Grenze International Journal of Engineering and Technology, January Issue

Echoes of Gender: Unveiling Voice Identity with Neural Networks

Dr. Adilakshmi Yannam[1], Dr.Shaik Salma Begum[2], Mr. Jami Sathvik[3], Ms. Javvadi Tulasi[4], Mr. K. Doondi Subhash[5] and Mr. V. Vigneswara Rao[6]
[1, 3-6]Dept. of CSE (AI & ML), S R Gudlavalleru Engineering College, Gudlavalleru, India
Email: laxmi072003@gmail.com, sathviksacchu@gmail.com, tulasijavvadi7@gmail.com, doondisubhash06@gmail.com, vvigneswararao0@gmail.com
[2]Dept. of CSE, S R Gudlavalleru Engineering College, Gudlavalleru, India
Email: shaiksalma.gec@gmail.com

***Abstract*—The broad applications of voice-based gender detection in fields including security systems, sociolinguistic research, and human-computer interaction have attracted a lot of attention in recent years. The objective of this work is to investigate the viability and efficiency of classifying gender using only audio data and neural network techniques. We do this by using deep learning techniques to identify gender-signalizing patterns in a neural network trained on a variety of datasets containing voices, both male and female. We show the effectiveness of our approach in reliably guessing gender from voice recordings through testing and assessment. We also address the possible uses, difficulties, and ramifications of voice-based gender recognition technology, emphasizing its importance in the modern world.**

***Index Terms*— Mel Frequency Cepstral Coefficients (MFCC), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Cross-cultural Analysis, Delta MFCC's, Train accuracy and Test Accuracy.**

I. Introduction

The human voice is a versatile communication tool that conveys socio-cultural indicators, including gender identity and linguistic information. Teaching machines to understand these nuanced signs is a challenge in the fields of languages, social psychology, and artificial intelligence. Recent advancements in machine learning techniques, particularly neural networks, have led to breakthroughs in voice-based gender recognition. These advanced algorithms can analyze vast information and identify meaningful patterns, inspired by the structure and function of the human brain.

This study aims to leverage the powers of neural networks to discern gender from speech recordings. The importance of this work extends beyond technological innovation, touching on issues of identity, representation, and social norms. Voice-based gender recognition has the potential to challenge binary gender constructs and foster a deeper appreciation for the diverse spectrum of voice identities.

The research adopts a comprehensive approach, aiming to develop reliable and accurate algorithms while engaging in critical discussions about the moral, societal, and cultural implications of gender recognition technology. By addressing these broader considerations, the researchers aim to contribute to a more inclusive and equitable technological landscape that respects the complexities of human identity and expression.

Grenze ID: 01.GIJET.11.1.376

The development and implementation of voice-based gender recognition technology involve ethical considerations and societal obligations. The research emphasizes the significance of openness, accountability, and inclusivity in algorithmic decision-making. The goal is to develop inclusive technologies that honor and enhance the diversity of human expression, bridging gender, cultural, and identity boundaries.

II. Related Works

Neural network classifiers are sophisticated machine learning models designed to categorize input data into distinct groups, mirroring the structure and functionality of the human brain Chachadi et al. (2022) [1]. Comprised of interconnected nodes or neurons organized into layers, neural networks process raw data through input layers Uddin et al. (2022) [2], gradually transforming it through one or more hidden layers to extract relevant features for classification. Activation functions such as ReLU, tanh, or sigmoid introduce nonlinearity to the model Adilakshmi.Y et al (2020) [14], enabling it to capture complex patterns in the data S. S. Begum et al. (2019) [3]. With their ability to learn and adapt to data structure without explicit feature engineering, neural network classifiers excel in handling unstructured data like text, audio, and images Fahmeeda et al.(2022) [4].
Multiple hidden layer deep neural networks have proven highly effective in learning hierarchical representations of complex data S. S. Begum et al. (2020) [16], making them particularly suitable for tasks involving the extraction and classification of high-dimensional features S. S. Begum et al.(2020) [5]. One key advantage of neural network classifiers lies in their capacity to automatically adjust to the underlying data structure, obviating the need for manual feature engineering Alnuaim et al. (2022) [6]. The selection of optimal attributes plays a crucial role in enhancing the performance of voice gender recognition models, with popular optimizers like the wolf, PSO Adilakshmi.Y et al (2020) [15], and EA techniques employed as wrapper selection algorithms Nirmala et al. (2023) [7]. These approaches leverage natural behaviour to explore potential solutions and identify the most effective ones for use in classifier evaluation MIZRAK et al. (2023) [8].
The quality of characteristics in the training data significantly impacts the performance of classification algorithms Kannapiran et al. (2023) [9]. Therefore, meticulous attribute selection is essential for refining voice gender recognition models Di Cesare et al. (2024) [10]. By employing wrapper selection algorithms, such as the wolf, PSO, and EA techniques, informed decisions can be made regarding feature inclusion, optimizing the model's efficacy M. A. Hossain et al. (2020) [11]. These methods harness natural behaviours to navigate the solution space, identifying attributes that enhance classifier performance, as demonstrated by the RF evaluator Vijayan, B et al. (2019) [12]. This iterative approach ensures that the model captures the most relevant information for accurate gender classification, Gupta, P et al. (2018) [13], thereby improving overall performance and applicability in real-world scenarios.

III. Our Methods

The vocal signal contains essential features for effective gender classification, such as fundamental frequency, formant frequencies, energy distribution, and spectral characteristics like Mel Frequency Cepstral Coefficients (MFCCs) and Delta MFCCs. To extract these features, a preprocessing step is necessary. This involves normalizing the audio data to account for recording levels and background noise, segmenting the signal into frames, and feeding it into feature extraction techniques. These techniques transform the raw audio into structured features that can be analyzed by machine learning models.
The training set for a classifier model identifies the speaker's gender using these extracted feature sets and corresponding voice gender labels. The implementation uses .py files instead of .ipynb files, enhancing modularity and reusability, improving development efficiency and making the codebase easier to maintain and scale.

A. Data Manager

The algorithm handles a gender recognition dataset by performing several key operations, including extraction, organization, and splitting into training and testing sets. It incorporates functions such as file extraction from compressed archives, creation of necessary folders, transfer of files to appropriate directories, and segmentation of the dataset based on gender labels. To streamline these processes, a DataManager class is employed. This class efficiently manages the dataset preparation workflow, ensuring that the data is well-organized and ready for machine learning tasks. By automating these steps, the DataManager class facilitates effective data handling, allowing researchers to focus on model development and analysis. This systematic approach enhances the overall efficiency and accuracy of gender recognition systems.

B. Feature Extraction

The FeaturesExtractor class streamlines the process of extracting voice features, particularly Mel Frequency Cepstral Coefficients (MFCC), from audio files using the python_speech_features package. It allows for precise configuration of the feature extraction process through parameters such as analysis window length, step size, number of cepstrum coefficients, number of filters in the filterbank, and FFT size. Additionally, the class applies Cepstral Mean Normalization (CMS) to the MFCC features and incorporates delta and double delta features to enhance the feature representation. The resulting feature matrix is then processed to retain a specific number of features. By encapsulating these functionalities, the FeaturesExtractor class offers a straightforward interface for extracting standardized voice features. This makes it particularly useful for various machine learning applications, including gender recognition, speech recognition, and speaker identification, by ensuring consistent and reliable feature extraction.

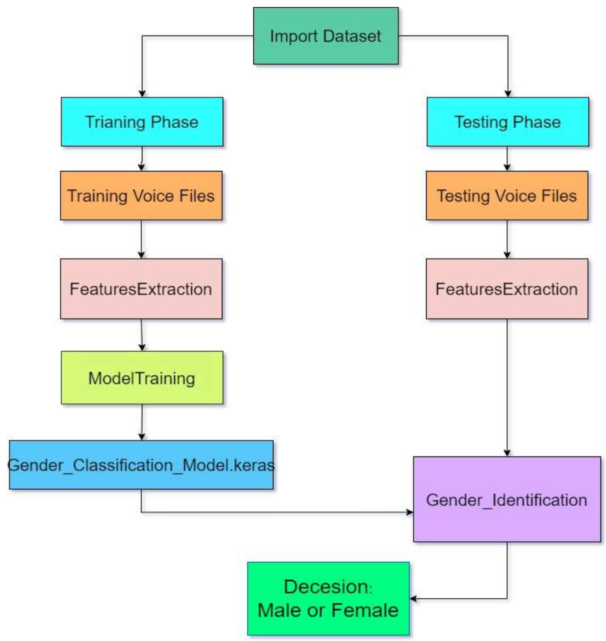

Figure 1: Architecture Flow

C. Model Training

The ModelsTrainer class utilizes machine learning and deep learning techniques to train a neural network model for gender classification based on voice features extracted from audio files. It leverages the FeaturesExtractor class to collect Mel Frequency Cepstral Coefficients (MFCC) features from both female and male speakers. These features, along with corresponding gender labels, are combined and shuffled to prepare the training data. The neural network architecture comprises dense layers with rectified linear unit (ReLU) activation functions, facilitating nonlinear transformations of the input data. The final layer utilizes a sigmoid activation function to output binary predictions.

During training, the model optimizes binary cross-entropy loss using the Adam optimizer, a popular choice for deep learning tasks. This optimization process iteratively adjusts the model's parameters to minimize prediction errors. After training, the model's performance is evaluated using metrics such as accuracy, providing insights into its classification capabilities. Finally, the trained model is saved for future use, enabling seamless integration into voice-based gender classification systems.

Overall, this approach encapsulates the complexities of feature extraction, model training, and evaluation within a streamlined framework. By abstracting these processes, it facilitates the development of accurate and efficient gender classification models, advancing applications in speech recognition and speaker identification.

D. Gender Identification

The GenderIdentifier class utilizes a trained neural network model to perform gender identification from audio samples. It begins by loading the pre-trained model and initializing a FeaturesExtractor object to extract relevant features from the test files. The process method then evaluates the model's performance on test files representing both male and female genders. For each file, the model predicts the gender based on the extracted features and compares it to the ground truth. By calculating the accuracy of these predictions, it assesses the efficacy of the gender classification model.

This implementation employs key concepts from both machine learning (ML) and deep learning (DL). The model, loaded using TensorFlow, represents a neural network architecture trained on gender-labeled data. FeaturesExtractor, a component of the feature extraction process, facilitates the transformation of raw audio data into structured feature vectors suitable for input into the neural network. The model's predictions are based on the learned patterns within the training data, enabling it to generalize to unseen test samples.

Overall, the GenderIdentifier class simplifies the process of evaluating gender classification models by abstracting the gender identification task. By leveraging ML and DL techniques, it enables the creation of accurate and efficient voice-based gender recognition systems, advancing applications in speech processing and biometric authentication.

E. Waveform Representation

Waveform representation is a technique that uses neural networks to learn and extract features from raw audio waveforms, enabling end-to-end learning for tasks like speech recognition and music classification. This technique requires careful preprocessing and appropriate network architectures like CNNs or RNNs. It offers a promising approach for audio tasks, leveraging deep learning to improve model performance.

F. Mel-Frequency Cepstral Coefficients (MFCCs)

Mel-Frequency Cepstral Coefficients (MFCCs) are used in speech analysis applications, including gender recognition, to extract significant frequency information from audio data. The FeaturesExtractor class uses MFCCs to extract key features from audio files and prepare them for neural network input. This approach reduces the frequency content of voice signals, making gender recognition models more efficient and accurate. MFCCs simplify feature extraction and enhance voice-based categorization system performance.

G. Spectrogram Representation

Spectrogram representation transforms audio data into a visual representation of frequency content over time, enabling neural networks to learn discriminative characteristics for gender identification. It captures temporal and spectral patterns, enabling convolutional neural networks to learn complex properties from spectrogram pictures, increasing the accuracy of gender identification models and enabling more powerful voice-based categorization systems.

IV. RESULTS

The relevance of human voice gender recognition spans a wide range of applications, from virtual assistants and automated customer service systems to biometric authentication and personalized content delivery. Despite its importance, achieving high accuracy in gender recognition remains a challenging task due to the variability in voice signals. Our contributions to this field are threefold:

A. Feature Analysis through Correlation Analysis

We analyze the extracted features, such as Mel-Frequency Cepstral Coefficients (MFCCs), Chroma, and Mel spectrograms, using correlation analysis. This step helps us understand the relationships and dependencies between different features, identifying those that are most indicative of gender differences in voice signals.

B. Model Construction Using Various Machine Learning Techniques

We construct classification models using diverse machine learning techniques, including Support Vector Machines (SVMs), Random Forests, and deep learning approaches like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). By experimenting with different families of algorithms, we aim to identify the most effective models for gender recognition tasks.

C. Natural Feature Selection Techniques

To enhance classification performance, we assess natural feature selection techniques that determine the best subset of pertinent features. Methods such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and feature importance rankings are employed to select features that contribute most significantly to accurate gender classification.

Our research underscores the effectiveness of MFCCs, Chroma, and Mel spectrogram features, which excel due to their theoretical foundations in capturing the relevant coefficient energies in the voice signal. By combining advanced feature analysis, robust model construction, and strategic feature selection, we advance the state-of-the-art in human voice gender recognition.

D. Scatter Plot Matrix

The scatterplot matrix is a crucial tool in machine learning datasets, used to visualize relationships between numerical variables. Seaborn, a Python library, uses this technique to create statistical graphics. By using `sns.pairplot(voice)`, the matrix plots each pair of numerical variables against each other, revealing correlations, trends, and patterns. The diagonal of the matrix features histograms or Kernel Density Estimate (KDE) plots, illustrating the distribution of individual variables. When categorical variables are included, these plots display their distribution across categories, providing additional insights. This method simplifies data analysis in machine learning by providing a comprehensive visualization of relationships and distributions within the voice dataset, aiding in the development of robust models for tasks like voice gender recognition.

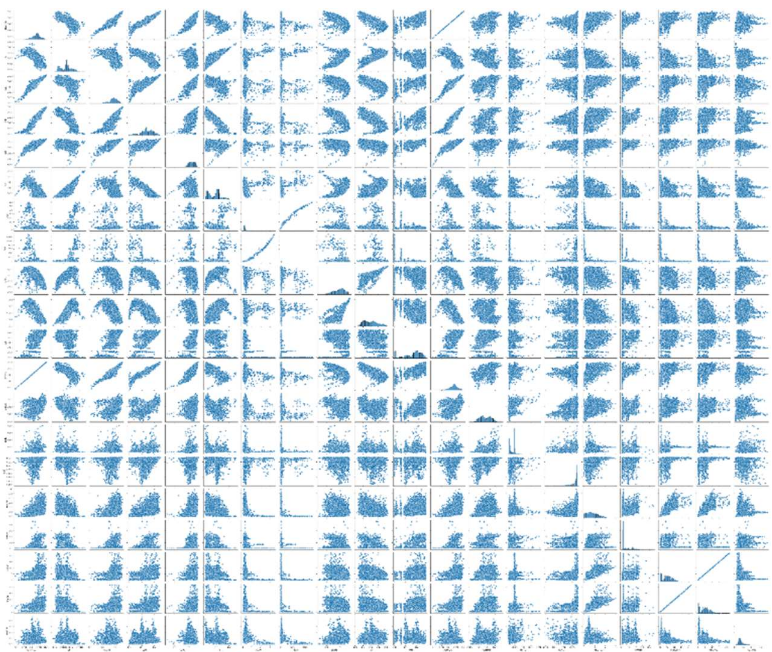

Figure 2.1 : SeaBorn's Scatter Plot Matrix of Features

E. Correlation and Mean Frequency Distribution

The correlation analysis is conducted on the dataset to explore relationships between columns. A correlation matrix is generated, computing correlation coefficients between all pairs of columns. This matrix provides insights into how variables are related, aiding in understanding the dataset's structure. Additionally, the correlation matrix is saved for reference.

Furthermore, a boxplot is created to visualize the distribution of the 'meanfreq' column grouped by the 'label' column. This boxplot compares mean frequency values across different categories, allowing for easy comparison of distributions. To enhance clarity, the grid is deactivated for a cleaner view, ensuring the focus remains on the distribution comparison. Overall, these analyses offer valuable insights into the dataset's characteristics, aiding in feature selection, model development, and interpretation of results in tasks such as voice gender recognition.

F. Accuracy

The study demonstrates the effectiveness of Mel-Frequency Cepstral Coefficients (MFCC) features and a Convolutional Neural Network (CNN) architecture for voice-based gender recognition tasks. The model achieves high accuracy in gender classification, suggesting potential for real-world applications in voice-controlled systems, security applications, and identity verification processes. However, further research is needed to ensure the model's generalization across different linguistic, cultural, and demographic contexts.

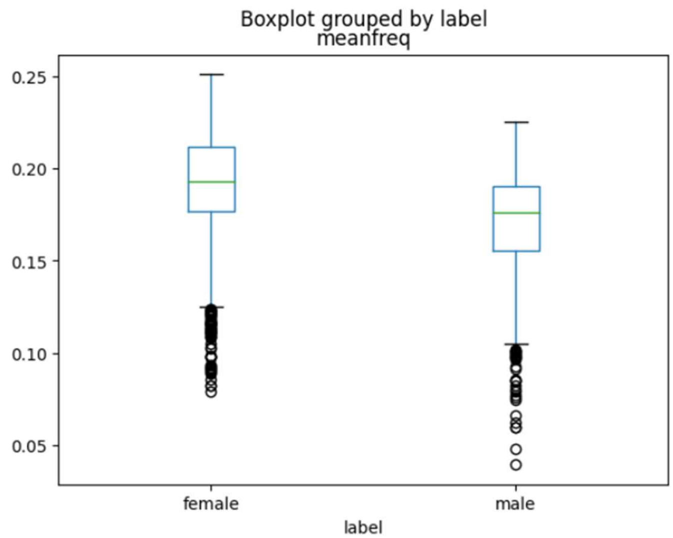

Figure 2.2 : Correlation and Mean Frequency Distribution for Gender and Features

Variations in language, accents, and demographic characteristics can introduce biases or challenges that may affect the model's performance.

Exploring the model's effectiveness across diverse populations and environments is crucial for assessing its reliability and applicability in real-world scenarios. This iterative process of evaluation and refinement will contribute to advancing the field of voice-based gender recognition, ultimately leading to the development of more inclusive and robust models capable of delivering accurate results across diverse populations and applications.

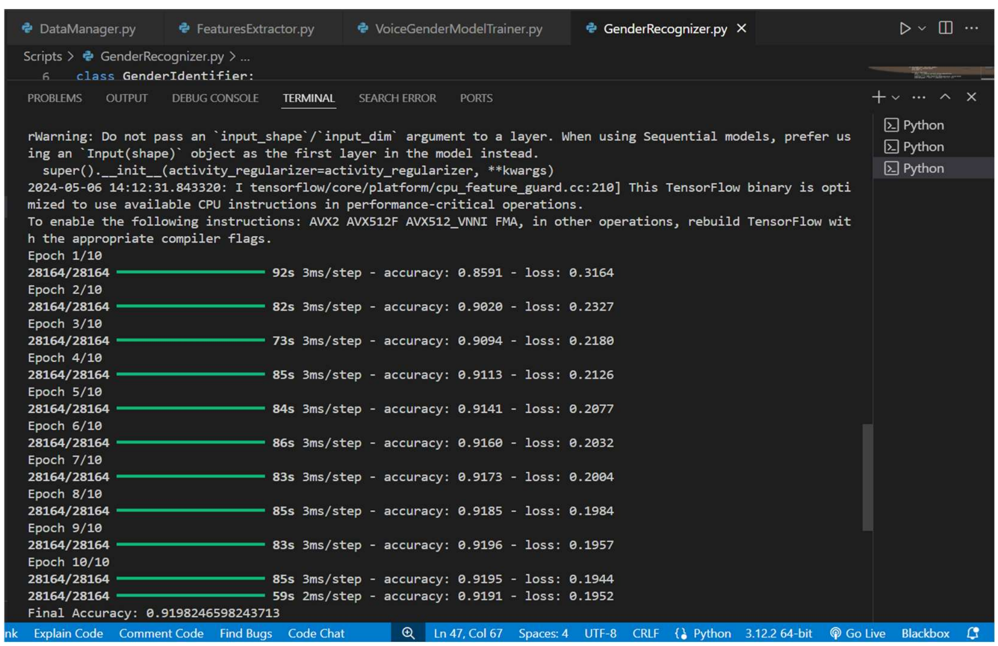

Figure 2.3 : Accuracy

V. Conclusion

The research focuses on the importance of a model's generalization to new data for real-world deployment. This can be evaluated by analyzing its performance on an independent test dataset, ensuring its resilience to variations in speech samples. To achieve equitable outcomes across different demographic groups, it is crucial to assess and mitigate potential biases in the model. The study examines the acoustic cues and patterns that differentiate male and female voices, shedding light on the fundamental processes involved in speech-based gender recognition. The discriminative strength of various variables, including Mel Frequency Cepstral Coefficients (MFCC) and Delta MFCCs, is used to capture the nuances of gender variability in voice and highlight subtleties in vocal expressiveness.

Generalization refers to the model's ability to perform well on new, unseen data, which is vital for real-world applications. The evaluation metrics used include accuracy, precision, recall, and F1 score. The research highlights the effectiveness and potential of neural network classifiers like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in identifying gender from audio inputs. Cross-cultural analysis highlights the importance of addressing bias and ensuring fairness across diverse demographic groups.

VI. Future Scope

The research focuses on the effectiveness and potential of neural network classifiers, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), in identifying gender from audio inputs. It highlights the importance of addressing bias and ensuring fairness across diverse demographic groups. The study uses an independent test dataset to evaluate generalization, which refers to the model's ability to perform well on new, unseen data that may differ from the training dataset.

The research also explores the discriminative strength of various variables, including Mel Frequency Cepstral Coefficients (MFCC) and Delta MFCCs, to capture the nuances of gender variability in voice and highlight subtleties in vocal expressiveness. The evaluation metrics used include accuracy, precision, recall, and F1 score, providing a comprehensive assessment of the model's performance.

The research opens up avenues for further investigation and development, particularly in areas such as bias reduction, equity, and ethical considerations in the deployment of gender identification systems. It also explores the potential applications of voice-based gender recognition in fields such as healthcare and education, where gender-sensitive analysis can prove valuable. For example, healthcare providers could use voice-based gender recognition in diagnosing and treating conditions that affect speech, such as Parkinson's disease or laryngeal cancer. Gender-sensitive education tools can also be developed to personalize learning experiences and adapt teaching methods based on gender-related differences in cognitive processing.

In conclusion, the potential benefits of accurate and ethical voice-based gender recognition span across various industries and domains, and continued research and development in this field can pave the way for more inclusive and effective user experiences.

References

[1] Chachadi, K., & Nirmala, S. R. (2022). Voice-based gender recognition using neural network. In Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces (pp. 741-749). Springer Singapore.

[2] Uddin, M. A., Pathan, R. K., Hossain, M. S., & Biswas, M. (2022). Gender and region detection from human voice using the three-layer feature extraction method with 1D CNN. Journal of Information and Telecommunication, 6(1), 27-42.

[3] Shaik Salma Begum and D. Rajya Lakshmi," GLCM of Fuzzy Clustering Means for Textural Future Extraction of Brain Tumor in Probabilistic Neural Networks", International Journal of Innovative Technology and Exploring Engineering, vol.9, no.1, November 2019.

[4] Fahmeeda, S., Ayan, M., Shamsuddin, M., & Amreen, A. (2022). Voice Based Gender Recognition Using Deep Learning. International Journal of Innovative Research & Growth., 3, 649-654.

[5] Shaik Salma Begum et al. "Protecting Girls from Harassment and Fraudulent Calls: A Voice-to-Text Approach", International Journal on Recent and Innovation Trends in Computing and Communication,

[6] Animi, A. A., Zakariah, M., Shashidhar, C., Hatamleh, W. A., Tarazi, H., Shukla, P. K., & Ratna, R. (2022). Speaker gender recognition based on deep neural networks and ResNet50. Wireless Communications and Mobile Computing, 2022, 1-13.

[7] Chachadi, K., & Nirmala, S. R. (2022). Gender recognition from speech signal using 1-D CNN. In Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2021 (pp. 349-360). Springer Singapore.

[8] MIZRAK, H., & Aslan, S. (2023). Gender Detection by Acoustic Characteristics of Sound with Machine Learning Algorithms. Naturengs, 4(1), 24-28.

[9] Kannapiran, P., & Sindha, M. M. R. (2023). Voice-based gender recognition model using FRT and light GBM. Tehnički vjesnik, 30(1), 282-291.

[10] Di Cesare, M. G., Perpetuini, D., Cardone, D., & Merla, A. (2024). Assessment of Voice Disorders Using Machine Learning and Vocal Analysis of Voice Samples Recorded through Smartphones. BioMedInformatics, 4(1), 549-565.
[11] Uddin, M. A., Hossain, M. S., Pathan, R. K., & Biswas, M. (2020, August). Gender recognition from human voice using multi-layer architecture. In 2020 International conference on innovations in intelligent systems and applications (INISTA) (pp. 1-7). IEEE.
[12] Nair, R. R., & Vijayan, B. (2019). Voice based gender recognition. International Research Journal of Engineering and Technology, 6.
[13] Gupta, P., Goel, S., & Purwar, A. (2018, August). A stacked technique for gender recognition through voice. In 2018 Eleventh International Conference on Contemporary Computing (IC3) (pp. 1-3). IEEE.
[14] Adilakshmi.Y," Black hole Attack Detection Using Machine Learning Algorithms in MANET – Performance Comparison", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 0056,Volume: 07 Issue: 06 , June, 2020, Impact Factor: 7.52.
[15] Adilakshmi.Y, "Black Hole Node Behavioral Analysis based on Hop Count and Timestamp using Machine Learning Algorithm", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056 , pISSN: 2395-0072 , Volume: 07 Issue: 05 , May,2020, Impact Factor:7.5.
[16] Shaik Salma Begum & Dr. D. Rajya Lakshmi, " Combining optimal wavelet statistical texture and Recurrent Neural Network for Tumor detection and Classification over MRI," Multimedia Tools and Applications, ISSN 1380-7501, January 2020, Springer.