



Archives available at journals.mriindia.com

**International Journal on Advanced Computer Engineering and
Communication Technology**

ISSN: 2278-5140

Volume 15 Issue 01, 2026

Voice Frequency-Based Gender Classification Using Convolutional Neural Network

¹Ms. Jifny S., ²Annuja J. S., ³Anns Harisha H. C.

¹M.E., Assistant Professor, Department of Computer Science and Engineering, Arunachala College of Engineering for Women, Nagercoil, Tamil Nadu, India

^{2,3} UG Student, Department of Computer Science and Engineering, Arunachala College of Engineering for Women, Nagercoil, Tamil Nadu, India

Peer Review Information	Abstract
<p><i>Submission: 08 March 2026</i></p> <p><i>Revision: 26 March 2026</i></p> <p><i>Acceptance: 05 April 2026</i></p> <p>Keywords</p> <p><i>Gender Classification, Voice Frequency, 1D CNN, Acoustic Features, Deep Learning.</i></p>	<p>Voice-based gender classification is a fundamental task in speech processing with applications in biometric systems, human-computer interaction, and demographic analytics. This study proposes an efficient deep learning approach for gender recognition using frequency-domain acoustic features. A structured dataset comprising 21 statistical voice features is employed to train a one-dimensional Convolutional Neural Network (1D CNN). Unlike conventional approaches that rely on raw audio or spectrogram representations, the proposed method operates directly on pre-extracted frequency features, thereby reducing computational complexity while preserving discriminative information. The dataset is preprocessed through label encoding and feature standardisation before being fed into the model. The trained network achieves a test accuracy of 97.00% and an AUC score of 1.00, demonstrating excellent classification capability. The results confirm that structured acoustic features, when combined with deep learning, provide a robust and scalable solution for voice-based gender classification.</p>

Introduction

Speech signals inherently encode biological and physiological attributes of a speaker, among which gender is one of the most distinguishable characteristics. Differences in vocal cord structure and vocal tract dimensions result in distinct frequency patterns between male and female voices. Typically, male voices exhibit lower fundamental frequencies ranging from 85 Hz to 180 Hz, whereas female voices lie between 165 Hz and 255 Hz. These variations form the basis for frequency-based gender classification. Traditional approaches to gender recognition rely on handcrafted features and classical machine learning algorithms such as Support Vector Machines and Logistic Regression. While these methods have demonstrated moderate success, they often struggle to generalise in real-

world scenarios due to noise, variability in speech patterns, and their inability to capture complex non-linear relationships.

Recent advancements in deep learning have addressed these limitations by enabling automatic feature extraction and hierarchical pattern learning. Convolutional Neural Networks (CNNs), in particular, have shown remarkable performance in processing structured and sequential data. This study adopts a one-dimensional CNN architecture that directly operates on frequency-based feature vectors, eliminating the need for computationally expensive transformations such as spectrogram generation.

The primary contribution of this work lies in demonstrating that high classification accuracy can be achieved using structured acoustic

features alone, without relying on raw audio inputs. This approach simplifies the processing pipeline while maintaining strong predictive performance.

Related Work

Several studies have explored the application of machine learning and deep learning techniques for voice-based gender classification. Deep neural networks utilising Mel-Frequency Cepstral Coefficients (MFCCs) have achieved high accuracy, often exceeding 98%, outperforming traditional classifiers [5]. Hybrid models integrating deep neural networks with Support Vector Machines have further improved robustness in noisy environments, achieving approximately 97% accuracy [4].

One-dimensional CNN architectures have been specifically designed for sequential voice data and have demonstrated effectiveness in capturing local dependencies within feature vectors, achieving accuracy levels above 92% [1]. Lightweight models such as Multilayer Perceptrons and Logistic Regression have also been investigated, achieving reasonable performance; however, they lack the ability to learn deep hierarchical representations [3], [7]. Recent large-scale implementations and web-based systems highlight the scalability and practical deployment potential of such models [6], [9]. These studies collectively indicate that CNN-based approaches offer superior performance and adaptability for gender classification tasks. These studies indicate that deep learning models, especially CNN-based approaches, are highly effective for voice-based gender classification.

Web-Based and Scalable Voice Classification Systems:

With the growing demand for practical and scalable gender recognition solutions, recent research has focused on integrating machine learning models with web-based platforms to enable real-time and large-scale deployment. These systems combine data processing, model inference, and user interface design to support real-world applications.

Sakshi Malhotra et al. [6] (2023) proposed a cross-platform voice-based gender classification system that allows users to upload audio files and receive predictions in real time. Their implementation utilised Python and R for backend data processing and employed Support Vector Machine (SVM) and XGBoost algorithms for classification. The modular architecture, integrating web technologies with machine learning models, demonstrated the feasibility of

deploying speech-based biometric systems on live platforms.

Similarly, Vinayak Sudhakar Kone et al. [9] (2023) developed a large-scale gender classification framework capable of handling more than 66,000 voice samples. Their approach combined dimensionality reduction techniques such as Principal Component Analysis (PCA), feature extraction methods including Mel-Frequency Cepstral Coefficients (MFCC) and Fast Fourier Transform (FFT), along with classification models such as Convolutional Neural Networks (CNN) and Logistic Regression. The system achieved an accuracy of 91%, highlighting the effectiveness of combining deep learning and traditional machine learning techniques for large-scale datasets.

These studies emphasise the importance of scalability and system integration in modern gender classification systems, particularly for applications in telecommunications, customer analytics, and voice-based biometric authentication.

Methodology

The dataset used in this study contains 3,168 voice samples, equally representing male and female speakers. Each sample includes 21 frequency-based acoustic features such as mean frequency, spectral properties, and statistical measures. During preprocessing, gender labels are converted into numerical form for model compatibility. The feature values are then normalized to ensure that all inputs contribute equally during training. The dataset is divided into training, validation, and testing sets in the proportions of 64%, 16%, and 20%, respectively. The input data is reshaped into a format suitable for the convolutional neural network.

The proposed model is a one-dimensional CNN comprising convolutional layers for feature extraction, activation functions to introduce nonlinearity, pooling layers for dimensionality reduction, and fully connected layers for classification. The final layer produces the output indicating the predicted gender.

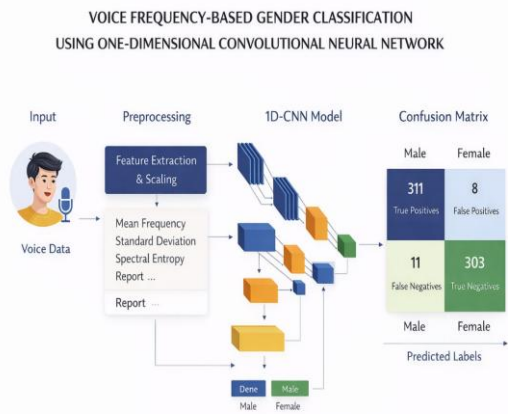
The dataset is divided into training, validation, and testing sets in the ratio of 64%, 16%, and 20%, respectively. This split enables effective model training while ensuring unbiased evaluation on unseen data. To facilitate compatibility with the CNN architecture, the input data is reshaped into a three-dimensional format of (samples, 21, 1), allowing the model to interpret the feature vector as a sequential input. The classification model is based on a one-dimensional Convolutional Neural Network designed to capture local dependencies within the feature sequence. Convolutional layers are

employed to extract relevant patterns, followed by Rectified Linear Unit (ReLU) activation functions to introduce non-linearity. Max-pooling layers reduce dimensionality and enhance feature robustness. The extracted features are flattened and passed through fully connected layers, enabling high-level abstraction. The final output layer utilises a sigmoid activation function to produce a probability score for binary classification.

Approach to Dynamic Threshold Update:

The classification decision is based on probability scores generated by the model. Instead of relying strictly on a fixed threshold, the system can adjust the decision boundary based on validation performance.

This dynamic threshold approach improves classification reliability, especially in cases where prediction confidence is close to the decision boundary. It helps reduce misclassification and enhances overall system performance.



Model Training and Evaluation

The model is trained using a batch size of 32 over a maximum of 50 epochs. To prevent overfitting and ensure optimal generalisation, an

EarlyStopping mechanism is employed, monitoring validation loss and halting training when performance stabilises.

Binary cross-entropy is selected as the loss function due to its suitability for binary classification problems. Model performance is evaluated using accuracy, precision, recall, F1-score, and the Receiver Operating Characteristic (ROC) curve. The Area Under the Curve (AUC) is used to assess the model’s ability to distinguish between classes.

The trained model achieves a test accuracy of 97.00% and an AUC score of 1.00, indicating excellent classification performance. Evaluation results confirm that the model effectively captures gender-specific acoustic patterns and generalises well to unseen data.

Model Architecture:

The classification model is based on a one-dimensional Convolutional Neural Network. The architecture consists of the following layers:

- **Input Layer:** Accepts feature vectors of shape (21, 1).
- **First Convolutional Layer:** 32 filters with kernel size 3 and ReLU activation, followed by MaxPooling with pool size 2.
- **Second Convolutional Layer:** 64 filters with kernel size 3 and ReLU activation, followed by MaxPooling with pool size 2.
- **Flatten Layer:** Converts multi-dimensional feature maps into a one-dimensional vector.
- **Dense Layer:** 64 neurons with ReLU activation for high-level abstraction.
- **Output Layer:** Single neuron with sigmoid activation producing a binary probability score.

The total number of trainable parameters is approximately 18,753. A threshold of 0.5 is applied to the sigmoid output to determine the final predicted class.

Table 1: Classification Report

Class	Precision	Recall	F1-Score	Support
Male	0.95	0.99	0.97	297
Female	0.99	0.96	0.97	337
Accuracy	–	–	0.97	634
Macro Avg	0.97	0.97	0.97	634
Weighted Avg	0.97	0.97	0.97	634

Table 2: Comparison of CNN with Other Neural Network Architectures

Architecture	Strengths	Limitations
RNN	Captures temporal dependencies	Prone to vanishing gradients in long sequences.
RNN	Captures temporal dependencies	Prone to vanishing gradients in long sequences.
LSTM	Handles long-term dependencies	High computational cost; unnecessary for fixed-length data.
MLP	Learns non-linear feature relationships	Does not capture local spatial patterns.
1D CNN (Proposed)	Pattern extraction, dimensionality reduction, generalizes well	Requires careful architectural tuning.

Results and Discussion

The experimental results demonstrate that the proposed 1D CNN model performs reliably across both training and testing datasets. The confusion matrix indicates minimal misclassification, reflecting balanced performance across both gender classes. The ROC curve further confirms the model's strong discriminative capability, with an AUC value close to unity.

A sample input evaluation yields a prediction probability of 0.9879, correctly classifying the input as male. This result highlights the model's robustness and its ability to generalise beyond the training dataset.

Compared to traditional approaches, the proposed method eliminates the need for manual feature engineering and complex preprocessing pipelines. The use of structured frequency features reduces computational overhead while maintaining high accuracy, making the model suitable for real-time and resource-constrained applications.

Conclusion

This study presents a deep learning-based approach for gender classification using frequency-domain acoustic features and a one-dimensional Convolutional Neural Network. The model achieves high accuracy and demonstrates strong generalisation capability, validating the effectiveness of structured feature inputs for voice-based classification tasks.

The proposed approach simplifies the overall pipeline by avoiding raw audio processing while maintaining competitive performance. These characteristics make it a practical solution for deployment in voice-driven applications such as virtual assistants, biometric systems, and communication platforms.

Future Work

Future work may focus on extending the model to multi-task learning scenarios, including age prediction, emotion detection, and speaker

identification. Expanding the dataset to include diverse languages and speaking conditions would further enhance model robustness. Additionally, integrating real-time processing capabilities and adaptive learning mechanisms could improve performance in dynamic environments.

References

A. Adhikary and D. Saha, "Gender and dialect classification from speech signals using one-dimensional convolutional neural networks," *International Journal of Speech Technology*, 2022.

M. K. Benkaddour, "Hybrid deep learning model combining convolutional neural networks and deep neural networks for gender classification," *International Journal of Computer Vision and Pattern Recognition*, 2020.

E. Buyukyilmaz and M. D. Cibikdiken, "Voice gender classification using machine learning techniques," *Procedia Computer Science*, vol. 102, pp. 199–204, 2016.

M. Mutiany and I. O. Herlistiono, "Hybrid DNN-SVM model for gender classification based on speech features," *Journal of Information and Communication Technology*, 2020.

F. Fahmeeda et al., "Deep neural network-based gender classification using MFCC features," *International Journal of Advanced Computer Science and Applications*, 2022.

S. Malhotra et al., "Web-based voice gender classification system using machine learning techniques," *International Journal of Emerging Technologies in Computer Science*, 2023.

S. Hizlisoy et al., "Gender classification from speech signals using logistic regression and spectral features," *Journal of Signal Processing Systems*, 2022.

K. Zhang et al., "Deep residual networks of residual networks for gender classification from facial images," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4913–4925, 2017.

V. S. Kone et al., "Large-scale voice-based gender classification using machine learning and deep

learning techniques," *International Journal of Data Science and Analytics*, 2023.

A. Greco et al., "Lightweight convolutional neural network for real-time gender classification," *IEEE Access*, vol. 8, 2020.