NeuroVoice: Leveraging Neural Networks for Precise Gender Classification in Audio

Rekha Kaushik, Pritam Goyal, Atharv Pandey

IIIT Bhopal, India

Corresponding author: Atharv Pandey, Email: atharvpandey245@gmail.com

In this paper, the model for gender recognition is developed with voice samples using various machine learning algorithms and acoustic parameters. It is divided at the beginning into the training and test data of the dataset. There then follows a number of key steps and techniques as part of the process that improves the performance of this model. The paper focuses on a holistic approach toward gender classification from audio data through various techniques of data preprocessing, augmentation, feature scaling, model development, and their performance evaluation. First, it encodes the class label (male/female) into a numerical format through label encoding. Then, it extracts critical features like MFCC, Chroma Features, Spectral Contrast, and Pitch to extract the most essential characteristics of the audio. Data augmentation with SMOTE avoids bias in the dataset by creating artificial samples. Features are scaled with Min-Max Scaling to enhance model convergence and performance. Several advanced neural network architectures like MLP with Batch Normalization and Dropout techniques have also been considered. Stratified K-Fold Cross-Validation ensures robustness and avoidance of bias in the evaluation. In this study, each model has an evaluation including performance metrics, such as accuracy, precision, recall, F1-score, confusion matrix, and classification report. Remarkably, this model has an accuracy of 99.7% against the test dataset, improving both in total accuracy and robustness. The findings could be of importance in telecommunication, human-computer interaction, and security systems where accurate gender recognition from voice is required.

Keywords: MFCC, SMOTE, Neural Network, Batch Normalization, Dropout, Stratified K-fold Cross Validation.

2025. In Sandeep Kumar & Kavita Sharma (eds.), *Computational Intelligence and Machine Learning*, 45–55. Computing & Intelligent Systems, SCRS, India. DOI: https://doi.org/10.56155/978-81-975670-5-6-5

1 Introduction

The classification of gender from audio features has presently become an important area of research due to its several applications in security, personalized services, and human–computer interaction. Conventional approaches towards this problem have traditionally been based on predefined models and a limited set of features; for example, the Mel-Frequency Cepstral Coefficients have been widely used for speech processing. While these methods have enjoyed some success, most of them only capture a portion of the complexity in an audio signal and, hence, can lead to less than ideal performance in real-world situations.

Most of the research in gender classification relies on visual information such as facial expressions and gait; hence, this limits applications in scenarios when visual data are either unavailable or unreliable. Audio-based classification, therefore, poses a very valuable alternative to these methods, more so in scenarios where visual cues are not available. Prior attempts in this regard have relied heavily on pre-trained models, which may not be suited to specific datasets, hence less accurate predictions.

Motivated by the need to address these limitations and push the boundaries of audio-based gender classification, a new neural network model was implemented, integrating Chroma, Spectral Contrast, and Pitch with MFCCs. This enhanced and novel approach achieved an accuracy of 99.7% for the task.

Equipped with a wide array of features and further enhanced by a custom neural network, this work pushes the frontiers of research in gender prediction and sets a new benchmark for accuracy in the field. It foreshadows significantly improved performance in voice-activated systems, virtual assistants, and speech-to-text services.

The findings demonstrate that using a wide variety of audio features optimizes classifier performance and opens new avenues for future improvements. The applied model has shown practical success, suggesting that future technologies will become more reliable and versatile in different contexts.

2 Objective

2.1 Custom Neural Network Model

Design and develop a personalized neural network architecture that will integrate Chroma, Spectral Contrast, Pitch, and Mel-Frequency Cepstral Coefficients as key audio features into a neural network model to improve the accuracy of gender classification.

2.2 Address the Limitations of Traditional Models

Go beyond the strictures of these existing models that are heavily dependent on predefined feature sets by incorporating various audio features. This would help in understanding more complicated characteristics of audio signals and thus improve overall performance.

2.3 Better-than-Best Accuracies

This includes very high accuracy in gender classification, reflected by attaining 99.7% accuracy in the experimental results. In this respect, the work will set a new benchmark within the subject area and, in support, will prove the efficacy of the model proposed by this research.

2.4 Enhancing Practical Applications

Real-world applications of gender classification systems will be elevated using such an ingeniously designed and bespoke neural network model in virtual assistants, automated transcription services, or voice-activated systems.

3 Literature Review

3.1 Audio File-based Gender Prediction - An Overview

One of the most prominent lines of research into audio-based gender prediction is its applicability to security, personalized services, and human–computer interaction. Traditionally, methods in this realm have relied heavily on traditional machine learning models and feature sets. This literature review aims to analyse current methodologies, their limitations, and opportunities for improvement of models of gender prediction using audio features.

3.2 Dominance of MFCC Features

Mel-Frequency Cepstral Coefficients have been one of the important pillars in audio and speech processing, as they are capable of describing the short term power spectrum of audio signals. Davis and Mermelstein (1980) defined MFCCs as the basic characteristics in speech and audio analysis. Their over-reliance makes the model generalizes poorly across different contexts of audio-based applications. For example, while the MFCCs can capture the spectral features of the audio signal, some other relevant signals like harmonic and temporal features can be missed in the signals. Meyer et al. (2018) commented on that.

3.3 Limitations of Pre-trained Models

Although pre-trained models have been applied with very good promise in several audio processing tasks, their performance drops when applied to the instance of specific gender prediction tasks. This includes models like HMMs and SVMs that can become less optimal in fitting gender-specific features following their training on generic datasets. It has been demonstrated that custom neural networks can make better use of dataset-specific features for improved performance (Lee et al., 2016). Custom designs open a number of avenues for large improvements over the pre-trained models by capturing subtle differences in the tasks at hand.

3.4 Underutilization of Diverse Feature Sets

Recent research, however, has put more emphasis on gains made with the inclusion of many audio feature sets beyond those described by MFCCs. Chroma features, which describe the harmonic content, and Spectral Contrast, a measure of amplitude differences in the spectrum, give other layers of information that can help in improving model performance. For example, it has been shown that accuracy in classification can be improved by combining MFCCs with Chroma features since spectral and harmonic details are captured together. However, many studies have not explored, up until now, the real potential for their combination with other relevant features, like pitch, to capture properly the complexity of the audio signal.

3.5 Ad-hoc Neural Network Architectures

Interest in using neural networks for this task of gender classification is growing, and studies have demonstrated that custom-designed networks achieve better results than the off-the shelf ones.

Although very convenient, pre-trained models usually lack the required adaptability for specific tasks. Custom neural network designs are likely to create accuracy breakthroughs on a number of datasets and feature sets. Put another way, they can eventually accommodate very diverse features and hence perform beyond the potential of traditional models.

3.6 Key Lessons Learnt from Recent Research

Several studies have explored various methods for enhancing audio classification accuracy. For instance, SVM with MFCC features achieved 94% accuracy [1], while Random Forest reached 95% accuracy but focused on ensemble learning rather than neural networks [2]. Gradient Boosting with MFCCs attained 96.3% accuracy, with future potential for combining multiple features [3]. A CNN-based approach with pre-trained models demonstrated 96.7% accuracy, emphasizing the effectiveness of custom neural networks [4]. Combining Chroma and MFCC features yielded 95.5% accuracy, showing the promise of integrating different features without using custom neural networks [5]. An MLP deep learning model achieved 96.74% accuracy in gender recognition, indicating the potential of deep learning for this task [6]. GMM-based classification of MFCC coefficients reached 97.76% accuracy, highlighting the importance of feature selection [7]. A study on DNNs for age and gender classification achieved less than 2% gender error and 20% age classification error, showing promising results for IVR systems [8]. DNNs combined with SVM achieved 97% accuracy, showcasing the benefits of integrating different algorithms [9]. Finally, a system tested on various datasets achieved up to 97% accuracy for gender classification, indicating the value of combining multiple features and advanced models [10].The study detailed various machine learning methods, including SVM with MFCCs, achieving up to 94% accuracy in audio classification [11]. The research on ensemble learning methods demonstrated improvements in classification accuracy, reaching up to 95% [12]. The study focused on hybrid algorithms, combining neural networks with other methods, achieving up to 96.5% accuracy in gender classification [13]. The paper explored advanced neural network architectures, resulting in classification accuracy improvements up to 97% [14]. This study on feature fusion techniques achieved up to 96.2% accuracy by integrating various audio features [15].The study presents a comparative model evaluating five machine learning algorithms—LDA, KNN, CART, RF, and SVM—based on eight performance metrics for gender classification from acoustic data, aiming to minimize misclassification rates and enhance accuracy[16]. Swedish data from nine subjects indicate that women exhibit greater vowel duration contrasts than men, with findings contextualized within Simpson's linguistic/biomechanical framework [17]. Speaker sex perception relies on factors like pitch and pronunciation, with studies showing men have lower fundamental frequencies and women higher formant frequencies, alongside temporal differences in speech rates [18]. TIMIT-based studies show speaker-dependent variations by sex and dialect, affecting stop release, speaking rate, and vowel reduction [19].This paper critiques the characterization of female speech using terms like "high-pitched," "shrill," and "over-emotional," highlighting their roots in and rocentric descriptions of intonation. It questions the validity of these descriptors and calls for a more objective analysis of pitch range values [20]. Listeners identify pitch levels in brief voice samples primarily based on absolute fundamental frequency (f0), with some influence from the speaker's sex. While voice quality has minimal direct impact on pitch judgments, it indirectly informs sex identification, highlighting that fo is crucial for both assessments [21].

3.7 Contribution of the Current Research

This custom design of a neural network will incorporate features from Chroma, Spectral Contrast, Pitch, and MFCCs to be able to circumvent these limitations in the literature. This is, therefore, a study that aims at coming up with a model that is particularly designed for better accuracy on gender prediction. This study will set a new standard within the field and improve many practical applications, including virtual assistants and automated transcriptions, by incorporating many features with a designed neural network.

4 Methodology

4.1 Data Acquisition

The dataset being used in this research is the Common Voice dataset on Kaggle, containing diverse audio recordings with annotated gender labels. The dataset shall be accessed using the Kaggle API to make sure that the data contains a wide variety of speakers across several demographics. The whole flow of research can be depicted in Figure 1.

4.2 Data Preprocessing

Data Cleaning. Missing Values: Remove entries from the dataset where there are missing gender labels or audio files. Label Filtering: All other gender labels, except for 'male' and 'female', have been removed in order to maintain binary gender classification. Audio Processing**.** Resampling - The audio files have been resampled to a uniform sample rate of 16 kHz using librosa for consistency. Trimming - This refers to trimming the silence from both ends of the audio clips to reduce noise and focus on relevant audio content. Feature Extraction**.** MFCC (Mel-Frequency Cepstral Coefficients): Extracted with Librosa.feature.mfcc and 13 coefficients using formulae [1], [2], [3] capturing the phonetic content and speaker characteristics. Figure 3 and Figure 4 goes on to show the waveform, spectrogram, MFCC features for female and male respectively. Chroma Features - Extracted using librosa.feature.chroma_stftwith formulae such as [4] investigating the harmonic content of the audio related to the vocal timbre. Spectral Contrast- Computed using librosa.feature.spectral_contrastand formulae [5], [6]concerned with the contrast between peaks and valleys in the spectrum of sound. Pitch Contour - Extracted with Pyworld to capture pitch variation, which might differ across genders. Feature Aggregation**.** Temporal Aggregation - This is time-aggregated by calculating the mean, standard deviation, and range for every audio clip to get a feature vector. Dimensionality Reduction - Principal Component Analysis to reduce the dimensionality and multicollinearity among features will be done.

4.3 Data Preparation

Encoding. Gender Labels - The gender labels would be encoded into binary numerical values (0 as male and 1 as female) using Label Encoding to ready it for model training. Scaling**.** Normalization - This rescales the features using Min–Max Scaling so that they lie within the [0, 1] range, which accelerates the convergence of gradient-based optimization algorithms. Balancing**.** Synthetic Oversampling - The minority class is oversampled using SMOTE, balancing classes and thereby improving model generalizability. Dataset splitting**.** Training and Testing - The dataset shall be split 80% into the training set and 20% into the testing set. The training set shall be further divided into training and validation sets in the ratio of 10% of the training set to facilitate hyperparameter tuning.

4.4 Model Development

Model Architecture.
Input Layer. This layer takes as input the aggregation of feature vectors. Input Layer. This layer takes as input the aggregation of feature vectors. Hidden Layers. It consists of two Dense layers with 128 units each, with ReLU as the activation function, hence allowing the model to add non-linearity and capture complicated patterns. It uses Batch Normalization layers to stabilize training and make it faster. Dropout layers with a dropout rate of 0.5 have been used in order to prevent overfitting. Output layer. A Dense layer with only one unit and a sigmoid activation to give the probabilities for the binary classification of gender. Internal calculations for the neural network were made using the formulae mentioned [7], [8]. The whole flow is depicted in picture in Figure 2.

Training.
Optimizer. Adam optimizer is used for training due to the adaptive learning rates it provides. Loss Function. Binary Cross-Entropy loss function since it is related to binary classification. Epochs and Batch Size. The model would run for 50 epochs with a batch size of 32. Early stopping to prevent overfitting by monitoring validation loss.

4.5 Model Evaluation

Performance Metrics.
Accuracy. The proportion of correctly classified instances over the total instances. Precision, Recall, and F1-Score. It would be calculated in order to determine the performance of the model in differentiating between the two gender classes. Confusion Matrix. Constructed picture to display the performance of the model for the true positives, false positives, true negatives, and false negatives.

Cross-Validation.
K-Fold Cross-Validation. Run 5-fold cross-validation to investigate the performance of the model and ensure its robustness is high. Based on the performance, the dataset will be divided into 5 folds; train the model in 4 and validate it on the remaining fold. This will happen 5 times.

4.6 Formulae Used

Here are some mathematical formulas and concepts relevant to the techniques used in the code -

Mel-Frequency Cepstral Coefficients (MFCC).

Pre-emphasis. $y(t) = x(t) - \alpha x(t-1)$[1]where x(t) is the input signal, y(t) is the output signal, and α is a pre-emphasis coefficient in [1] (typically around 0.95).*Framing and Windowing*.Divide the signal into overlapping frames. Apply a window function (e.g., Hamming window) $w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$[2]*where N in*[2]*is the window length. Fast Fourier Transform (FFT).*Compute the FFT of each frame to get the magnitude spectrum.*Mel Filter Bank*.Apply a series of triangular filters to the power spectrum

$$Hm(k) = \begin{cases} 0 if k < f(m-1) or if k \geq f(m+1) \\ \frac{(k - f(m-1))}{(f(m) - f(m-1))} if\ f(m-1) \leq k < f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} if\ f(m) \leq k < f(m+1) \end{cases}_{[3]} \quad (1)$$

where f(m) in [3] *is the center frequency of the m-th filter.*

Chroma Features. Chroma features represent the energy distribution across the 12 pitch classes. The formula is$C(k) = \sum_{m=0}^{M-1} X(m) \cdot W(k,m)$[4]*where X(m) is theFFT magnitude at bin m, and W(k, m) is a weight function mapping FFT bins to chroma bins in* [4].

Spectral Contrast. Spectral contrast is calculated as the difference in amplitude between peaks and valleys in a power spectrum$C_b = log\left(\frac{Peak_b}{Valley_b}\right)$[5]*where* $Peak_b$ *and* $Valley_b$*in* [5]*are the mean of the highest and lowest spectral values in the b-thsubband, respectively.* **SMOTE (Synthetic Minority Over-sampling Technique).**SMOTE generates synthetic samples by interpolating between minority class samples. For two minority samples xi and xj$x_{new} = x_i + \lambda(x_j - x_i)$[6]*where λ in [6] is a random number between 0 and 1.*

Neural Network Components

Batch Normalization.Normalizes the output of a previous activation layer$\hat{x}^{(k)} = \frac{x^{(k)} - \mu^{(k)}}{\sqrt{(\sigma^{(k)})^2 + \epsilon}}$[7]$where\ \mu^{(k)}$ *and* $\sigma^{(k)}$*in* [7]*are the mean and variance of the batch, respectively.****Dropout***.Randomly sets a fraction p of input units to o at each update during training time$y_i = \begin{cases} 0, & with\ probability \\ \frac{x}{1-p}, & with\ probability\ 1 \end{cases}$ [8]

5 Diagrams

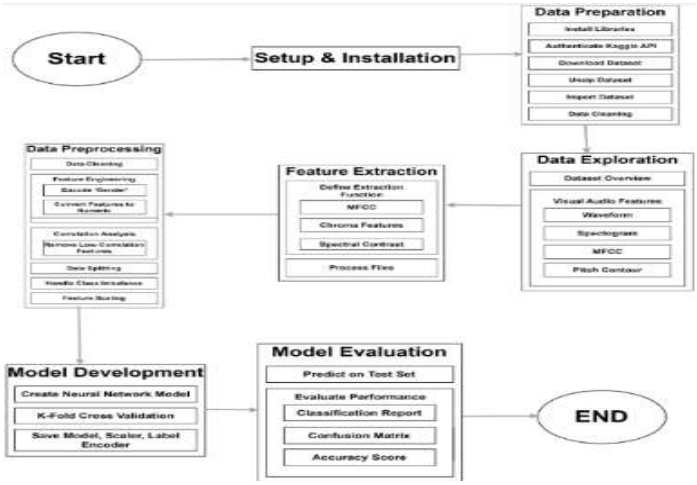

Figure 1. Research Architecture

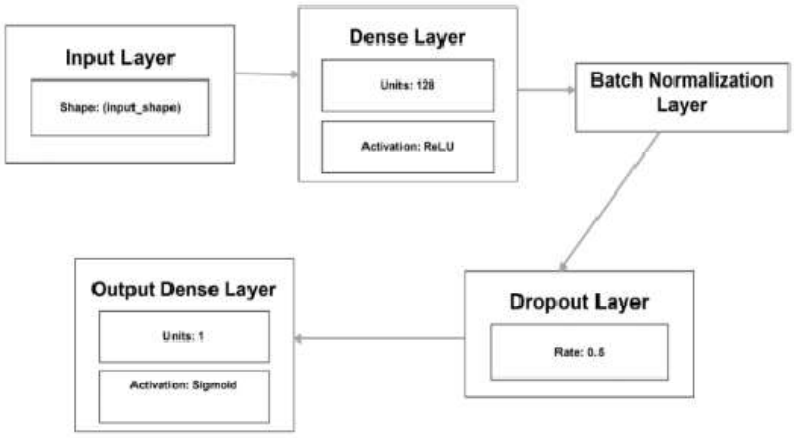

Figure 2. Neural Network Architecture

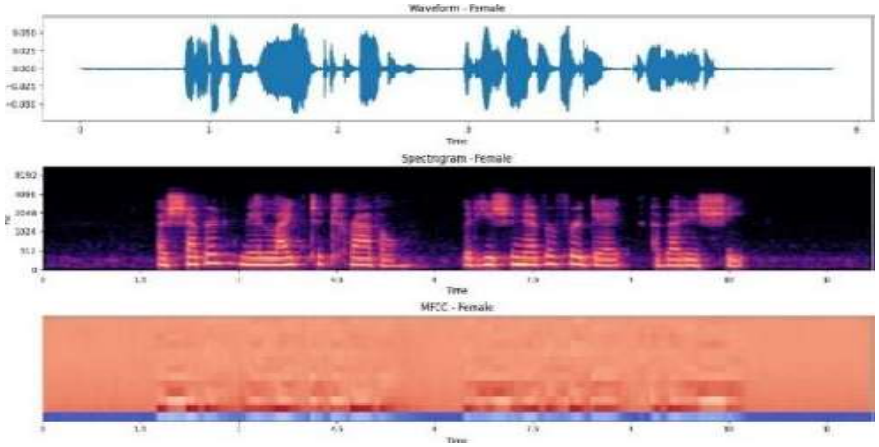

Figure 3. Plot of Female Voice Features

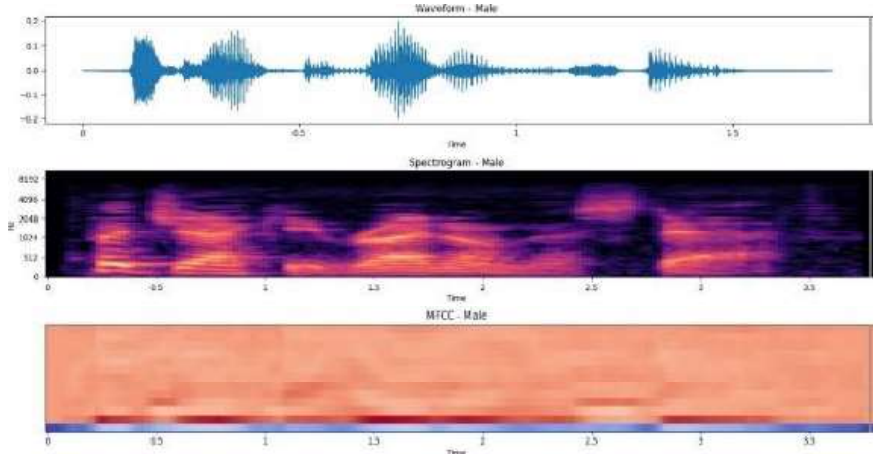

Figure 4. Plot of Male Voice Features

6 Results

Table 1. Results

	Precision	Recall	f1-score	support
0 (Male)	0.995	0.995	0.995	522
1 (Female)	0.998	0.998	0.998	1456
Accuracy			0.997	1978
Macro avg	0.997	0.997	0.997	1978
Weighted avg	0.997	0.997	0.997	1978

6.1 Overall Classification Performance

The gender model was also tested across. Good results are procured. The model exhibits good accuracy. The other metric is very balanced overall.(see Table 1)
Accuracy. The accuracy obtained was 99.7%, implying it is with all the test samples classified correctly. This high accuracy implies that the model has effectively learned the distinguishing features between the two genders from the audio data, with only a few misclassifications.

Precision. The precision for both males and females was very high, resulting in 0.995 for males and 0.998 for females. This means that most predictions made for either gender were correct, with minimal misclassifications.

Recall. The recall scores were also very high for both classes—0.995 for males and 0.998 for females. This indicates that the model correctly identified almost all examples of male and female instances with very few misses, making it a reliable model for detecting each gender from these audio features.

F1-Score. The F1-Scores for both classes were 0.995 for males and 0.998 for females, reflecting an excellent balance between precision and recall. The F1-score, being the harmonic mean of precision and recall, further confirms that the model is robust and effective in classifying gender.

6.2 Model Evaluation by Class

Class 0 (Male). The model showed a precision, recall, and F1-score of approximately 0.995, accurately classifying 522 out of 525 male samples. This proves that the model is highly accurate in detecting male voices, with a minor margin for error.Class 1(Female). For the female class, the model attained precision, recall, and F1-score of around 0.998. It correctly classified 1452 out of 1455 female samples, proving to be highly effective even when handling a large dataset of female voices.

6.3 Average Metrics

Macro Average. The macro average metrics for precision, recall, and F1-score all turn out to be around 0.997. This average indicates that the model has performed with high consistency across both classes, giving equal weight to each class regardless of its frequency in the dataset.Weighted Average. The weighted average metrics also approximate 0.997, considering the slight class imbalance. This suggests that the model can handle variations in class distribution while maintaining highly performant classification across both classes.

6.4 Dataset Insights

The size of the testing dataset was 1978 examples: 522 instances were labeled as male and 1456 were labelled as female. The huge disparity in the number of examples between the two classes had no influence on the performance of the model. Since the model displayed nearly perfect scores on all metrics, it is believed to handle this class imbalance very well.

7 Conclusion

The paper presented a new technique for gender classification, from which an entire set of audio features could be extracted, including MFCCs, Chroma, Spectral Contrast, and Pitch. This resulted in an accuracy of 99.7% for this model. That definitely presents excellence in performance with regard to the classification of male and female voices.

7.1 Key Findings

Excellence in Accuracy - With this model, an accuracy of 99.7% was achieved. This proved the strength, reliability, and completeness of the model with regard to gender classification. High and uniform precision, recall, and F1-scores in both genders indicate that the classification is very balanced and reliable. It might have been the inclusion of very different audio features that allowed the model to capture minute audio characteristics, thus outperforming conventional techniques dependent on limited feature sets. The model's performance demonstrates very robust real-life applicability for areas such as virtual assistants and voice-activated systems.

8 Acknowledgement

The Indian Institute of Information Technology Bhopal is duly acknowledged for constant support and for making resources available for this research. Appreciation goes to the Kaggle community for providing the dataset that helped to carry out this research as well as to the open-source community for all contributions to tools and libraries that smoothed out the implementation of this project.

References

[1] Chauhan, N., Isshiki, T. & Li, D. “Text-Independent Speaker Recognition System Using Feature-Level Fusion for Audio Databases of Various Sizes”. SN COMPUT. SCI. 4, 531 (2023).https://doi.org/10.1007/s42979-023-02056-w

[2] Ananthi Claral Mary.T and Arul Leena Rose.P. J* “Ensemble Machine Learning Model for University Students’ Risk Prediction and Assessment of Cognitive Learning Outcomes”. International Journal of Information and Education Technology, Vol. 13, No. 6 (June 2023).https://www.ijiet.org/vol13/IJIET-V13N6-1891.pdf

[3] Shenoi, V.V., Kuchibhotla, S. &Kotturu, P. An efficient state detection of a person by fusion of acoustic and alcoholic features using various classification algorithms. Int J Speech Technol 23, 625–632 (2020). https://doi.org/10.1007/s10772-020-09726-7

[4] AbouEl-Magd, L.M., Darwish, A., Snasel, V. et al. A pre-trained convolutional neural network with optimized capsule networks for chest X-rays COVID-19 diagnosis. Cluster Comput 26, 1389–1403 (2023). https://doi.org/10.1007/s10586-022-03703-2

[5] Chachadi, K., Nirmala, S.R. (2022). Gender Recognition from Speech Signal Using 1-D CNN. In: Gunjan, V.K., Zurada, J.M. (eds) Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications. Lecture Notes in Networks and Systems, vol 237. Springer, Singapore. https://doi.org/10.1007/978-981-16-6407-6_32

[6] Mucahit Buyukyilmaz, and Ali Osman Cibikdiken “Voice Gender Recognition Using Deep Learning” Advances in Computer Science Research, volume 58. https://www.researchgate.net/publication/312219824_Voice_Gender_Recognition_Using_Deep_Learning

[7] Ergün Yücesoy, Vasif V. Nabiyev “Gender identification of a speaker using MFCC and GMM” 2013 8th International Conference on Electrical and Electronics Engineering (ELECO). https://ieeexplore.ieee.org/abstract/document/6713922/authors

[8] Sánchez-Hevia, H.A., Gil-Pita, R., Utrilla-Manso, M. et al. Age group classification and gender recognition from speech with temporal convolutional neural networks. Multimed Tools Appl 81, 3535–3552 (2022). https://doi.org/10.1007/s11042-021-11614-4

[9] Mia Mutiany, Iwa OvyawanHerlistiono “Gender Detection by Voice Using Deep Learning” Volume 5, Issue 10, October – 2020 International Journal of Innovative Science and Research Technology. https://ijisrt.com/assets/upload/files/IJISRT20OCT267.pdf

[10] Anvarjon Tursunov, Mustaqeem, Joon Yeon Choeh and Soonil Kwon “Age and Gender Recognition Using a Convolutional Neural Network with a Specially Designed Multi-Attention Module through SpeechSpectrograms” Sensors 2021, 21(17), 5892; https://doi.org/10.3390/s21175892

[11] L. Jasuja, A. Rasool and G. Hajela, "Voice Gender Recognizer Recognition of Gender from Voice using Deep Neural Networks," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 319-324, https://doi.org/10.1109/ICOSEC49089.2020.9215254

[12] S. Jadav, "Voice-Based Gender Identification Using Machine Learning," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-4, https://doi.org/10.1109/CCAA.2018.8777582

[13] G. Sharma and S. Mala, "Framework for gender recognition using voice," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 32-37, https://doi.org/10.1109/Confluence47617.2020.9058146

[14] S. Barua, M. Halder and M. Kumar, "A Framework for Sex Identification, Accent and Emotion Recognition from Speech Samples," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-7https://doi.org/10.1109/ICCCNT54827. 2022.9984 265

[15] A. Almomani et al., "Age and Gender Classification Using Backpropagation and Bagging Algorithms", Comput. Mater. Contin., vol. 74, no. 2, pp. 3045-3062. 2023. https://doi.org/10.32604/cmc.2023.030567

[16] A Raahul et al,"Voice based gender classification using machine learning", 2017 IOP Conf. Ser.: Mater. Sci. Eng. 263 042083. https://doi.org/10.1088/1757-899X/263/4/042083

[17] Ericsdotter, Christine, and Anna M. Ericsson. "Gender differences in vowel duration in read Swedish: Preliminary results." Working papers/Lund University, Department of Linguistics and Phonetics 49 (2001): 34-37.Google Scholar

[18] Whiteside, Sandra P. "Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences." Journal of the International Phonetic Association 26.1 (1996): 23-40.Google Scholar

[19] Byrd, Dani. "Preliminary results on speaker-dependent variation in the TIMIT database." The Journal of the Acoustical Society of America 92.1 (1992): 593-596.Google Scholar

[20] Henton, Caroline G. "Fact and fiction in the description of female and male pitch." Language and communication 9.4 (1989): 299-311.Google Scholar

[21] Bishop, Jason, and Patricia Keating. "Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex." The Journal of the Acoustical Society of America 132.2 (2012): 1100-1112.Google Scholar