

Gender identification from tribal speech using several learning techniques

Subrat Kumar Nayak¹, Kumar Surjeet Chaudhury², Nirmal Keshari Swain³, Yugandhar Manchala³,
Ajit Kumar Nayak⁴, Smitaprava Mishra⁴, Nrusingha Tripathy¹

¹Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India

²School of Computer Engineering, Kalinga Institute of Industrial Technology (KIIT), Deemed to be University, Bhubaneswar, India

³Department of Information Technology at Vardhaman College of Engineering (Autonomous), Hyderabad, India

⁴Department of Computer Science and Information Technology, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India

Article Info

Article history:

Received Sep 20, 2024

Revised Jun 3, 2025

Accepted Jul 3, 2025

Keywords:

Deep learning

Gender identification

Gradient boosting

Low resourced language

Machine learning

MFCC

ABSTRACT

Language processing and linguistics researchers are interested in gender identification through audio, as human voices have many distinctive features. Although several gender identification algorithms have been developed, the accuracy and efficiency of the system can still be improved. Despite extensive studies on the topic in various languages, there aren't many studies on gender identification in the KUI language. Using a variety of machine learning (ML) and deep learning (DL) classifiers, including decision tree (DT), multilayer perceptron (MLP), gradient boosting (GB), linear discriminant analysis (LDA), recurrent neural networks (RNN), long short-term memory (LSTM), gated recurrent units (GRU), and transformer, the goal of this study is to assess the accuracy of gender identification among diverse KUI language speakers. To verify the effectiveness of the suggested model, several prediction evaluation metrics were calculated, such as the area under the receiver operating characteristic curve (AUC), F1-score, precision, accuracy, and recall. While the findings are compared to other learning models, the gradient-boosting strategy yielded better results with an accuracy rate of 97.0%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Subrat Kumar Nayak

Department of Computer Science and Engineering, Siksha 'O' Anusandhan (deemed to be University)

Bhubaneswar, India

Email: subratsilicon28@gmail.com

1. INTRODUCTION

The human voice contains so much information that it may be used to infer mental states, behavior, age, gender, and emotion. People in this modern century rely on technology and artificial intelligence to make their lives simpler. Many facets of contemporary life, such as computer-human contact, a wide range of commercial fields, automated question-answering devices, and more, use gender identification systems. It also occurs in other fields, such as robots and advanced security systems [1]. Gender may be inferred from an image, speech, fashion sense, and body language. However, in this investigation, we identified a person's gender by listening to their voice. The characteristics and modifications of the human voice tract establish an individual's gender. Consequently, several researchers made an effort to pinpoint the exact features of the vocal cords and separate them from noises utilizing a range of methods and models.

The human hearing system is a technological advance that uses speech recognition to determine a person's gender. Our brain can quickly determine the gender-specific frequencies and levels of loudness that

reach the human ear. However, using machine learning (ML) techniques, robots may do tasks that humans can't perform independently, such as identifying a person's gender. The most suitable and adequate elements from the human voice must be selected to determine the gender [2]. The machine's recognition rate relies on its ability to extract the most helpful information from speech. The gender was identified by the researchers utilizing a variety of learning techniques using a collection of attributes that were retrieved using the mel frequency cepstral coefficient (MFCC) [3].

For addressing different classification problems, deep learning (DL) and ML models are considered state-of-the-art options. DL is a subfield of ML that creates more comprehensive representations of the original data by using many layers in the models. This approach simplifies the process of creating classifiers for new tasks. Moreover, learning algorithms can generate features from raw input data by using the knowledge gained during training [4]. On the other hand, to get the best outcomes, standard ML algorithms need human-crafted features. Computational models consisting of various processing layers can learn multiple abstractions of data representations due to DL. These techniques have significantly advanced speech categorization beyond previous limits. Determining a speaker's gender from uttered utterances is one of the goals of spoken speech processing research. For many applications, gender categorization based on spoken voice signals is a crucial and challenging task [5].

This research aims to develop many models for classifying speaker gender in a tribal language with limited resources. The remaining portion of this article is structured as follows: the prior relevant works are included in section 2. Section 3 presents the experimental design, speech attributes, corpus, and performance evaluation protocols. A thorough explanation of the developed model is also provided. Section 4 discusses the suggested model. The results are given and discussed in section 5. Section 6 presents our conclusions and suggestions for more study, which concludes in the present article.

2. RELATED WORK

Nowadays, various equipment, such as computers, mobile phones, and security systems, utilize gender identification based on speech features. The published literature comprehensively documents current techniques and concepts. There has been relatively little study done on the KUI language. Gender identification from speech in other languages has been the subject of several studies, but KUI speech has not been studied. Table 1 summarizes the identification of gender for many languages using ML or DL approaches.

Table 1. Several methods are applied in gender identification in different languages

Dataset used	Methods	Year of research
Speech Corpus [2]	Decision trees (DT), gradient boosting (GB), random forest (RF), support vector machine (SVM)	2019
Common voice [5]	Multilayer perceptron (MLP)	2020
SHRUTI [6]	Tensor analysis	2020
Kannada dataset [7]	GMM	2021
Arabic speech [8]	Bidirectional long short-term memory (BiLSTM)	2021
Own dataset [9]	GB, linear discriminant analysis (LDA), logistic regression (LR)	2022
ELSDSR dataset [10]	LR, GB, GN	2023
Common voice [11]	Recurrent neural network (RNN) – BiLSTM	2023
Turkish speech [12]	Convolutional neural network (CNN)	2024
German speech [4]	Deep neural network (DNN)	2019
Low dataset [13]	BiLSTM	2020
Mozilla audio dataset [14]	GB, DT, RF	2021
Sepedi speech [15]	CNN, LSTM	2021
Common voice [16]	LDA, SVM	2022
Open-source data [1]	LDA, artificial neural network (ANN)	2022
Common voice [17]	BiLSTM	2023
Turkish dataset [18]	CNN 1D, CNN 2D	2024
Arabic speech [19]	Hybrid learning	2024

3. MATERIALS AND METHODS

3.1. Data creation

Audio signals are required to identify the genders of the speakers. The common Indian residents record the KUI voice signals to accomplish the suggested work's purpose. These speech samples were captured using a shared platform at 16 kHz [20]. The audio signals are recorded in .wav format since the suggested classification model can analyze the voice signals in this format. 2,000 samples total consisting of 1,125 samples for men and 875 samples for women are utilized in the proposed study to determine the gender

identity of the speakers. Two sets of all 2,000 voice samples are created. The classification model is trained using the first batch and tested using the second. Eighty percent of the voice samples for each gender group are utilized in the training set, while the remaining twenty percent are used in the second set. Transgender speech samples are not recorded in the current paper to analyze voice signals. Since there are no recognized resources that include such type datasets, these voice samples fall under the low-resource language category.

3.2. Technique used for feature extraction: mel-frequency cepstral coefficient

Gender identification is mainly based on traits extracted from voice sounds. The collected features that encapsulate the key attributes of the speech samples constitute an important input for the classification algorithms. The most significant technique for extracting speech-based characteristics in this domain is MFCC [21]. The MFCC plays a significant function because of its capacity to illustrate speech amplitudes succinctly. The following steps outline the process for gaining the MFCC functionalities. The diagrammatic representation of MFCC is shown in Figure 1.

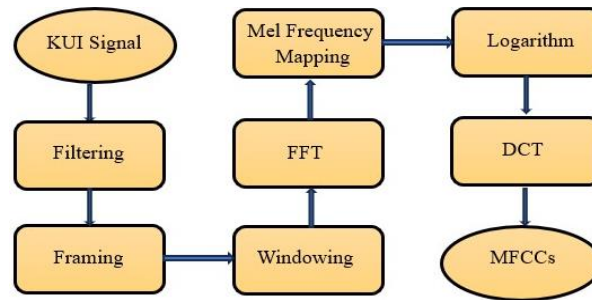


Figure 1. Diagrammatically representation of MFCC

Filtering: this is the initial stage of boosting the high frequencies for less prominent speech in the audio signal $s[n]$ by applying a pre-emphasis filter. The pre-emphasized signal $s_{emp}[n]$ is given in (1).

$$s_{emp}[n] = s[n] - \alpha \cdot s[n - 1] \quad (1)$$

Where α is typically set to a value between 0.95 and 0.98.

Framing: the signal is split up into overlapping frames with L samples in length. Every frame has a certain number of samples P that overlap with the frame before it. If the signal has K samples, the number of frames R is given in (2).

$$R = \left\lfloor \frac{K-L}{P} + 1 \right\rfloor \quad (2)$$

Each frame is denoted as $s_f[n]$, where f indexes the frame.

Windowing: a window function $c[n]$ is multiplied by each frame to decrease spectral leakage. The hamming window is given in (3).

$$c[n] = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{L-1}\right) \quad (3)$$

The windowed frame is given by $s_c[n] = s_f[n] \cdot c[n]$

Fast fourier transform (FFT): the windowed frame $s_c[n]$ is transformed into the frequency domain utilizing the FFT, as shown in (4).

$$S[a] = \sum_{n=0}^{N-1} s_c[n] \cdot e^{-m \frac{2\pi a n}{N}}, a = 0, 1, \dots, N - 1 \quad (4)$$

The resulting $S[a]$ is a complex value representing the amplitude and phase of the signal's frequency components.

Mel frequency mapping: a series of triangle filters $T_d[a]$ which are mel-scaled and applied to the power spectrum. Each filter is designed to capture the energy in a specific mel frequency band, as given in (5).

$$B_d = \sum_{a=0}^{N-1} Y[a] \cdot T_d[a], d = 1, 2, \dots, P \tag{5}$$

Where $Y[a]$ denotes the power spectrum. The mel scale f_{mel} is related to the linear frequency f is given in (6).

$$f_{mel} = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) \tag{6}$$

Logarithm: the logarithm of the filtered output is taken to compress the dynamic range of the signal is given by $\log(B_d)$.

Discrete cosine transforms (DCT): applying the DCT to the log mel spectrum is the last step to minimize dimensionality and decorrelate the filter bank coefficients, as shown in (7).

$$D_n = \sum_{d=1}^P \log(B_d) \cdot \cos \left[n \cdot \left(d - \frac{1}{2} \right) \cdot \frac{\pi}{P} \right], n = 0, 1, \dots, N_c - 1 \tag{7}$$

Here, D_n are the MFCCs, and N_c is the number of coefficients.

3.3. Classification techniques

One of the key elements in determining the speaker’s gender is a classification algorithm. Choosing a classification method with a high degree of gender identification accuracy is the most difficult challenge. The classifier is used to ascertain the speakers’ genders after the feature has been taken from the speech samples. Several learning approaches are used here as classification algorithms, including DT, MLP, GB, LDA, RNN, LSTM, gated recurrent units (GRU), and transformer.

3.3.1. Decision tree

DTs are a subset of bagging techniques useful for efficiently managing non-linear datasets. Using the criteria Gini index, we have instantiated the DT classifier model. A decision is represented by leaf nodes that have no edges on the exterior [2]. DT approximates a sine curve for decision-making purposes using a series of IF-THEN rules. The DT model can mathematically represent a series of recursive binary splits in (8).

$$f(x) = \sum_{n=1}^N \gamma_n I(x \in R_n) \tag{8}$$

Where $f(x)$ is the prediction for input x , N represents the leaf nodes, γ_n denotes the final value, I is an indicator function and R_n is the region of the input space.

3.3.2. Multilayer perceptron

Multi-layer perceptron’s, a kind of neural network commonly used for supervised learning tasks, are abbreviated as MLPs. It consists of several linked layers of nodes, where each neuron sends its output to the layer that comes after it and takes input from the one before it. The last layer, also referred to as the output layer, generates the estimates for the final output. It is trained using a supervised learning method known as backpropagation [18]. The output \hat{y} can be expressed in (9).

$$\hat{y} = \sigma(W_{out} h_m + b_{out}) \tag{9}$$

Where W_{out} is the weight matrix, b_{out} is bias and σ is the sigmoid function and is represented in (10).

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{10}$$

3.3.3. Gradient boosting

A ML technique called GB creates collective weak prediction models that work as classifiers. It creates an additive model step-by-step and is typically applied when accuracy is not achieved with individual classifiers. The training dataset’s X and Y columns are needed for the model to fit. On the testing data set, it is predicted once the model has been fitted. Both the loss function and the basic learner models in the GBs method are freely defined [9]. The final model after M iterations are shown in (11).

$$F_M(X) = F_0(X) + \sum_{m=1}^M \gamma_m h_m(X) \tag{11}$$

Where $X = [X_1, X_2, \dots, X_n]$ as the input feature vector, $F_M(X)$ as the model at the M th stage.

3.3.4. Linear discriminant analysis

For dimensionality reduction and classification, supervised learning techniques like LDA are used. LDA looks for the optimum linear feature combination to divide an object or event into two or more classes. LDA is often used to divide data into two or more labels. If there are two classes, labels are classified linearly using one hyperplane. To divide the classes in various ways, however, several hyperplanes are required [1]. The Scatter matrices are given in (12) and (13).

$$S_W = \sum_{i=1}^{N_m} (x_i^m - M_m)(x_i^m - M_m)^T + \sum_{i=1}^{N_f} (x_i^f - M_f)(x_i^f - M_f)^T \quad (12)$$

$$S_B = (M_m - M)(M_m - M)^T + (M_f - M)(M_f - M)^T \quad (13)$$

Where M is the overall mean vector of the features, S_W is the scatter matrix inside the class, and S_B is a scatter matrix between the classes.

3.3.5. Recurrent neural network

The functioning of an artificial neural network (ANN) resembles that of the human brain. RNN belongs to the ANN group. Time series signals, speech signals, and other signals are produced by combining sequential data. This type of data can be efficiently managed with the RNN classification method. RNN's memory is limited [11]. This disadvantage reduces the field's usefulness in gender identification. LSTM can help mitigate the impact of this limitation. An RNN maintains a hidden state s_{t-1} that contains data from earlier time steps as it goes through the sequence of input. The hidden state is computed at the time step t , in (14).

$$s_t = f(W_s s_{t-1} + W_x x_t + b_s) \quad (14)$$

Where x_t is the input feature vector at the current time step, W_s and W_x are weight matrices, b_s is a bias vector f is an activation function and s_{t-1} is the hidden state from the previous time step.

3.3.6. Long short-term memory

LSTM can expand the system's memory. However, LSTM only really works in one way. Bidirectional LSTM (BiLSTM) is utilized for the two-direction operation to improve the gender identification system's accuracy. The final production of a BiLSTM layer is created by concatenating the outputs from the two layers. Because it learns sequential patterns in both directions, a BiLSTM layer outperforms a single LSTM layer [22]. The output layer of the LSTM model may be defined as shown in (15).

$$y = \text{softmax}(W_y \cdot h_T + b_y) \quad (15)$$

Where W_y is the weight matrix of the output layer, b_y is the bias vector of the output layer, h_T is the last hidden state, softmax ensures the output is a probability distribution over classes

3.3.7. Gated recurrent units

One kind of RNN architecture utilized for sequence modeling applications is the GRU architecture. GRUs have similarities with LSTM networks, although they are more computationally efficient due to their more straightforward architecture. GRUs can perform similarly to LSTMs on various tasks, including gender identification, although having a simpler structure. GRUs are frequently less prone to overfitting and easier to train than LSTMs since they have fewer parameters [23].

3.3.8. Transformer

The self-attention mechanism is used when using a transformer model for gender identification to extract meaningful patterns from the input data, which may be auditory attributes that describe the voice. The encoder's output must be taken and transformed into a series of text tokens by the decoder [24]. Because of the multi-head attention capability, the model may analyze data from many representation subspaces at different times.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_o \quad (16)$$

Where each head is computed as $\text{head}_i = \text{Attention}(QW_Q^i, KW_K^i, VW_V^i)$ and W_o is the output weight matrix. The transformer model may handle sequential data for tasks like gender identification from speech data by utilizing this as shown in (16) [25].

4. PROPOSED MODEL

Gender identification system working model: since KUI has few resources, gathering data from the field is extremely difficult. Following data collection, we apply several learning models. Algorithm 1 shows the procedures in this suggested KUI gender identification system. The comprehensive process is shown in Figure 2.

Algorithm 1. Gender identification system model

1. In the first stage, the voice data has to be provided. For a single speaker, we entered 200 voice data points here. We then preprocessed the first phase's supplied input data. In this step, the voice data is processed and cleaned up. The speech data quality for gender identification is enhanced at this level.
2. After the previous stage, voice data extract various prosodic or acoustic speech properties. Pitch, energy, and intensity may be extracted from speech data using MFCC, one of the feature extraction techniques for voice data.
3. The gender is determined using the learning models.
4. Using various criteria, including accuracy, the effectiveness of gender identification is evaluated at the last stage. It also computed the learning models' accuracy.

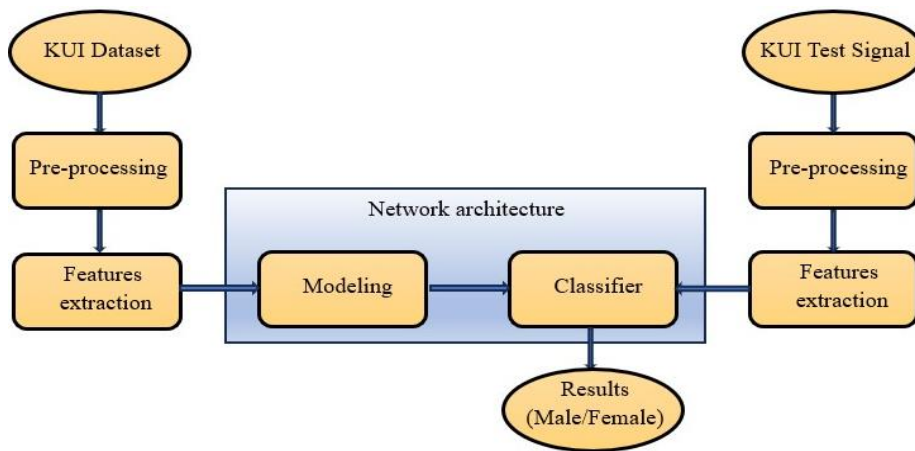


Figure 2. Workflow of KUI gender identification

5. RESULTS AND DISCUSSION

Plenty of studies have been done on gender identification in other languages using a variety of learning strategies, but not in the KUI language. Researchers have encountered numerous difficulties, including preprocessing and dataset preparation, during earlier studies; some of these issues are covered in related publications. Additionally, in past research, there have been issues with dataset development because it is challenging to build a dataset in a tribal language. The predictive model's performance measurements include F1-score, accuracy, precision, recall, and other metrics. Table 2 shows a detailed analysis of different parameters. Figure 3 displays the accuracy graphically, while Figure 4 displays the AUC value. The confusion matrix (CM) is used to compute these performance indicators. Every confusion matrix has an x-axis representing the expected labels and a y-axis representing the actual labels. Figures 5 through 12 show several models' CMs.

Table 2. Performance indicators for various techniques

Methods	Parameters			
	Precision	Recall	F1-score	Accuracy
DT	95	96	95	96
MLP	87	88	90	89
LDA	92	91	91	92
RNNs	89	90	90	90
LSTM	91	91	92	92
GRUs	92	91	92	93
Transformer	92	92	93	94
GB	96	97	96	97

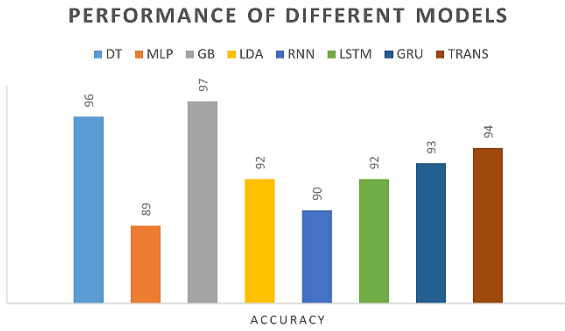


Figure 3. Accuracy of several learning methods

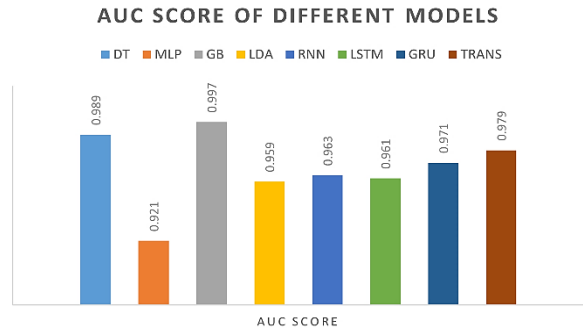


Figure 4. AUC score of several learning methods

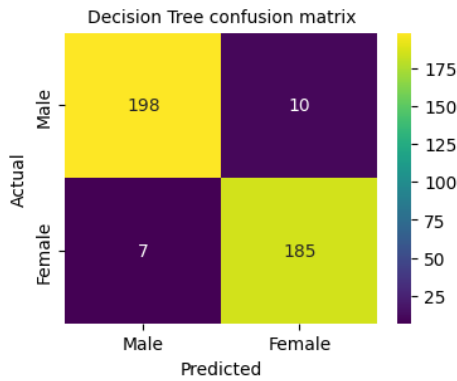


Figure 5. Performance matrix of DT

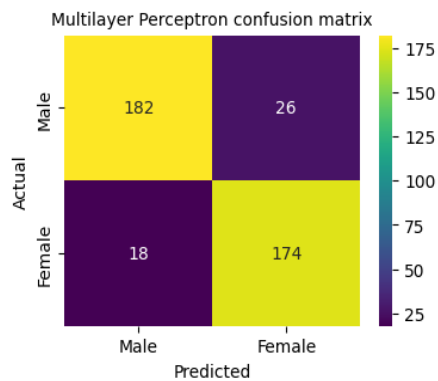


Figure 6. Performance matrix of MLP

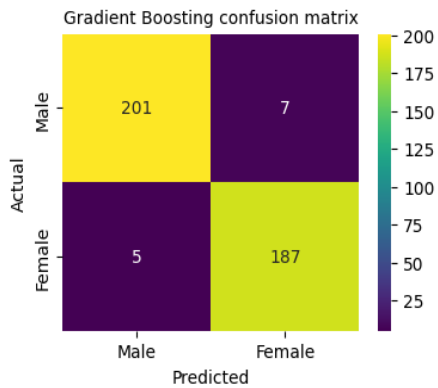


Figure 7. Performance matrix of GB

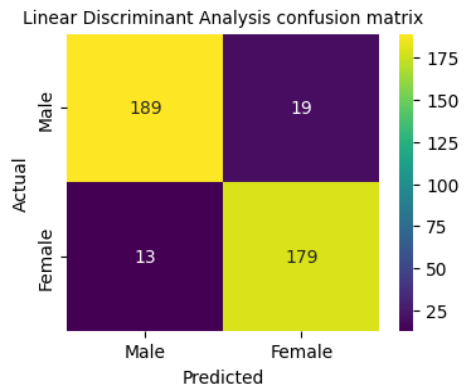


Figure 8. Performance matrix of LDA

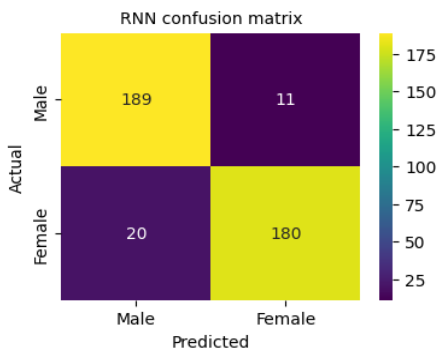


Figure 9. Performance matrix of RNN

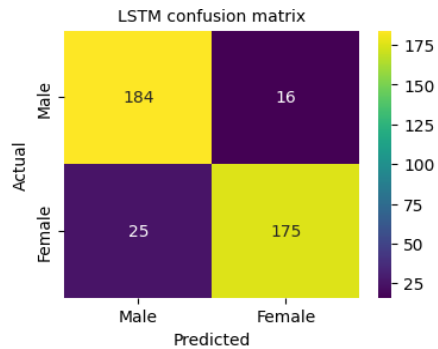


Figure 10. Performance matrix of LSTM

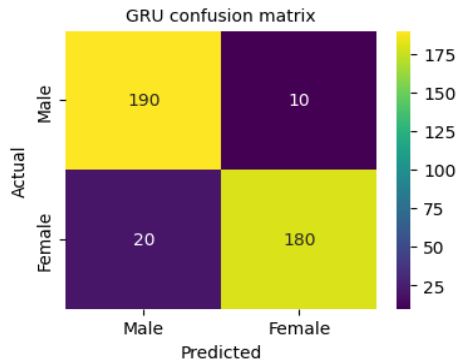


Figure 11. Performance matrix of GRU

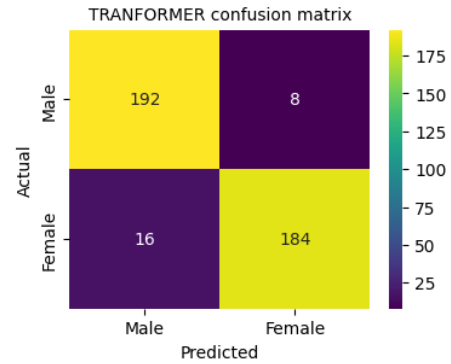


Figure 12. Performance matrix of transformer model

As noted earlier, we use the KUI dataset to evaluate many models. There can be up to 32 batches in each training session. We first train the model with various parameters having a learning rate equal to 0.01. Accuracy is the most often used and simple performance evaluation criterion for gender identity. Recall is calculated by dividing the total number of true members of the positive class by the percentage of all instances that were recognized correctly as belonging to the positive class. The accuracy and recall of the model determine the F1-score.

In Figure 7, the GB performance matrix is shown, which yields a noteworthy result when compared to other models utilizing our KUI dataset. Table 3 displays the accuracy of several models across various languages. It demonstrates that, when using the GB classifier, KUI yields the highest accuracy among the languages used for gender identification. The ROC curve of learning models is shown in Figure 13. We trained and tested the neural network using diverse speech samples. Eighty percent of the data were utilized for training, and twenty percent were used for testing. The neural network undergoes training over 500 epochs. Epochs are periods when a ML algorithm runs through the training data in a single cycle. The classifiers DT, MLP, GB, LDA, RNN, LSTM, GRU, and transformer were applied in conjunction with MFCC. The highest identification accuracy of 97.0% was obtained using our KUI dataset.

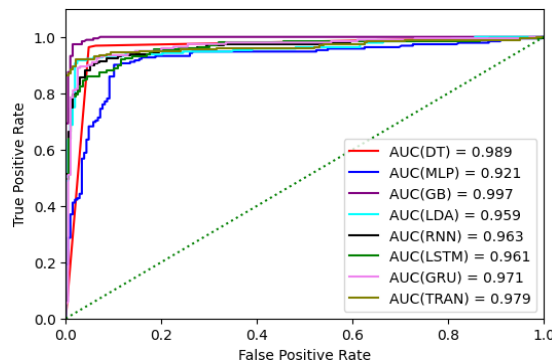


Figure 13. ROC curve of several learning techniques

Table 3. Comparison of different parameters using several learning techniques

Work	Models	Year	Assessment of performance parameters (in %)			
			Precision	Recall	F1-score	Accuracy
Hızlısoy <i>et al.</i> [9]	LR	2022	87	92	89	89
Singhal and Sharma [16]	RNN-BiLSTM	2022	-	92	-	90
Alashban and Alotaibi [8]	BLSTM	2021	-	-	-	91
Shabbir <i>et al.</i> [10]	SVM	2023	89	86	83	92
Nair et and Vijayan [2]	GB	2019	-	-	-	94
Zaman <i>et al.</i> [14]	GB	2021	95	96	96	96
Ali <i>et al.</i> [1]	ANN	2022	97	-	-	97
Sefara and Mokgonyane [15]	LSTM	2021	-	-	97	97
Proposed method	GB	2024	96	97	96	97

6. CONCLUSION AND FUTURE WORK

Gender identification from a tribal language remains a challenging task due to several factors. This paper concludes that firstly, it specifies various stages of gender identification and literature reviews of many published research papers that use many different approaches and have different datasets. We used several learning techniques to identify the gender from our low-resourced KUI dataset. The GB method can identify the gender with an accuracy of 97.0%. The performance of the classifier is demonstrated to be influenced by recall, precision, and F1-score. The corresponding values were 97%, 96%, and 96%. It is speculated that by using hybrid classifiers and other classifiers, these values may be enhanced. Furthermore, it is suggested that using distinct characteristics from our KUI dataset would increase the success rate of gender identification. The KUI gender identification system can be integrated into speech recognition and speaker recognition of the KUI language. We want to expand the number of speakers and would like to work with a variety of languages in the future, such as Odia and Santali.

FUNDING INFORMATION

The authors state no funding is involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Subrat Kumar Nayak	✓	✓	✓	✓	✓	✓		✓	✓	✓				✓
Kumar Surjeet Chaudhury		✓				✓			✓	✓	✓	✓		
Nirmal Keshari Swain	✓		✓	✓		✓			✓		✓			✓
Yugandhar Manchala														
Ajit Kumar Nayak					✓		✓			✓		✓		
Smitapraava Mishra			✓							✓		✓		✓
Nrusingha Tripathy					✓			✓				✓		✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.





REFERENCES

- [1] Y. M. Ali, E. Noorsal, N. F. Mokhtar, S. Z. M. Saad, M. H. Abdullah, and L. C. Chin, "Speech-based gender recognition using linear prediction and mel-frequency cepstral coefficients," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 28, no. 2, pp. 753–761, Nov. 2022, doi: 10.11591/ijeecs.v28.i2.pp753-761.
- [2] R. R. Nair and B. Vijayan, "Voice based gender recognition," *International Research Journal of Engineering and Technology*, vol. 6, no. 5, pp. 2109–2112, 2019.
- [3] S. K. Nayak, A. K. Nayak, S. Mishra, P. Mohanty, N. Tripathy, and S. Prusty, "Improving KUI digit recognition through machine learning and data augmentation techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 35, no. 2, pp. 867–877, Aug. 2024, doi: 10.11591/ijeecs.v35.i2.pp867-877.
- [4] M. Markitantov and O. Verkholyak, "Automatic recognition of speaker age and gender based on deep neural networks," in *Speech and Computer: 21st International Conference (SPECOM)*, 2019, pp. 327–336, doi: 10.1007/978-3-030-26061-3_34.
- [5] L. Jasuja, A. Rasool, and G. Hajela, "Voice gender recognizer recognition of gender from voice using deep neural networks," in *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, Sep. 2020, pp. 319–324, doi: 10.1109/ICOSEC49089.2020.9215254.




- [6] P. Roy, P. Bhagath, and P. Das, "Gender detection from human voice using tensor analysis," in *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, 2020, pp. 211–217.
- [7] V. G. Nandan, S. Shivakumar, J. Sangeetha, M. Pandurang Nayak, and N. S. K., "A comparative study of deep learning and machine learning approaches in speech emotion and gender recognition system," *NVEO-Natural Volatiles & Essential Oils Journal*, vol. 8, no. 5, pp. 12261–12273, 2021.
- [8] A. A. Alashban and Y. A. Alotaibi, "Speaker gender classification in mono-language and cross-language using BLSTM network," in *2021 44th International Conference on Telecommunications and Signal Processing, TSP 2021*, Jul. 2021, pp. 66–71, doi: 10.1109/TSP52935.2021.9522623.
- [9] S. Hızlısoy, E. Çolakoğlu, and R. S. Arslan, "Speech-to-gender recognition based on machine learning algorithms," *International Journal of Applied Mathematics Electronics and Computers*, vol. 10, no. 4, pp. 84–92, Dec. 2022, doi: 10.18100/ijamec.1221455.
- [10] M. Shabbir, A. Hussain, and M. M. Khan, "Age and gender estimation through speech: a comparison of various techniques," in *18th IEEE International Conference on Emerging Technologies, ICET 2023*, Nov. 2023, pp. 228–233, doi: 10.1109/ICET59753.2023.10374670.
- [11] A. Singhal and D. K. Sharma, "Low resource language analysis using deep learning algorithm for gender classification," *ACM Transactions on Asian and Low-Resource Language Information Processing*, Aug. 2023, doi: 10.1145/3614427.
- [12] T. M. Taha, Z. Ben Messaoud, and M. Frikha, "Convolutional neural network architectures for gender, emotional detection from speech and speaker diarization," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 3, pp. 88–103, Feb. 2024, doi: 10.3991/ijim.v18i03.43013.
- [13] R. D. Alamsyah and S. Suyanto, "Speech gender classification using bidirectional long short term memory," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, Dec. 2020, pp. 646–649, doi: 10.1109/ISRITI51436.2020.9315380.
- [14] S. R. Zaman, D. Sadekeen, M. A. Alfaz, and R. Shahriyar, "One source to detect them all: gender, age, and emotion detection from voice," in *Proceedings - 2021 IEEE 45th Annual Computers, Software, and Applications Conference, COMPSAC 2021*, Jul. 2021, pp. 338–343, doi: 10.1109/COMPSAC51774.2021.00055.
- [15] T. J. Sefara and T. B. Mokgonyane, "Gender identification in Sepedi Speech Corpus," in *2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, Aug. 2021, pp. 1–6, doi: 10.1109/icABCD51485.2021.9519308.
- [16] A. Singhal and D. K. Sharma, "Estimation of accuracy and recall values for the different age groups based on voice signals using different classifiers," in *2022 8th International Conference on Signal Processing and Communication, ICSC 2022*, Dec. 2022, pp. 364–369, doi: 10.1109/ICSC56524.2022.10009564.
- [17] A. Singhal and D. K. Sharma, "Precision value, error rate and accuracy of human gender identification based on randomized voice signals datasets," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 4, pp. 348–361, 2023, doi: 10.22266/ijies2023.0831.28.
- [18] E. Yücesoy, "Speaker age and gender recognition using 1D and 2D convolutional neural networks," *Neural Computing and Applications*, vol. 36, no. 6, pp. 3065–3075, Nov. 2024, doi: 10.1007/s00521-023-09153-0.
- [19] A. R. Khan, "Automatic gender authentication from arabic speech using hybrid learning," *Journal of Advances in Information Technology*, vol. 15, no. 4, pp. 532–543, 2024, doi: 10.12720/jait.15.4.532-543.
- [20] S. K. Nayak *et al.*, "Speech data collection system for KUI, a low resourced tribal language," *Journal of Autonomous Intelligence*, vol. 7, no. 1, Oct. 2023, doi: 10.32629/jai.v7i1.1121.
- [21] S. K. Nayak, A. K. Nayak, S. Mishra, and P. Mohanty, "Deep learning approaches for speech command recognition in a low resource KUI language," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 377–386, Oct. 2023.
- [22] N. Tripathy, P. Satapathy, S. Hota, S. K. Nayak, and D. Mishra, "Empirical forecasting analysis of bitcoin prices: a comparison of machine learning, deep learning, and ensemble learning models," *International Journal of Electrical and Computer Engineering Systems*, vol. 15, no. 1, pp. 21–29, Jan. 2024, doi: 10.32985/ijeces.15.1.3.
- [23] N. Tripathy, S. Parida, and S. K. Nayak, "Forecasting stock market indices using gated recurrent unit (GRU) based ensemble models: LSTM-GRU," *International Journal of Computer and Communication Technology*, vol. 9, no. 1, pp. 85–90, Jul. 2023, doi: 10.47893/ijcct.2023.1443.
- [24] S. K. Nayak, A. K. Nayak, S. Mishra, P. Mohanty, N. Tripathy, and K. S. Chaudhury, "Exploring speech emotion recognition in tribal language with deep learning techniques," *International Journal of Electrical and Computer Engineering Systems*, vol. 16, no. 1, pp. 53–64, Jan. 2025, doi: 10.32985/ijeces.16.1.6.
- [25] X. Wang, M. Thakker, Z. Chen, N. Kanda, S. E. Eskimez, S. Chen, and T. Yoshioka, "SpeechX: neural codec language model as a versatile speech transformer," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 32, pp. 3355–3364, 2024, doi: 10.1109/TASLP.2024.3419418

BIOGRAPHIES OF AUTHORS






Subrat Kumar Nayak     received the degree in MCA from Biju Patnaik University of Technology, Odisha, India in 2010, M.Tech. in computer science from Utkal University, Bhubaneswar, Odisha in 2012. He is currently pursuing his Ph.D. in CSE at SOA University, Bhubaneswar, India. He has published 7 papers in various international journals and international conferences. He qualified for UGC Net in the year 2012. He has more than 6 years of academic experience and 3 years of govt experience. He can be contacted at email: subratsilicon28@gmail.com.






Dr. Kumar Surjeet Chaudhury    received M.E. in computer science and engineering from Jadavpur University, Kolkata in 2008 and Ph.D. degree in computer science from Fakir Mohan University, Odisha in 2022. He is currently working as an assistant professor in the School of Computer Engineering, Kalinga Institute of Industrial Technology (KIIT), Deemed to be University, Bhubaneswar, Odisha. He has published 20 research papers in various international journals and international conferences. He has more than 19 years of academic experience. He can be contacted at email: surjeet.chaudhuryfcs@kiit.ac.in.






Nirmal Keshari Swain    received an MCA degree in computer science in 2004, and an M.Tech. degree in computer science and engineering in 2011, both the degree from Biju Patnaik University of Technology, Rourkela, Odisha, India. He is currently working as an assistant professor in the Department of Information Technology at Vardhaman College of Engineering (Autonomous), Hyderabad, India. He is having more than 20 years of academic experience. His research interests include machine learning, and deep learning. He has published ten conference papers and twelve journal papers. He can be contacted at email: swain.nirmal6@gmail.com.






Yugandhar Manchala    received M.Tech. (software engineering) from JNTU, Kakinada in the year 2012. He is currently working as an assistant professor in the Department of Information Technology at Vardhaman College of Engineering (Autonomous), Hyderabad, India. He is having more than 12 years of academic experience. His research interests include internet of things (IoT), machine learning, deep learning, network security. He can be contacted at email: yugandhar1230@gmail.com.






Prof. Dr. Ajit Kumar Nayak    is the Professor and Head of the Department of Computer Science and Information Technology, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha. He received degree electrical in engineering from the Institution of Engineers, India in the year 1994, M. Tech and Ph.D. degree in computer science from Utkal University in 2001 and 2010 respectively. He has published about 70 research papers in various journals and conferences. Also co-authored a book 'Computer Network Simulation using NS2'. Ten Ph.D. scholars have been awarded Ph.D. under his supervision. He can be contacted at email: ajitnayak@soa.ac.in.



Prof. Dr. Smitaprava Mishra    is currently working as Professor in Department of Computer Science and Information Technology at Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha. She has 18 years of teaching and research experience in the current organization. She has published 30 numbers of research papers in various reputed journals and conferences. 15 M.Tech. scholars produced under her supervision in the research area of data mining, machine learning, and natural language processing. 4 Ph.D. scholars pursuing research under her supervision. She has contributed several academic activities at organization level. She can be contacted at email: smitamishra@soa.ac.in.



Nrusingha Tripathy    received an MCA degree in computer science from Ravenshaw University, Cuttack, Odisha, India, in 2018, and an M.Tech. degree in computer science from Utkal University, Bhubaneswar, Odisha, India, in 2020. Currently, he is pursuing a Ph.D. in computer science and engineering at the Institute of Technical Education and Research (ITER) at Siksha 'O' Anusandhan (Deemed to be) University, Bhubaneswar, India. He has published twenty conference papers and twenty-one journal papers. With over five years of teaching experience. He can be contacted at email: nrusinghatripathy654@gmail.com.