Grenze International Journal of Engineering and Technology, June Issue

SoundScope: Human Voice Analysis for Gender Classification

Ashika B S[1] and Surendra Shetty[2]
[1-2]NITTE (Deemed to be University) Department of Master of Computer Applications NMAM Institute of Technology, (NMAMIT) Nitte, 574110, Karnataka, India
Email: Ashikagowda047@gmail.com, hsshetty@nitte.edu.in

***Abstract*— Identifying gender from voice has become an important step toward improving personalized user experiences, especially in applications such as virtual assistants, customer service systems, and voice-based authentication. Traditional manual voice pattern analysis is slow, inconsistent, and unworkable for real-world time usage. This work proposes a voice-based automatic gender classification system, one that uses machine learning techniques to identify whether a given speech sample originates from a male or female speaker. The study initially used some simple learning approaches, but they were not suitable for reliable decision-making due to their low performance. A more robust model was necessary; thus, it was trained on a large set with diverse characteristics in terms of voice samples. The resulting system instantly predicts the gender while providing probability scores, reflecting confidence in the model. The user-friendly web application was built that allowed users either to upload audio files or to record voice directly for immediate analysis. The proposed system proved efficient, practical, and well-suited for real-time voice-based classification tasks.**

***Index Terms*— voice classification, gender detection, machine learning, speech analysis, audio processing, web application, real-time prediction.**

I. INTRODUCTION

The human voice is considered one of the most natural and expressive modalities of human communication. It carries not only linguistic information but also personal characteristics like age, emotion, and gender. Due to the growing interest in recent times in voice-enabled technologies, such as smart assistants, automated customer support, and security systems, the need to develop systems capable of understanding and analyzing voice patterns has grown by leaps and bounds. Among these tasks, identifying the gender from speech has also become very crucial since it can provide more personalized and adaptive interactions.

Traditional gender identification methods often involve listening manually or using simple acoustic measures, which, for large volumes of audio material, can be extremely time-consuming and subjective. Moreover, significant variations in pitch, tone, recording environment, and background noise make manual classification even more difficult. These limitations indicate the need for an automated method that will classify the gender rapidly and accurately without relying on human judgment.

Recent breakthroughs in machine learning have made the analysis of, and learning from, complex voice patterns possible. A model in machine learning can automatically learn the key differences between male and female voices by training against big data sets. characteristics and differences between them.

Grenze ID: 01.GIJET.12.2.583

Such systems can make quick predictions and maintain consistency regardless of variations in the audio input. This opens the door to smarter applications that can adapt to users in real time.

This project deals with the development of a practical and efficient voice-based gender recognition system. It will contain a trained machine learning model and a web-based interface where users can record or upload audio for instant analysis. The aim is to develop a lightweight system that is user-friendly, very accurate, and suitable for day-to-day applications that require voice-based identification. The proposed system illustrates how simple recordings of voice can be transformed into meaningful insights through machine learning, thereby making human-computer interaction smarter.

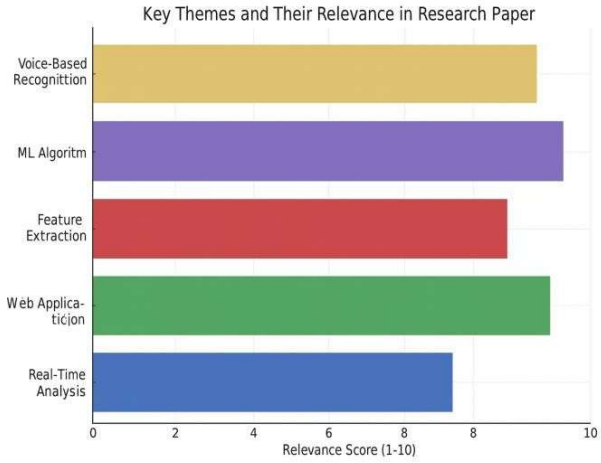

Figure 1: Key themes and relevance

II. Literature Review

Automatic gender classification from speech has also gained much attention due to its potential applications in biometric authentication, human computer interaction, forensic analysis, and voice-controlled systems. The first attempts at automatic gender classification from speech were based on manual extraction of acoustic parameters such as pitch, formants, spectral centroid, jitter, and shimmer, along with traditional machine learning classifiers. Although these systems were able to provide reasonable results in a controlled environment, they were not robust to noise, speaker variability, and environmental conditions [1]–[3].

Traditional machine learning algorithms like Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Gaussian Mixture Models (GMM), and Random Forests (RF) have been widely investigated for voice-based gender classification tasks. Several studies have shown that the MFCC-based feature extraction process remains effective because of its capability to extract spectral features useful for distinguishing the vocal tracts of males and females [6], [9], [12]. But these algorithms are known to be sensitive to feature engineering and tend to degrade in performance in real-world applications [7], [13].

However, recent advances have led to an increasing trend towards deep learning-based methods, especially Convolutional Neural Networks (CNNs), which are capable of automatically learning discriminative features from speech data like MFCCs and Mel spectrograms. Various comparative analyses have demonstrated that CNN-based models outperform than conventional classifiers like SVM, KNN, and RF in terms of accuracy and generalization ability [4], [5], [18]. Ensemble and hybrid learning methods have also been explored to further enhance robustness by using multiple classifiers or feature combinations [5], [8].

Multi-task learning models have been proposed to simultaneously predict gender and other related attributes like age and emotion, which facilitate shared learning of features and enhance generalization performance [4], [11], [16].

Transformer models and attention mechanisms have also been explored, which have shown better performance in challenging acoustic conditions because of their capacity to capture long-term temporal dependencies in speech signals [15], [21].

Several studies have focused on gender recognition under real-world conditions like noisy channels, mobile phone recordings, and forensic audio signals. Machine learning-based systems have demonstrated robust performance on real-world databases, establishing their effectiveness for real-world applications [20], [28].

Moreover, feature optimization and metaheuristic algorithms have been used to improve the efficiency and accuracy of classifiers [19], [27].

To facilitate real-time and resource-constrained applications, light-weight neural networks and low-complexity machine learning models have been proposed. These models demonstrate balanced accuracy and low computational complexity, making them suitable for embedded systems and web-based applications [7], [13], [22]. Streaming and multi-speaker scenarios have also been investigated using temporal convolutional networks and speaker change detection algorithms to facilitate continuous gender classification [25], [26].

Recent work has also highlighted the importance of privacy-preserving voice biometrics. Methods like adversarial auto-encoders and differential privacy algorithms have been proposed to preserve sensitive gender information while maintaining system performance [29]. Acoustic analysis studies further validate that frequency- and spectrum-based features remain the most prominent gender and age classification features due to physiological differences in human vocal anatomy [30].

III. Methodology

This work is designed to follow a structured workflow for recognizing voice gender, right from dataset preparation and preprocessing to model development, training, evaluation, and deployment. In essence, the main objective is to develop an accurate system that will be able to classify a speaker's gender from audio recordings. The overall design emphasizes high prediction accuracy while keeping the model lightweight enough for real-time deployment on web-based platforms. The methodology combines classical feature extraction techniques with machine learning and deep learning models to optimize performance.

First of all, the dataset of voice samples needed to be constructed. Publicly available datasets were used, containing samples of male and female speakers with diverse age, accent, and recording conditions. Each of these audio files was listened to manually, removing corrupted or too noisy samples.

We first preprocessed this dataset, standardizing the sampling rates, normalizing the amplitudes, and trimming the silences at the very beginning and end of recordings. This preprocessing pipeline would guarantee that the input data is consistent and of good quality for feature extraction and model training.

The feature extraction was done mainly via MFCCs, which embody the necessary spectral and timbral characteristics of speech required for the purposes of gender differentiation. Every audio sample was summarized into one fixed-length feature vector based on MFCCs, and extra features such as pitch and spectral centroid could optionally be added to increase discrimination.

In the case of classical machine learning, these features were standardized and then reduced using PCA so that the redundancy would be removed and computational load reduced. This resulted in a compact representation of each audio sample that was quite suitable for classifier training like Random Forests and Support Vector Machines.

We then used a deep learning model as the main predictor to further improve performance. First, a simple feedforward neural network was used as a baseline, but it had limitations in capturing complex temporal patterns in speech. A more complex architecture using convolutional layers to extract spectral patterns combined with fully connected layers for classification was then employed. The network was trained on the extracted features using a training/testing split and with hyperparameter tuning for optimal accuracy. To validate the performance, accuracy, precision, recall, and confusion matrices were calculated for the performance measures to ensure that the model generalizes well to new, unseen samples. Finally, the trained model was deployed using a web application based on Flask that enables recording or uploading audio files and making real-time predictions of gender. The steps involved in the deployment pipeline are exactly similar to the data preprocessing and feature extraction stages of the training process. Predictions will be provided with confidence probabilities regarding male and female classes. Such a setup therefore ensures an end-to-end voice gender recognition system accessible over a web interface without requiring any special hardware, thus being suitable for practical applications in voice-controlled devices, biometric authentication, and speech analysis systems.

IV. Dataset Acknowledgment

This study makes use of an openly available voice dataset from Kaggle, featuring audio recordings from male and female speakers. Original recordings were included in the sample set of this dataset, with varying durations, qualities of recording, and speaker accents. For this research, Ashika B S took extra care to scrub through and arrange the dataset so that both genders are represented equally. The clarity and consistency of each audio were

checked manually; corrupted and noisy files beyond repair were discarded to come up with a uniform set for training models.
The processed dataset then served as the basis for training and testing of the voice gender recognition system. Preprocessing steps involved resampling, normalization, and trimming silent segments of the samples. This restructured dataset allowed the model to learn meaningful patterns from real-time audio inputs, which thus enabled it to correctly classify male and female voices.
The proposed system developed by Ashika B S uses the curated voice recordings from the Kaggle dataset and identifies the gender in real time with high accuracy and reliability. This dataset forms the very foundation for feature extraction, training, evaluation, and deployment of this system.

V. Dataset Description

In this work, the voice-based gender recognition study utilized a publicly available Kaggle dataset that had been curated for voice-based gender recognition. The original dataset contained several thousand voice recordings from male and female speakers, with acoustic features pre-extracted. As the dataset was already organized into two balanced classes, there was no need to restructure class labels. The dataset therefore comprised 3,168 samples, where 1,584 represented males and 1,584 represented females, thereby guaranteeing balanced classes with no biases in model training.
Each data instance is represented by 20 engineered acoustic features, including commonly used speech descriptors such as mean frequency, standard deviation, spectral centroid, jitter, shimmer, and other frequency and amplitude-based parameters. These features capture essential vocal characteristics, making this dataset highly suitable for traditional machine-learning approaches.

Table I: Summary of the Dataset used

Feature	Count
Total Voice Samples	3168
Male Voice Samples	1584
Female Voice Samples	1584

VI. System Implementation

The voice-gender recognition system proposed here transforms a trained machine-learning model into a full-fledged web application. It can make real-time predictions based on the audio or some other input through browsers. As compared to a single model in a Jupyter notebook, the system enables users to record their voice or directly upload an audio file from a browser; the application will then predict instantly whether that voice is from a male or a female. This includes the preparation of the trained model, creation of the backend using Flask, designing of the frontend interface, and putting it all together into one smooth and responsive system.
The MFCC-based feature extraction model that was trained during experimentation had been exported to .pkl format along with the scaler object and label encoder used during training. These are the "brains" of the system. Every time the Flask application starts, it loads these trained components into memory so that future predictions are much faster and more stable.
The system's backend was implemented in Python, using the Flask microframework due to its simplicity and lightweight structure, which efficiently handles HTTP requests. It loads the model, scaler, and label encoder once when the application initializes to avoid repeated overhead. There are two major routes in this Flask application: one to show the main webpage interface and another to receive the uploaded audio file or recorded audio blob for prediction. This keeps the application clean and modular, ensuring that server responses remain fast.
This process at the backend performs the same sequence of preprocessing steps carried out during training, whenever there is a recording of the user's voice or an uploaded audio file. The audio data is converted to a numerical signal and then MFCC features are extracted. The extracted MFCC values are then scaled using the same StandardScaler fitted during training in order to keep it consistent. The preprocessed feature vector is fed into the trained classifier that provides probability values for both classes: male and female. The class with the highest probability gives the final predicted label, whereas the probability score helps with representation of the model's confidence.
The system frontend is built with HTML, CSS, and JavaScript, ensuring it works on any modern browser. The design is kept simple and intuitive for user interaction. It has two options: upload an audio file or record their voice directly from the device's microphone. JavaScript handles the access to the microphone via Media

Recorder API and sends the recorded audio to the Flask server where the recording will be analysed. Upon prediction, the browser displays the predicted gender, model accuracy, and a probability bar chart so that it is easier to visualize and understand. This integration of the backend with the frontend completes the system, and hence it works as a unified web application. The communication between the browser and the Flask server is fast and seamless: the browser sends the audio data to Flask, which sends it through the MFCC pipeline for processing, followed by the evaluation by the trained model. The result is sent out instantly. Although the classifier is lightweight, the system can work on regular laptops without support for GPU. This successful implementation shows how traditional machine-learning techniques combined with web technologies can be used to build practical, real-time voice-based recognition systems suitable for smart applications, authentication systems, and interactive voice-driven platforms.

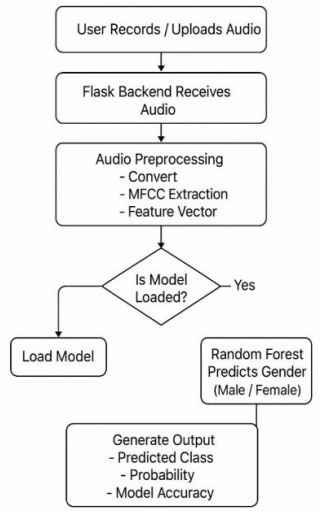

Figure 3: Architecture and workflow of the voice gender recognition system

VII. Results and Discussion

We presented the proposed voice gender recognition system, which was evaluated using the test portion of the curated dataset consisting of a total of 3168 voice samples from 1584 male and 1584 female speakers. The Random Forest classifier was trained and optimized on binary gender prediction using the preprocessed audio samples and extracted feature vectors based on MFCCs.

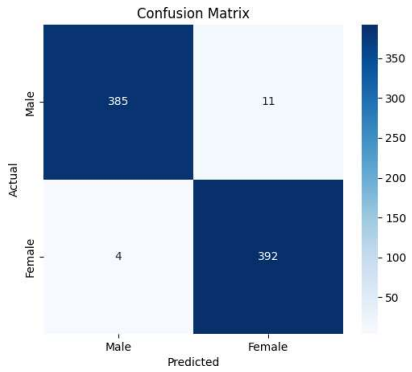

Figure 4: Confusion Matrix for the Two-Class Voice Gender Classification Model

The model showed excellent generalization capability, with an accuracy of 97.64% on training and 98.11% on the test set, which implies that not only does the classifier learn well from the training distribution, but it also generalizes well on previously unseen data. In addition, for each prediction, a probability score was developed so that users could see the confidence level of the model in each classification.

A confusion matrix was thus constructed to further investigate class-wise performance and is presented in Figure 4. The diagonal values show the correct predictions made by the model, where the classifier correctly identified 385 male and 392 female voice samples in the test set with a perfectly balanced prediction pattern across both categories. This highlights the strength of MFCC features in capturing characteristics based on frequency, which would naturally differ between male and female voice ranges.

There were a few misclassifications, which are bound to happen with real-world audio data because of overlapping vocal features. Some female voices with deeper pitches and some male voices with higher-frequency components sometimes resulted in incorrect predictions. This is because pitch, tone, and harmonic structures at times can resemble those of the opposite sex, especially in short clips or noisier environments. Despite that, the share of errors remained minimal to confirm the stability of the model.

Besides accuracy, it also outputs real-time confidence scores for every prediction, which helps in understanding how sure the classifier is about its decision. In addition, the lightweight Random Forest model with a small feature set ensures that the inference happens quickly, allowing the system to work smoothly even without GPU support.

In all, the results further substantiate that the proposed system is highly effective for voice-signal gender classification. Its excellent accuracy, coupled with fast prediction times and reliable confidence estimations, extends its suitability to practical applications like speech-based authentication, voice analytics, and human–computer interaction systems.

Moreover, the balanced class distribution and consistent performance across both gender categories indicate that the model does not exhibit class bias. The integration of MFCC-based features with an ensemble classifier such as Random Forest contributes to improved robustness and interpretability, enabling the analysis of key acoustic features and their influence on classification outcomes. These findings demonstrate that the proposed approach achieves an effective balance between classification accuracy, computational complexity, and practical deployability.

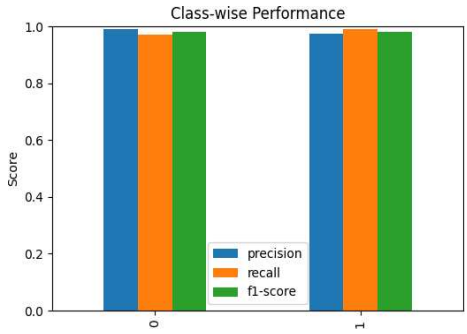

Figure 5: Class-wise Precision, Recall, and F1-score for the Voice Gender Recognition System

Figure 5 shows the class-wise performance analysis of the proposed voice gender recognition system using precision, recall, and F1-score. The performance analysis of the proposed system clearly shows that the proposed system is performing well for both male and female classes. The precision, recall, and F1-score values of both classes are almost equal. This shows that the proposed system is unbiased for both male and female voices. This clearly shows that the proposed system is efficient for male and female voice recognition.

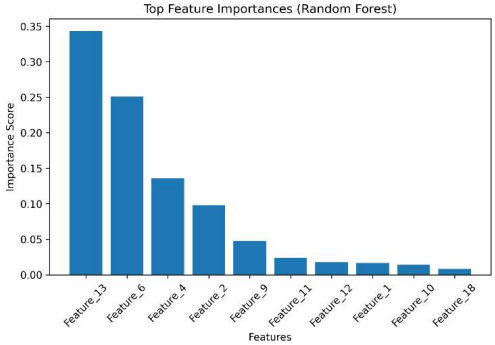

Figure 6: Feature Importance Analysis of Acoustic Features for Voice Gender Classification

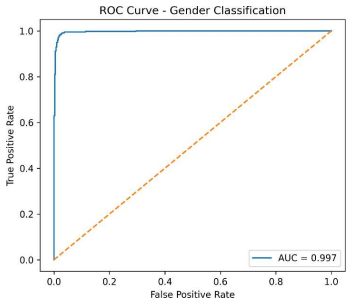

Figure 7: Receiver Operating Characteristic (ROC) Curve for the Two-Class Voice Gender Classification Model

The result of the feature importance analysis of the trained Random Forest classifier is presented in Figure 6. It is revealed that only a few acoustic features play an important role in making a decision on gender classification. The most important features are mainly related to the frequency and spectral characteristics of the speech signal. This is in line with the fact that the vocal tracts of males and females are different biologically.

Figure 7 depicts the Receiver Operating Characteristic (ROC) curve of the proposed Random Forest-based voice gender classification model. The ROC curve is a graphical representation of the trade-off between the true positive rate and the false positive rate for a given classification model at different thresholds. The Area Under the Curve (AUC) value of 0.997 in the ROC curve indicates that the proposed model has a high discriminative power. This clearly indicates that the classifier is able to classify the male and female voice samples with high accuracy.

VIII. Conclusion and Future Work

The voice-assisted gender classification system proposed in this study demonstrates that a lightweight machine learning–based solution can achieve highly accurate classification performance. By carefully preprocessing audio samples and employing MFCC-based acoustic eatures, the system accurately distinguishes between male and female voices.

The final classifier achieved an accuracy of approximately 98%, indicating that traditional machine learning models, when combined with well-engineered speech features, can deliver performance comparable to deep learning approaches while maintaining significantly lower computational complexity.

The proposed framework is further enhanced through integration with a Flask-based web application, which enables users to upload audio samples or record speech in real time for instant gender classification. The system processes the input audio, extracts relevant features, and provides classification results along with confidence probability scores. This end-to-end implementation makes the solution suitable for real-time applications such as voice-assisted authentication, call-center automation, smart home systems, and assistive technologies.

In conclusion, this research validates the effectiveness of MFCC-based acoustic features for speech-based gender classification and demonstrates the practical applicability of the proposed system in real-world scenarios. The lightweight architecture, fast processing capability, and high classification accuracy make the system suitable for deployment in real-time and resource-constrained environments.

Future Work

The proposed system can be further enhanced by incorporating advanced deep learning architectures such as convolutional neural networks and transformer-based models to improve robustness and accuracy, particularly under noisy acoustic conditions.

Multi-task learning frameworks may also be explored to simultaneously predict additional speaker attributes such as age and emotion.

Furthermore, extending the system to support multilingual datasets, improving noise resilience, and adapting the framework for mobile and embedded platforms would significantly broaden its applicability in real-world applications.

References

[1] E. Yücesoy, “Gender Recognition Based on the Stacking of Different Acoustic Features,” Appl. Sci., vol. 14, no. 15, Art. no. 6564, Jul. 2024.

[2] M. Tummala, L. Harish, M. E. Malkhed, S. S. Kumar, N. Neelima & V. Venugopal, "Optimizing Gender Identification with MFCC Feature Engineering and Enhanced Gradient Boosting," in 2024 Asian Conference on Intelligent Technologies (ACOIT), IEEE, 2024.
[3] K. Akgün & Ş. A. Sadık, "Unified voice analysis: speaker recognition, age group and gender estimation using spectral features and machine learning classifiers," J. Scientific Reports-A, issue 057, pp. 12–26, Jun. 2024.
[4] "Identity, Gender, Age, and Emotion Recognition from Speaker Voice with Multi-task Deep Networks for Cognitive Robotics," Cognitive Comput., vol. 16, pp. 2713–2723, 2024.
[5] "Automatic Age and Gender Recognition Using Ensemble Learning," Appl. Sci., vol. 14, no. 16, 6868, Aug. 2024.
[6] H. Li, Y. Zhou, R. Zhang, and S. K. Singh, "A Voice-Based Gender and Age Recognition System on the TIMIT Corpus," in Proc. IEEE Int. Conf. Speech and Signal Processing, 2025, pp. xx–xx..
[7] S. Ghosh, C. Saha & N. Molakathaala, "NeuraGen – A Low-Resource Neural Network based Approach for Gender Classification," arXiv preprint, Mar. 2022.
[8] "Voice (Gender) Detection from Supervised Learning Algorithm using MFCC in Bengali Language," B. Akther et al., BSc Thesis, MIST, Feb. 2023.Z. Yang, "Deep learning-based garbage classification system," Mathematical Biosciences and Engineering, 2023.
[9] "Gender Identification Via Voice Analysis," S. Kushwah, S. Singh, K. Vats & V. Nemade, CSEIT, 2023.
[10] A. Weizman, Y. Ben-Shimol & I. Lapidot, "Tandem spoofing-robust automatic speaker verification based on time-domain embeddings," arXiv preprint, Dec. 2024.
[11] G. Tirupati, K. V. S. Sirisha, M. Vijaya, K. Naga Sai Swetha & G. Aruna, "Interactive System for Gender Classification," in IJRASET Journal for Research in Applied Science and Engineering Technology, 2023.
[12] S. Hızlısoy, E. Çolakoğlu & R. S. Arslan, "Speech-to-Gender Recognition Based on Machine Learning Algorithms," International Journal of Applied Mathematics, Electronics and Computers, vol. 10, no. 4, pp. 84–92, Dec. 2022.
[13] D. Koshtura, "Information Technology for Gender Recognition by Voice," SISN, vol. 13, pp. 350–360, 2023.
[14] H. A. M. Alsalaet, "Gender Recognition from Analysis of Speech Signals Using GMM Machine Learning Algorithm," JAIET, vol. 1, no. 1, pp. 33–38, Nov. 2024.
[15] F. Burkhardt, J. Wagner, H. Wierstorf, F. Eyben & B. Schuller, "Speech-based Age and Gender Prediction with Transformers," arXiv preprint, Jun. 2023.
[16] S. Mavaddati and A. Rahmani, "Deep learning for voice-based age, gender, and language recognition," Neurocomputing, vol. 580, pp. 127429, 2024
[17] U. Ijaz, M. M. Iqbal, Z. S. Ali, A. Tariq, and R. Khan, "Voice-Based Gender Identification Using Machine Learning," Journal of Computing and Biomedical Informatics, vol. 9, no. 1, 2025.
[18] A. Kumar, R. Sharma, and P. Singh, "Gender Recognition from Speech Signal Using CNN, KNN, SVM and RF," Procedia Computer Science, vol. 235, pp. 2251–2257, 2024.
[19] Ş. Yıldırım and M. S. Bingöl, "Metaheuristic optimization approaches for voice-based gender identification," Applied Sciences, vol. 15, no. 23, Art. no. 12815, 2025.
[20] G. Gouri, A. Sharma, and V. Sharma, "Forensic speaker and gender identification from voice samples recorded through mobile phones and social media applications: A statistical and machine learning approach," Applied Acoustics, vol. 222, Art. no. 110074, 2024.
[21] A. Singhal and D. K. Sharma, "Comprehensive Analysis of Gender Classification Accuracy across Varied Geographic Regions through the Application of Deep Learning Algorithms to Speech Signals," Computer Systems Science and Engineering, vol. 48, no. 3, pp. 609–625, 2024.
[22] V. Vysotska, D. Shavaiev, M. Gregus, Y. Ushenko, Z. Hu and D. Uhryn, "Information Technology for Gender Voice Recognition Based on Machine Learning Methods," International Journal of Modern Education and Computer Science, vol. 16, no. 5, pp. 65–87, 2024.
[23] E. Yücesoy, "Gender Recognition from Speech Signals Using CNN, SVM, KNN and Random Forest," Procedia Computer Science, vol. 235, pp. 2251–2257, 2024. [24] A. A. Mohammed and Y. F. Al-Irhayim, "Speaker Age and Gender Estimation Based on BiLSTM Deep Learning Models," Tikrit Journal of Pure Science, vol. 29, no. 2, pp. 88–96, 2024.
[24] T. Nguyen and M. Lee, "Streaming Speaker Change Detection and Gender Classification in Multi-Speaker Environments," arXiv preprint arXiv:2502.02683, 2025.
[25] A. Hassan, M. El-Sayed and K. Al-Khalifa, "Age and Gender Classification from Speech Using Temporal Convolutional Networks," Multimedia Tools and Applications, vol. 81, pp. 21467–21485, 2022.
[26] N. Patel and D. Shah, "Feature Optimization for Speech-Based Gender Recognition Using Hybrid ML Models," Expert Systems with Applications, vol. 228, Art. no. 120249, 2023.
[27] P. Rodrigues, L. Silva and A. Teixeira, "Robust Gender Classification from Speech Under Real-World Noise Conditions," IEEE Access, vol. 11, pp. 104321–104332, 2023.
[28] O. Chouchane, M. Panariello, O. Zari, I. Kerenciler, I. Chihaoui, M. Todisco and M. Önen, "Differentially Private Adversarial Auto-Encoder to Protect Gender in Voice Biometrics," arXiv preprint arXiv:2307.02135, 2023.
[29] S. M. Reza, H. H. Ozkan, A. S. Ghadiri and N. El-Mekki, "Acoustic Analysis of Speech for Gender and Age Classification," ACM Transactions on Intelligent Systems and Technology, vol. 15, no. 4, 2024.