gender cannot be eliminated for the sake of applications in everyday life. The process can be broken down into steps: The first step is to create features from your audio file in pre-works. Next, we will use this set of features to train the model in feature extraction. Finally, we test it with a CSV file containing some other features in order to see how well the model predicts their "true" value. It is part of **Artificial Intelligence**(A.I.). We used Gradient boosting, Random Forest, KNN, Decision Tree, Naive Bayes, XGBoost, SVC and Linear regression. We achieved 96.02% accuracy in a dataset containing 3,200 data points from 250 different speakers. 850 are male, 850 are female and there are 1500 of the third gender.

Keywords: MFCC, MFCCS, Feature Extraction

TABLE OF CONTENTS

| CONTENTS | PAGE |
|--------------------------------------|-------------|
| Board of examiners | i |
| Declaration | ii |
| Acknowledgements | iii |
| Abstract | iv |
| CHAPTER | |
| CHAPTER 1: Introduction | 1-6 |
| 1.1 Introduction | 1 |
| 1.2 Motivation | 2 |
| 1.3 Rationale of the Study | 2 |
| 1.4 Research Questions | 2 |
| 1.5 Expected Output | 3 |
| 1.6 Project Management and Finance | 3 |
| 1.7 Report Layout | 3-4 |
| CHAPTER 2: Background | |
| 2.1 Terminologies | 5 |
| 2.2 Related Works | 5 |
| 2.3 Comparative Analysis and Summary | 6 |
| 2.4 Scope of the Problem | 6 |

| | |
|---|-------|
| 2.5 Challenges | 6 |
| CHAPTER 3: Research Methodology | |
| 3.1 Research Subject and Instrumentation | 7 |
| 3.2 Data Collection Procedure | 7 |
| 3.3 Statistical Analysis | 8-9 |
| 3.4 Proposed Methodology | 10 |
| 3.4.1 Mel-frequency cepstral coefficients (MFCC) | 10-11 |
| 3.4.2 Preemphasis | 12 |
| 3.4.3 Windowing | 12 |
| 3.4.4 DFT (Discrete Fourier Transform) | 12 |
| 3.4.5 Mel-Filter Bank | 12 |
| 3.4.6 Applying Log | 12 |
| 3.4.7 IDFT (Inverse Discrete Fourier Transform) | 13 |
| 3.4.8 Dynamic Features | 13 |
| 3.4.9 Train Test Split | 14 |
| 3.4.10 Supervised learning | 14 |
| 3.5 Implementation Requirements | 15 |
| CHAPTER 4: Experimental Results and Discussion | |
| 4.1 Experimental Setup | 16 |
| 4.2 Experimental Results and Analysis | 16-17 |
| 4.3 Discussion | 18 |
| CHAPTER 5: Impact on Society, Environment and Sustainability | |

| | |
|---|-----------|
| 5.1 Impact on Society | 19 |
| 5.2 Impact on Environment | 19 |
| 5.3 Ethical Aspects | 19 |
| 5.4 Sustainability Plan | 19-20 |
| CHAPTER 6: Summary, Conclusion, Recommendation and Implication for Future Research | |
| 6.1 Summary of the Study | 21 |
| 6.2 Conclusions | 21 |
| 6.3 Implication for Further Study | 21 |
| REFERENCES | 22 |

LIST OF FIGURES

| FIGURES | PAGE NO |
|---|----------------|
| Figure 3.1.1: Workflow | 7 |
| Figure 3.2.1: dataset ratio (male data, female data and third gender data) | 8 |
| Figure 3.3.1: dataset ratio (male data, female data and third gender data) | 9 |
| Figure 3.3.2: dataset ratio for test, train (male, female and third gender) | 10 |
| Figure 3.4.1.1: Workflow of MFCC | 11 |
| Figure 3.4.1.2: Analog signal | 11 |
| Figure 3.4.1.3: Digital signal | 11 |
| Figure 3.4.7.1: Normalize Frequency | 13 |
| Figure 3.4.7.2: Sample Of voice | 13 |
| Figure 3.4.8.1: 20 MFCC features | 14 |
| Figure 3.4.10.1: ML Classifier | 15 |
| Figure 4.3.1: Confusion matrix | 19 |

LIST OF Table

| FIGURES | PAGE NO |
|---|----------------|
| Table 4.2.1: Model Evaluation | 16 |
| Table 4.2.2: Difference between Random Forest & Logistic Regression | 17 |
| | |

CHAPTER 1

Introduction

1.1 Introduction

Voice is one of the most widely-used methods of communication, so it's no surprise that voice recognition is a common topic in NLP. After all, human beings interact through speech much more than any other method. This verbal expression or speech is created by human's several body parts. Our brain can identify gender differences by hearing voices. Devices can not do this work without commands. In conducting social interaction and communication, there is a necessity of gender. So, many technologies are required for classifying gender, which play a vital role. The previous era of working on gender identification investigated the statistical properties of the speaker's slope features, using phantom features in conjunction with slope features for identifying gender.[1] Voice detection technologies are required to detect gender by reading word sequencing. Many famous voices like singers we can easily detect by using their personal information which are stored.[2] In this research paper, we show how gender is identified by voice using Bengali Language. There are many paths of gender identifying through voice. Some of them are:

In crime detection, people try to commit different types of crime in various ways like through phone calls, voice messages and so on. Most of the time, criminals try to hide their identity intentionally or knowingly. This could be very detrimental to the citizens. So, through this detection system, nation security can surveillance the criminals. Categorizing gender could be the solution to these crimes using voice detection systems. Different type of voice identify by different parameter.[3]

Helping blind, surroundings are enriched with many modern technologies. Voice recognition is one of them. Blind people can communicate with others through voice. Sometimes unscrupulous people want to betray them over devices. So investigators can catch them by matching their voice. In the mobile healthcare system, this voice detection can play a vital role. It can make the professional's work easier in the time of prescribing patients more accurately. It can detect the disease which is specifically found in a specific gender.

1.2 Motivation

Work and everything else has their own motivations to succeed. Voice recognition technology dissertation the long-standing issue of capturing complex and rapid sequences of behavior.[4] The researcher should clearly acknowledge the goals of the research. The basic hypothesis of detection is that blind and sighted subjects would have significantly different voice-recognition accuracy.[5]. The objectives of the research must be informed by the respondents, as it is a matter of Concern. When it is a matter of involvement in research development then it does not mean only researchers but also conscious respondents. The respondents can involve themselves in identifying, thinking, implementing and evaluating in critically various research programs. We choose to take this work as it our personal interest. This work can help people in many ways, and we can develop voice detection systems. We can do a lot of research on it in the future.

1.3 Rationale of the Study

Voice recognition is based on speech. It is a great research detection of speech or voice signal processing and a branch of pattern recognition. Voice recognition has gained prominence and is used with AI and intelligent assistants. Voice is becoming more common and powerful, it will have an increasing impact on our daily lives. Let us work together to create a world in which everyone's voice is clearly heard. [6]

1.4 Research Questions:

WHO is speaking?

The main theme of voice recognition is to determine specific speech. The goal of speech recognition is to understand and comprehend WHAT was said. It analyzes a person's tone, voice slope, and accent to identify them as male, female, and others. Voice identification appears to be similar across species and emerges at a young age, findings raise a number of new questions about this evolutionarily conserved process.[7]

1.5 Expected Output:

We use different types of algorithms to detect voices which are male, female and third gender, using those algorithms we expect our outcome to come 99.9%, but we got around 96.2%. It is one of the highest accuracy voice detections. So our detection system takes voice sample, matches them with a dataset. If the sample matched with the dataset and compared with the array, then it reveals the gender.

1.6 Project Management and Finance:

There are three types of voice classification, male, female and third gender. The Government of Bangladesh provides different types of financial aid to every layer of people. Male and female groups can easily collect their aids. In the term of third gender, they face many kinds of hassle because of their identity crisis. Sometimes their aids are snatched out by either man or woman who pretend to be a third gender. In this perspective, our recognition system could be the solution.

1.7 Report Layout:

- Chapter 1 of the study included an introduction, outlining the purpose and motivation, research questions asked, expected outcomes, project management & financial details.
- Chapter 2 of the book included several technical terms, as well as a review of existing literature, comparison of different approaches and a brief summary. It also covered the scope of the issue and highlighted issues that may arise.
- Chapter 3 of the paper enumerated the research subject, measuring instruments, collection of data procedure, statistical evaluation approach, recommended method and precise prerequisites for execution.
- Chapter 4 of the book included details regarding the experimental setup, results of the experiments and an analysis of them as well as a discussion surrounding it.

- Chapter 5 delved into the impact of this venture on societal, ecological and ethical aspects. Moreover, a plan for sustainability was also crafted.
- Chapter 6 was composed of summarizing the research, reaching conclusions, proposing further studies and providing relevant references. Appendices were also included.

Chapter 2

Background

2.1 Terminologies:

During the research process, we gathered raw data. In Voice Recognition, we gathered more than 850 male samples, 850 female samples and 1500 third gender samples. These information samples are gathered through interviews. These interviews were conducted over the phone questionnaire, face-to-face and gathering census data. The typed record of interviews, which are obtained from audio and video recordings. In the social situation chosen for the study, observation is the methodical noticing and recording of events, behaviors, and objects. Field notes, which are in-depth, impartial explanations of what is being observed, are the official name for the observational record. Give more attention to ethnography, which is direct observation, representation and analysis of the activities of sample members.

2.2 Related Works:

Francesco Asci et al. analyze their paper about voice samples collected through smartphones using Support Vector Machine algorithm, the highest accuracy 95.4% [8]. In Arabic voice detection the accuracy is very poor, around 40.[9] Support Vector Machine is the most accurate algorithm to detect voice among Linear Discriminant Analysis, K-Nearest Neighbor, Classification and Regression Trees, and Random Forest. With it, we can achieve higher precision in our results.[10] Semi-supervised self-labeled algorithm helps to detect gender and voice recognition.[11] Physiologically, unclasp and undertaking the vocal cords causes pattern changes in the voice and emotion.[12] Using KNN recognizer models simulated using MATLAB to identify gender and age from speech, this system is divided into two parts: one pre-processing and future interlineage second one is classification. [13] Utilizing a Multilayer Perceptron deep learning model, voice recognition is possible based on gender. This algorithm provides 96.74% accuracy when using around 4000 data points.[14]

2.3 Comparative Analysis and Summary:

Speech recognition and voice recognition are two different technologies that serve distinct purposes. The former is used to recognize spoken words, while the latter utilizes biometric technology to identify an individual based on their unique voice. The announcer recognition system could evaluate probability distributions, the preceding facet vectors could be evaluated by the announcer recognition system with a high-dimensional space, classifying each sharp vector turnout way of training is unsustainable, human voice has many extraordinary attributes, gender voice recognition is the foremost research zone in audial and speech processing.[15]

2.4 Scope of the Problem:

Inconsistency and inappropriate interpretations. It's not always possible for speech recognition technologies to precisely transcribe spoken words. If a speech recognition system (SRS) is to be useful, it must have a high level of accuracy. The difficulty of covering every language, accent, and dialect, the complexity of privacy and security, the feasibility of deployment and expense. Encousing prosopagnosia caused by ulceration limited to the fusiform cortex had normal voice recognition.[16]

2.5 Challenges:

Misinterpretation and improper interpretations, It will be interesting to observe how new innovations can retain the momentum of explosive development, as well as how current speech recognition issues will be resolved. It's not always possible for speech recognition technologies to accurately translate spoken words. Due to computers' inferior ability to comprehend how words and sentences relate to one another in context, misinterpretations of what the speaker intended to say or accomplish result. There are numerous obstacles with voice recognition today. The two main reasons for the current voice detection issues are rich and loud settings. This calls for even more sophisticated systems that can handle the trickiest ASR use-cases. Consider real-time interviews, speech recognition at a boisterous family supper, or large-scale meetings. These are the impending challenges for next-generation voice recognition.

Chapter 3

Research Methodology

3.1 Research Subject and Instrumentation:

Here, we follow the workflow outlined in figure 3.1.1

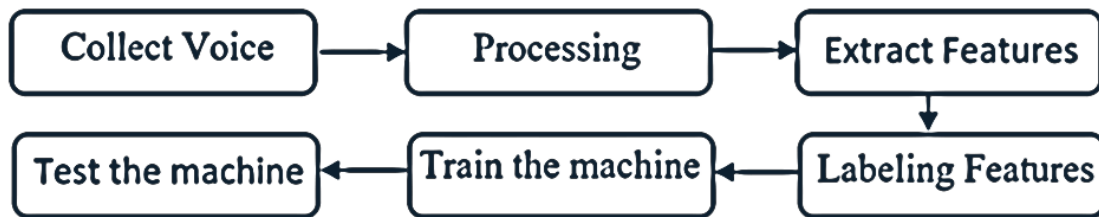


Figure 3.1.1: Workflow

Researchers have used machine learning to classify a diverse range of objects from voice, and it's also been applied to natural language processing. For example, when we work with natural voices from natural language processing, we need features of voice. It's important to understand what elements can distinguish male, female and third gender voices. For example, the intonation patterns and pitch range of each may affect how they pronounce words or phrases. We made a plan for how to reach the goal, given in figure 3.1.1.

3.2 Data Collection Procedure:

We have a dataset with three types of data: Male, Female and Third Gender. There were a total of 3200 data of them. Furthermore, we collected data for both male and female participants and extended them. Usually, the speakers are aged 20-50. We have collected data across Bangladesh by making use of various sources like call recordings, videos and voice records. The dataset duration was 3–7 seconds. There were 1700 data (for male 850 and for female 850 also). For third gender voice dataset, there were 1500 audio data of third gender. We got them from recordings of people speaking, videos on YouTube, conversation, news portal and interview. We took the video in video editing software Camtasia 2019 and removed all unnecessary audio, noise and silence that allows us to

make clean voice videos. Then we used Filmora Software to remove noise and make it audio. After that, we used WavePad 17.13 to make same duration small size audios, and it made an organized dataset. The duration of the datasets of the third gender was 5 seconds each. In figure 3.2.1, we compare the ratio of male, female and third gender voices over time.

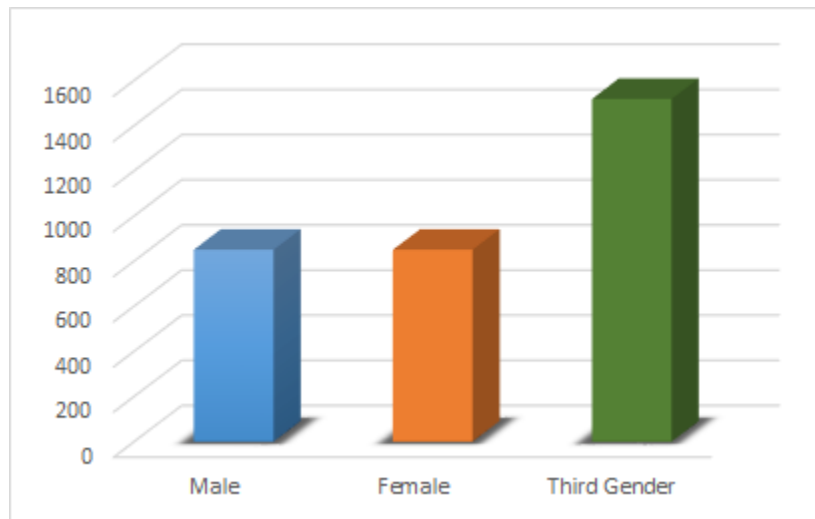


Figure 3.2.1: dataset ratio (male data, female data and third gender data)

3.3 Statistical Analysis:

In the present day, gender diversity is becoming increasingly important in our society. To gain insight into how gender impacts various aspects of life, it is important to explore the differences between male and female experiences, we conducted a survey that collected data from 3200 participants across three genders: male, female and third gender. We analyzed the data to identify patterns and trends among the three genders and gain insights into how each group can improve their individual circumstances. Our findings provide valuable information for businesses and organizations to ensure they are meeting their goals of promoting diversity and providing equal opportunities for everyone. With the advent of technology, data collection in Bangladesh has become easier and more reliable than ever before. By making use of various sources such as call recordings, videos and voice records, we have been able to collect a vast amount of data across Bangladesh. This data is being used to gain insights into the culture, economy and

population of Bangladesh. With this data, we are able to identify trends and develop strategies that can be implemented to improve the country's overall development. The dataset duration was 3–7 seconds. There were 1700 data (for male 850 and for female 850 also). For third gender voice dataset, there were 1500 audio data of third gender. We got them from recordings of people speaking, videos on YouTube, conversation, news portal and interview. Features extraction visualization are shown in figure 3.3.1.

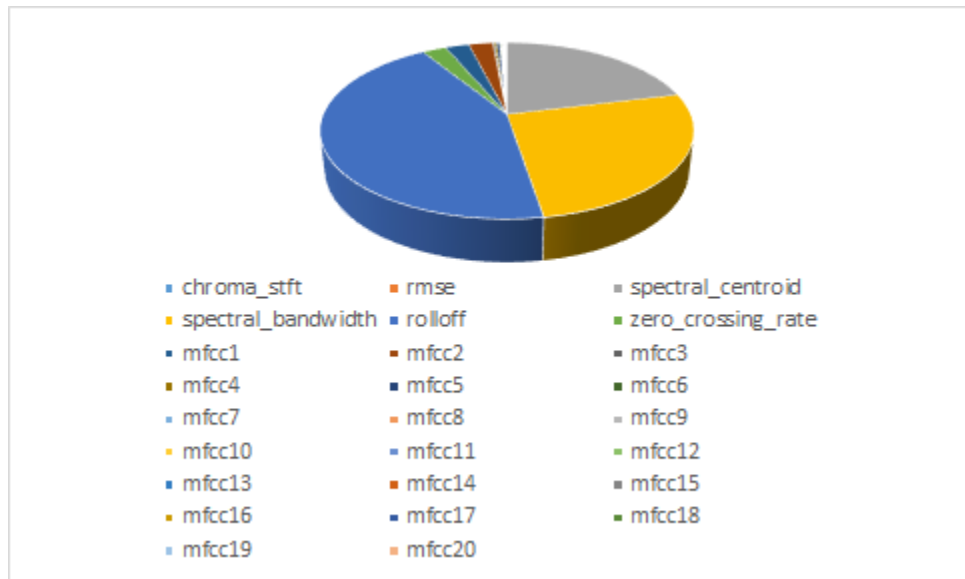


Figure 3.3.1: dataset ratio (male data, female data and third gender data)

We get 20% data for test from male, female and third gender. And we get 80% data for training. Visualization given in figure 3.3.2.

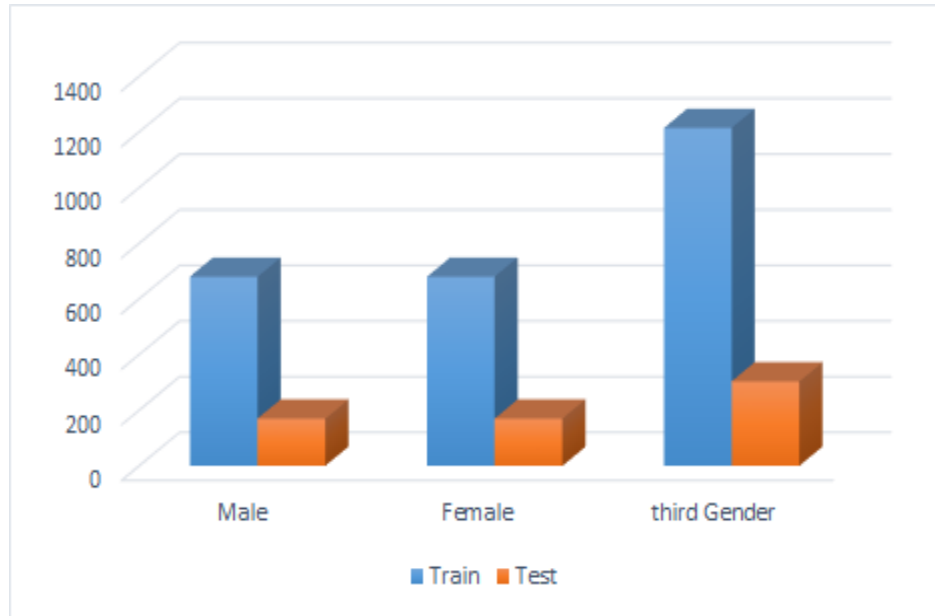


Figure 3.3.2: dataset ratio for test, train (male data, female data and third gender data)

3.4 Proposed Methodology:

Machines can't easily detect the voice. Humans and other species are in need of the capacity to acknowledge other people. Person-identity recognition has become more evident in the last decade.[17]. We use MFCCs to extract information from voice signals. The Discrete Fourier Transform, Mel-spaced filter bank energies, and log filter bank energies then provide us with a number of features that can be used for speech recognition. Speech recognition is a form of supervised machine learning, where an audio signal is fed into a model, which then outputs text. The reason why the raw audio cannot be used is because it can contain a considerable amount of noise. Feature extraction from audio signals has been found to facilitate a more efficient model and consequently, deliver better performance in comparison to raw signal analysis. Studies have revealed that if you preprocess and extract features from the original audio signal before feeding it into the base model, you can expect better results compared to using just the audio input. MFCC stands for Mel Frequency Cepstral Coefficients, and it's a technique widely used to decode information from audio signals.

3.4.1 Mel-frequency cepstral coefficients (MFCC):

The MFCC technique, outlined below, is an effective method of a type of speech recognition, shown in figure 3.4.1.1

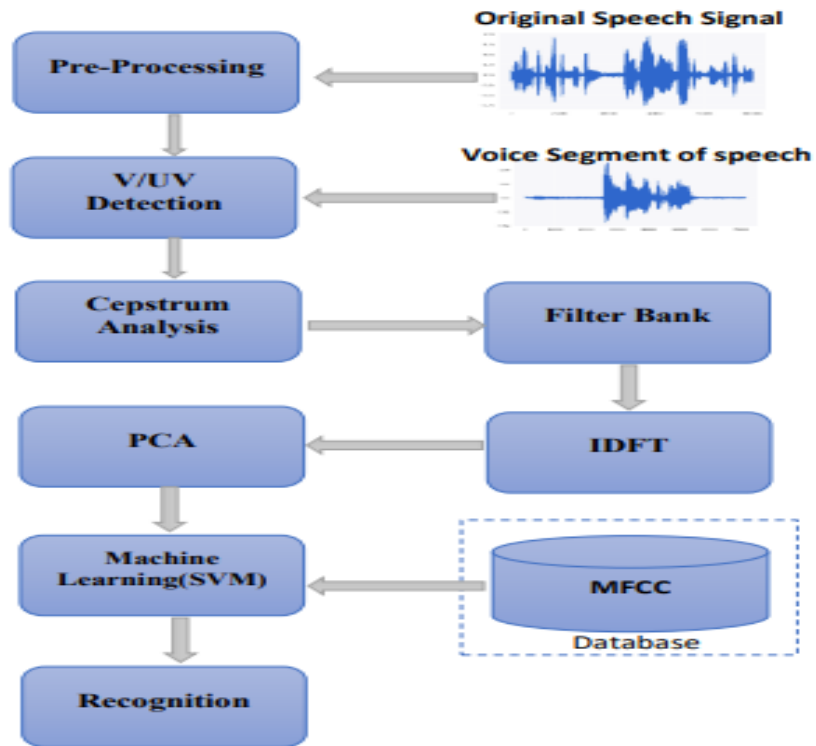


Figure 3.4.1.1: Workflow of MFCC

To analyze our signal, we'll need to convert it from analog to digital format with a sample rate of 8 kHz as shown in Figure 3.4.1.2 and Figure 3.4.1.3 respectively.

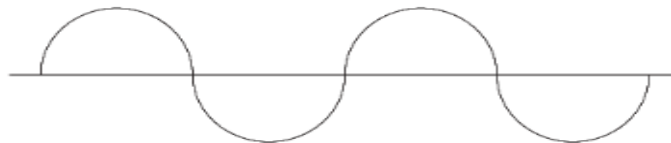


Figure 3.4.1.2: Analog signal



Figure 3.4.1.3: Digital signal

The MFCC features from the signal include windowing, followed by a discrete fourier transform (DFT). After getting the magnitude in log form, we can warp the frequency to the Mel scale. Once this has been done, it will be back transformed in order to match with the DFT of that particular frequency.

3.4.2 Preemphasis:

It increases the magnitude of power in the frequency.

3.4.3 Windowing:

MFCC provides the way to develop features from voice signals. Given the audio signal, a total of N phones can be expected. Frame width will be 25ms and each frame will be offset 10ms from the next. Illustrated below: We encounter our awareness in just a couple of milliseconds. The average person speaks three words per second and they tend to experience emotional shifts with each word. So, in just one second of conversation, there can be a multitude of changes in state. The result is 36 states per second or less than one-quarter second per state, which means we're usually over the threshold before we start to become aware. Each segment is made up of about 39 features.

3.4.4 DFT (Discrete Fourier Transform): We can convert a signal from the time domain to the frequency domain through the application of a discrete Fourier transform. Unfortunately, when analyzing waves in the audio time domain, it is much more difficult to determine frequencies and phases of the frequency spectrum - but in other domains this isn't so hard.

3.4.5 Mel-Filter Bank:

We chose to use the mel scale because it corresponds to our ability to hear pitches.

The formula for this is: $\text{mel}(f) = 1127 \ln(1 + f/700)$

3.4.6 Applying Log:

To provide the user with a natural experience, we applied a logarithm to the output of the Mel-filter.

3.4.7 IDFT: This is the opposite process to what was done in the prior step. To understand why we need to do this first, we'd have to explain how sound is made with human vocal cords. In figure 3.4.7.1 we see the normalize voice frequency and sample voice in figure 3.4.7.2.

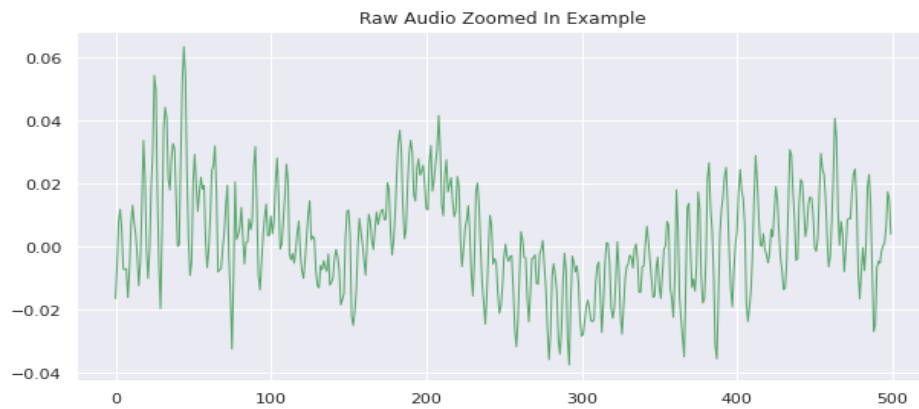


Figure 3.4.7.1: Normalize Frequency

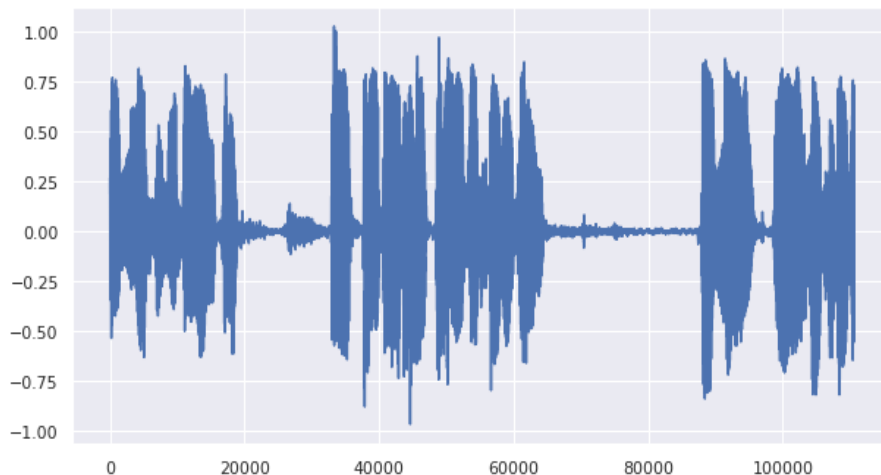


Figure 3.4.7.2: Sample Of voice

The MFCC model is calculated by extracting the first 12 coefficients of a signal after the IDFT operation.

3.4.8 Dynamic Features:

By utilizing 13 features and the MFCC technique, which takes into account both first-order & second-order derivatives, there are a total of 26 features. MFCCs are great technology for speech recognition by generating 39 features.

Lastly, we calculated 20 MFCC features and displayed them in figures 3.4.8.1.

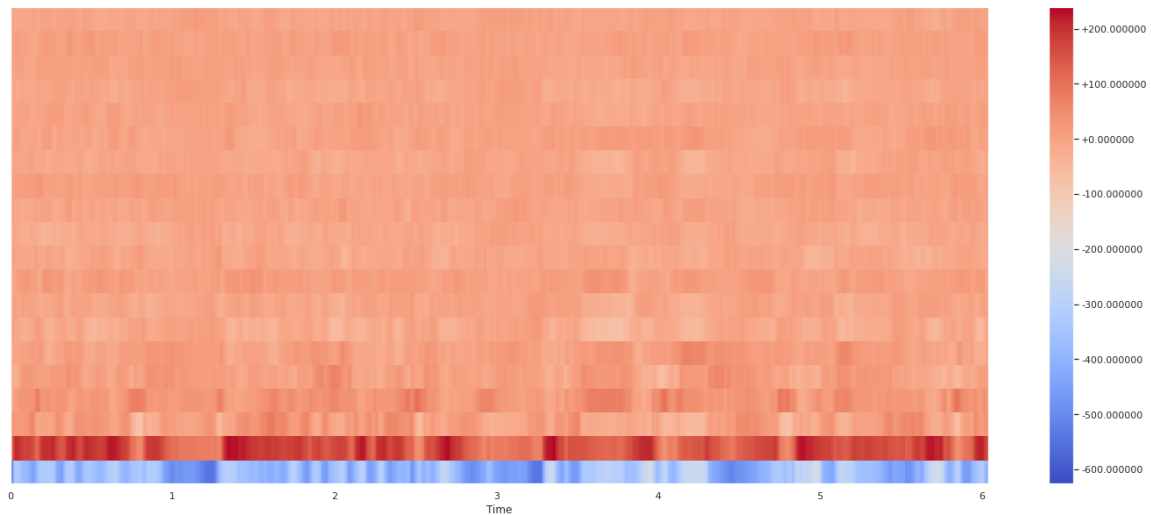


Figure 3.4.8.1: 20 MFCC features

To classify the audio files, MFCC features were gathered using the python "librosa" module [18] and imported into a numpy array [19]. The data was then labeled with 0, 1 and 2 for male, female and third gender respectively. The 'sklearn' module of python [20] was used to implement the test_train_split method. The dataset was split with a ratio of 80:20, wherein 80% was for training and for testing 20%.

3.4.9 Train Test Split:

The dataset, following the feature extraction and labeling process, was split into “train” and “test” sets for further analysis.

3.4.10 Supervised learning:

Machine learning models are fed labeled datasets to help them make predictions and classify outputs. This training goes on automatically as part of the process, with adjustments being made as you feed more data into the model. The goal is for it to provide the simplest, most accurate solution possible when classifying inputs. The procedure in figure 3.4.10.1.

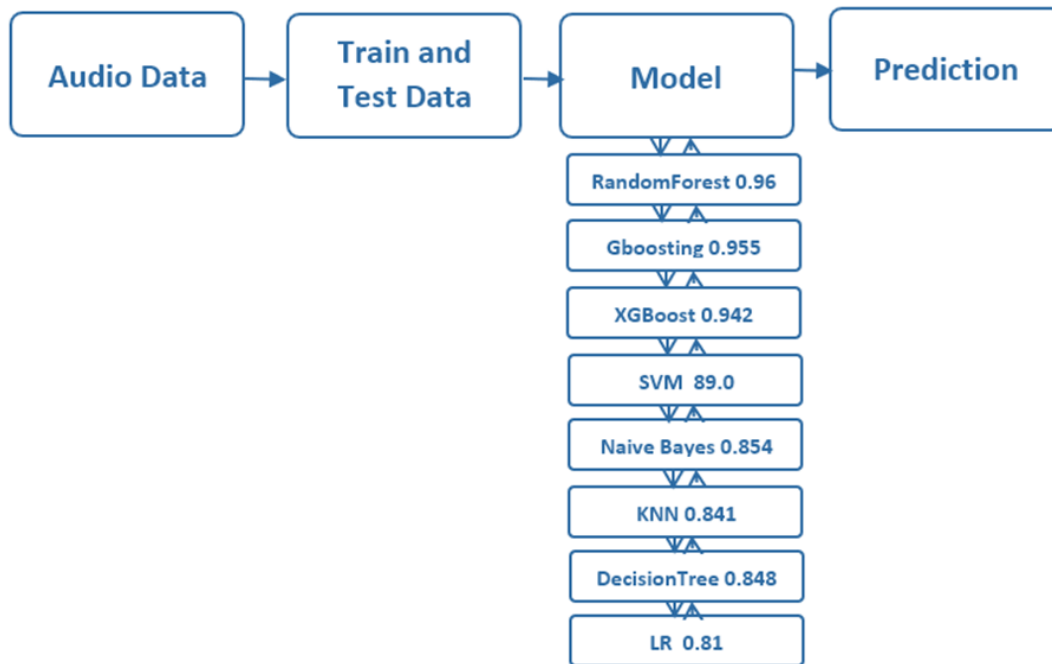


Figure 3.4.10.1: ML Classifier

3.5 Implementation Requirements:

The most important requirement is to collect the dataset for implementing code.

- Collecting Bangle Audio
- Pre-process the audio file, convert into CSV using audio converting tools
- Google Colab
- Machine Learning Algorithm implementation with proper coding
- CPU of google Colab was Intel(R) Xeon(R) CPU @ 2.20GHz

Chapter 4

Experimental Results and Discussion

4.1 Experimental Setup:

Collect the dataset, then convert the audio file into CSV. Every day, a huge amount of data is given rise to and let go on the Internet. It is proper first foremost to grow the number of news. CSV files are not always simple to embrace. Their structure can be hard with many columns and rows, an uneven number of columns in each row, different types of data in every part. We increase the dormant use of CSV datasets by providing better tools.

4.2 Experimental Results & Analysis:

There are seven types of algorithm using here:

- Random Forest
- Gradient boosting
- XGBoost
- SVM
- Naive Bayes
- KNN
- Decision Tree
- Logistic Regression

In table 4.2.1, there are the model evaluation value that help to find the best model and their accuracy.

Table 4.2.1: Model Evaluation

| Algorithms | Label | Precision | Recall | f1-score | Accuracy (%) |
|---------------|-------|-----------|--------|----------|--------------|
| Random Forest | 0 | 0.98 | 0.91 | 0.94 | 0.962 |
| | 1 | 0.95 | 0.94 | 0.95 | |

| | | | | | |
|-------------------|---|------|------|------|--------|
| | 2 | 0.96 | 0.99 | 0.98 | |
| KNN | 0 | 0.71 | 0.60 | 0.65 | 0.8414 |
| | 1 | 0.80 | 0.61 | 0.69 | |
| | 2 | 0.80 | 0.95 | 0.87 | |
| XGBoost | 0 | 0.95 | 0.90 | 0.92 | 0.94 |
| | 1 | 0.96 | 0.89 | 0.92 | |
| | 2 | 0.94 | 0.99 | 0.96 | |
| SVM | 0 | 0.89 | 0.88 | 0.89 | 0.89 |
| | 1 | 0.90 | 0.86 | 0.88 | |
| | 2 | 0.92 | 0.94 | 0.93 | |
| Naive Bayes | 0 | 0.87 | 0.84 | 0.85 | 0.854 |
| | 1 | 0.80 | 0.83 | 0.81 | |
| | 2 | 0.90 | 0.89 | 0.90 | |
| Gradient boosting | 0 | 0.98 | 0.91 | 0.94 | 0.96 |
| | 1 | 0.95 | 0.94 | 0.94 | |
| | 2 | 0.95 | 0.98 | 0.97 | |
| Decision Tree | 0 | 0.79 | 0.80 | 0.80 | 0.849 |
| | 1 | 0.79 | 0.85 | 0.82 | |
| | 2 | 0.91 | 0.87 | 0.89 | |

| | | | | | |
|---------------------|---|------|------|-------|-------|
| Logistic Regression | 0 | 0.70 | 0.72 | 0.701 | 0.809 |
| | 1 | 0.81 | 0.70 | 0.75 | |
| | 2 | 0.87 | 0.93 | 0.90 | |

- Highest accuracy 96.2% using Random forest
- Lowest accuracy 81% using Logistic Regression

In our research we found the best accuracy for Random forest and Logistic regression. So, some difference will help to better know the models in table 4.2.2.

Table 4.2.2: Difference between Logistic Regression and Random Forest

| Random Forest | Logistic Regression |
|--|---|
| Provide the high accuracy | Fast at classifying unknown records. |
| Automatically balance with big dataset | Observation number is lesser than the number of features, |
| Categorical | Overfitting |

4.3 Discussion:

We got the highest accuracy of 96.2% using the Random Forest algorithm because It is an indispensable lot of computational power as well as funds as it puts up countless trees to match up their outputs. We got the lowest accuracy on 81% of Logistic regression algorithms because it degrades their performance on big or complicated datasets. The confusion matrix in figure 4.3.1 is a great way to evaluate the accuracy of an algorithm. It contains true positives, true negatives, false negatives, and false positives, providing a concise explanation of how the model is performing.

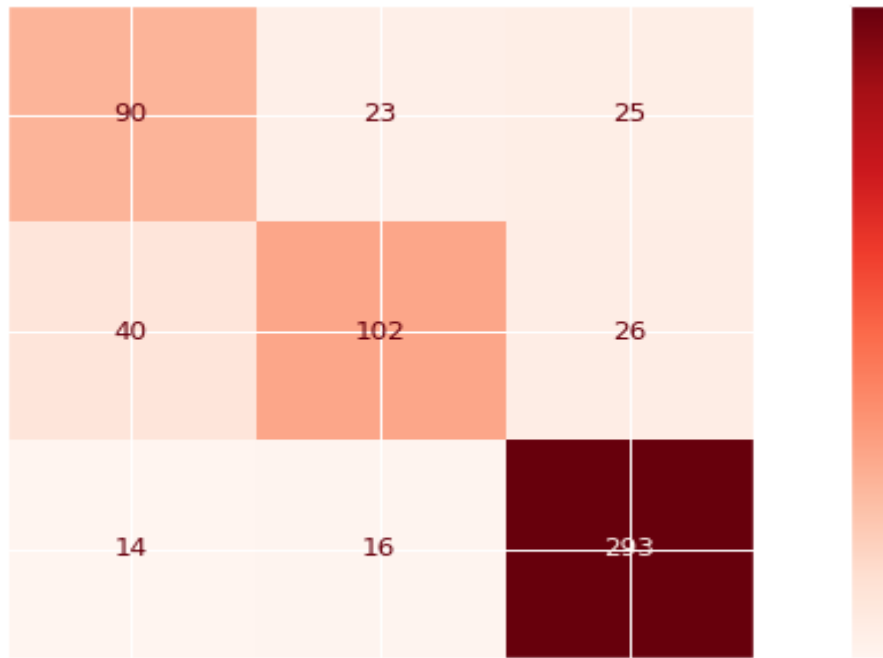


Figure 4.3.1: Confusion matrix

Chapter 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society:

The usage of speech recognition technology is uneven. Voice recognition is currently being used in a wide range of industries to facilitate routine tasks. For contrast, voice recognition has had a substantial beneficial impact on the legal sector. Lawyers use it to record key meetings, so they can later write the transcripts down in documents. By doing this, they not only save time, but also guarantee that all data is accurately captured. There is a significant amount of promise for voice recognition. The interface between users and devices can be improved, business operations can be greatly improved, and the machine-human relationship can be elevated to a completely new level.

5.2 Impact on Environment:

We come here with a system that is not new, but an updatable system. As the 20th century came to an end, voice recognition systems had discovered a wide and broad range of applications. Automated identification; Institutions that may choose to employ voice recognition to authenticate the identities of their clients in order to avoid disclosing sensitive and potentially dangerous personal information.

5.3 Ethical Aspects:

We understand that ethics is the term of doing something in a positive or right way. So the ethical use of technologies is a great concern nowadays. As these technologies help us a lot, we need to maintain them properly. There are many ethical aspects of voice recognition. Voice recognition has ethical aspects in detecting fraud activities, in forensic work, and in crime investigation.

5.4 Sustainability Plan:

Utilizing cutting-edge voice technologies has advantages beyond aiding environmental responsibility. Security is an essential consideration. No need to worry if a paper file was

lost, destroyed by a shredder, thrown in the trash, or encrypted in real time using native security capabilities. Furthermore, by speeding up procedures and cutting back on time-consuming administrative manual activities, workflows may be made simpler and more productive employing strong yet user-friendly digital voice solutions.

Chapter 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study:

Identifying a person's gender as male or female, based on a sample of their voice, appears to initially be an easy task for a human being. KNN algorithm interaction helps to react properly. This study focuses on the censorious features to extricate from a voice signal in order to build a successful gender and insider state detection system and showing the advantages of combining the total algorithm on the total approximate recognition score.

6.2 Conclusions:

We get 96% accuracy using KNN algorithm, smoothly detecting the voice which contains male, female or third gender. It has recently received a lot of attention in order to create more natural voice recognition. Voice recognition displays all the technical background and sequence fulfillment to create a system, the digital form of speech, speech improvement methods, classification methods, and relevant working procedure. The main theme of this project, all three types of gender identification systems were built, and their performance was judged in a variety of tests and parameters. The whole mashup algorithm could be thinking about success.

6.3 Implication for Further Study:

The ability to accurately identify gender based on voice is a powerful tool that can be used in a variety of fields. However, the accuracy of this system depends heavily on the quality of the microphone and other conditions such as background noise and reverberation. As such, it is important to ensure that these conditions are optimal in order to ensure accurate results when using AI-based gender identification systems. Furthermore, it is based on a continuous approach to Fast Fourier Transformation for a more rousting system.

Reference:

1. Levitan, Sarah Ita, Taniya Mishra, and Srinivas Bangalore. "Automatic identification of gender from speech." *Proceeding of speech prosody*. Semantic Scholar, 2016.
2. Van Lancker, Diana, Jody Kreiman, and Karen Emmorey. "Familiar voice recognition: Patterns and parameters part I: Recognition of backward voices." *Journal of phonetics* 13.1 (1985): 19-38.
3. Van Lancker, Diana, Jody Kreiman, and Thomas D. Wickens. "Familiar voice recognition: Patterns and parameters Part II: Recognition of rate-altered voices." *Journal of phonetics* 13.1 (1985): 39-52.
4. White, David J., Andrew P. King, and Shan D. Duncan. "Voice recognition technology as a tool for behavioral research." *Behavior Research Methods, Instruments, & Computers* 34 (2002): 1-5.
5. Bull, Ray, Harriet Rathborn, and Brian R. Clifford. "The voice-recognition accuracy of blind listeners." *Perception* 12.2 (1983): 223-226.
6. Bajorek, Joan Palmiter. "Voice recognition still has significant race and gender biases." *Harvard Business Review* 10 (2019).
7. Roswadowitz, Claudia, et al. "Two cases of selective developmental voice-recognition impairments." *Current Biology* 24.19 (2014): 2348-2353.
8. Ascì, Francesco, et al. "Machine-learning analysis of voice samples recorded through smartphones: the combined effect of ageing and gender." *Sensors* 20.18 (2020): 5022.
9. Alsharhan, Eiman, and Allan Ramsay. "Investigating the effects of gender, dialect, and training size on the performance of Arabic speech recognition." *Language Resources and Evaluation* 54 (2020): 975-998.
10. Raahul, A., et al. "Voice based gender classification using machine learning." *IOP Conference Series: Materials Science and Engineering*. Vol. 263. No. 4. IOP Publishing, 2017.
11. Livieris, Ioannis E., Emmanuel Pintelas, and Panagiotis Pintelas. "Gender recognition by voice using an improved self-labeled algorithm." *Machine Learning and Knowledge Extraction* 1.1 (2019): 492-503.
12. Huang, Kuo-Liang, Sheng-Feng Duan, and Xi Lyu. "Affective Voice Interaction and Artificial Intelligence: A research study on the acoustic features of gender and the emotional states of the PAD model." *Frontiers in Psychology* 12 (2021): 664925.
13. Abdulsatar, Assim Ara, et al. "Age and gender recognition from speech signals." *Journal of Physics: Conference Series*. Vol. 1410. No. 1. IOP Publishing, 2019.
14. Buyukyilmaz, Mucahit, and Ali Osman Cibikdiken. "Voice gender recognition using deep learning." *2016 International Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA2016)*. Atlantis Press, 2016.
15. Shareef, Mustafa Sahib, Thulfiqar Abd, and Yaqeen S. Mezaal. "Gender voice classification with huge accuracy rate." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 18.5 (2020): 2612-2617.
16. Liu, Ran R., et al. "Voice recognition in face-blind patients." *Cerebral Cortex* 26.4 (2016): 1473-1487.
17. Roswadowitz, Claudia, et al. "Two cases of selective developmental voice-recognition impairments." *Current Biology* 24.19 (2014): 2348-2353.
18. Learn about librosa.feature.mfcc — librosa 0.7.0 documentation, available at << <https://librosa.github.io/librosa/generated/librosa.feature.mfcc.html> >>, last accessed on 02-06-2022 at 12:00 PM.
19. Learn about numpy.append, NumPy v1.17 Manual, available at << <https://docs.scipy.org/doc/numpy/reference/generated/numpy.append.html> (accessed 9. 30. 19). >>, last accessed on 06-06-2022 at 16:00 PM.
20. Learn about sklearn.model_selection.train_test_split, available at << https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html>>, last accessed on 03/07/2022 at 19:00 PM.

ReportFormatBScCSEUpdated

ORIGINALITY REPORT

27 %
SIMILARITY INDEX

26 %
INTERNET SOURCES

10 %
PUBLICATIONS

20 %
STUDENT PAPERS

PRIMARY SOURCES

1 dspace.daffodilvarsity.edu.bd:8080 11 %
Internet Source

2 Submitted to Daffodil International University 3 %
Student Paper

3 www.analyticsvidhya.com 2 %
Internet Source

4 Xueping Hu, Xiangpeng Wang, Yan Gu, Pei Luo, Shouhang Yin, Lijun Wang, Chao Fu, Lei Qiao, Yi Du, Antao Chen. "Phonological experience modulates voice discrimination: Evidence from functional brain networks analysis", Brain and Language, 2017 1 %
Publication

5 www.frontiersin.org 1 %
Internet Source

6 Submitted to Jacksonville University 1 %
Student Paper

7 etd.lsu.edu <1 %
Internet Source
