Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 3, June 2026, pp. 2308~2315
ISSN: 2302-9285, DOI: 10.11591/eei.v15i3.10985

Cross-lingual deep learning model for gender detection

Kavita Jhajharia[1], Ginika Mahajan[2], Dhondi Samrudh[2], Koustubh Patel[2]
[1]Department of Information Technology, Faculty of Science, Technology, and Architecture, Manipal University Jaipur, Jaipur, India
[2]Department of Data Science and Engineering, Faculty of Science, Technology, and Architecture, Manipal University Jaipur, Jaipur, India

Article Info

Article history:

Received Jul 8, 2025
Revised Apr 6, 2026
Accepted Apr 19, 2026

Keywords:

Deep learning
Gender detection
Machine learning
Neural networks
Voice recognition

ABSTRACT

Speech recognition is transforming the way humans interact with technology and automatic gender recognition is an essential part of this evolution. This study develops a multilingual deep learning (DL) model for gender detection using three audio datasets: RAVDESS (English), Berlin EmoDB (German), and IITKGP-SEHSC (Hindi). These datasets provide linguistic diversity, enabling the development of a multi-lingual gender identification model. The mel-frequency cepstral coefficients (MFCC) and VGGish embeddings and other audio features were used to process raw audio data into something meaningful. The findings show the machine learning (ML) models (random forest (RF) and extreme gradient boosting) achieved good results in the monolingual (98.26% using Hindi and 96.90% using cross-lingual) setup. In DL models, convolutional neural network (CNN) outperformed other models in both monolingual and cross-lingual scenarios, with 99.33% accuracy for Hindi and 98.11% accuracy in cross-lingual setup. These findings show how well DL works for gender detection in multilingual and emotionally complex settings. It outperforms traditional models. The experiment describes the potential of DL in speech-based artificial intelligence (AI) applications, which enhances the performance in real-life scenarios.

Corresponding Author:

Ginika Mahajan
Department of Data Science and Engineering, Faculty of Science, Technology, and Architecture
Manipal University Jaipur
Dehmi Kalan, Off Jaipur-Ajmer Expressway, Jaipur, Rajasthan, India
Email: ginika.mahajan@jaipur.manipal.edu

1. INTRODUCTION

Human beings can communicate effectively through speech. It conveys information such as language, gender, emotion, age, and accent. Whereas humans detect those through instinctive patterns very easily, machines find it hard to cope with variations in speech. Thanks to rapid advancements in artificial intelligence (AI), detecting gender through speech has become essential for uses in security, personalized services, targeted marketing, and accessibility [1], [2]. Banks, call centers, voice bots, and digital assistants all rely on this technology [3], [4]. Tech giants like Google and Amazon use voice-based gender detection to enhance user experiences, make their services more adaptive and user-friendly [5].

This study introduces a system that identifies whether a speaker is male or female, working across several languages. The drive behind this research comes from the increasing demand for language-agnostic, gender-aware systems. To make this happen, researchers used three different datasets, including RAVDESS (English) [6], Berlin EmoDB (German) [7], and IITKGP-SEHSC (Hindi) [8]. These datasets present a variety of patterns of speech, which allows evaluating gender detection holistically, precision in monolingual and cross-lingual. The amount of audio data was 13,975 samples, gender stratified and divided into 70% training

and 30% testing. This produced 9,782 training samples and the training samples (4,193 testing samples) and the classes per half of the sample were evenly divided to clear out bias. The most important acoustic features were acquired: mel-frequency cepstral coefficients (MFCC), chroma, mel spectrogram, and VGGish embeddings [9], [10]. The integrity of the data was maintained by no data augmentation. In order to enhance precision, we tested machine learning (ML), deep learning (DL), and hybrid models. ML models were decision tree (DT), random forest (RF), XGBoost, K-nearest neighbors (KNN), support vector machine (SVM), logistic regression, and ensemble soft voting [11]-[13]. Convolutional neural network (CNN), long short-term memory (LSTM), and a hybrid CNN+LSTM+GRU were used as DL models [14], [15].

These models have been tested on monolingual and cross-lingual data to test their strength. This research preconditions the introduction of AI applications that are gender-sensitive, such as personal assistants, sentiment analysis, and language learning applications [13], [16]. The vast majority of the existing cross-lingual techniques are not very good due to the differences in accents, and the disproportionate nature of the data, which makes it unable to generalize between languages.

In recent years, there has been a lot of research done on speech-based gender recognition, implementing strategies from modern day DL to basic signal processing approaches. A summary of previous research is presented in Table 1.

Table 1. Comparative summary of prior gender detection studies

Authors and year	Dataset	Methods	Reported accuracy (%)	Limitaions
Uddin *et al*. (2021) [13].	TIMIT, RAVDESS, and BGC.	Multi-output 1D CNN with MFCC + LPC.	93.01	Limited cross-lingual scope and smaller dataset.
Doukhan *et al*. (2018) [17].	French audiovisual corpus.	GMM, i-vectors, and CNN.	96.52	Focused only on French and limited generalization.
Ali *et al*. (2022) [16].	American English vowel-emphasized dataset.	ANN with MFCC.	97.07	Monolingual and vowel-biased dataset.
Jasuja *et al*. (2020) [18].	3,168 voice samples.	Time-variant multilayer perceptron (MLP).	96.00	Basic MLP, did not handle cross-lingual variation.

The paper has the following organization: section 1 presents the introduction of the work and summarizes the related research, section 2 provides the description of the methods, section 3 contains result and discussion, section 4 presents conclusion of the work.

2. METHOD

The method describes how a multilingual gender detection model will be built with the help of ML and DL algorithms as illustrated in Figure 1. It entails processing of data, feature extraction, selection of models and performance testing. Both the monolingual and cross-lingual models were used in training and testing the proposed models to determine their effectiveness.

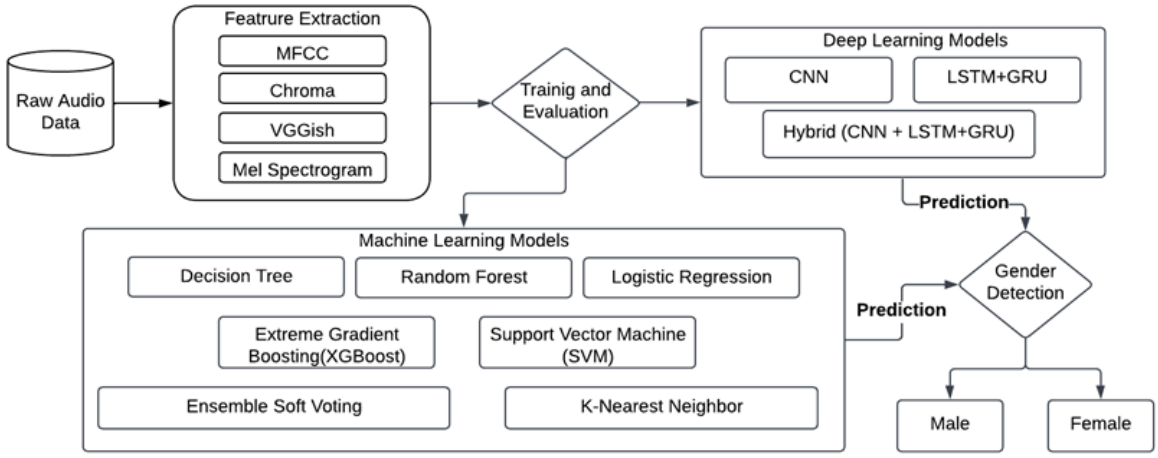

Figure 1. Architecture diagram of the proposed gender detection model

2.1. Feature extraction

In order to convert raw audio data into meaningful representations that ML and DL models may use to classify gender, feature extraction is essential. In this work 40 MFCCs were extracted for every audio. MFCCs capture pitch, timber and give timely features differentiating the genders [9], [11], [13], [16], [19]. Extracted 12 chroma features that represents spectral energy distribution across 12 pitch classes that capture tonal patterns [12], [13], [16], [20]. Mel spectrogram demonstrates the absolute frequencies in an audio signal which extracts 128 Mel bands. It represents frequency energy distribution [16], [21]. Spectral contrasts isolate the sound texture and dynamics. The Spectral centroid identifies gender through pitch. The spectral roll-off distinguish between speakers by vocal power. Zero crossing rate (ZCR) measures signal changes in pitch and energy. RMSE measures loudness [11], [15]. A VGGish embedding extracts deep audio embeddings of 128 dimensions. All audio files are resampled to 16 kHz [10].

The above extracted features were normalized and concatenated into a unified vector. Then resized into an array of the shape (20, 450) and saved as .npy files with differentiated gender. The above extracted features were normalized and concatenated into a unified vector. Then resized into an array of the shape (20, 450) and saved as .npy files with differentiated gender. Finally, the data is combined into one file for training and testing. MFCCs, VGGish embeddings were the most crucial, and CNN accuracy decreased by was eliminated. Chroma and spectral features made 2% or less difference. Figures 2 and 3 represent the MFCC features and mel spectrogram of energy distribution.

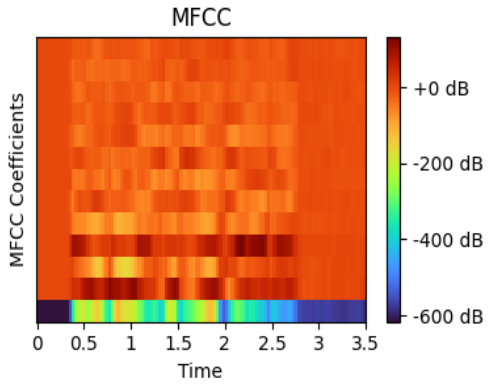

Figure 2. MFCC feature extraction for an audio sample

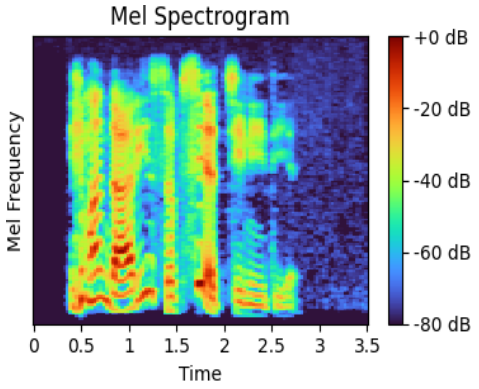

Figure 3. The mel spectrogram shows energy distribution across frequencies

2.2. Machine learning models

This study developed gender classification models using different ML methods for both monolingual and cross-lingual data. Audio features were extracted and transformed into a 2D format. The data was split into 70% training and 30% testing, and the evaluation is done based on accuracy and F1-score Koolagudi *et al*. [8].

- DT: utilized with a fixed random state with probability-based log loss.
- RF: implemented with 100 estimators and a max depth of 20.
- XGBoost: applied to improve computational efficiency and reduce overfitting [22].
- KNN: used with k=5 for classification [21].
- SVM: applied with a linear kernel on standardized features [22].
- Logistic regression: used with the 'lbfgs' solver for multinomial classification.

An ensemble soft voting model combined DT, RF, SVM, KNN, XGBoost, and logistic regression, using probability-based voting to improve accuracy [11], [21]. Normalization of SVM and logistic regression and other models was done. All models were tested in both monolingual (trained and tested on the same language) and cross-lingual (trained on one language, tested on others) conditions.

2.3. Deep learning models convolutional neural network

On gender classification, a CNN model was developed. Audios were restructured to normalized 2D input. The CNN consisted of three convolutional layers, rectified linear unit (ReLU), batch normalization, max-pooling, and dropout. Filter sizes were enlarged in a gradual manner of 64, 128, and 256 [14]. The progression with the greatest trade off of overfitting and recall produced a more accurate result of 1.8% and

best generalization. This was flattened and fed into a dense layer of 512 neurons and dropout (0.5) after which it was fed into a sigmoid output layer to be classified into binary. Adam optimizer (learning rate equal to 0.001) and binary cross-entropy loss were incorporated in the model. Training was optimized by the use of callbacks such as ReduceLROnPlateau, EarlyStopping, and ModelCheckpoint [8], [11], [15], [17], [21], [23]. To measure the performance of the models, accuracy, F1-score, and classification reports were measured and the results were graphically represented in Figure 4.

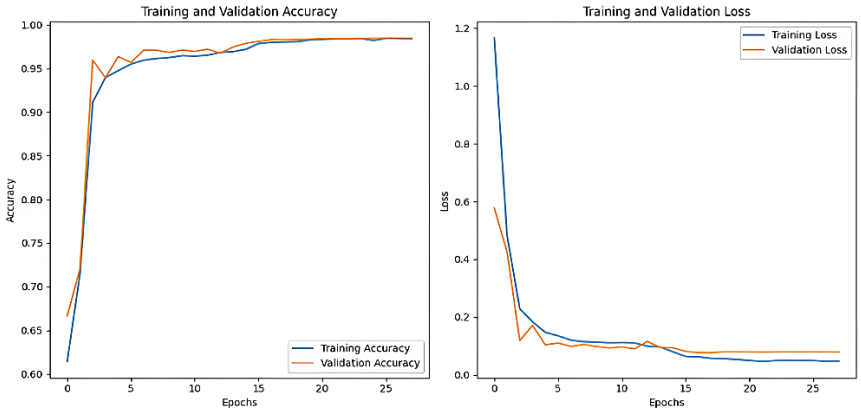

Figure 4. Training and validation performance of CNN model

2.4. Long short-term memory+gated recurrent unit model

An LSTM+GRU model was developed for gender classification using precomputed audio features in a 3D tensor format (samples, time steps, and features) to capture temporal dependencies [13]. The architecture includes an LSTM layer (256 units) with tanh activation and 0.4 dropout, followed by a GRU layer (128 units) to capture sequential patterns. A Bidirectional LSTM (128 units) enhances context learning. Outputs pass through a dense layer (128, ReLU) and a final sigmoid layer for binary classification [24]. The model uses Adam (lr=0.001) and binary cross-entropy, with training optimized by EarlyStopping, ModelCheckpoint, and ReduceLROnPlateau. The model effectively captures long- and short-term dependencies, making it robust for cross-lingual gender prediction [12], [15]. Although temporal models are more advantageous, CNN was competitive across languages because of better timbral feature learning. The performance of the LSTM+GRU model is shown in Figure 5.

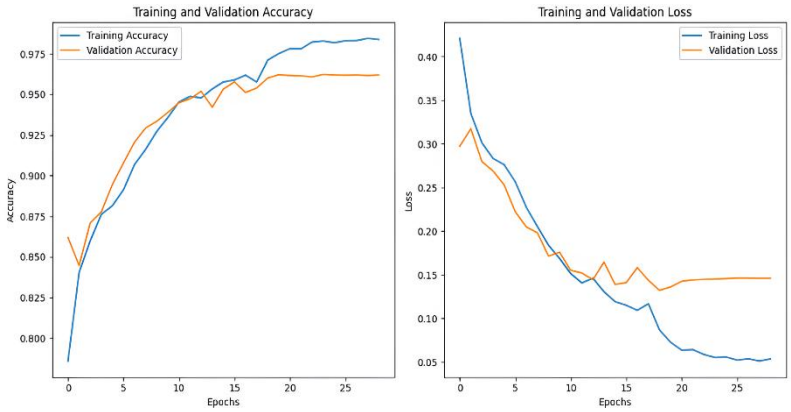

Figure 5. Training and validation performance of LSTM+GRU model

2.5. Hybrid convolutional neural network+long short-term memory+gated recurrent unit model

Apart from CNN and LSTM+GRU, we have combined CNNs and LSTM/GRU/bidirectional LSTM for gender detection from the precomputed features. This data is then rearranged to a four dimensional compatible with the passage through CNN. The CNN layers include three layers of Conv2D to detect spatial

patterns that are present in the input, followed by BatchNormalization, MaxPooling2D, and Dropout to decrease the over-fitting issue. It was the output that recast and reoccurred on the time step 20 that rendered it amenable to sequential processing with CNN [12], [13], [18], [21].

The repetition is to adjust 2D spectral frames to a known length sequence but produces redundancy. The sequential layers represent a 128-unit LSTM which captures the long-term dependency structures, 128-unit GRU that captures the short-term structures and a Bidirectional LSTM layer that should be able to capture both the long and short-term structures in the data. For feature extraction, a dense layer with 'same padding' is added to the last layer of the model. For gender prediction, a fully connected layer with sigmoid as an activation function is applied as the output layer. It uses the Adam optimizer with binary cross-entropy loss; thus they incorporate efforts like ReduceLROnPlateau, EarlyStopping, and ModelCheckpoint to enhance the model and to avoid overfitting [14]-[16], [25]. It underperformed due to over-parameterization relative to dataset size and the above redundancy, simplifying with temporal pooling or attention in place of repetition is recommended. The results of a Hybrid (CNN+LSTM+GRU) model are given in Figure 6.

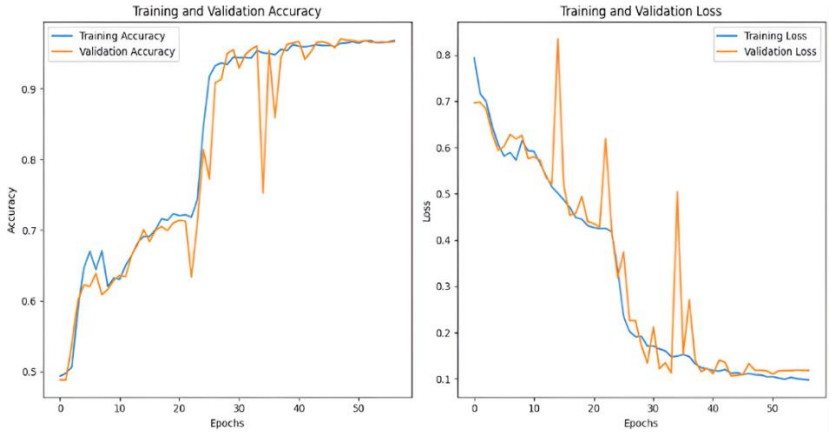

Figure 6. Training and validation performance of hybrid (CNN+LSTM+GRU) model

3. RESULTS AND DISCUSSION

The paper compares the performance of different ML and DL models on gender detection on monolingual and cross-lingual basis on the RAVDESS, EmoDB, Hindi, and combined datasets. In order to compare the performance of different systems under different conditions, multiple ML algorithms and DL architectures were deployed. The performance of the model is provided in Table 2.

Table 2. Performance comparison of ML and DL models for gender detection

Lingual approach	Algorithm type	Algorithm	Dataset			
			RAVDESS (%)	EmoDb (%)	Hindi (%)	Combined (%)
Monolingual	ML	DT	82.64	66.98	85.59	N/A
		RF	94.56	82.24	97.14	N/A
		XGBoost	94.91	83.18	98.26	N/A
		K-NN	85.07	66.98	85.34	N/A
		SVM	88.77	71.65	95.58	N/A
		Logistic regression	91.67	74.45	97.79	N/A
		Ensemble soft voting classifier	94.10	85.98	97.96	N/A
	DL	CNN	74.65	75.08	99.33	N/A
		LSTM+GRU	91.55	75.08	97.37	N/A
		Hybrid (CNN+LSTM+GRU)	66.67	57.32	91.00	N/A
Cross-lingual	ML	DT	N/A	N/A	N/A	83.56
		RF	N/A	N/A	N/A	96.12
		XGBoost	N/A	N/A	N/A	96.90
		KNN	N/A	N/A	N/A	84.83
		SVM	N/A	N/A	N/A	93.64
		Logistic regression	N/A	N/A	N/A	94.96
		Ensemble soft voting classifier	N/A	N/A	N/A	96.57
	DL	CNN	N/A	N/A	N/A	98.11
		LSTM+GRU	N/A	N/A	N/A	96.22
		Hybrid (CNN+LSTM+GRU)	N/A	N/A	N/A	96.61

3.1. Monolingual approach

In the monolingual approach, ML algorithms generally performed well on individual datasets. Models such as DT, RF, KNN, XGBoost, logistic regression, SVM, and ensemble soft voting were evaluated. As shown in Table 2, RF achieved strong performance, especially on the Hindi dataset with 97.14% accuracy outperformed [16]. The ensemble soft voting classifier also performed well with 94.10% on RAVDESS, 85.98% on EmoDB, and 97.96% on Hindi. Among all ML models, XGBoost provided the best accuracy of 94.91% on RAVDESS, 83.18% on EmoDB, and 98.26% on Hindi.

In DL models, CNN achieved 99.33% on Hindi but showed lower results on RAVDESS (74.65%) and EmoDB (75.08%). The LSTM+GRU model outperformed CNN on RAVDESS (91.55%) and Hindi (97.37%). On Hindi, the Hybrid CNN+LSTM+GRU model achieved 91.00% accuracy. The large performance gap between Hindi, English, and German datasets is mainly due to dataset size, and stronger emotional variation in RAVDESS and EmoDB.

3.2. Cross-lingual approach

In the cross-lingual approach using the combined dataset, DL models outperformed ML models. Among ML algorithms, XGBoost achieved the highest accuracy of 96.90%, followed by the ensemble soft voting classifier (96.57%) and RF (96.12%). DL models demonstrated even stronger performance. CNN achieved the highest accuracy of 98.11%, which outperforms prior studies such as Uddin *et al*. [13], represented in Table 2, the hybrid model (96.61%) outperformed Jasuja *et al*. [18] and LSTM+GRU (96.22%). They also provided cross-language matching with different errors and suggested further experiments with multilingual embeddings/transformer models. Detailed results are provided in Table 2.

4. CONCLUSION

The paper introduces a multilingual DL gender detection model based on RAVDESS: to compare ML and deep, EmoDB: dataset, and IITKGP-SEHSC (Hindi): dataset learning styles of monolingual and cross-lingual. Results from the monolingual experiments show that XGBoost and RF were very high-accurate among ML. Whereas LSTM+GRU excelled in the classification of the DL-based models, it was able to handle the complicacy of speech data. Nevertheless, the hybrid architecture (CNN+LSTM+GRU) has not been successful. Good performance, perhaps because of the complexity of the model and the nature of the data. In cross-lingual models, experiments were tested on the joint dataset. Particularly ML models. When they do well, however XGBoost, and the ensemble methods, but the DL models, especially. CNN was rather useful in deriving gender specific features across languages. CNN was the most accurate therefore, it is the most resistant to gender detection. These findings show that cross-lingual datasets may be used to develop effective gender detection. systems. Weaknesses: noise sensitivity. The following working day can be spent on the work on the improvement of hybrid. building bigger datasets that have more languages and accents and exploring, more complicated models such as attention mechanisms and transformers. It restricted to dualistic gender classification and may be culturally biased therefore the future expansion should provide, non-binary representation, and use in real-world systems like voice assistants and accessibility devices.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses 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
Kavita Jhajharia	✓	✓		✓		✓	✓	✓		✓		✓		
Ginika Mahajan	✓	✓		✓		✓		✓		✓		✓		
Dhondi Samrudh		✓	✓			✓	✓	✓	✓					
Koustubh Patel		✓	✓			✓	✓	✓	✓					

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the first author, [initials: KJ].

REFERENCES

[1] S. Amith, M. D. Chinmayi, K. H. Kshama, H. G. Kartik, and C. K. Shashank, "Speech Emotion Detection System," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 14, no. 5, May 2025, doi: 10.17148/IJARCCE.2025.14502.

[2] M. L. Mura and P. Lamberti, "Human-Machine Interaction Personalization: A Review on Gender and Emotion Recognition Through Speech Analysis," in *IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT 2020 - Proceedings*, 2020, pp. 319–323, doi: 10.1109/MetroInd4.0IoT48571.2020.9138203.

[3] D. Weerabahu, A. Gamage, C. Dulakshi, G. U. Ganegoda, and T. Sandanayake, "Digital Assistant for Supporting Bank Customer Service," in *Communications in Computer and Information Science*, 2019, vol. 890, pp. 177–186, doi: 10.1007/978-981-13-9129-3_13.

[4] A. K. Somani, R. S. Shekhawat, A. Mundra, S. Srivastava, and V. K. Verma, *Smart Systems and IoT: Innovations in Computing*, vol. 141. Singapore: Springer Singapore, 2020, doi: 10.1007/978-981-13-8406-6.

[5] G. A. Amiri, S. Shahidi, M. Mehri, F. A. Darmel, J. A. Niazi, and M. A. Anwari, "Decoding Gender Representation and Bias in Voice User Interfaces (VUIs)," *International Journal of Computer Science and Mobile Computing*, vol. 13, no. 5, pp. 76–88, 2024, doi: 10.47760/ijcsmc.2024.v13i05.007.

[6] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north American english," *PLoS ONE*, vol. 13, no. 5, 2018, doi: 10.1371/journal.pone.0196391.

[7] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of German emotional speech," in *9th European Conference on Speech Communication and Technology*, pp. 1517–1520, 2005, doi: 10.21437/interspeech.2005-446.

[8] S. G. Koolagudi, R. Reddy, J. Yadav, and K. S. Rao, "IITKGP-SEHSC: Hindi speech corpus for emotion analysis," in *2011 International Conference on Devices and Communications, ICDeCom 2011-Proceedings*, 2011, doi: 10.1109/ICDECOM.2011.5738540.

[9] Z. K. Abdul and A. K. Al-Talabani, "Mel Frequency Cepstral Coefficient and its Applications: A Review," *IEEE Access*, vol. 10, pp. 122136–122158, 2022, doi: 10.1109/ACCESS.2022.3223444.

[10] M. Diwakar and B. Gupta, "The robust feature extraction of audio signal by using VGGish model," *Internatioinal Journal of Computer Science and Information Security (IJCSIS)*, vol. 21, no. 6. pp. 1–18, Jun. 13, 2023, doi: 10.21203/rs.3.rs-3036958/v1.

[11] R. S. Alkhawaldeh, "DGR: Gender Recognition of Human Speech Using One-Dimensional Conventional Neural Network," *Scientific Programming*, vol. 2019, pp. 1–12, Sep. 2019, doi: 10.1155/2019/7213717.

[12] D. Kwasny and D. Hemmerling, "Gender and age estimation methods based on speech using deep neural networks," *Sensors*, vol. 21, no. 14, 2021, doi: 10.3390/s21144785.

[13] M. A. Uddin, R. K. Pathan, M. S. Hossain, and M. Biswas, "Gender and region detection from human voice using the three-layer feature extraction method with 1D CNN," *Journal of Information and Telecommunication*, vol. 6, no. 1, pp. 27–42, 2022, doi: 10.1080/24751839.2021.1983318.

[14] P. Damacharla, H. Rajabalipanah, and M. H. Fakheri, "LSTM-CNN Network for Audio Signature Analysis in Noisy Environments," in *Proceedings - 2023 International Conference on Computational Science and Computational Intelligence, CSCI 2023*, 2023, pp. 78–82, doi: 10.1109/CSCI62032.2023.00019.

[15] A. Mohanty and R. C. Cherukuri, "Whispered Speech Emotion Recognition with Gender Detection using BiLSTM and DCNN," *Journal of Information Systems and Telecommunication*, vol. 12, no. 2, pp. 152–161, 2024, doi: 10.61186/jist.43703.12.46.152.

[16] 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*, vol. 28, no. 2, pp. 753–761, 2022, doi: 10.11591/ijeecs.v28.i2.pp753-761.

[17] D. Doukhan, J. Carrive, F. Vallet, A. Larcher, and S. Meignier, "An Open-Source Speaker Gender Detection Framework for Monitoring Gender Equality," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, vol. 2018, pp. 5214–5218, doi: 10.1109/ICASSP.2018.8461471.

[18] 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*, 2020, pp. 319–324, doi: 10.1109/ICOSEC49089.2020.9215254.

[19] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, "Deep neural network framework and transformed MFCCs for speaker's age and gender classification," *Knowledge-Based Systems*, vol. 115, pp. 5–14, 2017, doi: 10.1016/j.knosys.2016.10.008.

[20] G. R. Nitisara, S. Suyanto, and K. N. Ramadhani, "Speech Age-Gender Classification Using Long Short-Term Memory," in *2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020*, 2020, pp. 358–361, doi: 10.1109/ICOIACT50329.2020.9331995.

[21] S. Bhattacharya, B. K. Mishra, S. Borah, N. Das, and N. Dey, "Cross-lingual deep learning model for gender-based emotion detection," *Multimedia Tools and Applications*, vol. 83, no. 9, pp. 25969–26007, 2024, doi: 10.1007/s11042-023-16304-x.

[22] O. Andronati, S. Antoshchuk, A. Nikolenko, O. Arsirii, O. Babilunha, and D. Petrosiuk, "Ensemble Classifiers of Audio Data for Speech Emotions Recognition," in *Proceedings - International Conference on Advanced Computer Information Technologies, ACIT*, 2023, pp. 623–626, doi: 10.1109/ACIT58437.2023.10275604.

[23] V. K. Chauhan, J. Zhou, P. Lu, S. Molaei, and D. A. Clifton, "A brief review of hypernetworks in deep learning," *Artificial Intelligence Review*, vol. 57, no. 9, 2024, doi: 10.1007/s10462-024-10862-8.

[24] S. M. Al-Selwi *et al.*, "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 5, 2024, doi: 10.1016/j.jksuci.2024.102068.

[25] A. A. Alnuaim *et al.*, "Speaker Gender Recognition Based on Deep Neural Networks and ResNet50," *Wireless Communications and Mobile Computing*, 2022, doi: 10.1155/2022/4444388.

BIOGRAPHIES OF AUTHORS

Kavita Jhajharia is an Assistant Professor in the Department of Information Technology, Manipal University Jaipur, India. She earned her Ph.D. in Information Technology from the same institution in 2023. Her research primarily focuses on the application of machine learning and deep learning techniques in agriculture. She has published extensively in high-impact journals. Her recent works address challenges in agricultural prediction, biomedical acoustic feature classification, and software engineering. She can be contacted at email: Kavita.chaudhary@outlook.com.

Ginika Mahajan is senior Assistant Professor working at the Department of Information Technology, Manipal University Jaipur, India. Her main teaching and research interests include software engineering, security, machine learning, NLP, and IR. She has published several research articles in international journals and conferences of software engineering. She can be contacted at email: Ginika.mahajan@jaipur.manipal.edu.

Dhondi Samrudh is a B.Tech. final year student in Data Science and Engineering at Manipal University Jaipur. He has hands-on experience in Python, machine learning, and deep learning. He is passionate about building systems that bridge the gap between data and real-world solutions. His current research focuses on leveraging these technologies to address challenges in various fields. Alongside his technical skills, he is growing his expertise in web development and data structures, with a keen interest in solving real-world problems through innovation and code. He can be contacted at email: samrudhdhondi@gmail.com.

Koustubh Patel is a B.Tech. student in Data Science and Engineering at Manipal University Jaipur. With a strong foundation in machine learning, and deep learning, and is especially interested in building smart applications. His current activity is to use these tools to address real-life issues in various fields. Koustubh is passionate about constant learning and innovation and strives to create solutions that have an impact and integrate technology and daily needs. He can be contacted at email: koustubhpwork@gmail.com.