Hindawi
Journal of Engineering
Volume 2022, Article ID 9291099, 9 pages
https://doi.org/10.1155/2022/9291099

Research Article

Estimation of Accuracy in Human Gender Identification and Recall Values Based on Voice Signals Using Different Classifiers

Abhishek Singhal and Devendra Kumar Sharma

Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Delhi NCR Campus, Ghaziabad, UP, India

Correspondence should be addressed to Abhishek Singhal; abhisheksinghal.srm@gmail.com

Received 21 June 2022; Revised 18 November 2022; Accepted 30 November 2022; Published 15 December 2022

Academic Editor: Fan Yang

This paper presents the estimation of accuracy in male, female, and transgender identification using different classifiers with the help of voice signals. The recall value of each gender is also calculated. This paper reports the third gender (transgender) identification for the first time. Voice signals are the most appropriate and convenient way to transfer information between the subjects. Voice signal analysis is vital for accurate and fast identification of gender. The Mel Frequency Cepstral Coefficients (MFCCs) are used here as an extracted feature of the voice signals of the speakers. MFCCs are the most convenient and reliable feature that configures the gender identification system. Recurrent Neural Network–Bidirectional Long Short-Term Memory (RNN-BiLSTM), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA) are utilized as classifiers in this work. In the proposed models, the experimental result does not depend on the text of the speech, the language of the speakers, and the time duration of the voice samples. The experimental results are obtained by analyzing the common voice samples. In this article, the RNN-BiLSTM classifier has single-layer architecture, while SVM and LDA have a k-fold value of 5. The recall value of genders and accuracy of the proposed models also varied according to the number of voice samples in training and testing datasets. The highest accuracy for gender identification is found as 94.44%. The simulation results show that the accuracy of the RNN is always found at a higher value than SVM and LDA. The gender-wise highest recall value of the proposed model is 95.63%, 96.71%, and 97.22% for males, females, and transgender, respectively, using voice signals. The recall value of the transgender is high in comparison to other genders.

1. Introduction

Voice signals are the basic and essential medium to interchange views from one person to another for human beings. The voice signals carry the message's information and the speaker's characteristics during the conversation. These characteristics of the voice signals are variable. The variation in the characteristics of the voice signal depends on the plasticity of the vocal cord and the external and internal parameters. The voice signal also carries the paralinguistic details of the speaker, such as gender, emotion, and the status of health. The quality of the voice signal is degraded due to the noise mixing in the clean speech. To reduce the noise from the noisy voice signals, voice signal enhancement techniques are applied. In the enhancement techniques, silence and noise are removed from the voice signal to enhance the quality of the voice signals [1, 2].

Several applications are implemented with the help of a voice recognition system. Some important applications are recognition of gender, age, health information, sociolect, emotional state, attentional state, language, dialect, and accent. The characteristics of voice signals also have many applications in forensic, human-robot interaction, law enforcement, language learning, call routing speech translation, and intelligent workspaces. Analyzing the voice signals, the characteristics are also identified even when the speaker is hidden or during a conversation on the telephone. Gender is one of the important characteristics of the speakers. Gender identification of the speaker through the analysis of the voice signals is a technique that is used to determine the sex of the speaker. Gender identification is a very tough task in voice signal analysis because the gender identification system has several problems, such as variations in the voice signals and environmental noises. Several applications are available which are dependent on the identification of gender [3, 4]. Gender recognition systems are categorized into two categories as gender-dependent and gender-independent. Gender-independent systems are provided with less accuracy for recognition than gender-dependent systems [5]. The accuracy of the identification of the gender is increased by limiting the search space, which is developed by the analysis of the voice signals [2, 6]. The identification of the gender from the voice signals is also utilized in criminal cases as available evidence in the recorded form because the characteristics of the voice signal are unique for the specific gender or speaker [3, 7, 8]. Considering the security issues affecting the whole world, the gender classification shows excellent attention towards the researcher. A gender recognition system is a mandatory requirement in the current rapid development of the computerized environment [9].

In this article, first time in the field of the identification of gender, the identification of the transgender is introduced with the identification of the male and female based on the voice analysis. The gender identification system has voice samples, voice signal acoustic characteristics, and classification algorithms. Extracting the features from the voice signals is challenging because the acoustic parameters of the voice signals show variation in nature [4, 10, 11].

This article aims to calculate the recall value of the male, female, and "first-time" transgender. This article also compared the accuracy of the proposed models, which is obtained after analyzing all three gender voice signals by using BiLSTM-RNN, LDA, and SVM with MFCCs with 12 coefficients. The SVM and LDA algorithms have the k-fold value of 5. An optimization technique named "ADAM" is used in the RNN-BiLTSM algorithm to achieve the goal of the proposed work. RNN-BiLSTM shows the highest accuracy percentage among the three classifiers with 94.44% with single-layer architecture. These voice samples are recorded in a noisy environment and are text-independent. The simulation result is carried out by analyzing the voice signals with several datasets.

The recall value of each class, i.e., male, female, and transgender, is obtained by the confusion matrix. The accuracy of the proposed model is also calculated with the help of an analysis of the voice signals. The recall value of the third gender is calculated for the first time in the field of gender identification. Similarly, the accuracy comparison of the proposed models is reported for the first time, which is based on all three genders. The remanent part of the article is arranged as follows: Section 2 contains the information about the reported work in the available literature. The details of the experimental setup are available in Section 3. This section contains information about datasets, the parameter of the voice signals, and the classification algorithm. The experimental results are discussed with the help of the confusion matrix in Section 4. The conclusion of the article is available in Section 5.

2. Related Work

The research on the identification of gender has been rapidly analyzed in the recent era. The basic and initial element of the voice processing system is the identification of gender [1]. Several approaches are compared to find the identification of the gender by using Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), and SVM with the analysis of the voice signals which are recorded via telephone channels [12–14]. The accuracy of the speech identification can be reduced due to the noise [11, 15]. The accuracy of the text-dependent system is greater than the text-independent system. The accuracy of the identification of the gender also depends on the age of the speaker. The classification of the gender for young speakers is more convenient in comparison to the old speaker [2, 16, 17].

The MFCC was developed and utilized to identify gender in 2012. After the development of MFCC, several modifications have taken place to improve the system's performance. MFCC has also been examined for the identification of gender in different domains [18]. SVM is generally used for binary classification to identify gender. It shows the best accuracy in the class separation technique. An article reported that the Gaussian radial basis function SVM has the best accuracy compared to other kernels of SVM [3].

These acoustic parameters are generally used in classification algorithms to train and test voice samples [19]. After that, an approach of inserting the variable-length voice signal into a fixed-sized embedding vector is made. Probabilistic linear discriminant analysis (PLDA) has been combined after applying the vector in an external classifier [20, 21]. Before the deep learning era, I–vectors were the basis of several popular embedding methods [20]. The accuracy of the LSTM model is 98.4% which is very high compared to the others [22]. A new semisupervised approach, iCST-Voting, also achieved 98.4% classification accuracy for the audio signals [23].

The author showed the significant difference between deep learning and traditional machine learning models in the participant teams evaluated in the Third Workshop on NLP for similar languages, varieties, and dialects [24]. The bidirectional RNN with the gated recurrent unit is used to classify, which is combined with the feature of the input [25]. By using this model, the accuracy of the gender classification is achieved by 79% [26]. The classification accuracy is 62.3% and 78.8% for naïve Bayes and RNN, respectively [27, 28].

The identification system is used to segregate the same group of speakers into two groups of gender, i.e., male and female [1]. Support vector machine (SVM) classifier can be used to predict, and the accuracy is achieved by more than 90% [21]. The result can be enhanced using the deep neural network to learn the data sequence instead of classical machine learning techniques. Several pieces of research also claimed that deep learning is an excellent choice for classification. Using the deep learning model for the classification, the accuracy is achieved by 95.4% [23]. In the literature, the DNN model provided a good result compared to other classification models [29, 30]. DNN, CNN, and RNN models are generally used to classify voice signals in the present era [31].

3. Materials and Methods

Lungs can control the airflow due to the movement of the muscular action. Lungs also work as an activator that can generate the source to produce the vocal fold vibration to generate the voice signals. The plasticity of voice signals is varied and controlled according to the movement of vocal folds. The characteristics of the voice signals also depend on the body's resonators (Chest, Sinus, and face) [32]. The voice signals are divided into small spans of time to find stability in the voice signal. To characterize the voice signals, short-time spectral analysis can be utilized. In this experiment, the frequency of the recorded voice samples is 44.1 kHz. In the training phase, several numbers of the voice samples in each database group are used to extract the feature MFCCs for training the system. With the help of classification techniques, the system shows the result with the help of a confusion matrix for the identification of genders. These classifiers act as the decision-maker. The final output of the classifier provides the gender of the speakers. This article uses RNN–BiLSTM, LDA, and SVM as a classifier to classify gender.

The main principle of the gender identification system is to extract the features from the voice signals and provide the decision after comparing the extracted features with stored feature vectors. The gender identification system has two phases: (1) the training phase and (2) the testing phase, as shown in Figure 1.

In the first phase, several numbers of the voice samples are available in the database for the training. In the training process, the extraction of the feature is started with preprocessing. In the preprocessing, the energy level of the high-frequency voice signals increases. After the preprocessing, the feature of the voice signals is extracted. The extracted features are stored in the database known as a feature vector. In the testing phase, the features' preprocessing and extraction are similar to the training phase. After extracting the features, the classification algorithm is compared with the stored database. After comparing the features, the classification algorithm provides the decision in terms of gender.

3.1. About the Database. The common voice dataset is utilized for the identification of gender. Voice signals are recorded in the mp3 format. The duration of the voice samples varied from 3 seconds to 9 seconds. The voice signals are recorded at 44 kHz. These voice samples belong to three genders, i.e., male, female, and transgender. The voice sample datasets have ordinary person voice samples in both languages, i.e., Hindi and English. The sets of voice samples are created to clarify the result better. Every set has different numbers of testing and training samples, as shown in Table 1. The recording is done in different indoor and outdoor environments, i.e., public places and houses. In the proposed study, first time, several transgender samples are recorded to analyze voice signals. Table 1 shows the details of the number of samples used to analyze the voice signals.

3.2. Feature Extraction of Voice Samples. For the gender identification system's high accuracy, the extracted feature selection shows a very important role. The extracted features are the important and essential input for the classifier because the extracted feature contains beneficial information about the speakers. The main motive for collecting and extracting the features from the voice signals is to reduce the search space for the classifier. The short-term spectral voice signals are required to analyze the voice signals because the function of the human ear is similar to the quasi-frequency analysis. The analysis of the auditory nerve also depends on the mel frequency scale. The frequency-domain features of the voice signals have less noise than the time domain features of the voice signals. So, MFCC is used as a frequency-domain feature in the proposed model. The extraction of the MFCC is possible if the voice signals seem stationary for a short-time [33].

3.3. Mel Frequency Cepstral Coefficients (MFCC). MFCC is developed by Davis and Mermelstein [11]. MFCC has information about the several parameters of the speakers. So, MFCC is widely used as a feature of voice signals to identify gender with the help of voice signal analysis [17]. In the proposed model, MFCC is used as a unique feature to identify the gender of the speakers. Mel-frequency cepstrum lies on an equally spaced frequency band on the mel scale and shows a response similar to the human auditory system [34]. The extraction of MFCC is illustrated in Figure 2.

To extract the MFCC, the pre-emphasis is done in the first step. In this step, the energy of the high-frequency voice signals is amplified because the high-frequency voice signal's energy is more diminutive than the low-frequency voice signal's energy. So, gender identification accuracy is enhanced by this process. It can be assessed by (1), where B is the output signal, A is the input signal, C is constant, and m is the signal samples [9].

$$B(m) = A(m) - C * A(m-1). \tag{1}$$

Framing is the next step in the extraction of the MFCC. In this process, the voice signals are divided into small segments. These segments are useful to represent the stationary features of the voice signals. These voice samples are segregated into N samples. In the next step, the windowing removes the discontinuities from the samples. After implementing the window function, the voice signals can be calculated as equation (2).

$$B(m) = A(m) * W(m), \tag{2}$$

where $W(m)$ is the hamming window [9].

Now, the edges of the signals are smooth. Fast Fourier transform (FFT) is used to identify the spectrum of the voice signals. This step converts time-domain voice signals into frequency-domain voice signals. The output of the process of FFT is applied to the mel filter bank because the human auditory response is always logarithmic. The output of the mel filter bank can be obtained with the help of equation (3).

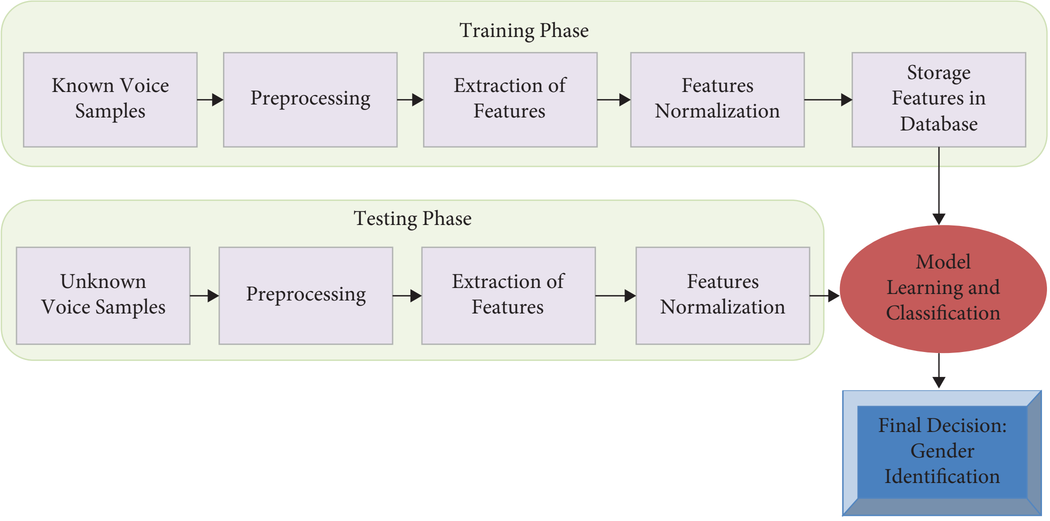

Figure 1: Model for voice signal processing.

Table 1: Datasets for error rate and precision values.

	Data sets	Male	Female	Transgender	Total
Set–1	Training	80	80	80	240
	Testing	20	20	20	60
Set–2	Training	40	40	40	120
	Testing	20	20	20	60
Set–3	Training	88	83	82	253
	Testing	12	17	18	47
Set–4	Training	40	40	40	120
	Testing	12	17	18	47
Set–5	Training	80	80	80	240
	Testing	12	17	18	47
Set–6	Training	40	40	40	120
	Testing	20	20	20	60

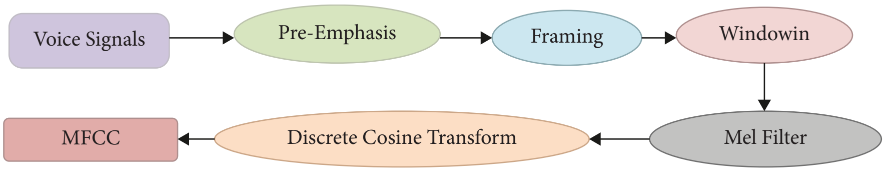

Figure 2: Extraction process of MFCC.

$$F\,(\text{Mel}) = 2595 * \ \ln\left(1 + \frac{f}{100}\right), \tag{3}$$

where f is the input frequency in Hz and $F(\text{Mel})$ is the frequency of the output signal [50].

Discrete cosine transform (DCT) is used in the last step to extract the MFCC from the voice signals. The frequency-domain voice signal is again converted into the time domain voice signals in this process. The outcome of the metamorphosis is known as MFCC. The collection of the MFCCs is called the acoustic vectors.

3.4. Classification Algorithm. The process of classification is similar to the process of supervised learning. The classification algorithms are used to segregate the class of the genders of the speakers. The classifier selection is the most challenging task for achieving high gender identification accuracy. The classifier mapped the features of the tested voice signals with the stored features of the training voice signals to decide the gender of the speakers. Several classification algorithms are available to identify gender, such as SVM, GMM, LDA, RNN, and HMM. The proposed work uses the RNN-BiLSTM, LDA, and SVM algorithm as a classifier.

3.5. Support Vector Machine (SVM). SVM is a very powerful algorithm to identify gender with the help of voice signals. The main aim of the SVM is to fix the hyperplane according to the features that differentiate the genders. With the help of the hyperplane, the SVM can execute binary classification [35]. The data points, which lie near the hyperplane, are known as support vectors. The separation of the support vector is complicated from the other data points available near the hyperplane. The margin value decides the classification of the unknown samples. The margin is the perpendicular line from the hyperplane [3, 36, 37]. To classify the nonlinear data, SVM can be utilized as suitable kernels such as polynomial, radial basis function, and multilayer perceptron. The kernel develops an apparent expansion into SVM feature space [37, 38].

3.6. Linear Discriminant Analysis (LDA). LDA is generally used to segregate between two labels or many labels. If the classes are two, then labels are classified linearly with the help of one hyperplane. On the other hand, several hyperplanes are required to segregate the classes in multiple discrimination. The hyperplane is developed according to the following principles: (i) the distance between the two labels is maximum and (ii) the variation of the values of the features should be minimum in both labels [39].

3.7. Recurrent Neural Networks. The artificial neural network (ANN) is a nonlinear classification algorithm that acts as a human brain. In the training phase of the process, the biases and weights are regularly adjusted according to the input signals. This process is continuously executed until the variation in the values of the bias and consequences are negligible [40–42]. The conventional ANN combines three layers: an input layer, one hidden layer, and an output layer. RNN is the family member of the ANN classification algorithms. RNN has the special ability to process sequential data such as voice signals and time series. That output of the RNN unit forwards to the next unit of the RNN and loops back to it. The inputs of the RNN algorithm are of two types: (a) present input and (ii) previously applied input. To predict the following input, the previous input sequence is the essential component of the RNN algorithm [43, 44]. The drawback of the RNN is that it has a short memory. By using the long-short term memory (LSTM), RNN can enhance long-term memory capability. LSTM-RNN is the combination of several LSTM cells. With the help of these cells, the movement of the information in the network can be controlled more efficiently. LSTM has three types of gates: (i) input gate, (ii) forget gate, and (iii) output gate. With the operation of the gate, the LSTM cell can decide the movement of the information. An LSTM layer can perform its operation only in the forward direction, while on the other hand BiLSTM layer can perform its operation in both forward and backward directions. BiLSTM is the combination of two LSTM layers. The first LSTM layer can operate in the forward direction, while the second layer can perform its operation in the reverse direction of the first one. The main objective of the BiLSTM is to capture the future and past input features for the specific time setup. The network's behavior depends on two conditions, namely, (i) the current input and (ii) the output of the recent past input.

4. Results and Discussion

The voice signal is the combination of several characteristics and information about the speaker, such as age, gender, and emotions. The gender identification of the speaker has several applications with the analysis of voice signals. In this article, the MFCC feature with LDA, RNN-BiLSTM, and SVM classifiers is used to determine the gender identification accuracy and the recall values of each gender, i.e., male, female, and transgender. The confusion matrix is utilized to examine the model's degree of gender identification. The recall value of the genders can be calculated as the true values divided by the total values, as shown in (4). The identification accuracy of the classifiers is obtained by the total true values divided by the total values, as shown in equation (5) [1].

$$\text{Recall Values} = \frac{\text{TP}}{\text{TP} + \text{TN}}, \tag{4}$$

$$\text{Overall Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100, \tag{5}$$

where TP represents True Positive. TN is True Negative. FN indicates the value of False Negative, and FP is False Positive.

A confusion matrix is used to compare the decision made by the classifier. The classifier's accuracy also depends on the quality of the voice samples. The accuracy of the proposed model depends on the number of voice samples in the training and testing data sets. Table 2 shows the male recall values for the different types of classifiers and the datasets. Table 2 is designed with the help of a confusion matrix. The first row of Table 2 contains the male recall values of the proposed LDA model. Similarly, the following rows contain the male recall values of the proposed RNN-BiLSTM and SVM model. Table 2 also shows the average male recall values of the all-proposed models.

Table 3 shows the female recall values for the different types of datasets of the voice samples. LDA, RNN-BiLSTM, and SVM classification algorithms are used to compute the female recall values. The values of Table 3 are calculated with the help of a confusion matrix. Table 3 also shows the average female recall values of the proposed models according to the recall values computed for the different datasets.

Table 4 shows the transgender recall values for the different classifiers and the datasets. Table 4 is designed with the help of a confusion matrix. The transgender recall values for the LDA classification algorithm are available in the first row of Table 4 for the different datasets of the voice samples. Similarly, the following rows contain the transgender recall values of the proposed RNN-BiLSTM and SVM model. Table 4 also shows the average transgender recall values of the all-proposed models.

Figure 3 shows the average recall values of all three classification algorithms to identify the male voices for the different training numbers and Test three gender voice samples.

Table 2: Average and recall values for proposed models to identify the male voice samples.

Algorithms	Male						Average
	Data sets						
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	
LDA	77.11	76.51	75.66	74.95	68.68	68.80	73.62
RNN-BiLSTM	90.15	89.41	95.63	93.29	88.87	86.39	90.62
SVM	75.60	75.00	78.50	75.81	67.63	68.51	73.51

Table 3: Average and recall values for proposed models to identify the female voice samples.

Algorithms	Female						Average
	Data sets						
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	
LDA	70.04	69.37	62.33	59.03	61.67	61.17	63.94
RNN-BiLSTM	96.71	94.62	94.86	91.43	92.20	90.48	93.38
SVM	72.63	71.34	65.05	62.21	66.32	64.72	67.05

Table 4: Average and recall values for proposed models to identify the transgender voice samples.

Algorithms	Transgender						Average
	Data sets						
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	
LDA	77.35	78.13	81.62	83.95	79.18	78.93	79.86
RNN-BiLSTM	97.19	94.97	97.22	96.35	95.06	93.31	95.68
SVM	74.77	74.74	81.88	83.98	76.61	76.76	78.12

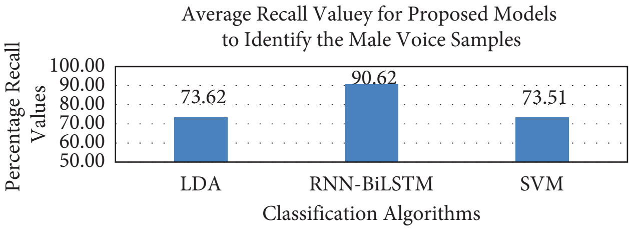

Figure 3: Average recall value for proposed models to identify the male voice samples.

Similarly, the average recall values for all three classification algorithms are calculated from the confusion matrix to identify the female speakers. Figure 4 shows the average recall values of the female speakers.

Figure 5 represents the average values of all three classification algorithms to identify the transgender voices from the testing datasets containing all three gender voice samples.

The accuracy of the identification of gender and the performance of the system depends on the types of classifiers. The accuracy result varies according to the changes in the classification techniques. The recall values also vary according to the variation in the number of training and testing voice samples. Three classification algorithms are used to compare the accuracy of the identification of the gender with MFCC. The RNN-BiLSTM algorithm demonstrates the maximum accuracy and recall values in differentiating the three gender groups. The accuracy result is achieved 93.41% by the RNN-BiLSTM. The worse accuracy is achieved 72.42% by the LDA in comparison with other classifiers. On the other hand, the identification accuracy reached 73.36% by SVM, as shown in Figure 6. Recall values of the transgender are computed as 95.68%, the highest in comparison with other genders. The MFCCs, extracted feature of the voice signal, plays a significant and vital role in achieving this milestone. The function of the MFCC is to represent the envelope of the short-term spectrum of the voice signals, which is the appearance of the shape of the vocal tract. Transgender people have unique and different vocal tract, so the performance of the system for identification of the transgender can achieve a good result. Table 5 shows the comparison status for gender identification. The recall values for several numbers of the transgender voice samples through the analysis of the voice signals are reported first time in this literature which is achieved based on the analysis of the voice signals. The identification accuracy of all three classification algorithms is compared with the reported literature.

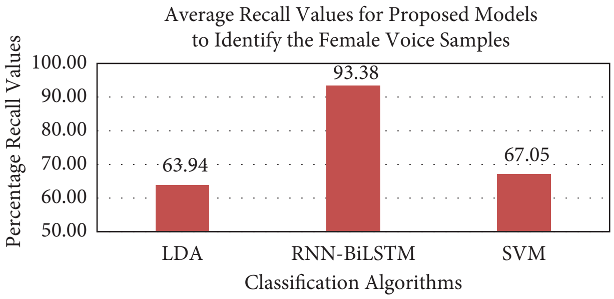

Figure 4: Average recall value for proposed models to identify the female voice samples.

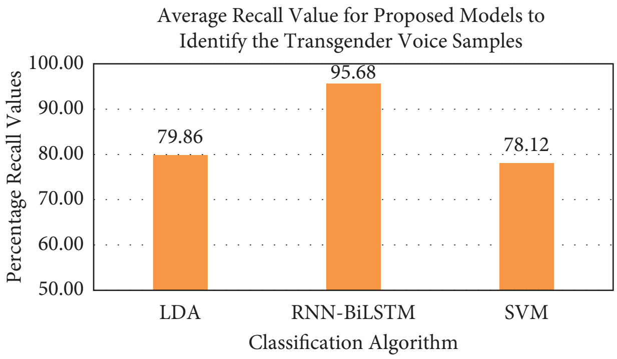

Figure 5: Average recall value for proposed models to identify the transgender voice samples.

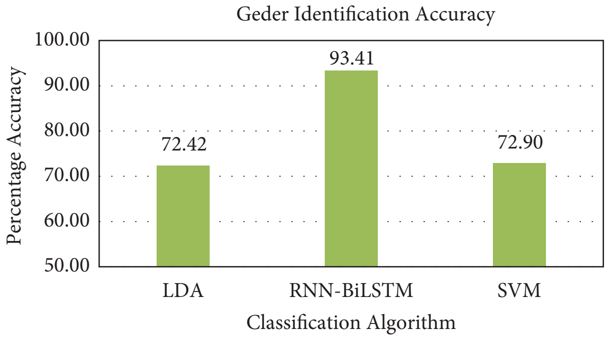

Figure 6: Percentage accuracy comparison for the different classifiers.

Table 5: Comparison of the recall values and identification accuracy with reported in the literature.

	Comparison with reported models							
	LDA	RNN-BiLSTM	SVM	LVQ [45]	BiLSTM [46]	CNN [47]	Neural network [48]	CNN & LSTM [49]
Recall values of male	73.62	90.62	73.51	94.60	—	97.00	89.7	—
Recall values of female	63.94	93.38	67.05	94.60	—	91.00	88.1	—
Recall values of transgender	79.86	95.68	78.12	—	—	—	87.3	—
Identification accuracy	72.42	93.41	72.90	94.60	86.70	96.00	—	97.00

5. Conclusion

The main objective of this article is to estimate the accuracy of the proposed models and recall values for the identification of male, female, and transgender. For this purpose, three classifiers and one feature are considered. The identification of the transgender through live recorded voice samples is done for the first time with the help of voice analysis. The common voice sample dataset for the experiment consists of male speakers, female speakers, and transgender speakers. The recall values for the three genders are calculated with the voice samples recorded at 44.1 kHz. The accuracy for SVM, RNN-BiLSTM, and LDA is calculated after analysis of the voice samples and comparison. The transgender gender achieves the best recall value for identifying the gender, and the male shows the lowest recall value for the same purpose. The MFCCs are used as a feature of the voice signals. The RNN-BiLSTM has achieved an accuracy of 93.41%, higher than the other classifiers. The LDA classifier shows the worst accuracy. The proposed RNN-BiLSTM model achieved the recall values of 90.62%, 93.38%, and 95.68% for male, female, and transgender classification, respectively, for the common voice samples datasets. The performance and accuracy result of the classifier are also varied when the number of samples of the voice signals is also varied. Hence, the classifier's performance is enhanced by using a large number of samples. The accuracy of gender identification is varied according to the type of classifier. For better accuracy in identifying the gender, the emotion of the speakers is also included because the emotion of the male and females has extensive and different characteristics. The accuracy of the system may also be increased by using the majority rule in the confusion matrix. The accuracy of the model also depends on the quality of the voice samples. So, the noise should be significantly less while recording the voice samples. Health condition with age is an essential parameter of the voice signals of the speaker. A new hybrid model can also be designed to classify gender with age, emotion, and health condition [51].

Data Availability

The Common Voice dataset is recorded for the identification of the gender. The voice samples data, which is used to identify the accuracy, are not available within the any published article, publish research papers, websites, and other formats of the file(s).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Jchaudhari and R. M Kagalkar, "Automatic speaker age estimation and gender dependent emotion recognition," *International Journal of Computer Application*, vol. 117, no. 17, 2015.

[2] A. Singhal and D. K. Sharma, "Analysis of classifiers for gender identification using voice signals," in *Proceedings of the 2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, pp. 1–4, Mathura, India, October 2021.

[3] E. Ramdinmawii and V. K. Mittal, "Gender identification from speech signal by examining the speech production characteristics," in *Proceedings of the 2016 International Conference on Signal Processing and Communication (ICSC)*, pp. 244–249, Noida, India, December 2016.

[4] V. Y. Mali and B. G. Patil, "Human gender classification using machine learning," *International Journal of Engineering Research and Technology*, vol. 8, pp. 474–477, 2019.

[5] A. Acero and X. Huang, "Speaker and gender normalization for continuous-density hidden Markov models," in *Proceedings of the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, USA, May 1996.

[6] R. Djemili, H. Bourouba, and A. Korba, "A combination approach of Gaussian mixture models and SVMs for speaker identification," *The International Arab Journal of Information Technology*, vol. 6, no. 5, pp. 490–497, 2009.

[7] V. Passricha and R. K. Aggarwal, "Convolutional support vector machines for speech recognition," *International Journal of Speech Technology*, vol. 22, no. 3, pp. 601–609, 2019.

[8] V. Patil, R. Vineetha, S. Vatsa et al., "Artificial neural network for gender determination using mandibular morphometric parameters: a comparative retrospective study," *Cogent Engineering*, vol. 7, Article ID 1723783, 2020.

[9] O. S. Faragallah, "Robust noise MKMFCC–SVM automatic speaker identification," *International Journal of Speech Technology*, vol. 21, no. 2, pp. 185–192, 2018.

[10] M. K. Mishar and A. K. Shukla, "A survey paper on gender identification system using speech signal," *International Journal of Engineering and Advanced Technology*, vol. 6, pp. 165–167, 2017.

[11] M. Gupta, S. S. Bharti, and S. Agarwal, "Gender-based speaker recognition from speech signals using GMM model," *Modern Physics Letters B*, vol. 33, no. 35, Article ID 1950438, 2019.

[12] K. H. Lee, S. I. Kang, D. H. Kim, and J. H. Chang, "A support vector machine based gender identification using speech signal," *IEICE - Transactions on Communications*, vol. E91-B, no. 10, pp. 3326–3329, 2008.

[13] R. R. Rao, "Source feature based gender identification system using GMM," *International Journal on Computer Science and Engineering*, vol. 3, no. 2, pp. 586–593, 2011.

[14] R. Rajeshwara Rao and A. Prasad, "Glottal excitation feature based gender identification system using ergodic HMM," *International Journal of Computer Application*, vol. 17, no. 3, pp. 31–36, 2011.

[15] S. Agrawal and S. Rathor, "A robust model for domain recognition of acoustic communication using Bidirectional LSTM and deep neural network," *Neural Computer & Application*, vol. 33, pp. 11223–11232, 2021.

[16] C. Muller, F. Wittig, and J. Baus, "Exploiting Speech for Recognizing Elderly Users to Respond to Their Special Needs," in *Proceedings of the Europspeech 2003 Geneva*, pp. 1305–1308, Geneva, Switzerland, 2003.

[17] H. Kim, K. Bae, and H. Yoon, "Age and gender classification for a home-robot service," in *Proceedings of the RO-MAN 2007 - the 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 122–126, Jeju, Korea, August 2007.

[18] M. A. Nasr, M. Abd-Elnaby, A. S. El-Fishawy, S. El-Rabaie, and F. E. Abd El-Samie, "Speaker identification based on normalized pitch frequency and Mel Frequency Cepstral Coefficients," *International Journal of Speech Technology*, vol. 21, no. 4, pp. 941–951, 2018.

[19] D. Mahmoodi, H. Marvi, M. Taghizadeh, A. Soleimani, F. Razzazi, and M. Mahmoodi, "Age estimation based on speech features and support vector machine," in *Proceedings of the 2011 3rd Computer Science and Electronic Engineering Conference (CEEC)*, pp. 60–64, Colchester, UK, July 2011.

[20] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[21] J. Villalba, N. Chen, D. Snyder et al., "State-of-the-Art speaker recognition for telephone and video speech: the JHU-MIT submission for NIST SRE18," in *Proceedings of the Interspeech 2019*, pp. 1488–1492, Graz, Austria, September 2019.

[22] F. Ertam, "An effective gender recognition approach using voice data via deeper LSTM networks," *Applied Acoustics*, vol. 156, pp. 351–358, 2019.

[23] I. E. Livieris, E. Pintelas, and P. Pintelas, "Gender recognition by voice using an improved self-labeled algorithm," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 492–503, 2019.

[24] S. Malmasi, "Discriminating between similar languages and Arabic dialect identification: a report on the third DSL shared task," in *Proceedings of the Third Workshop on {NLP} for Similar Languages, Varieties and Dialects*, pp. 1–14, Osaka, Japan, December 2016.

[25] B. Bsir and M. Zrigui, *Bidirectional LSTM for author gender identification*, Springer, Berlin, Germany, pp. 393–402, 2018.

[26] L. Stout, R. Musters, and C. Pool, "Author profiling based on text and images," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association*2018 Evaluation Labs and Workshop - Working Notes Papers, Avignon, France, September 2018.

[27] B. Bsir and M. Zrigui, *Gender Identification: A Comparative Study of Deep Learning Architectures*, pp. 792–800, Springer, Berlin, Germany, 2018.

[28] D. Bhardwaj and R. K. Galav, "Identification of speech signal in moving objects using artificial neural network system," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 4, pp. 418–424, 2020.

[29] M. Buyukyilmaz and A. Osman, "Voice gender recognizer using deep learning," *Advances in Computer Science Research*, vol. 58, pp. 409–411, 2016.

[30] R. V. Sharan and T. J. Moir, "Robust acoustic event classification using deep neural networks," *Information Sciences*, vol. 396, pp. 24–32, 2017.

[31] Ö. B. Ḋinler and N. Aydin, "An optimal feature parameter set based on gated recurrent unit recurrent neural networks for speech segment detection," *Applied Sciences*, vol. 10, no. 4, p. 1273, 2020.

[32] Saloni, R. K. Sharma, and A. K. Gupta, "Classification of high blood pressure persons vs normal blood pressure persons using voice analysis," *International Journal of Image, Graphics and Signal Processing*, vol. 6, no. 1, pp. 47–52, 2013.

[33] S. G. Koolagudi, Y. V. S. Murthy, and S. P. Bhaskar, "Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition," *International Journal of Speech Technology*, vol. 21, no. 1, pp. 167–183, 2018.

[34] A. A. Malode and S. Sahare, "Advanced speaker recognition," *International Journal of Advances in Engineering & Technology*, vol. 4, no. 1, pp. 443–455, 2012.

[35] B. Jena, A. Mohanty, and S. K. Mohanty, "Gender recognition of speech signal using KNN and SVM," in *Proceedings of the International Conference on IoT based Control Networks and Intelligent Systems*, pp. 548–557, January 2021.

[36] S. Jchaudhari and R. Kagalkar, "Automatic speaker age estimation and gender dependent emotion recognition," in *International Journal of Computer Application*, vol. 117, no. 17, pp. 548–557, Kottayam, India, December 2020.

[37] M. Gupta, S. S. Bharti, and S. Agarwal, "Support vector machine based gender identification using voiced speech frames," in *Proceedings of the 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 737–741, Waknaghat, India, December 2016.

[38] S. V. N. Vishwanathan and M. NarashimaMurty, "SSVM: a simple SVM algorithm," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, vol. 3, pp. 2393–2398, Honolulu, HI, USA, May 2002.

[39] C. Castaldelloa, A. Guberta, F. Galvaninb et al., "A model-based support for diagnosing von Willebrand disease," *Computer Aided Chemical Engineering*, vol. 27, pp. 2779–2784, 2017.

[40] M. K. Reddy and K. S. Rao, "Excitation modelling using epoch features for statistical parametric speech synthesis," *Computer Speech & Language*, vol. 60, Article ID 101029, 2020.

[41] Y. Liu, L. He, J. Liu, and M. T. Johnson, "Introducing phonetic information to speaker embedding for speaker verification," *Journal on Audio, Speech, and Music*, vol. 19, 2019.

[42] A. Greco, A. Saggese, M. Vento, and V. Vigilante, "A convolutional neural network for gender recognition optimizing the accuracy/speed tradeoff," *IEEE Access*, vol. 8, pp. 130771–130781, 2020.

[43] L. Jasuja, A. Rasool, and G. Hajela, "Voice Gender Recognizer Recognition of Gender from Voice Using Deep Neural Networks," in *Proceedings of the 2020 International Conference on Smart Electronics and Communication*, pp. 319–324, Trichy, India, September 2020.

[44] V. Pratap, Q. Xu, A. Sriram, G. Synnaeve, and R. Collobert, "MLS: A Large-Scale Multilingual Dataset for Speech Research," 2020, https://arxiv.org/abs/2012.03411.

[45] R. Djemili, H. Bourouba, and M. C. A. Korba, "A speech signal based gender identification system using four classifiers," in *Proceedings of the 2012 International Conference on Multimedia Computing and Systems*, pp. 184–187, Tangiers, Morocco, 2012.

[46] R. D. Alamsyah and S. Suyanto, "Speech Gender Classification Using Bidirectional Long Short Term Memory," in *Proceedings of the 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 646–649, Yogyakarta, Indonesia, December 2020.

[47] A. Tursunov, J. Y. Choeh, and S. Kwon, "Age and gender recognition using a convolutional neural network with a specially designed multi-attention module through speech spectrograms," *Sensors*, vol. 21, no. 17, p. 5892, 2021.

[48] G. Yasmin, O. Mullick, A. Ghosal, and A. K. Das, "Gender recognition inclusive with transgender from speech classification," *Emerging Technologies in Data Mining and Information Security*, Springer, Singapore, 2019.

[49] T. J. Sefara and T. B. Mokgonyane, "Gender identification in sepedi speech corpus," in *Proceedings of the 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1–6, Durban, South Africa, August 2021.

[50] A. Bala, A. Kumar, and N. Birla, "Voice Command Recognition System Based On mfcc And dtw," *International Journal of Engineering Science and Technology*, vol. 2, no. 12, pp. 7335–7342, 2010.