INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 06 | JUNE - 2023 SJIF RATING: 8.176 ISSN: 2582-3930

Different Circumstances Human Detector

Diwakar Kumar Singh
Department of Information Technology & Engineering(IT)
Noida Institute of Engineering & Technology
Greater Noida, India
diwakar.raj52@gmail.com

Vikas Tiwari
Department of Information Technology & Engineering(IT)
Noida Institute of Engineering & Technology
Greater Noida, India
vikas20t@gmail.com

Aditya Singh Chauhan
Department of Information Technology & Engineering(IT)
Noida Institute of ngineering & Technology
Greater Noida, India
adityachauhan8265@gmail.com

Sumit Shukla
Department of Information Technology & Engineering
Noida Institute of Engineering & Technology
Greater Noida, India
sumitsanjayshukla@gmail.com

Ms. Neetu kumari rajput(project guide)
Department of Information Technology & Engineering
Noida Institute of Engineering & Technology
Greater Noida, Inida

***Abstract*— Numerous applications, including human-computer interaction, security systems, and customised services, highlight the growing significance of automatic gender and age identification through speech signals. This study extensively examines the development and assessment of a reliable system for gender and age identification by extracting acoustic information from speech signals. To guarantee a balanced representation across different age groups and genders, a wide range of speech samples from various sources is employed in this study. The dataset is partitioned into separate subsets for training, validation, and testing, enabling a comprehensive evaluation of the proposed system's performance. Rigorous cross-validation techniques are employed to ensure the reliability and practicality of the findings. In summary, this research introduces an innovative and effective approach for accurately determining gender and age using voice cues. The system demonstrates a remarkable level of precision and resilience, rendering it applicable in diverse domains such as human-computer interaction, security systems, and specialized services. Future studies will focus on enhancing the system's performance by incorporating supplementary features and exploring advanced machine learning techniques.**

Keywords— ML,CSV,KNN,support vector Regression, MFCC,CNN,SVM(support vector machine).

I. INTRODUCTION

Voice-based age detection and gender identification pose significant challenges in telephone speech processing when determining an individual's identity. A novel approach is proposed for gender and age recognition, utilizing generative adversarial models, to address these challenges. These bases are learned using non-negative matrix factorization under the sparsity constraint on data of male and female individuals.In the identification process, various biometric features are applied, such as fingerprint, forensic verification, iris scanning, signature, face geometry, and voice components. An examination of several telephone speech gender recognition methods is looked into.When the observed speech is long, the first recognizer system produced exceeds other approaches.We're taking a quick look at the suggested starting setup.

The goal of this exciting area of research and development is to analyse and forecast an individual's age and gender based on the peculiarities of their spoken language. Through the analysis of different acoustic and linguistic variables found in speech signals, machine learning algorithms can be effectively trained to precisely identify a speaker's age and gender. The main objective of employing speech-based age and gender recognition is to automate the process of gathering demographic data, offering a wide range of practical applications in real-life situations. For instance, using voice-controlled systems to modify responses based on the user's age and gender can be effective in contact centres for customer segmentation, personalised marketing campaigns, or even in those systems.

There are various processes involved in the age and gender detection procedure. A sizable collection of labelled speech samples is initially gathered, often representing a range of ages and gender identities. Following processing, these samples are examined for pertinent acoustic characteristics like pitch, intensity, spectral content, and temporal patterns. Language factors including tone, speech pattern, and word usage may also be taken into account.Machine learning techniques are used to train models that can categorise voice samples based on age and gender once the features have been collected. Support Vector Machines (SVM), Random Forests, Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) are commonly employed algorithms for this task. The models are trained on a dataset with labeled samples to acquire the necessary knowledge.

The models are trained on a labeled dataset to understand the patterns and correlations between speech features and corresponding age and gender labels. During the testing or inference phase, new speech samples are inputted into the trained model, which then predicts the most probable age and gender group of the speaker. The training dataset's quality and representativeness, the efficiency of the feature extraction methods, and the machine learning algorithm of choice all affect how accurate the predictions turn out to be.It's crucial to keep in mind that the field of age and gender detection through speech is still developing, and despite major advancements, it still has problems. The precision of forecasts may be impacted by elements such changes in accent, dialect, linguistic ability, or the impact of emotional states. The robustness and generalisation abilities of the models are still being worked on in order to make speech-based age and gender detection more precise and dependable.

A method called age and gender detection by voice seeks to identify a person's age and gender from their speech traits. It falls within the larger category of biometric identification and voice processing. Algorithms can determine a person's age and gender by examining several acoustic aspects in their voice, such as pitch, intonation, and speech patterns.

In order to develop automatic and non-intrusive ways for detecting demographic information, age and gender recognition using speech has as its main objective. It has uses in market research, consumer profiling, voice-based user interfaces, and tailored user experiences, among other fields. The capability to recognize the age and gender of speakers based on speech is an invaluable tool with diverse applications across various fields where customized interactions and targeted communication are essential.

II. PROBLEM DESCRIPTION

In the current data-driven era, client demographic traits such as age, language, and gender hold significant value as they can aid businesses in enhancing their services and making informed decisions. While humans possess the ability to distinguish between male and female voices and estimate age based on available information, achieving comparable accuracy with computer code can present more complexities. Here, gender and age estimate using audio samples is the issue that we are seeking to solve.

Numerous applications of gender identification and age estimation have been demonstrated. By eliminating options for apparel and accessories, one can simplify the process of making purchases online. These forecasts have made it easier for emergency dispatch centres to determine the victim's gender. Criminal behaviour and legal processes will become more open and individualised. Computer systems may readily employ speech recognition for gender identification if the proper inputs and methods are used.

III. TECHNOLOGY USED

PYTHON: Python is a high-level, interpreted programming language that prioritizes code readability and uses significant whitespace. Its object-oriented approach and multiple programming paradigms make it easy to write code for various project sizes. Python has robust scientific computing tools like NumPy, Pandas, and SciPy that reduce coding needs and speed up development. Although Python allows pseudocode-like programming, its dynamically typed nature and ambiguous return types in documentation can lead to compatibility issues with packages and require more testing when using new libraries.

JUPYTER NOTEBOOK: A Jupyter Notebook file is a JSON document that follows a standardized schema and consists of a sequential arrangement of input/output cells. These cells can contain code, text, mathematical equations, plots, and multimedia elements. Typically, these files are identified by the ".ipynb" extension.

LIBROSA : It is a Python package for obtaining audio features. Both Soundfile and Audioread are used internally by the programme.

NUMPY : It is employed to obtain statistical features such as mean, median, and mode, as well as the Fourier transform of frequencies. This library is also used to calculate Skew and Kurtosis.

IV. SYSTEM ANALYSIS

As stated in the introduction, the suggested model follows the customary hierarchical structure, in which data is first entered, then pre-processed, fed to various machine learning models, split into training and testing datasets, and then evaluated.

The primary objective of this model is to determine the gender of a person based on various speech qualities when provided with input in CSV format. The model should be taught to categorise age into several age categories in addition to gender.

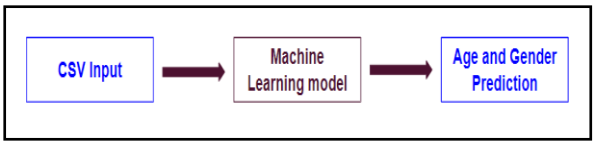

Fig.1 Representation of Functional Requirements

Datasets used : Age and gender have been identified via speech study utilising a variety of datasets. Here are a few datasets that are frequently used in this field:

1. TIMIT: The TIMIT dataset is a frequently utilised tool for research on speaker and speech recognition. It includesrecordings of native American English speakers from various locations, together with available demographic data like age and gender. Despite not being created with age and gender detection in mind particularly, it has been used as a standard dataset in related investigations.

2. VCTK Corpus: This multi-speaker collection contains speech recordings from 109 English speakers of varying ages and genders. It contains useful information for age- and gender-related analysis and covers a wide range of accents.

3. EmoDB: EmoDB is a database that specialises in emotive speech. It has a broad collection of emotional speech samples from male and female speakers of various age groups, despite not being specifically built for age and gender recognition, making it ideal for researching voice-related traits.

4. VoxCeleb: VoxCeleb is a sizable dataset made up of speech recordings made by thousands of famous people. Due to the available metadata, such as birthdates and gender labels, it can be used for age and gender analysis even if its primary purpose is speaker recognition research.

Model architecture : The first module of this model consists of data collection, visualisation, and labelling. The other two continuous modules are for gender prediction and age estimation. It also covers data transformation, such as scaling data to normalise it, and data cleaning using imputation techniques. Applying feature is the focus of the following module.

The project aims to predict gender by utilizing feature extraction techniques and evaluating various Machine Learning and Deep Learning models. The model that exhibits the highest accuracy on the test data is chosen and subjected to comprehensive analysis. In the subsequent phase, various age groups are categorized in the final module. To select the optimal model for age prediction, a comparative analysis is performed on various feature extraction techniques, highlighting their similarities and differences. Subsequently, the selected model undergoes testing and evaluation to determine its performance.

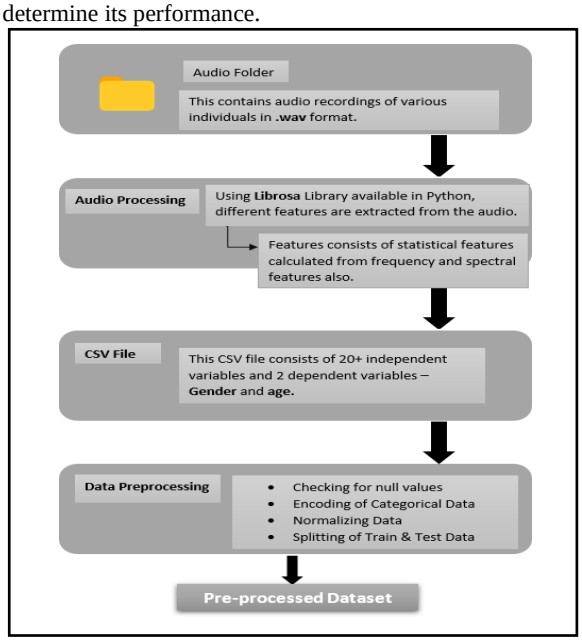

Fig. model architecture

Data Preparation :

In order to advance the project, we performed Audio Processing on the dataset to extract the necessary acoustic characteristics essential for the two models. As mentioned previously, the data for our models is sourced from three different origins, predominantly in the form of .wav files.

Data Visualization :

Using the BVC Gender & Age Estimation dataset, kernel density estimate plots were created to explore the association between the independent variables and the dependent variable (Gender - male and female).These visualizations assist in analyzing the associations between the variables.

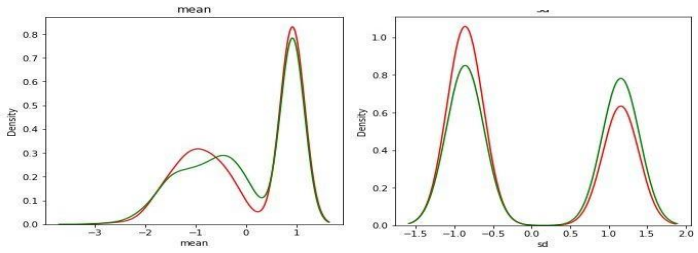

Fig. Mean of Frequencies Fig. Standard Deviation

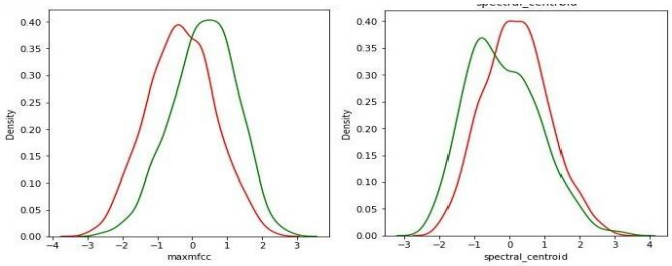

Fig. Max MFCCs Fig. Spectral Centroid

Data Preprocessing : The following preprocessing procedures were carried out by importing libraries like Pandas and Scikit-Learn (sklearn. preprocessing).

1. The Spyder environment is loaded with the dataset.
2. The dependent and independent variables are split apart.
3. The features were thoroughly examined to identify and remove any null values. If any null values were detected, they could be addressed and filled using imputation techniques.
4. The categorical data underwent coding, specifically for the Gender variable, where it was transformed into binary values of 0s and 1s within our dataset.
5. Testing and training data are then separated from the dataset.
6. Using the Feature Scaling Technique, the features are then scaled.
7. The dataset is now prepared for usage in the testing, evaluation, and training of several machine learning algorithms.

Gender Detection : In order to identify the optimal model for Gender Prediction, a comparative study was conducted using different classification machine learning models. The dataset underwent feature extraction and dimensionality reduction

Fig. Gender detection

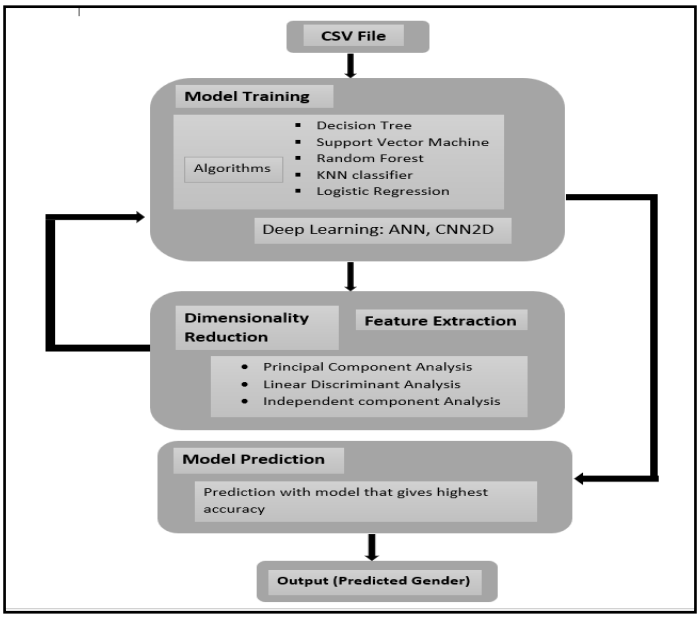

Fig. Gender detection

techniques, followed by the computation of accuracies. The employed Machine Learning Classifiers included Naïve Bayes, Decision Tree, Support Vector Machine, and Artificial Neural Networks.

Age Estimation : The "Common Voice" dataset, which contains speech data that website visitors have read, is the one utilised for age prediction. This dataset was generated by the Common Voice Team. Approximately 73,000 audio files were processed to extract audio features using a consistent formula. The extracted features consist of 22 measurements, including mean, median, mode of frequencies, skewness, kurtosis, tempo, mel, quantile 1 and 3, interquartile range, and spectral characteristics such as centroid, bandwidth, flatness, roll-off, and energy. Before delving into additional features, it is important to gain an understanding of the human voice and its traits that undergo modifications with age. One noticeable change in the voice is a significant decrease in pitch accompanied by vocal range alterations.

As individuals grow older, it is common for the spectral and temporal variability to decrease.

Mel Frequency Cepstral Coefficients (MFCCs), which take into account the characteristics that help one person's age from another, were determined for each audio recording.

The cepstrum is acquired by performing a Fourier analysis on the logarithm of the recorded signal spectrum. This technique is utilized to examine frequency spectra for periodic patterns. By incorporating the power cepstrum, the processing of human speech can better align with human auditory perception. The MFCCs (Mel Frequency Cepstral Coefficients) are the coefficients that form the MFCC representation of an audio file, which is derived from the cepstral properties of the audio.

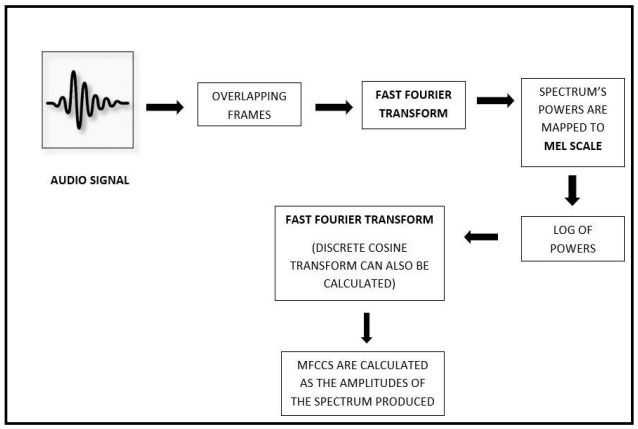

Fig. Age Estimation

V. CONCLUSION

An developing technology called age and gender detection by voice makes use of speech analysis to determine an individual's age and gender. Acquiring acoustic data from speech samples and training machine learning models enables the creation of automated and unobtrusive predictions regarding the demographic characteristics of speakers.

While age and gender recognition using voice holds immense potential in various fields such as market research, consumer profiling, and voice-based user interfaces, it is crucial to acknowledge and understand its limitations. Prediction accuracy can be impacted by variables such as accent, language, emotions, and individual differences. Therefore, it is essential to include this technology into a larger analysis and take other contextual information into account for a more thorough understanding.

VI. FUTURE SCOPE

The growing accessibility of huge datasets and the advancement of more complex machine learning algorithms have fueled the rapid advancement of age and gender detection in recent years. Future research and development in this field could go in a number of different ways, including:
Greater accuracy:There is always opportunity for improvement, even though the age and gender detection algorithms used today have quite high levels of accuracy. To further improve the accuracy of these systems, researchers can investigate novel machine learning approaches such as deep learning and ensemble methods.
Processing in real-Time: At the moment, the majority of age and gender detection algorithms take some time to analyse an image or video stream. Real-time processing capabilities would expand the possibilities for these devices, allowing for use in areas like security and surveillance.
Cross-cultural and demographic analysis: The majority of the age and gender recognition algorithms in use today are built on data sets gathered from certain populations, including those in western nations. In order to reliably identify age and gender across various cultural and demographic groupings, including non-binary and transgender people, algorithms are required.
Multi-modal age and gender detection: Sources such as voice, body type, and facial traits can all be used to determine an individual's age and gender. Combining these many modalities could result in detection algorithms that are more precise and reliable.

VII. REFERENCES

1. *Schuller, B.; Steidl, S.; Batliner, A.; Burkhardt, F.; Devillers, L.; MüLler, C.; Narayanan, S. Paralinguistics in speech andlanguage—State-of-the-art and the challenge. Comput. Speech Lang.* ***2013****, 27, 4–39. [CrossRef]*
2. *Panek, D.; Skalski, A.; Gajda, J.; Tadeusiewicz, R. Acoustic analysis assessment in speech pathology detection. Int. J. Appl. Math.Comput. Sci.* ***2015****, 25, 631–643. [CrossRef]*
3. *Techmo. Available online: https://www.techmo.pl (accessed on 12 February 2021).*
4. *Zazo, R.; Sankar Nidadavolu, P.; Chen, N.; Gonzalez-Rodriguez, J.; Dehak, N. Age Estimation in Short Speech Utterances Based on LSTM Recurrent Neural Networks. IEEE Access* ***2018****, 6, 22524–22530. [CrossRef]*
5. *Mahmoodi, D.; Marvi, H.; Taghizadeh, M.; Soleimani, A.; Razzazi, F.; Mahmoodi, M. Age Estimation Based on Speech Features and Support Vector Machine. In Proceedings of the 2011 3rd Computer Science and Electronic Engineering Conference (CEEC), Colchester, UK, 13–14 July 2011; pp. 60–64.*
6. *Dehak, N.; Kenny, P.J.; Dehak, R.; Dumouchel, P.;*

Ouellet, P. Front-End Factor Analysis for Speaker Verification. IEEE Trans. Audio Speech Lang. ***2011****, 19, 788–798. [CrossRef]*

7. *Villalba, J.; Chen, N.; Snyder, D.; Garcia-Romero, D.; McCree, A.; Sell, G.; Borgstrom, J.; Richardson, F.; Shon, S.; Grondin, F.; et al. State-of-the-Art Speaker Recognition for Telephone and Video Speech: The JHU-MIT Submission for NIST SRE18. In Proceedings of the INTERSPEECH 2019, Graz, Austria, 15–19 September 2019; pp. 1488–1492.*
8. *Snyder, D.; Garcia-Romero, D.; Sell, G.; Povey, D.; Khudanpur, S. X-vectors: Robust dnn embeddings for speaker recognition. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5329–5333.*
9. *McLaren, M.; Lawson, A.; Ferrer, L.; Castan, D.; Graciarena, M. The speakers in the wild speaker recognition challenge plan. In Proceedings of the Interspeech 2016 Special Session, San Francisco, CA, USA, 8–12 September 2015.*
10. *Wan, L.; Wang, Q.; Papir, A.; Moreno, I.L. Generalized end-to-end loss for speaker verification. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4879–4883.*
11. *Jasuja, L.; Rasool, A.; Hajela, G. Voice Gender Recognizer Recognition of Gender from Voice using Deep Neural Net- works. In Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India,10–12 September 2020; pp. 319–324.*
12. *Djemili, R.; Bourouba, H.; Korba, M.C.A. A speech signal based gender identification system using four classifiers. In Proceedings of the 2012 International Conference on Multimedia Computing and Systems, Tangiers, Morocco, 10–12 May 2012, pp. 184–187.*