Received: February 27, 2023. Revised: May 28, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Precision Value, Error Rate and Accuracy of Human Gender Identification Based on Randomized Voice Signals Datasets

Abhishek Singhal[1]* **Devendra Kumar Sharma[1]**

[1]*Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Delhi – NCR Campus, Ghaziabad, UP, India*
* Corresponding author's Email: abhisheksinghal.srm@gmail.com

Abstract: Voice signals are the most convenient and essential method to establish the communication between human beings. The linguistic characteristics of the human being can be recognized with the help of voice signal analysis such as the ascent of language, health status, emotions, age, gender, *etc*. Gender is an unprecedented peculiarity of the speakers. This peculiarity can be recognized through the analysis of the voice signals. This paper presents the precision value of gender, i.e., transgender, female and male. The live voice samples for females and males are recorded as well as the live voice samples for the transgender in India are recorded first time because the voice samples of the transgenders are not available at the authenticated sources. The error factor and accuracy is also calculated for the proposed model, which is utilized for the gender identification of the speakers by analyzing the voice signals. In the proposed work, bidirectional long short-term memory (BiLSTM) architecture with recurrent neural network (RNN) model is used as a classification algorithm with mel frequency cepstral coefficients (MFCCs). The MFCC is utilized as a trait of the signals used to identify the gender of the human beings. The outcomes of the proposed model are not dependent on the content of the articulation and the dialect of the speakers. The precision values for female, male and transgender, the accuracy and error rate of the proposed model, depend on applied quantities of voice signals in testing and training datasets. The average precision value is 93.91%, 94.70%, 97.43% and 95.35%for the male, transgender, female and proposed model, respectively. It is found that the female precision value is the highest, and the male has the lowest. Further, the average error rate of the proposed model is 0.045. The gender identification accuracy is 95.52%.

Keywords: Recurrent neural network – bidirectional long short-term memory, Mel frequency cepstral coefficients, Accuracy, Precision value, Gender identification, Error rate, Common voice signals.

1. Introduction

Vox signals are an essential and vital path to establish the conversation between human beings. Voice signals are evoked due to the synchronized motions in the vocal tract. Linguistic parameters of speech signals can vary due to the movement, shape, and position of the vocal tract. So, the voice signals can provide the characteristics of the human beings, such as health status, age, gender, affective factors, *etc.* Due to these traits of the human being, voice signal analysis has multifarious usage, such as identification of the gender, age recognition, emotion identification, *etc.*

Gender of the speaker is a crucial and essential characteristic. In the field of voice signal analysis, gender identification is a methodology utilized to identify biological status of speaker. The model for the gender identification experiences many challenges, such as changes in the environmental noises, voice signals, the area of the recording, *etc.* So, gender identification through the voice signal is a very typical and challenging task. The algorithms for identifying gender have a variety of applications in several fields, such as computerized educational systems, medical fields, criminal cases, *etc.* The voice signal analysis benefits healthcare systems, disabled people, military systems, *etc*. Presently, researchers pay excellent attention towards the identification of gender because the gender identification system is an essential requirement in the rapidly developing computerized world. In the area of the voice signal analysis, several types of identification systems for genders are acquirable as: gender-dependent, gender-dependent, voice recognition systems, *etc.* The gender identification model can be split into two processes: testing process and training process.

In the training process, the gender identification system has to be trained with the voice signals of the speakers. The feature extraction from the voice signals is performed at the initial stage of the training process. The voice signal features are accumulated at one point, called the feature vector. These feature vectors are basically utilized to train classification algorithms to differentiate the speakers according to gender. In the testing process, feature extraction from the unknown voice samples is just like the process of training. The extracted traits of unknown voice signals are applied to the trained classification algorithm, and the algorithm will segregate the gender of the human beings. It means that the identification model combines classification algorithms, voice signals acoustic characteristics, and recorded voice samples. In this process, the feature extraction is a very challenging and typical task because the acoustic features of the speech signals are varied pursuant to the variation in the status of the health, emotions of the human beings, environment, *etc.*

The gender identification system has several types of classification algorithms. In the present scenario, machine learning algorithms are highly recommended in the field of research. The machine learning algorithms such as artificial neural networks (ANN) and RNN are used in several fields such as business, institutions and industries. RNN is the unique technique of the ANN algorithm with a minimum of three layers. The function of the RNN is similar to the function of the human brain. In this article, the machine learning algorithm, i.e., RNN is used to identify all three genders of the speakers. The proposed model can predict the gender of human beings with the help of acoustic parameters of voice signals, i.e., MFFCs, bidirectional long short-term memory with RNN and live recorded voice samples. The voice samples are recorded with the help of normal common speakers in a noisy environment. These recorded voice samples are text-independent. An ADAM optimization technique in classification algorithm is utilized to obtain the objective of the proposed work.

Precision value for identification of gender based on voice signal analysis through the execution of several classification algorithms such as genetic algorithms (GA), fuzzy logic with neural network (FL with NN), neural network (NN) and naïve bayes (NB) were computed [1]. The average error factor for the gender identification systems was also reported in the literature by the researcher [2, 3]. Several classification techniques, such as deep neural network (DNN) with three layers, CNN with full weight sharing (FWS), support vector regression (SVR-baseline system), *etc.*, was used to recognize the gender to achieve the gender identification accuracy. The error factor, accuracy and precision value of identification model for gender of speakers was computed by utilizing voice signals analysis for only male and female in the reported literature. The researchers always ignore the voice signals of the third gender. In the field of voice signal analysis, the transgender identification is acquainted for the ‘first time’ with the identification of others, *i.e.* male and female, with help of voice signals that are recorded from the common human beings. The primary purpose of this article is to compute the precision values of the male, female, and transgender. The error factor and accuracy of the proposed model is also computed. All values of the precision values, error factor and accuracy are computed after analyzing the voice signals of the common human being by using BiLSTM-RNN with MFCCs.

The simulation is conducted, and results are analyzed using several datasets and epochs. The precision values for female, male and transgender are computed by the predicted value obtained from the confusion matrix after the execution of the classification algorithm. The accuracy and error factor are also computed after the execution of the proposed algorithm. The precision value for the transgender is also computed first time with the help of live recorded voice samples. Similarly, the accuracy and error factor for the proposed model is also calculated based on female, male and transgenders. 97.43% is the maximum precision average for the female class, while 93.91% is the minimum average for the male class. The average error rate is 0.045 for the proposed model. Similarly, the gender identification accuracy is 95.52%. The simulated results are also compared with the reported literature. The remaining part of the present article is distributed in four sections. The algorithms, features and outcomes which are available literature is summarised and briefly explained in section 2. Section 3 briefly describe the particulars used to simulate and analyse the voice signals. This section also provides detailed information about voice sample datasets. The results after the simulation and analysis of the voice signal with the help of classification are explained and discussed in section 4. The proposed work is summarized in the followed section.

2. Literature review

Gender identification in field of voice signal analysis behaves as a hot cake in the present area of the research. The features, such as MFCC, energy, pitch, *etc.*, have been extracted from the voice signals to recognize the gender of the speaker. These acoustic parameters can be used in both phases of the gender identification system [4]. MFCC was introduced to identify gender in the year 2012. After the evolution of MFCC, several changes in the property and characteristics of the feature have been taken place in the field of analysis of voice signals to improve the efficiency of the gender identification system. Gender identification has been efficiently done in the different domains with the help of MFCC [5] because these parameters vary according to the language and text of the speakers [6].

The gender identification system is utilized to segregate male and female voice samples after the unknown voice samples analysis [7]. Several classification algorithms, such as GMM, HMM, SVM, *etc.*, are utilized to recognize the gender of speakers and compare the calculated result by using a method of analysis of the voice signals [8-10]. For the binary classification of gender, the SVM classification algorithm is commonly used. SVM showed good identification accuracy. Some kernels are added to the SVM algorithm to improve the identification accuracy to increase accuracy. Gaussian radial basis function SVM shown remarkable identification accuracy compared to other kernels of SVM [11]. Another gender classification algorithm is the Gaussian mixture model (GMM) [12]. Gender identification accuracy for the classification system was 97.5% which was computed after simulation the algorithms through the voice sample analysis. For gender identification accuracy, voice signals are recorded in noisy environments [13]. 96.4% identification accuracy for the GMM classification system was calculated. The accuracy was computed for the different voice sample datasets [14].

The dynamic time warping (DTW) technique is utilized for feature matching [6, 15]. The gender identification accuracy is 93% for the text-independent gender identification system. The gender identification accuracy can be increased with the help of more utilizing spectral features of the voice signals. The gender identification accuracy can be enhanced by 5% for the VQ classification algorithm by using spectral features of the voice signals. The system accuracy can be degraded using noisy voice signals [16, 17]. The text-independent identification system has less accuracy in comparison with text dependent system. The accuracy of the system can be affected by the age of human beings. Gender classification of the young speaker shows more accurate outcome than the old speaker [18, 19].

Contour of the male shows a lower value in compare to the female speaker [11]. The feature ‘Energy’ of the voice also varied as per the variation the gender of the speakers. The energy of female shows a high value compared to male speakers. The identification accuracy for multilayer perceptrons (MLP) classification algorithm is computed as 96.4%.

Similarly, the identification accuracy for vector quantization (VQ) was also the same as the identification accuracy of MLP. A hybrid algorithm was developed with the combination of neural networks and VQ. The resultant algorithm is known as learning vector quantization (LVQ). LVQ has also achieved 96.4% accuracy in the system. In comparing GMM, MLP, VQ and LVQ, LVQ shows worst identification accuracy of the system [14].

I – vectors algorithm was the baseline for several renowned embedding methods before developing deep learning algorithms. The universal background and the projection matrix can be learned in the probabilistic linear discriminant analysis (PLDA) [20]. X – vector algorithm is a deep neural network (DNN) based embedding algorithm [21]. The LSTM model with DNN achieves 98.4%, the highest gender identification accuracy compared with other classification algorithms [22]. The exact value of the system accuracy is achieved with the help of iCST-Voting, a semi-supervised approach for identifying gender by using the voice signal analysis [23]. The gender identification accuracy is calculated at 78.8% and 62.3% for the RNN and native bayes algorithms, respectively [24, 25].

Gender identification accuracy is achieved by more than 90% using the SVM classifier [26]. This accuracy can be enhanced by using the DNN. The working of the DNN is based on learning through the input signals instead of conventional learning algorithms. Many reported kinds of literature are available to decide that deep learning algorithms are the best to achieve excellent accuracy results. The accuracy of the system is achieved at 95.4% by using a deep learning algorithm [23]. The reported literature describes that the DNN shows the highest gender identification accuracy by analyzing the voice signals [27, 28]. In the present scenario, the

Table 1. Datasets for precision values and error rate

	Data sets	Male	Female	Transgender	Total
Set – 1	Training	92	91	95	278
	Testing	8	9	5	22
Set – 2	Training	88	83	82	253
	Testing	12	17	18	47
Set – 3	Training	86	87	88	261
	Testing	14	13	12	39
Set – 4	Training	83	88	80	251
	Testing	17	12	20	49
Set – 5	Training	87	86	85	258
	Testing	13	14	15	42
Set – 6	Training	90	93	89	272
	Testing	10	7	11	28

CNN and RNN models are commonly used for gender identification through voice signals [29]. This is the main reason for selecting the RNN classifier to discriminate the genders of speakers.

3. Materials and methods

Airflow from the lungs, a source of voice signals, can be controlled by the muscular action. The voice signals have several variations, and the activity of the vocal folds controls these variations. Resonators of the body are also responsible for generating variations in the characteristics of the voice signals. The voice signals should have stable characteristics to gender identification of the human beings by utilization of the analysis of the voice signals. Stability of the characteristics can be achieved by dividing the voice signals into a tiny portion of time. Short-time spectral analysis is used to extract the characteristics of the voice signals. In general conditions, the small starting time has approximately the same characteristics as the voice signals for several speakers. The unique and useful tidings of the human beings is available in the remaining part of the voice signals. Live voice samples are recorded at 44.1 kHz for the proposed work.

The fundamental purpose of the gender identification model is basically the combination of two process. The unique feature of the voice signals is extracted and stored as a feature vector in the initial step. After this process, the feature of the voice samples for the unknown speakers is compared with stored feature vectors with the help of a classification algorithm. The gender identification model provides the decision about the genders of the unknown speakers. In the presented article, the proposed model for the gender identification is utilized MFCC as a feature of the voice samples to identify the gender. The RNN-BiLSTM classification algorithm is also used as a decision-maker. The result of the proposed model is in the form of a confusion matrix with information about the genders of the speakers.

3.1 Introduction to common voice dataset

For the proposed work, 300 voice samples in each dataset are utilized to recognize the gender of the speakers. These voice samples are distributed into six datasets. Every set has different training and testing samples to clarify the better result. The voice samples of all genders are randomly distributed in all six datasets, as shown in Table 1. In analyzing the voice samples, for the first time, the transgender voice samples are used to identify the gender. The voice samples for all three genders are recorded in a natural environment, i.e., outdoor and indoor and.mp3 format. The live voice signals are recorded from the ordinary person in India. These live voice samples are also recorded in different languages, i.e., English and Hindi.

3.2 Feature extraction

The extracted traits from signals play a significant role in the process of the identification of gender. These features have unique characteristics about the speakers. The identification accuracy of the classification algorithm depends on selecting the feature from the number of features. The feature is extracted from the voice signals and collected in one place, called a feature vector. The feature vector is basically used to reduce the search space for the classification algorithm. The operation of the ear is the same as the operation of the quasi-frequency signals. So, the voice signals are converted into short-term spectral voice signals. In the present article, MFCC is utilized as a feature because time domain features have noisier than the frequency. The feature can be extracted from a concise time span of the voice signal [30].

The frequency domain feature, MFCC, was introduced by Davis and Mermelstein [16]. The researcher widely utilizes MFCC to identify gender because it contains several pieces of information about the speaker [19-32].

The pre-emphasis is the initial step in the process of extraction of the MFCC. The process of pre-emphasis is performed to enhance the outcome of the identification model. Voice signals area combination of high-frequency and low-frequency signals. So, in this process, the vigour of high-frequency signals is augmented. The mathematical representation for pre-emphasis process is shown by Eq. (1).

$$y(n)=x(n)-0.95x(n-1) \qquad (1)$$

After the process of pre-emphasis, framing is done in second step. The voice signals should be contain the stationary property during the process of feature extraction. Using the framing process, the signals are converted into small pieces. These pieces can represent a stationary signal and are useful to represent the voice signal feature. The edges of these pieces are not smooths. These pieces do not seem continuous. So, the windowing function is utilized to smooth edges of pieces. After the windowing operation, fast fourier transform (FFT) converts from time domain to frequency domain signals. The spectrum of the signal is easily identified in frequency domain. The auditory model response is similar to the response of the logarithmic model. So, the Mel filter bank is used to achieve the same response. To extract the MFCCs, the output of Mel filter bank is again converted into the time domain from the frequency domain signal. This process is completed with the help of discrete cosine transform (DCT). The outcome of the process is known as MFCC.

3.3 Classification algorithm

The gender of the speakers is categorised by using the classification algorithms. It is an arduous task to select the classification algorithms because accuracy of the gender identification model also depends on the classification algorithm. The functions of the classification algorithms follow the same steps as the functions of the supervised learning algorithms. The classification algorithms always compare the features of the unknown voice samples with the feature vectors and provide the decision regarding the gender of the speakers. SVM, GMM, LDA, and RNN are examples of the several classification algorithms to segregate the genders of the speakers. The present article uses the RNN-BiLSTM algorithm with the ADAM optimization technique as a classification algorithm.

3.3.1. Recurrent neural networks

ANN is a non–linear classifier to identify the gender of speakers. The function of the ANN performs similar operations just like a human brain. The weights and biases continuously change as per the applied input signals in the training. The values of the weights and biases are continuously executed until a negligible variation in the values is achieved [33-35]. The architecture of ANN is the combination of three layers. The first and initial layer is the input layer. The hidden layer is the second layer. The output layer is the last layer of the architecture. RNN is one of the ANN classification algorithms. Voice signals, time series signals, *etc.*, combine sequential data. Such type of data can be efficiently processed by the RNN classification algorithm. The operation of the RNN is totally depend on the inputs. Two types of inputs are required to execute the RNN algorithm: (a) previously applied input and (ii) present input. The prediction of the upcoming input depends on the trailing applied input series, which is the important factor of the RNN algorithm [36, 37]. Short memory is the major drawback of the RNN algorithm. The classification accuracy can be increased by the improvement in the memory capacity of the classification algorithms. The long-term memory capability of classification algorithms can be increased by the utilization of BiLSTM.

3.3.2. BiLSTM layer

The BiLSTM layer is a set of two layers. The first layer performs the specified operation in one flank, while the other layer can operate in opposite of previous flank. These characteristics of BiLSTM have advantage over the long-short term memory (LSTM). The key function of the BiLSTM algorithm is to extract and store past and future input traits from the voice signals within the specified time period. So, the operation of gender identification model depends on the outcome of the recent past inputs and value of the current inputs.

One layer of the LSTM produces the hidden and cell states for the input, and the other layer produces the reverse order of the hidden and cell states. Both are combined to produce an output sequence for the BiLSTM layer, as shown in Eq. (2). Similarly, both are combined to produce an output sequence for the BiLSTM layer, as shown in Eq. (3) [38].

$$y_t = W_{\underset{h}{\rightarrow}t} \underset{h}{\rightarrow}_t + W_{\underset{h}{\leftarrow}t} \underset{h}{\leftarrow}_t + b_y \quad (2)$$

$$y_t = W_{\underset{c}{\rightarrow}t} \underset{c}{\rightarrow}_t + W_{\underset{c}{\leftarrow}t} \underset{c}{\leftarrow}_t + b_y \quad (3)$$

4. Results and discussion

The voice signals carry vast knowledge about the speakers. So, the voice samples have several characteristics of the speakers. In this article, the precision values for the male, female and transgender are computed with the help of MFCC as a feature of the voice signals and RNN-BiLSTM classification algorithm. The error factor of the proposed model is also calculated considering of the third gender first time.

The precision of each gender and the error rate of the proposed model can be computed by using the outcome values of the confusion matrix. The mathematical representation for the precision of the class, error factor and accuracy are shown in Eqs. (4), (5) and (6), respectively [40].

$$Precision = \frac{TP\ for\ the\ perticular\ class}{Predicted\ valuse\ for\ the\ perticular\ class} * 100 \quad (4)$$

$$Error\ factor\ = \frac{TN+FN}{TP+TN+FP+FN} * 100 \quad (5)$$

$$\text{Accuracy}\ = \frac{TP+FP}{TP+TN+FP+FN} * 100 \quad (6)$$

Where, TP = True Positive
TN=True Negative
FN=False Negative
FP=False Positive.

The voice sample dataset is divided into six sets during the experiment. Every set has more than 80% of the total samples for the training process. The remanent voice samples are available for the testing purpose, i.e., less than 20% of the total voice samples are available in the testing datasets.

Table 2.a express the male precision values for set 1 to set 3 of the voice samples at different epochs. Similarly, Table 2.b contains the male precision values for the remaining three datasets at different epochs. The values of these tables are computed by the output values of a confusion matrix.

Table 2.a Male precision values

No of Epoch	Set 1	Set 2	Set 3
	Precision (%)	**Precision (%)**	**Precision (%)**
Epoch 2	89.34	92.07	92.79
Epoch 10	95.15	92.05	94.58
Epoch 25	95.15	90.76	93.67
Epoch 50	94.83	90.61	95.57

Table 2.b Male precision values

No of Epoch	Set 4	Set 5	Set 6
	Precision (%)	**Precision (%)**	**Precision (%)**
Epoch 2	90.51	92.11	98.50
Epoch 10	94.65	94.75	95.17
Epoch 25	93.49	96.76	97.18
Epoch 50	93.50	95.20	96.15

Table 3.a Female precision values

No of Epochs	Set 1	Set 2	Set 3
	Precision (%)	**Precision (%)**	**Precision (%)**
Epoch 2	99.59	95.87	98.23
Epoch 10	98.48	98.24	98.07
Epoch 25	98.47	98.24	98.30
Epoch 50	98.47	98.75	98.55

Table 3.b Female precision values

No of Epochs	Set 4	Set 5	Set 6
	Precision (%)	**Precision (%)**	**Precision (%)**
Epoch 2	91.20	97.70	95.45
Epoch 10	97.25	98.55	96.75
Epoch 25	96.73	96.69	97.53
Epoch 50	96.76	96.93	97.93

Fig. 1 shows the average precision values for the male gender for all six datasets, which is calculated from the value of Tables 2.a and 2.b. Fig. 1 shows the graphical representation of the precision values for the male gender.

Tables 3.a and 3.b contains precision values according to the predicted and true positive values for the female gender class. The true positive and predicted values are received form the confusion matrix after simulation the classification algorithm for the different datasets of the same voice samples at different epochs. The precision values for the female class are calculated with Eq. (4).

Fig. 2 has the average female precision value for all six datasets, and these average precision values can be calculated with the help of Table 3.a and 3.b. The graphical representation of the precision values for the female gender is shown in Fig. 2

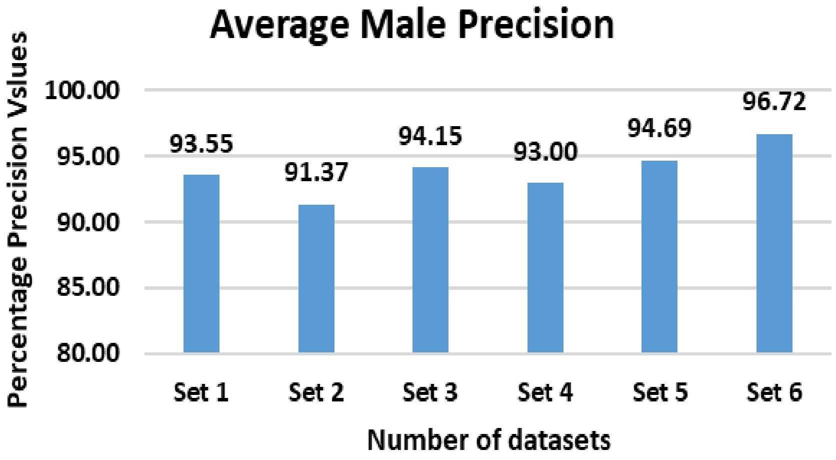

Figure. 1 Average male precision

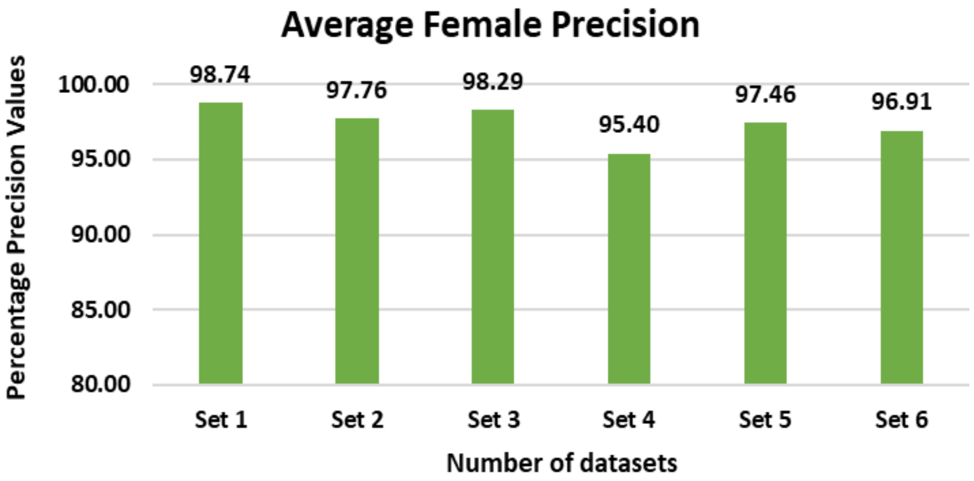

Figure. 2 Average female precision

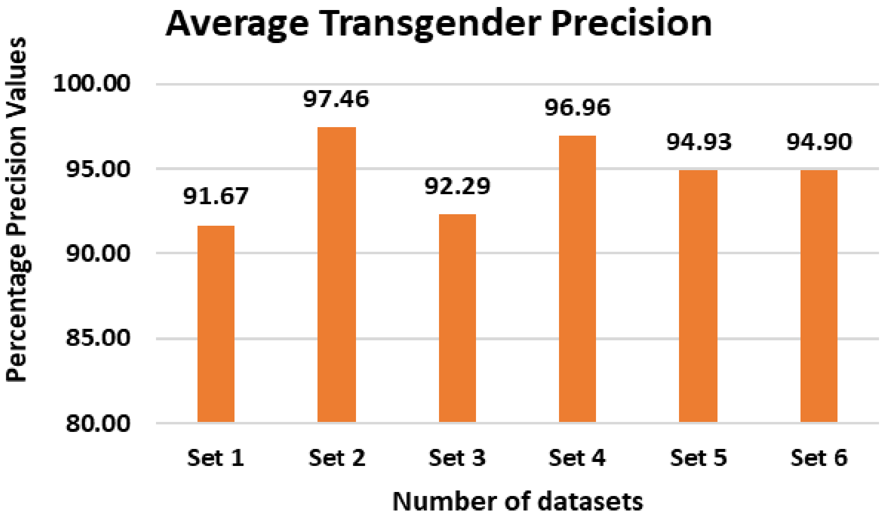

Figure. 3 Average transgender precision

Table 4.a. Transgender precision values

No of Epochs	Set 1	Set 2	Set 3
	Precision (%)	Precision (%)	Precision (%)
Epoch 2	91.46	97.54	88.89
Epoch 10	91.41	97.89	93.13
Epoch 25	90.50	97.89	93.09
Epoch 50	93.33	96.55	94.25

Table 4.b. Transgender precision values

No of Epochs	Set 4	Set 5	Set 6
	Precision (%)	Precision (%)	Precision (%)
Epoch 2	91.46	97.54	88.89
Epoch 10	91.41	97.89	93.13
Epoch 25	90.50	97.89	93.09
Epoch 50	93.33	96.55	94.25

Table 5. Final average precision value for all three genders

Gender	Average Precision Value (%)
Male	93.91
Female	97.43
Transgender	94.70

The first time, the precision value for the transgender is calculated by using the voice signals. The true positive and predicted values are noted for calculating transgender precision values for the different datasets of the same voice samples at different epochs. Table 4.a and 4.b contains the predicted and true positive values for the transgender class. The precision values for the transgender class are calculated with equation (4).

The value of Fig. 3 shows the average precision values for the transgender, which are calculated with the help of the values of Tables 4.a and 4.b. Fig. 3 shows the graphical representation of the precision values for the transgender.

The average precision value for the male class is achieved by 93.91%. The precision value for the female demonstrates the highest accuracy compared to other gender precision values. Precision value for the female is 97.43%. While for the transgender, the precision value reached 94.70%. Table 5 contains the average precision value for males, females, and transgender.

The average precision of the proposed model is 95.35% in consideration of all three genders. Four datasets with the same number of data samples are used in the reported literature to calculate the precision value for the different classification models. These datasets contain only male and female voice samples. The average precision values are 77.00%, 61.75%, 63.53, and 61.00% for Genetic Algorithms (GA), Fuzzy Logic with Neural Network (FL with NN), Neural Network (NN) and Naïve Bayes (NB), respectively. So, the proposed model shows the highest precision values, and NB has the worst. Table 6 compares the precision value of the proposed model with different classification models as reported in the literature [1].

The first time, the error factor for the proposed model for identifying for all three genders is calculated. Tables 7 -12 shows the error rate for the different voices samples datasets, i.e., dataset 1 to dataset 6 at the different epochs. The error rate is varied according to the variation of epochs.

The average error rate for the proposed model is varied from 0.039 to 0.051. The overall error rate for the proposed model considering all three genders is 0.045. Graphical representation for the average error rate for the proposed model, which is calculated after analyzing the voice signals of male, female and transgender speakers, is shown in Fig. 4.

Table 6. Comparison of precision value for the proposed model

Classification Algorithms	Proposed Model	GA [1]	FL with NN [1]	NN [1]	NB [1]
Types of Gender	Male Female Transgender	Male Female	Male Female	Male Female	Male Female
Precision Value (%)	95.35	77.00	61.75	63.53	61.00

Table 7. Error factor for the dataset 1

Samples Set 1	Error Factor
Epoch 2	0.06
Epoch 10	0.05
Epoch 25	0.05
Epoch 50	0.04

Table 8. Error factor for the dataset 2

Samples Set 2	Error Factor
Epoch 2	0.04
Epoch 10	0.03
Epoch 25	0.04
Epoch 50	0.04

Received: February 27, 2023. Revised: May 28, 2023.

Table 9. Error factor for the dataset 3

Samples Set 3	**Error Factor**
Epoch 2	0.07
Epoch 10	0.05
Epoch 25	0.05
Epoch 50	0.04

Table 10. Error factor for the dataset 4

Samples Set 4	**Error Factor**
Epoch 2	0.06
Epoch 10	0.04
Epoch 25	0.04
Epoch 50	0.04

Table 11. Error factor for the dataset 5

Samples Set 5	**Error Factor**
Epoch 2	0.06
Epoch 10	0.03
Epoch 25	0.04
Epoch 50	0.04

Table 12. Error factor for the dataset 6

Samples Set 6	**Error Factor**
Epoch 2	0.05
Epoch 10	0.04
Epoch 25	0.03
Epoch 50	0.04

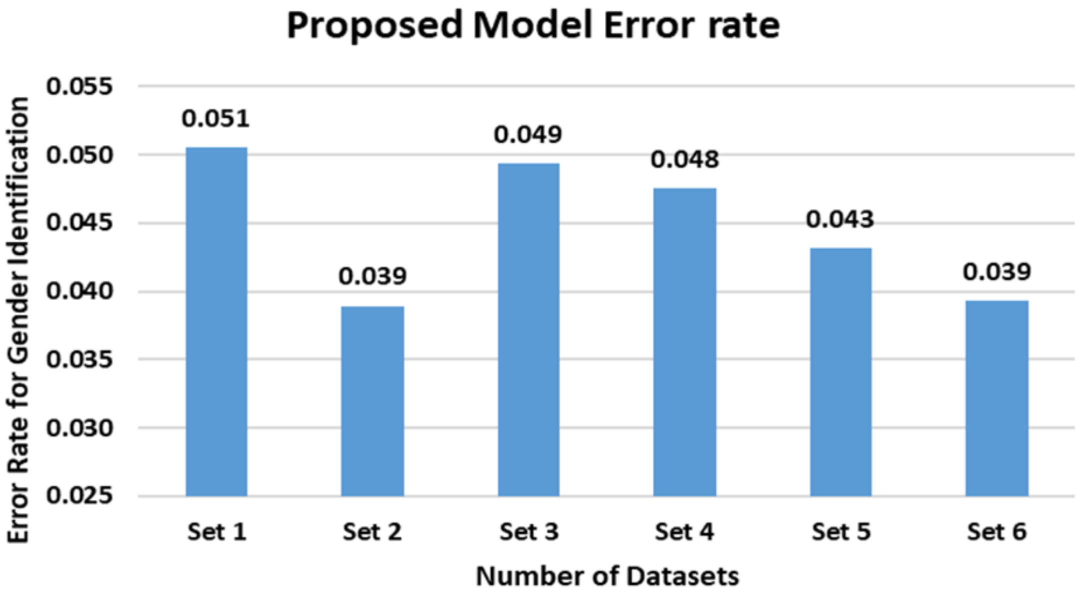

Figure. 4 Average error rate for the proposed model

The calculated average value of the error factor for the proposed model is 0.045 using the voice sample of all three genders. In the reported work in the literature, the error factors are calculated using only male and female voice samples. The average error factors were achieved 0.22, 0.21, 0.201 and 0.203 for deep neural network (DNN) with three layers, DNN with five layers, conventional neural network (CNN) with limited weight sharing (LWS) and CNN with full weight sharing (FWS), respectively [2]. Similarly, the average value of the error factor for temporal conventional neural network (TCNN-back level), support vector regression (SVR-baseline system) and random forest are 0.03, 0.13 and 0.07, respectively [3]. TCNN shows the minimum error factor, and SVR has the worst error factor value. Table 13 represents the average error rate comparison for the different classification models by analyzing the different voice sample datasets with types of genders.

The identification accuracy of the proposed system is computed after the execution of the classification algorithm. The gender identification accuracy for the proposed model is calculated for the live recorded voice signals of females, males and transgenders. Several researchers computed the gender identification accuracy for the females and males, but the gender identification accuracy is computed first time in the present article with the help of live recorded voice samples for females, males and transgenders. Tables 14-19 contains the computed values of the gender identification accuracy at different epochs. Gender identification accuracy for dataset 1 is available in Table 14. Similarly, the gender identification accuracies for datasets 2 – 6 are available in Tables 15-19,

Received: February 27, 2023. Revised: May 28, 2023.

Table 13. Comparison of error factor for the proposed model

Classification Algorithms	**Proposed Model**	**DNN with Three Layer [2]**	**DNN with Five Layers [2]**	**CNN-LWS [2]**	**CNN-FWS [2]**	**TCNN [3]**	**SVR [3]**	**Random Forest [3]**
Types of Gender	Male Female Transgender	Male Female	Male Female	Male Female	Male Female	Male Female	Male Female	Male Female
Precision Value (%)	0.045	0.22	0.21	0.201	0.203	0.03	0.13	0.07

Table 14. Gender identification accuracy for the dataset 1

Set 1	**Percentage Accuracy**
Epoch 2	93.60
Epoch 10	95.35
Epoch 25	95.06
Epoch 50	95.78

Table 15. Gender identification accuracy for the dataset 2

Set 2	**Percentage Accuracy**
Epoch 2	95.61
Epoch 10	96.61
Epoch 25	96.27
Epoch 50	95.94

Table 16. Gender identification accuracy for the dataset 3

Set 3	**Percentage Accuracy**
Epoch 2	93.33
Epoch 10	95.36
Epoch 25	95.11
Epoch 50	96.20

Table 17. Gender identification accuracy for the dataset 4

Set 4	**Percentage Accuracy**
Epoch 2	93.69
Epoch 10	95.85
Epoch 25	95.58
Epoch 50	95.85

Table 18. Gender identification accuracy for the dataset 5

Set 5	**Percentage Accuracy**
Epoch 2	93.77
Epoch 10	96.53
Epoch 25	96.37
Epoch 50	96.06

Table 19. Gender identification accuracy for the dataset 6

Set 6	**Percentage Accuracy**
Epoch 2	94.96
Epoch 10	96.13
Epoch 25	96.83
Epoch 50	96.37

respectively. It is observed that the gender identification accuracy of the proposed model is changed according to the variations in number of testing and training voice samples and the number of epochs.

The gender identification accuracy shows the variation according to the variations in the set. The average gender identification accuracies for the different sets varied from 94.95 % to 96.11 %. The gender identification accuracy for the proposed model is 95.52%, computed after the voice signals analysis of female, male and transgender. Fig. 5 explains the graphical gender identification accuracy for datasets 1 – 6. These accuracies are computed by analysing the live recorded voice signals of females, males and transgender.

The gender identification accuracy for the proposed model is 95.52%. This accuracy is achieved by analysing the live recorded voice samples for females, males and transgender. The identification accuracy was calculated by several researchers in the reported works of literature, but the analysis of female and male voice signals mainly computes these accuracies. In the published literature, the identification accuracy was 76.27% for the RNN classification algorithm [5]. Similarly, 96.00%, 99.60%, 80.00%, 96.40% and 94.60% were the gender identification accuracy for the conventional neural network (CNN), deep neural network (DNN), back-end system, Gaussian mixture model (GMM) and learning vector quantization (LVQ), respectively [14, 39-42]. The gender identification accuracy for neural networks (NN) was achieved by 88.37% after analysing all three genders [43]. This accuracy was computed by analysing the small number of voice samples, and

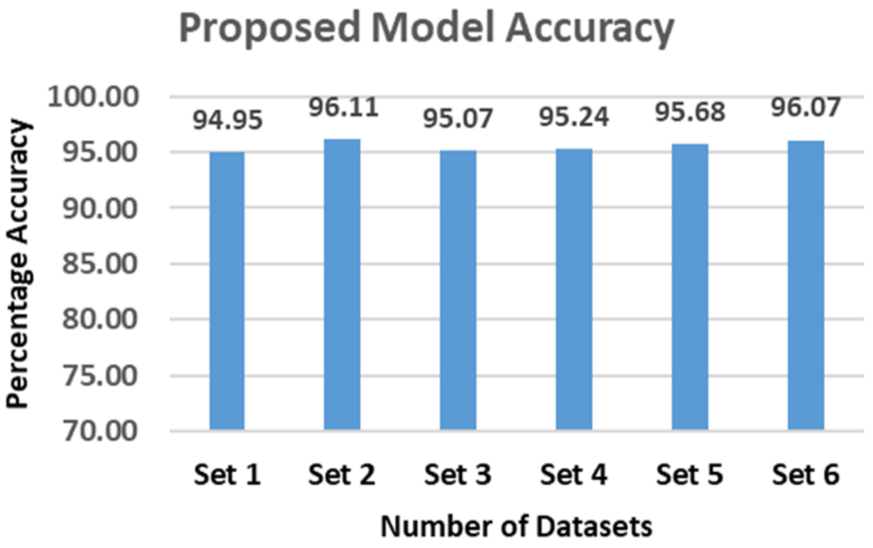

Figure. 5 Percentage accuracy for different datasets

Table 20. Comparison of percentage accuracy for the different classification algorithm

Classification Algorithms	**Proposed Model**	**RNN [39]**	**CNN [40]**	**DNN [41]**	**Back-end System [42]**	**GMM [14]**	**LVQ [14]**	**NN [43]**
Types of Gender	Male Female Transgender	Male Female	Male Female	Male Female	Male Female	Male Female	Male Female	Male Female Transgender
Percentage Accuracy	95.52	76.27	96.00	99.60	80.00	96.40	94.60	88.37

the source of the voice samples is not mentioned in the literature. The proposed model achieves high gender identification accuracy compared to the other reported literature considering all three genders. Table 20 represents the proposed model percentage accuracy compared with the different classification models reported in the literature after analysis for the different voice samples of genders.

5. Conclusion

The estimation of precision value for the male, female, transgender, and error factor for the proposed model are computed in the presented article. The common live voice samples are used to calculate the error factor of the proposed system. The precision values for female, male and transgender are computed. The live voice samples of the female, male and transgender are recorded at 44.1 kHz. The voice signal analysis is performed with the help of the RNN-BiLSTM classification algorithm and MFCC as extracted features. The precision value for the transgender is computed first time by using the live recorded voice samples, which are unavailable at any recognized sources. The classification model achieved the average precision value of 97.43%, 93.91%, and 94.70% for female, male and transgender classification, respectively. The average precision value for the proposed model is 95.35%. The computed precision value is the highest compared with the reported literature. The average error factor of the proposed model is computed as 0.045. After the analysis of the live recorded voice samples, it is observed that the female gender shows the highest precision value compared to the other genders. On the other hand, the lowest precision value is computed for the male gender. It is also observed that the error rate and precision value can be improved if the number of training samples is increased. The value of the error rate for the proposed model can be decreased by improving the quality of the recorded voice samples.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The authors have been contributed as follows:

Conceptualization -Abhishek Singhal and Devendra Kumar Sharma, Methodology- Abhishek Singhal, Software- Abhishek Singhal, Analysis – Abhishek Singhal, Investigation- Abhishek Singhal, Writing-Original draft Preparation-Abhishek Singhal, Writing-review and Editing – Abhishek Singhal and Devendra Kumar Sharma.

References

[1] T. Jayasankar, K. Vinothkumar, and A. Vijayaselvi, "Automatic Gender Identification in Speech Recognition by Genetic Algorithm", *Applied Mathematics & Information Sciences*, Vol. 11, No. 3, pp. 907-913, 2017.

[2] O. A. Hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional Neural Networks for Speech Recognition", *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* Vol. 22, No. 10, pp. 1533-1545, 2014.

[3] H. A. S. Hevia, R. G. Pita, R. M. U. Manso, and M. R. Zurera, "Age group classification and gender recognition from speech with temporal convolutional neural networks", *Multimedia Tools and Applications,* Vol. 81, No. 3, pp. 3535–3552, 2022.

[4] D. Mahmoodi, H. Marvi, M. Taghizadeh, A. Soleimani, F. Razzazi, and M. Mahmoodi, "Age Estimation Based on Speech Features and Support Vector Machine", In: *Proc. of 3rd Computer Science and Electronic Engineering Conference, Colchester,* UK, pp. 60–64, 2011.

[5] M. A. Nasr, M. A. Elnaby, A. S. E. Fishawy, S. E. Rabaie, and F. E. A. E. 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.

[6] S. J. Chaudhari and R. M. Kagalkar, "Methodology for Gender Identification, Classification and Recognition of Human Age", In: *Proc. of National Conference on Advances in Computing, A part of International Journal of Computer Applications,* Vol. 975, p. 8887, 2015.

[7] S. J. Chaudhari and R. M. Kagalkar, "Automatic Speaker Age Estimation and Gender Dependent Emotion Recognition", *International Journal of Computer Applications,* Vol. 117, No. 17, pp. 5 – 10, 2015.

[8] K. H. Lee, S. Kang, D. H. Kim, and L. 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.

[9] R. R. Rao and Nagesh, "Source feature based gender identification system using GMM", *International Journal on Computer Science and Engineering*, Vol. 3, No. 2, pp. 586-593, 2011

[10] R. R. Rao and A. Prasad, "Glottal excitation feature based gender identification system using ergodic HMM", *International Journal of Computer Applications*, Vol. 17, No. 3, pp. 31-36, 2011.

[11] E. Ramdinmawii and V. K. Mittal, "Gender identification from speech signal by examining the speech production characteristics", In: *Proc. of International Conference on Signal Processing and Communication*, Noida, India, pp. 244-249, 2016.

[12] J. Pribil, A. Pribilova, and J. Matousek, "GMM-based speaker age and gender classification in Czech and Slovak", *Journal of Electrical Engineering*, Vol. 68, No. 1, pp. 3-12, 2017.

[13] T. Maka and P. Dziurzanski, "An analysis of the influence of acoustical adverse conditions on speaker gender identification", In: *Proc. of XXII Annual Pacific Voice Conference*, Krakow, Poland, pp. 1–4, 2014.

[14] R. Djemili, H. Bourouba, and M. C. A. Korba, "A speech signal based gender identification system using four classifiers", In: *Proc. of International Conference on Multimedia Computing and Systems*, Tangiers, Morocco, pp. 184-187, 2012.

[15] M. A. Hossan, S. Memon, and M. A. Gregory, "A novel approach for MFCC feature extraction", In: *Proc. of 4th International Conference on Signal Processing and Communication Systems*, Gold Coast, QLD, Australia, pp. 1-5, 2010.

[16] 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.

[17] S. Rathor and S. Agrawal, "A robust model for domain recognition of acoustic communication using Bidirectional LSTM and deep neural network", *Neural Computer & Application*, Vol. 33, No. 17, pp. 11223–11232, 2021.

[18] C. Muller, F. Wittig, and J. Baus, "Exploiting Speech for Recognizing Elderly Users to Respond to their Special needs", In: *Proc. of Eighth European Conference on Speech Communication and Technology,* Geneva,

Switzerland, pp. 1305-1308, 2003.

[19] H. Kim, K. Bae, and H. Yoon, “Age and Gender Classification for a Home-Robot Service”, In: *Proc. of RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, Jeju, Korea (South), pp. 122–126, 2007.

[20] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, “Front-End Factor Analysis for Speaker Verification”, *IEEE Transactions Audio, Speech and Language*, Vol. 19, No. 4, pp. 788–798, 2011.

[21] D. Snyder, D. G. Romero, G. Sell, D. Povey, and S. Khudanpur, “X-vectors: Robust dnn embeddings for speaker recognition”, In: *Proc. of International Conference on Acoustics, Speech and Signal Processing*, Calgary, AB, Canada, pp. 5329–5333, 2018.

[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 & Knowledge Extraction*, Vol. 1, No. 1, pp. 492-503, 2019.

[24] B. Bsir and M. Zrigui, “Gender Identification: A Comparative Study of Deep Learning Architectures”, *Intelligent System Design and Applications*, Vol. 941. pp. 792-800, 2018.

[25] 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.

[26] R. S. Alkhawaldeh, “DGR: Gender Recognition of Human Speech Using One-Dimensional Conventional Neural Network”, *Scientific Programming*, Vol. 2019, Article ID 7213717, pp. 1-12, 2019.

[27] M. Buyukyilmaz and A. O. Cibikdiken, “Voice Gender Recognizer using Deep Learning”, In: *Proc. of International Conference on Modeling, Simulation and Optimization Technologies and Applications, Atlantis Press*, pp. 409-411, 2016.

[28] R. V. Sharan and T. J. Moir, “Robust acoustic event classification using deep neural networks”, *Information Sciences: An International Journal*, Vol. 396, pp. 24–32, 2017.

[29] Ö. B. Dinler 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, Article ID 1273, pp. 1-23, 2020.

[30] 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.

[31] A. A. Malode and S. Sahare, “Advanced speaker recognition”, *International Journal of Advances in Engineering and Technology*, Vol. 4, No. 1, pp. 443–455, 2012.

[32] E. Benmalek, J. Elmhamdi, and A. Jilbab, “Multiclass classification of Parkinson’s disease using different classifiers and LLBFS feature selection algorithm”, *International Journal of Speech Technology*, Vol. 20, No. 1, pp. 179–184, 2017.

[33] Y. Liu, L. He, J. Liu, and M. T. Johnson, “Introducing phonetic information to speaker embedding for speaker verification”, *EURASIP Journal on Audio, Speech, and Music*, Vol. 19, No. 1, pp. 1-17, 2019.

[34] 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. 13077-130781, 2020.

[35] M. K. Reddy and S. K. Rao, “Excitation modelling using epoch features for statistical parametric speech synthesis”, *Computer Speech & Language*, Vol. 60, pp. 101029, 2020.

[36] L. Jasuja, A. Rasool, and A. G. Hajela, “Voice Gender Recognizer Recognition of Gender from Voice using Deep Neural Networks”, In: *Proc. of International Conference on Smart Electronics and Communication*, Trichy, India, pp. 319-324, 2020.

[37] V. Pratap, Q. Xu, A. Sriram, G. Synnaeve, and R. Collobert, “MLS: A Large-Scale Multilingual Dataset for Speech Research”, *ArXiv*, Shanghai, China. abs/2012.03411, 2020.

[38] G. Keren and B. Schuller, B., “Convolutional RNN: An enhanced model for extracting features from sequential data”, In: *Proc. of International Joint Conference on Neural Networks*, Vancouver, BC, Canada, pp. 3412-3419, 2016.

[39] E. Rejaibi, A. Komaty, F. Mériaudeau, S. Agrebi, and A. Othmani, "MFCC-based Recurrent Neural Network for Automatic Clinical Depression Recognition and Assessment from Speech", *Biomedical Signal Processing and Control*, Vol. 71, No. A, p. 103107, 2022.

[40] A. Tursunov, Mustaqeem, 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, pp. 5892, 2021.

[41] D. Kwasny and D. Hemmerling, "Gender and Age Estimation Methods Based on Speech Using Deep Neural Networks", *Sensors*, Vol. 21, No. 14, pp. 4785, 2021.

[42] M. S. Ali, M. S. Islam, and M. A. Hossain, "Gender Recognition System Using Speech Signal", *International Journal of Computer Science, Engineering and Information Technology*, Vol. 2, No. 1, pp. 1-9, 2012.

[43] G. Yasmin, O. Mullick, A. Ghosal, and A. K. Das, "Gender Recognition Inclusive with Transgender from Speech Classification", *Conference on Emerging Technologies in Data Mining and Information Security, Part of the Advances in Intelligent Systems and Computing Book Series*, Vol. 755, pp. 89-98, 2019.