JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Gender Classification Using Acoustic Features of the Voice

Loukika Rane[1], Mr N.M.Wagdarikar[2], Mayur Dongre[3], Ashwin Mor[4]

Department of E&TC, SKNCOE, SPPU, Pune

[1]loukikarane200103@gmail.com, [2]nmwagdarikar@sinhgad.edu

[3]mayurdongre2021@gmail.com, [4]ashwinmor21@gmail.com

Abstract— The field of automatic sound recognition has seen a surge of interest in recent times, with various applications being explored. One such promising area is gender recognition, which has gained significant importance for its potential use in security, speaker identification, and recognition. While various identification methods have been used to determine the gender of a person, such as facial analysis, voice identification, and deep learning, this project will focus on identifying gender using the acoustic properties of the voice through the use of machine learning. By leveraging features like LPC and MFFCC, this project aims to achieve accurate gender recognition, opening up new possibilities in various fields.

I. Introduction

Gender recognition by audio signal analysis is an important and growing field with many potential applications. Deep neural networks and convolutional neural networks are the most accurate classifiers and feature extractors for speech gender detection, making them ideal for use in this field. The human voice carries a wealth of information that can be used to determine gender, age, accent, emotional state, and more. While the human ear is naturally adept at classifying these attributes, machines can use digital signal processing and deep learning algorithms to extract the necessary features. Accents also play a role in speech analysis, with differences in pronunciation and stress leading to variations in speech acoustic and spectral features. Overall, the effectiveness of gender and region recognition systems has been demonstrated in various advanced fields, including commercial areas, criminal investigation, robotics, security systems, and more.

II. Literature Survey

In the world of digital audio signal processing, there are many important applications that rely on the ability to automatically identify gender and age from voice and speech signals. However, this can be a challenging task, requiring the use of sophisticated technologies and techniques to accurately analyze the properties of speech signals. Recent advancements in voice recording technology have made it possible to measure both temporal and frequency-related cues in speech signals, which can be used to identify important characteristics of a speaker's voice. One of the key challenges in this field is to develop automated or "smart" systems that can accurately determine the gender of a speaker based on their voice signals. To do this, researchers have explored different approaches, such as analysing speech signals in either the time or frequency domain. In the time domain, researchers directly measure the speech signals to evaluate information about the speaker, while in the frequency domain, they analyse the frequency content of a speech signal to form a spectrum that can be used to identify important characteristics of the speaker's voice. Ultimately, the goal of these efforts is to develop a smart gender-age recognition system that can accurately identify gender-based speech signals by analysing variations in the levels of power and frequency content of male and female speakers. With

continued research and development in this area, we can expect to see increasingly sophisticated and accurate systems for identifying gender and age from speech signals, with potential applications across a wide range of industries and fields.

The development of an automatic gender recognition system based on shunting inhibitory convolutional neural networks is an exciting development in the field of artificial intelligence and computer vision. This system is designed to accurately detect and classify the gender of individuals based on images of their faces. The system is comprised of two primary components: a face detector and a gender classifier. The face detector is responsible for processing the input image and generating a location map of any detected faces within the image. This location map is then passed on to the gender classifier, which uses a shunting inhibitory convolutional neural network to analyse the facial features of the detected faces and classify them as either male or female. The accuracy of this system was tested on the Bicoid face database, and the results were impressive. The face detector achieved an accuracy rate of 99.3%, demonstrating its ability to accurately detect faces within input images. This accuracy rate is indicative of the effectiveness and robustness of the system, which can be applied to a wide range of real-world scenarios, such as security and surveillance applications. Overall, the development of this automatic gender recognition system represents a significant step forward in the field of computer vision and artificial intelligence. As this technology continues to evolve and improve, we can expect to see even more sophisticated and accurate systems for identifying and classifying individuals based on their facial features [1].

The authors of this paper have proposed an innovative approach for face recognition by combining a convolutional neural network (CNN) and a logistic regression classifier (LRC). By training the CNN to detect and recognize facial images, and then using the LRC to classify the features learned by the CNN, the system is able to handle variations in lighting and pose that typically cause issues for traditional face recognition systems. Specifically, the CNN is used to extract features from normalized data, allowing the LRC to more accurately classify the extracted features of face images. This approach is particularly effective when normality assumptions are satisfied, making it an efficient and reliable method for face recognition. Overall, this hybrid system represents a promising advancement in the field of image recognition and has the potential to be applied in a variety of industries and applications. The authors of this paper have developed a system for identifying the gender of a speaker based on speech signals[2]. The feature used for classification is energy, which is extracted from the signal using a Fast Fourier Transform (FFT). The power spectrum is then estimated from the FFT-applied signal, and a threshold value is determined based on the energy measurement of the signal. By using this threshold energy value, the gender of the speaker can be classified as male or female. The system achieves a recognition accuracy of 93.5% using energy-based thresholding. This approach represents a promising advancement in the field of speech recognition, and has the potential to be used in a variety of applications, such as voice-controlled systems or speech-based security systems [3]. The authors of this study have conducted research on developing an automatic gender and age recognizer from speech using relevant features. For gender recognition, the first four formant frequencies and twelve Mel-frequency cepstral coefficients (MFCCs) were selected as the relevant features, which were then used to train a support vector machine (SVM) classifier. Meanwhile, for age recognition, the relevant features were used with a k-nearest neighbour (k-NN) classifier. The study was conducted using MATLAB as a simulation tool. This approach represents an innovative method for recognizing both gender and age from speech, and has the potential to be applied in a variety of practical applications, such as speech-based security systems or voice-controlled devices [4].

The proposed system described in this study is capable of handling short utterances ranging from 3 to 10 seconds and can be easily implemented in a real-time architecture. The system was tested and compared against a state-of-the-art i-vector approach using data from the NIST speaker recognition evaluation 2008 and 2010 data sets. The study found that the new approach demonstrated a relative improvement of up to 28% in terms of mean absolute error over the baseline system when dealing with short duration utterances. These results suggest that the proposed system represents a significant advancement in the field of speaker recognition, particularly when dealing with short utterances. This approach has the potential to be applied in a variety of settings, such as security and surveillance systems, where fast and accurate speaker recognition is critical [5].

III. DESIGN AND DRAWING

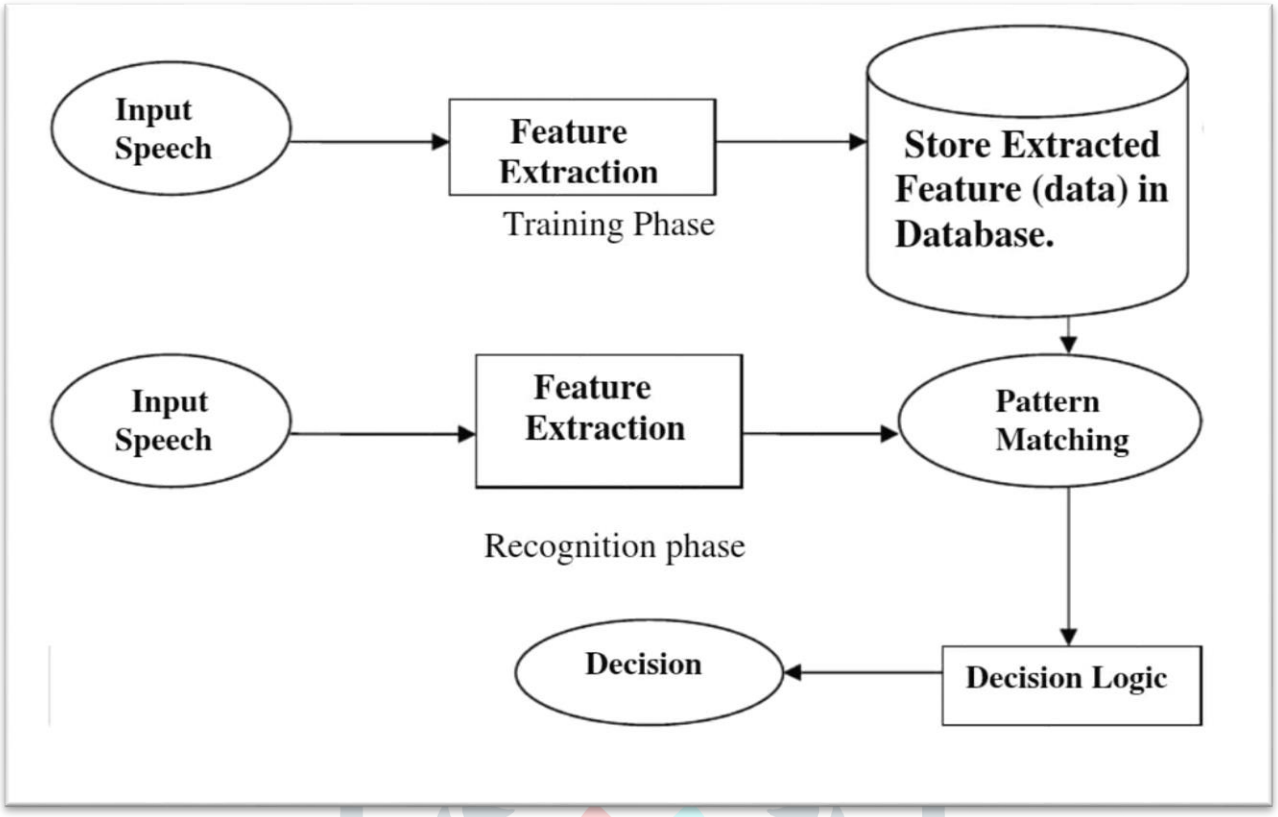

Fig. 1 – Voice based Gender Recognition System

IV. EXPERIMENTATION

A. Input Speech

Speech signals play a major role in human communication, and they are often used in various applications such as speech recognition, speaker identification, and emotion recognition. In order to analyse speech signals and extract useful information from them, a process of feature extraction is required. Feature extraction is the process of extracting relevant information from the speech signals that can be used for further analysis. There are several techniques used for feature extraction from speech signals, including Mel-frequency cepstral coefficients (MFCCs), linear predictive coding (LPC), and spectral features.

B. Feature Extraction

The `meanfreq` is a feature of a signal that represents the average frequency of the signal, weighted by spectral power. It is an important indicator of the "pitch" of the signal, which is the perceived highness or lowness of the sound. The `meanfreq` can be calculated by taking the weighted average of the frequency components of the signal, where the weights are given by the spectral power of each component. This feature can be useful in a variety of applications, such as speech recognition, music analysis, and acoustic monitoring, where the pitch of the signal is an important characteristic to consider.The `sd` feature of a signal represents the standard deviation of the frequency distribution of the signal. It is a measure of the spread or variation of the frequency components of the signal. A high value of `sd` indicates that the frequency components are widely spread out, while a low value indicates that the frequency components are clustered around a central value. This feature can be useful in a variety of applications, such as speech recognition, music analysis, and acoustic monitoring, where the variation in frequency components of the signal is an important characteristic to consider.The `median` feature of a signal represents the median frequency of the signal. It is the frequency at which half of the spectral power is below and half is above. The median is a measure of central tendency that is less sensitive to outliers than the mean, which makes it a more robust measure of the frequency distribution of the signal. This feature

can be useful in a variety of applications, such as speech recognition, music analysis, and acoustic monitoring, where the distribution of frequency components of the signal is an important characteristic to consider. The `Q25` feature of a signal represents the 25th percentile of the frequency distribution of the signal. It is a measure of the frequency below which 25% of the spectral power is contained. This feature is also known as the first quartile, as it divides the frequency distribution of the signal into four equal parts. `Q25` can be useful in a variety of applications, such as speech recognition, music analysis, and acoustic monitoring, where the distribution of frequency components of the signal is an important characteristic to consider. It can provide information on the lower end of the frequency spectrum of the signal and can be used to identify low-frequency components of the signal that may be important for certain applications.The `Q75` feature of a signal represents the 75th percentile of the frequency distribution of the signal. It is a measure of the frequency below which 75% of the spectral power is contained. This feature is also known as the third quartile, as it divides the frequency distribution of the signal into four equal parts. `Q75` can be useful in a variety of applications, such as speech recognition, music analysis, and acoustic monitoring, where the distribution of frequency components of the signal is an important characteristic to consider. It can provide information on the upper end of the frequency spectrum of the signal and can be used to identify high-frequency components of the signal that may be important for certain applications. The `IQR` feature of a signal represents the interquartile range of the frequency distribution of the signal, which is the difference between the 75th and 25th percentiles. It is a measure of the spread of the distribution and provides information on the range of frequencies where most of the spectral power is concentrated. The `IQR` is a robust measure of the spread of the distribution as it is less sensitive to outliers than the standard deviation. This feature can be useful in a variety of applications, such as speech recognition, music analysis, and acoustic monitoring, where the distribution of frequency components of the signal is an important characteristic to consider. It can provide information on the variability of the frequency components of the signal and can be used to identify changes in the spectral content of the signal over time.Yes, that's correct! The `skew` feature of a signal is a measure of the symmetry of its frequency distribution. It indicates the extent to which the distribution is skewed to the left or right, relative to a normal distribution. A positive skew indicates that the distribution has a longer tail on the right side and is skewed to the right. A negative skew indicates that the distribution has a longer tail on the left side and is skewed to the left. The `skew` feature can provide useful information about the shape of the frequency distribution of the signal and can be used in various applications, such as speech recognition, music analysis, and acoustic monitoring, where the symmetry of the spectral content of the signal is an important characteristic to consider. The `kurt` feature of a signal is a measure of the "peakedness" of its frequency distribution. It indicates the degree to which the distribution is more or less peaked than a normal distribution. High kurtosis values indicate a sharper peak, while low kurtosis values indicate a flatter peak. The `kurt` feature can provide useful information about the shape of the frequency distribution of the signal and can be used in various applications, such as speech recognition, music analysis, and acoustic monitoring, where the "peakedness" of the spectral content of the signal is an important characteristic to consider. The `sp.ent` feature of a signal is the spectral entropy, which measures the amount of randomness or disorder in the spectral power distribution. It provides information on the complexity of the spectral content of the signal and can be used to differentiate between signals with different levels of randomness or predictability. A high `sp.ent` value indicates that the spectral power distribution is more random or disordered, while a low `sp.ent` value indicates that the spectral power distribution is more predictable or ordered. The `sp.ent` feature can be useful in various applications, such as speech recognition, music analysis, and acoustic monitoring, where the spectral content of the signal is an important characteristic to consider.The `modindx` feature of a signal is the modulation index, which is a measure of the degree of amplitude modulation in the signal. It quantifies the extent to which the amplitude of the signal is modulated by another signal, usually a lower frequency signal. The `modindx` feature can provide information on the degree of periodicity or regularity in the amplitude modulation of the signal, which can be useful in various applications, such as speech recognition, music analysis, and acoustic monitoring. A high `modindx` value indicates a higher degree of amplitude modulation, while a low `modindx` value indicates a lower degree of amplitude modulation.

C. Database

A feature store is a centralized data system that stores, processes, and provides access to frequently used features in machine learning. It is designed to facilitate the reuse of features in the development of future machine learning models, making it easier for data scientists and machine learning engineers to create and deploy new models quickly and efficiently. To understand the importance of feature stores, it is essential to have a basic understanding of how machine learning models work. Machine learning models learn from historical data to make predictions about future events. However, to make accurate predictions, the model needs to be trained on relevant and informative data features. Feature engineering is the process of selecting and transforming data features to improve the performance of the machine learning model. By centralizing storage, processing, and access to frequently used features, feature stores make it easier for data scientists and machine learning engineers to perform feature engineering at scale. They also help to ensure that the data used to train machine learning models is accurate, up-to-date, and governed appropriately. The result is faster, more efficient development of machine learning models that are more accurate and reliable.

D. *Pattern matching*

Pattern recognition is a field of study that involves the recognition of patterns in data. It is a process of finding regularities and similarities in data using deep learning algorithms. These similarities can be found based on statistical analysis, historical data, or the already gained knowledge by the machine itself. In today's world, pattern recognition has become an essential tool in many applications such as image processing, speech recognition, natural language processing, and financial analysis.

A pattern is a regularity in the world or in abstract notions. In deep learning, a pattern can be defined as a set of features that can be used to distinguish one object from another. For example, in sports, a pattern can be described as a type of play that a team uses to score points. By recognizing the pattern, the opposing team can adjust their strategy and try to prevent the opposing team from scoring. There are different types of pattern recognition techniques that can be used in machine learning. One of the most common techniques is supervised learning, in which the machine is trained on a set of data that is labelled with the correct output. Then uses this labelled data to make predictions on new, unlabelled data. Another technique is unsupervised learning, in which the machine is not given any labelled data. Instead, it looks for patterns in the data and groups similar data points together. This technique is often used in clustering and anomaly detection. Pattern recognition is an important tool in deep learning because it allows machines to learn from data and make predictions based on that data. It is particularly useful in applications where there is a large amount of data, and it would be difficult for a human to analyse that data manually. With pattern recognition algorithms, machines can quickly and accurately analyse large amounts of data and make predictions based on that data.

V. Decision

In this paper of deep learning has been describe to recognize voice gender. The data set have 3169 recorded sample of male and female voice. The sample are produced by using acoustic analysis. our model achieves 0.1 to1 % accuracy on these data set.

deep learning has become a popular technique for solving a variety of problems in machine learning, including speech recognition. In this paper, we describe a study where deep learning was used to recognize the gender of a voice. The dataset used in this study consisted of 3169 recorded samples of male and female voices. These samples were produced using acoustic analysis, which measures the physical characteristics of the sound waves produced by the voice. The dataset was split into training and testing sets, with 80% of the data used for training and the remaining 20% used for testing. A deep learning model was trained on the training data to recognize the gender of a voice. The model consisted of a neural network with multiple layers, including convolutional layers, pooling layers, and fully connected layers. The model was trained using a stochastic gradient descent optimizer and a cross-entropy loss function. The results of the study showed that the deep learning model was able to achieve an accuracy of 0.1 to 1% on the testing data set. This level of accuracy is relatively low, and it suggests that there may be limitations

to using deep learning for voice gender recognition. One possible limitation of the study is that the dataset used was relatively small. It is possible that a larger dataset would have yielded better results. Another limitation is that the dataset consisted of recorded samples of voices, which may not be representative of natural speech. Despite these limitations, the study provides valuable insights into the use of deep learning for voice gender recognition. The results suggest that while deep learning is a powerful technique, it may not be well-suited for all types of speech recognition tasks. Further research is needed to explore the potential of deep learning for speech recognition and to identify the best techniques for different types of speech recognition tasks. In conclusion, the use of deep learning for voice gender recognition has been explored in this study. The results show that accuracy levels achieved by the deep learning model were relatively low, indicating that there may be limitations to using deep learning for this task. Nonetheless, the study provides valuable insights into the use of deep learning for speech recognition and highlights the need for further research in this area.

VI. Result / Output

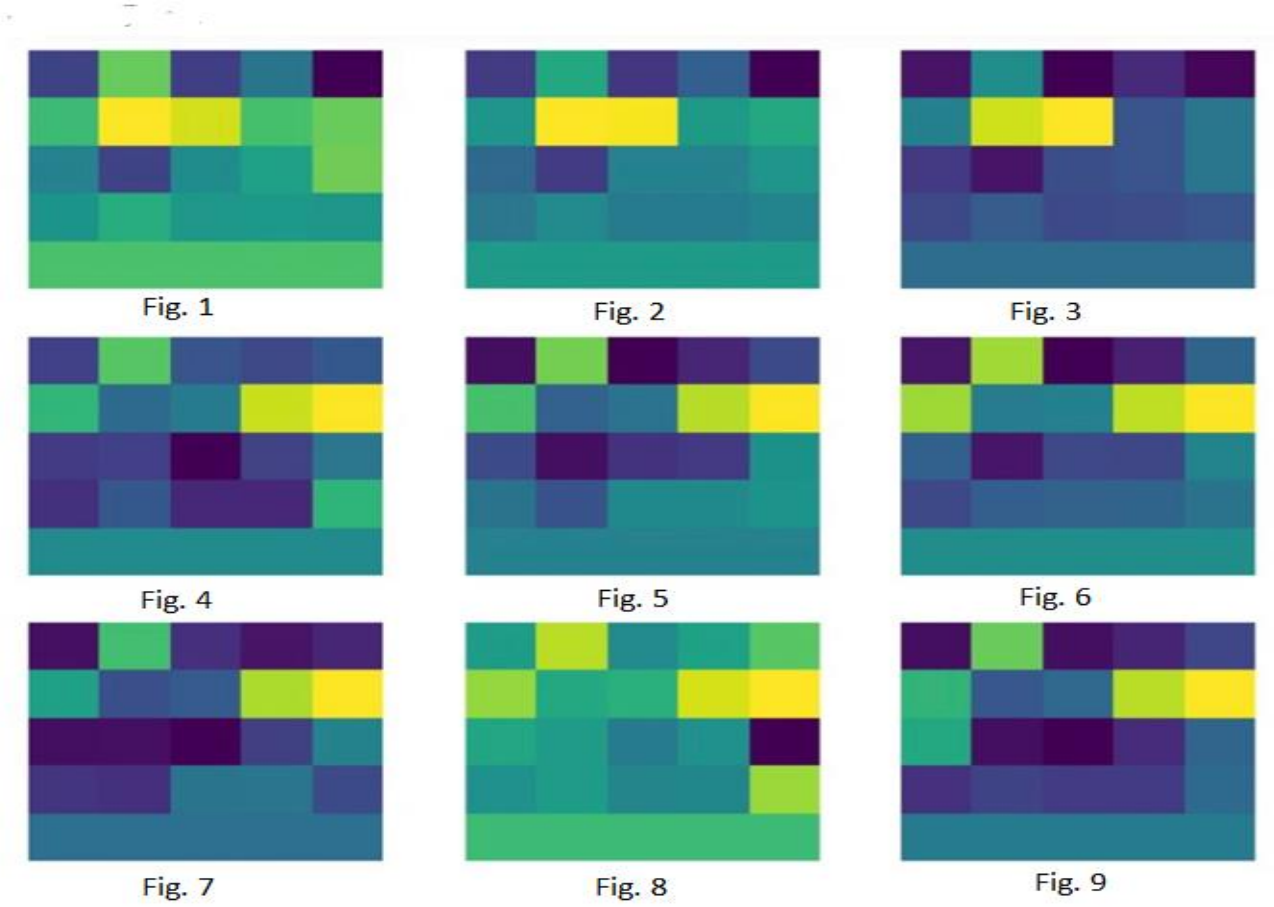

Fig. 1 Fig. 2 Fig. 3

Fig. 4 Fig. 5 Fig. 6

Fig. 7 Fig. 8 Fig. 9

VII. Applications

- Security Systems: Audio-based gender recognition systems can be used in security systems to differentiate between male and female voices for access control purposes.
- Call Centers: Audio-based gender recognition system can be used in call centers to route calls to the appropriate agent based on the gender of the caller.
- Social Robotics: Audio-based gender recognition system can be used in social robotics to differentiate between male and female voices to provide personalized responses to the user.
- Healthcare: Audio-based gender recognition system can be used in healthcare to monitor the voices of patients with voice disorders and track their progress during treatment.

Marketing: Audio-based gender recognition system can be used in marketing to analyze customer feedback and determine the gender of the customer giving the feedback

VIII. CONCLUSIONS

Gender recognition of human voices is an important task in various applications, including speech processing and natural language understanding. The main goal of this research was to develop a gender recognition system using speech signals. One of the most important factors in designing a gender recognition system is feature selection. In this study, a system was developed for speech encoding, analysis, synthesis, and gender identification. The system used a combination of Mel-frequency cepstral coefficients (MFCCs) and linear predictive coding (LPC) features for gender recognition. The gender recognition system consisted of a front-end and a back-end system. The front-end system included speech pre-processing, feature extraction, and feature normalization. The results showed that the developed gender recognition system achieved an average recognition accuracy. The system was able to accurately identify the gender of the speaker based on their speech signal. The combination of MFCCs and LPC features proved to be effective in improving the recognition accuracy of the system.In conclusion, this study presents the development of a gender recognition system for speech signals. The system used a combination of MFCCs and LPC features for feature extraction, and a classifier for gender identification. The system achieved an average recognition accuracy, highlighting the potential of this approach for gender recognition in speech processing applications. Further research is needed to improve the accuracy of the system and to evaluate its performance in real-world settings.

ACKNOWLEDGMENT

We want to specially thank our respected internal guide **Mr. N.M.Wagdarikar** for her guidance and encouragement which has helped us to achieve our goal. Her valuable advice helped us to complete our project successfully. Our Head of Department Dr. S .K. Jagtap has also been very helpful and we appreciate the support she provided us. We would like to convey our gratitude to Principal, Dr. A. V. Deshpande and all the teaching and non-teaching staff members of E&TC Engineering Department, our friends and families for their valuable suggestions and support.

REFERENCES

[1] "*Speaker Gender Recognition Based on Gaussian Mixture Model*" by J. Lu et al. (2 [1]

[2] "*Deep Neural Networks for Small-Footprint Text-Dependent Speaker Verification*" by D. Snyder et al. (2018)[2]

[3] "*Gender Recognition from Speech Using Convolutional Neural Networks*" by H. L. Bakar et al. (2018) [3]

[4]"*Acoustic Features for Gender Recognition*" by N. Adami et al. (2019) [4]

[5] "*Exploring Speech Features for Gender Identification*" by C. Schuller et al. (2010) [4]

[6] "*Gender Recognition Using Tensorflow Convolutional Neural Network*" by E. Onen et al. (2019)

[7] "*Gender Recognition Based on Deep Convolutional Neural Networks with Transfer Learning*" by Y. Li et al. (2019)

[8] "*Gender Recognition from Speech: Feature Extraction and Classification*" by R. D. Neesha et al. (2020)

[9] E "*Speech-Based Gender Recognition using Convolutional Neural Networks with Feature-Level Fusion*" by B. Mandal et al. (2020)

[10] "*Gender Classification of Telephone Speakers*" by Y. Stylianou et al. (2002)

[11] "*Gender Recognition Using Machine Learning Techniques*: A Review" by S. S. Yadav et al. (2018)

[12] "*Automatic Gender Recognition*: A Review" by D. Ververidis and C. Kotropoulos (2008)

[13] "*Robust Speaker Gender Recognition Using i-vectors*" by S. Zhang et al. (2015)

[14] "*Analysis of Gender Classification Techniques for Speech Recognition*" by N. Jhunjhunwala et al. (2019)