Communications in Mathematics and Applications
Vol. 15, No. 1, pp. 315–327, 2024
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v15i1.2450

http://www.rgnpublications.com

Research Article

Analysis of Speech Features for Gender Identification in Tai Language

Kankana Dutta*, Rizwan Rehman and Ankumon Sarmah

Centre for Computer Science and Applications, Dibrugarh University, Dibrugarh, Assam, India
*__Corresponding author:__ kankanadutta@dibru.ac.in

Received: October 5, 2023 **Accepted:** January 16, 2024

Abstract. The vast number of information packed into the human speech signal makes analysis a tough undertaking. This intricacy is notably noticeable in tasks like speaker recognition, especially when it comes to gender distinction. In this paper, we address this issue by conducting a thorough examination of the effectiveness of various speech features, namely Pitch, Formant Frequency, MFCC (Mel Frequency Cepstral Coefficients), and Chroma, in the context of gender identification in the Tai Language, which is spoken by the Tai people of Assam. In this study, we use machine learning (SVM, KNN, Decision Tree, Neural Network) to analyze speech features (Pitch, Formant Frequency, MFCC, Chroma) for gender identification in the Tai language spoken by the Tai people of Assam. Our results show that MFCC consistently outperforms other features, delivering the highest accuracy rates across all approaches. This demonstrates MFCC's ability to extract gender information from Tai Language speech signals, suggesting more accurate gender identification systems. Beyond gender identification, our study extends voice analysis in linguistics and improves the application of spoken language data, allowing for improved communication systems and linguistic insights. In summary, our findings highlight the critical significance of MFCC in gender identification in the Tai language, with ramifications that extend far beyond its local context, promising advances in voice analysis and improving our understanding of language and human communication.

Keywords. Machine learning methods, Neural networks, Gender identification, MFCC, Pitch, Formant frequency, Chroma

Mathematics Subject Classification (2020). 68T10

1. Introduction

With the growth of Artificial Intelligence, several systems are implemented using voice-based technologies that involve human-computer interaction. Gender identification is a significant part of speech-related technology. A significant component of speech-related technology is gender identification. Recognizing the speaker's gender is useful for different personal assistance in producing male and female-focused results. The process of recognizing the speaker or the spoken command sent to the computer gets easier with the identification of the speaker's gender (Alnuaim *et al*. [3]). It also increases the performance of other applications involving human-computer interaction. Gender recognition can be useful in speech-based biometric systems for authentication and identification purposes. In health care, gender identification can help by providing personalized care based on the patient's gender. Thus, gender identification plays an important role in several fields related to technology as well as other fields such as health care, research, etc.

The speech signal carries a lot of information about the speaker as well as the information conveyed. The initial method of speech processing is to study and identify the information or features, which are relevant to the purpose. Identification of these features is a critical task as it can determine the efficiency of the system. Features like pitch, formant frequency, and MFCC (Ramdinmawii and Mittal [14], and Ramteke *et al*. [15]) are extracted and used in various research works to acquire information about the speaker.

This study uses the Tai language spoken by the Tai community in Assam. This is a low-resource language and very few people practice this language in Assam. The language originates from the Thai language group in southern China and South-East Asia. Six Tai tribes practice Tai languages in Assam and North-East India. They are — Tai Ahom, Tai Phake, Tai Khamti, Tai Khamyang, Tai Iton, and Tai Turung. The Tai Ahom tribe founded the Ahom Kingdom and ruled Assam for about 600 years. These six tribes have tone differences in their respective languages but the basic characteristics of the Tai language can be found the same. Today, there are very few people left to practice the language, and Tai has been supplanted by the Assamese language. There are extensive Tai language manuscripts that can help find a variety of undiscovered information about the social, political, and cultural aspects of ancient Assam and the Tai peoples. Therefore digitization of the Thai language is considered important for preserving these manuscripts and also encourages young people to learn the language (Gohain and Gohain [7]).

In the field of gender identification, many successful studies have been conducted on different languages around the world. However, for low-resource languages, this is challenging due to the unavailability of data. Usually, very few datasets are available for conducting experiments for low-resource languages. For the Tai language, there is no publicly available Tai dataset for experimentation. Data collection is also difficult due to the lack of speakers. Also, to the best of our knowledge, no research work has been carried out in the field of gender identification to date for this language.

In this paper, we have attempted to find speech features that are useful for the Tai language in the case of gender identification. The speech features have been compared using different supervised machine-learning methods. The primary contribution of this paper is:

(i) Development of a dataset of voice data containing the vowels of a low-resource language Tai language.

(ii) Study of the Tai language speech features that are useful for gender identification.

(iii) Development of a feed forward neural network model for gender identification.

The rest of the paper is organized as follows. Section 2 shows related works. Section 3 describes the motivation and contribution. Section 4 describes the methodology used to perform the study. Section 5 describes the findings of the study. Finally, Section 6 concludes the paper.

2. Literature Review

Numerous researches have been carried out to determine the technology and speech features useful for gender identification. Some of the recent works relevant to the study are presented here.

A study was performed by Alnuaim *et al*. [3] for a perfect classification model for language-independent gender identification using Deep Neural Network and ResNet50 with features pitch, MFCC, chroma, and Tonnenz and obtained an accuracy of 98.57%.

A system for gender identification for under-resourced African language (Sefara and Mokgonyane [17]) was developed using Multilayer Perceptron, Convolutional Neural Network, and Long Short Term Memory with speech features energy, ZCC, MFCC, Chroma, Spectral values and obtained an accuracy of 97%.

Uddin *et al*. [19] proposed a model for gender and geographical region detection using CNN with three-layer feature extraction using the features' Fundamental frequency, spectral entropy, spectral flatness, and mode frequency in the first layer, Mel Frequency Cepstral Coefficient in the second layer and linear predictive coding in the third layer and the model has successfully predicted the gender with 93.01% of accuracy.

The performance of different deep neural networks to jointly identify age and gender has been analyzed (Sánchez-Hevia *et al*. [16]) using MFCC speech features and found that all types of neural networks are good in gender classification.

Khanum and Sora [9] proposed a method for gender identification using a feed-forward neural network with MFCC speech features and found that the proposed method learns and analyses faster and better.

3. Motivation and Objectives

Gender identification from speech is an essential and practical application with a wide range of potential uses, including human-computer interaction, voice-based authentication, and personalized services. Over the years, researchers have explored various methodologies and speech features such as MFCC to improve the accuracy of gender identification systems (Sefara

Table 1. Comparison of related work

S. No.	Paper name	Methods used	Features used	Pros and cons
1	Speaker Gender Recognition Based on Deep Neural Networks and ResNet50 [3]	DNN, REsNet50	Pitch, MFCC, chroma, and Tonnenz	Good accuracy but calculation complexity and time of calculation increase with and increase in the number of layers in the network
2	Gender Identification in Sepedi Speech Corpus [17]	Multilayer Perceptron, Convolutional Neural Network, and Long Short Term Memory	energy, ZCC, MFCC, Chroma, Spectral values	Models showed good results except for MLP, whose results could not be trusted since the model was not stable
3	Gender and region detection from human voice using the three-layer feature extraction method with 1D CNN [19]	CNN	Fundamental frequency, spectral entropy, spectral flatness, mode frequency, Mel Frequency, Cepstral Coefficient	Works well in the combined dataset. The accuracy rate can be improved
4	Age group classification and gender recognition from speech with temporal convolutional neural networks [16]	DNN	MFCC	Good accuracy but performance is dependent on the network size
5	Speech-based Gender Identification using Feed Forward Neural Networks [9]	Feed-Forward Neural Network	MFCC	Works well for noisy data

and Mokgonyane [17]). It is also found that in recent times researchers have used different neural network models to increase the accuracy rate of gender identification (Alnuaim *et al*. [3], Khanum and Sora [9], Sánchez-Hevia *et al*. [16], Sefara and Mokgonyane [17], and Uddin *et al*. [19]).

However, most of the existing studies focus on widely spoken languages, leaving low-resource languages, such as the Tai language, largely unexplored in this field. The Tai language is an example of a low-resource language with unique phonetic characteristics, making gender identification in this language challenging. By conducting research in the Tai language, we aim to fill this gap in the literature and provide valuable insights into the effectiveness of different speech features and machine learning techniques for gender identification in low-resource languages.

Motivated by the growing significance of gender identification in various applications and the lack of research in this field for the Tai language, this paper aims to explore and enhance the understanding of gender identification techniques specifically tailored to low-resource languages. The objectives of this study are as follows:

(i) To conduct a comprehensive investigation of the performance of distinct speech features, including pitch, formant frequency, chroma, and MFCC, in the context of gender identification for the Tai language. By assessing the strengths and weaknesses of each feature, we aim to gain insights into their suitability for gender classification in this specific linguistic context.

(ii) To evaluate the effectiveness of four supervised machine learning methods, namely kNN, Decision Tree, SVM, and a Feed Forward Neural Network model, for gender identification in the Tai language. By comparing these models, we intend to identify the most appropriate approach for accurate gender classification in this low-resource language.

(iii) To explore the potential benefits of feature fusion and combinations of speech features to enhance gender identification accuracy. By investigating different fusion strategies, we aim to leverage the complementary information provided by multiple features and improve the overall performance of gender classifiers.

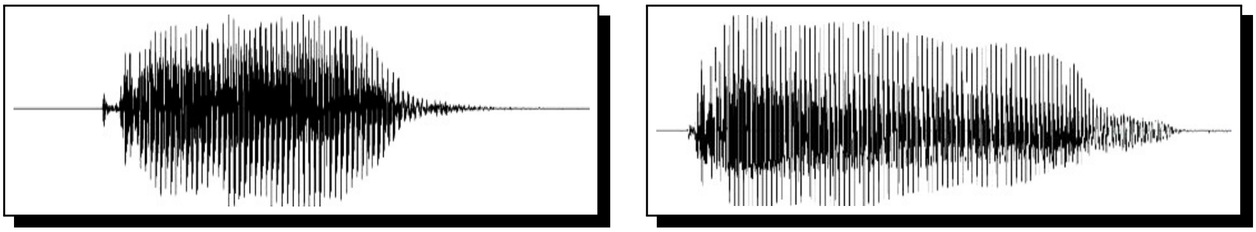
(iv) To establish a baseline performance for gender identification in the Tai language using the widely studied MFCC features. This will allow us to compare the results of other features and models against a common reference point.

4. Methodology

A set of experiments has been conducted to evaluate the performance of four common speech features that are relevant to gender identification. For that, speech features were extracted first and evaluation was done using different machine learning methods. This section presents the features, the dataset, and the experimental settings used.

4.1 Features

Audio signals are composed of lots of information. To extract the information or the features, the speech signal is represented in the frequency domain. Figure 1 represents voice samples of male and female speakers in the frequency domain.



(a) Male voice

(b) Female voice

Figure 1. Voice sample of male and female speakers

In this experiment, we have considered the following four types of speech features:

Pitch: Pitch is the perceived fundamental frequency of voiced speech. It is one of the most used features in speech processing. Every individual has a pitch range constrained by his or her larynx. For men, the pitch range is in the range of 50 to 250 hertz, and for women, it is between 120 to 500 Hz which is higher compared to men (Rabiner and Schafer [13]).

Formant frequency: Formants are distinctive frequency components of the acoustic signal produced by speech. There are four types of formant frequencies F1, F2, F3 and F4. Research shows that information about gender is concentrated on the lower frequency part of the speech signal and therefore F1 and F2 play a crucial role in gender identification (Medhi [11]).

MFCC: MFCC or Mel-frequency cepstral coefficients (MFCCs) is one of the most used features for gender identity as well as speaker identification. MFCC is a cepstral method that converts speech into parameters according to the Mel scale. Different research has been performed with different numbers of MFCC values. Research also shows that increasing the number of MFCCs yields better classification metrics with a lower computational burden (Hasan *et al*. [8]). Therefore, in this experiment, 40 number of MFCC features were used.

Chroma: In speech processing, chroma features represent 12 different pitch classes, this is a powerful tool in speech processing in determining gender and emotion detection (Alkhawaldeh [2]).

These speech features were extracted using the Python Librosa library, a tool for music and audio analysis. Table 2 shows the features with the number of parameters used for the experiments.

Table 2. Number of parameters for each feature

Features	Number of parameters
Pitch	01
Formant frequency	02
MFCC	40
Chroma	12

4.2 Dataset

The dataset used for this experiment is a Tai language dataset, which contains speech samples of both male and female speakers. There is a total 23 number of speakers, out of which 15 are male, and 8 are female. The gender distribution is unbalanced as we can see in Figure 2. A speech sample from each speaker is collected by speaking the vowel sounds of the Tai language. There are a total of 7329 speech samples present in the dataset. The age of the speakers ranges between 20 to 36 years. The audio files are in .wav format with a sampling rate of 22050. The audio files are of good quality and were recorded in a noise-free environment.

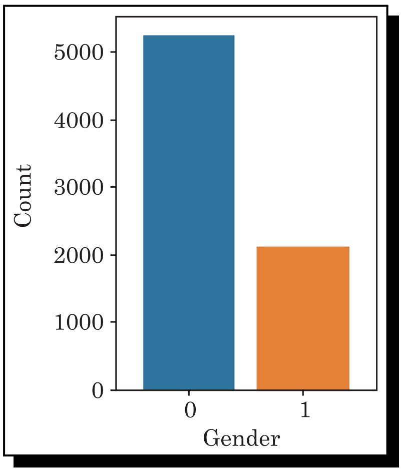


Figure 2. Dataset gender distribution

4.3 Supervised Machine Learning Techniques Used

Some of the popular methods used in gender identification are:

SVM: Support Vector Machine is a supervised machine learning algorithm used for classification and regression. Based on some training data, this method can classify non-familiar data or test data. SVM is a binary classifier that assigns a given input in one of the two possible classes, and therefore this method is suitable for gender identification which is defined as a two-class problem. It is also possible to use SVM in multi-class classification. One advantage of SVM is its ability to handle high-dimensional data while avoiding over-fitting (Dagher and Azer [4], and Ghosh and Bandyopadhyay [6]).

KNN: K Nearest Neighbour is a supervised machine learning algorithm that can be used in classification and regression tasks. It is a simple yet powerful algorithm. The basic idea behind KNN is to find the K nearest neighbor to a given data point and then classify the data point based on the majority class of its neighbors. KNN is a multi-class classification problem. KNN is used in gender identification due to its simplicity. It is known as a lazy learner because it does not perform any calculations on the training data to create a compact model that can be used later. Instead, it just stores the data. The calculation is done only when unseen data is applied to the model (Abdulsatar *et al*. [1], and Raahul *et al*. [12]).

Decision tree: The decision tree method is a supervised machine learning algorithm that is used for both classification and regression. It is a tree structure classifier. The tree has a root node which is considered an initial feature or the superior decision-making node of the system. Based on a question, the tree is further split into subtrees. A decision is represented via a leaf node. The decision tree is simple to understand as it mimics the human decision-making process. Models have been developed using a decision tree for gender identification (Zhong *et al*. [20]).

Neural network: Neural networks mimic how the human brain functions. It allows the computer to recognize patterns and solve common problems in the field of artificial intelligence. A neural network or Artificial neural network consists of 3 sets of nodes — input layer, hidden layer, and output layer. These ANNs are an effective tool to perform the classification task (Mamyrbayev *et al*. [10]). There are different types of ANN based on the architecture of the network. Some of them are — Feed-forward Neural Networks, Multi-layer perceptron, Long-Short Term Memory (LSTM), Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), etc. These Neural Networks are used widely for gender identification from speech data.

The methods SVM, KNN, and Decision Tree are used in this experiment along with a feed-forward neural network. These classifiers are chosen due to the non-linear property of data. These are the classifiers for gender identification that are most frequently used and have high classification accuracy.

The Feed Forward Neural Network used in this experiment was designed with 1 input layer, 2 hidden layers, and 1 output layer. The model was trained with 100 epochs with the ADAM optimizer. The activation function used in the system was RELU (Rectified linear unit). The network was tested with various batch sizes and the best result was achieved with a batch size of 10 samples.

For the experiments, the dataset is divided into three parts viz. training, testing, and validation. The training part is used to train the models used in the experiments. The 80 percent of the data in the dataset are considered training data. The testing part of the data is used to test the model after it is trained with the training data. 15 percent of the data are used for testing purposes and the rest 5 percent of the data are used as validation data for evaluating the results of the experiments.

5. Result and Discussion

This section shows and explains the findings of the experiments. The results of the experiment performed are shown in Table 3.

The best feature among the investigated features is MFCC which shows a high accuracy rate for all four examined methods. The second best feature is Chroma which also shows a high accuracy rate for all four examined methods. On the other hand, the pitch feature is also showing similar results for all four methods considered for the experiment. Formant frequency is in last place for SVM, KNN, and, the Decision Tree method but a good accuracy rate of 94.2% was obtained for the Feed Forward Neural Network model.

Several experiments were performed to identify the speech features suitable for gender identification and a similar type of result with a 98.9% accuracy rate can be found with MFCC features for the Indian languages Hindi, English, Sanskrit, and Telugu using a back propagation neural network (Sharma *et al*. [18]). Other experiments for the Indian language also showed similar results using MFCC, Pitch, and energy features (Deiv *et al*. [5], and Hasan *et al*. [8]).

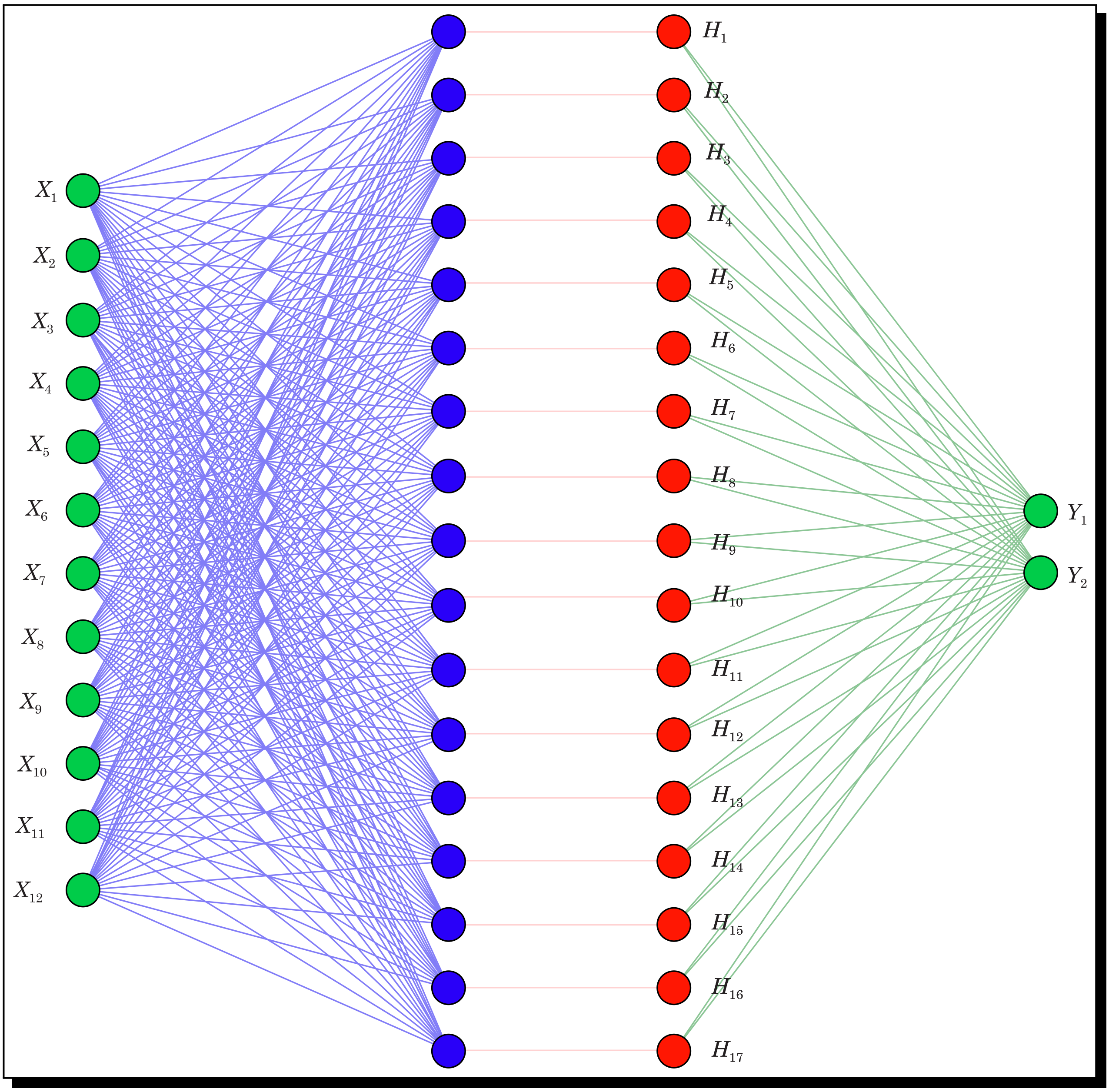


Figure 3. Feed Forward Neural Network with (1-2-1) architecture

Our experiments also showed that MFCC features are good in identifying the gender of the speaker for the Tai language and suggested that these features could be the prominent speech features to be considered in gender identification. However, chances of over-fitting in the experiments were possible due to the small size of the dataset used, as well as the small number of speakers providing the speech samples.

Table 5 represents the different statistical parameters obtained for the experiments.

The training and validation performance of the feed-forward neural network is shown in Figure 4.

Table 3. Accuracy rate of different methods applied

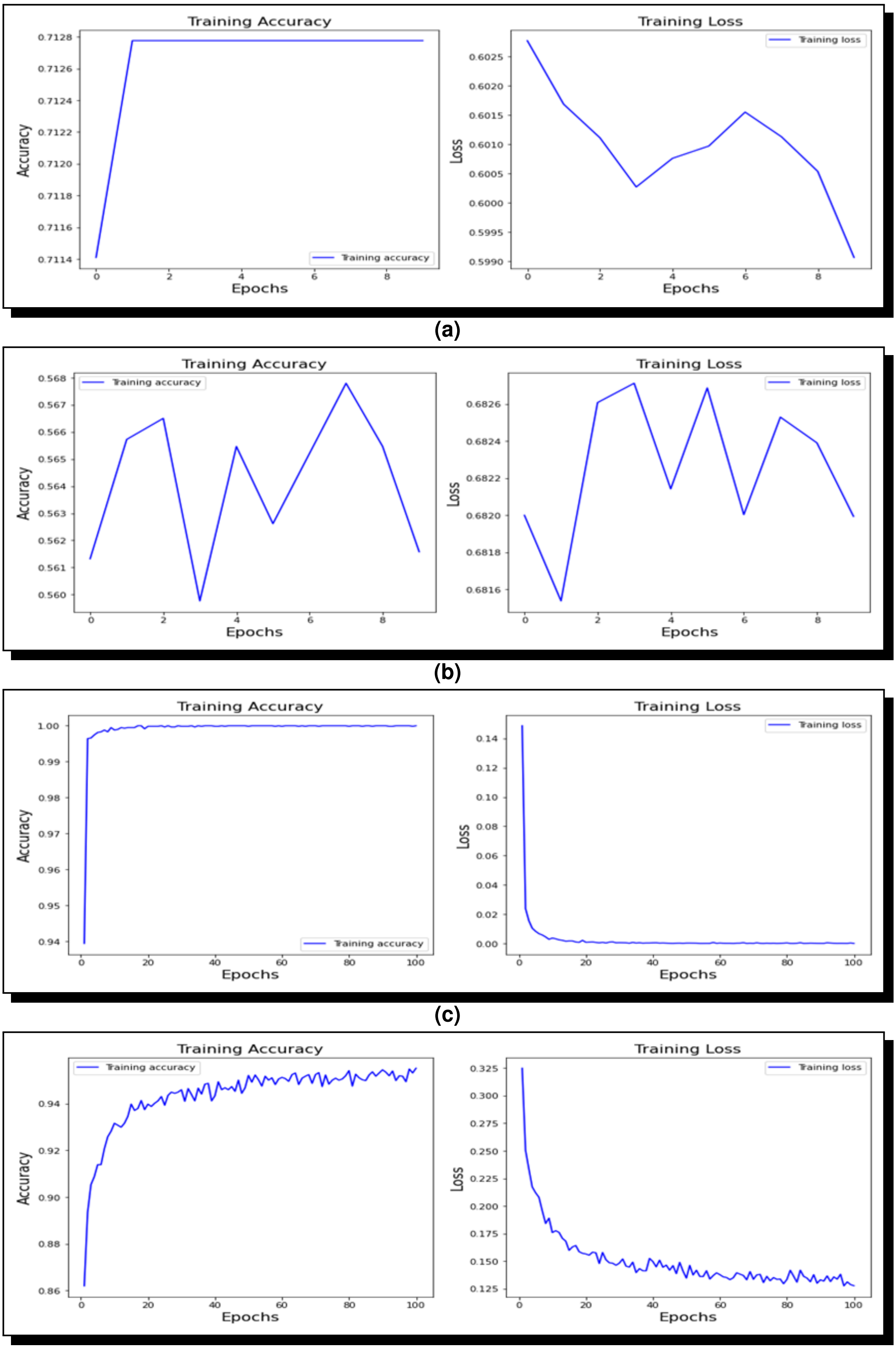
Method name	Features used	Accuracy rate
SVM	Pitch	72.23
KNN		74.76
Decision Tree		71.96
FFNN		71.50
SVM	Formant Frequency (F1, F1)	59.38
KNN		54.00
Decision Tree		62.79
FFNN		94.20
SVM	MFCC	99.55
KNN		99.91
Decision Tree		98.04
FFNN		99.80
SVM	Chroma	95.65
KNN		97.60
Decision Tree		92.99
FFNN		96.10

Table 4. A comparison of results with existing similar studies is in a tabular form in the results section

S. No.	Classifiers used	Features used	Accuracy
1	Back-propagation Neural Network [18]	MFCC	98.9%
2	HMM [5]	MFCC	100% using Hindi vowels
3	DNN with ADAM optimizer [8]	MFCC (24-25 number)	99% using vowels and 91% using words

Table 5. Error analysis of different methods

Feature	Method	Precision	Recall	F1-Score
Pitch	kNN	0.71	1.00	0.83
	SVM	0.72	0.72	0.72
	Decision tree	0.71	0.72	0.72
	FFNN	0.70	1.00	0.82
Formant Frequency	kNN	0.54	0.54	0.54
	SVM	0.59	0.59	0.59
	Decision tree	0.62	0.62	0.63
	FFNN	0.49	1.00	0.66
MFCC	kNN	1.00	1.00	1.00
	SVM	0.99	1.00	1.00
	Decision tree	0.98	0.98	0.98
	FFNN	1.00	1.00	1.00
Chroma	SVM	0.96	0.95	0.95
	kNN	0.97	0.96	0.96
	Decision tree	0.93	0.92	0.92
	FFNN	0.96	0.98	0.97



(a)

(b)

(c)

(d)

Figure 4. Training and validation performance with features: (a) Pitch, (b) Formant Frequency, (c) MFCC, (d) Chroma

6. Conclusion

This paper discussed the performance of different speech features in gender identification using machine learning methods SVM, KNN, Decision Tree, and a feed-forward neural network. We have used one low-resource language Tai and four experiments were performed using four types of speech features pitch, formant frequency, MFCC, and chroma which are mostly used in gender identification. It was found that MFCC features gave the best result for all the methods used. The overall best accuracy found was 99.8%. Other features also give a satisfactory result. The accuracy rate can be further improved by combining different features or using other Neural Network architectures. Also, an optimized set of features from the 40 MFCC features can be selected and analyzed to see how they perform. This can reduce the time as well as computational complexity. One drawback of this experiment is that due to the small number of data, there is a chance of model over-fitting. Therefore, more experiments can be performed using a bigger dataset. Also, there is a scope to use other neural network architecture. For future work, we intend to use an optimized set of speech features with different neural network architectures to analyze their significance in gender identification for the Tai language.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] A. A. Abdulsatar, V. V. Davydov, V. V. Yushkova, A. P. Glinushkin and V. Y. Rud, Age and gender recognition from speech signals, *Journal of Physics: Conference Series* **1410** (2019), 012073, DOI: 10.1088/1742-6596/1410/1/012073.

[2] R. S. Alkhawaldeh, DGR: Gender recognition of human speech using one-dimensional conventional neural network, *Scientific Programming* **2019**(1) (2019), 7213717, DOI: 10.1155/2019/7213717.

[3] A. A. Alnuaim, M. Zakariah, C. Shashidhar, W. A. Hatamleh, H. Tarazi, P. K. Shukla and R. Ratna, Speaker gender recognition based on deep neural networks and ResNet50, *Wireless Communications and Mobile Computing* **2022** (2022), 4444388, 13 pages, DOI: 10.1155/2022/4444388.

[4] I. Dagher and F. Azar, Improving the SVM gender classification accuracy using clustering and incremental learning, *Expert Systems* **36**(3) (2019), e12372, DOI: 10.1111/exsy.12372.

[5] D. S. Deiv, Gaurav and M. Bhattacharya, Automatic gender identification for hindi speech recognition, *International Journal of Computer Applications* **31**(5) (2011), 1 – 8.

[6] S. Ghosh and S. K. Bandyopadhyay, SVM classifier for human gender classification, *International Journal of Applied Research on Information Technology and Computing* **7**(2) (2016), 100 – 105, DOI: 10.5958/0975-8089.2016.00010.5.

[7] N. P. Gohain and A. Gohain, *Tai Bhashar Kathopakathan (Prathamik Path)*, *Publication Centre for Studies in Language*, Dibrugarh University, Assam, India (2023).

[8] M. R. Hasan, M. M. Hasan and M. Z. Hossain, How many Mel-frequency cepstral coefficients to be utilized in speech recognition? A study with the Bengali language, *The Journal of Engineering* **2021**(12) (2021), 817 – 827, DOI: 10.1049/tje2.12082.

[9] S. Khanum and M. Sora, Speech based gender identification using feed forward neural networks, in: *Proceedings of the National Conference on Recent Trends in Information Technology*, Foundation of Computer Science USA, Vol. NCIT2015 (2016), pp. 5 – 8.

[10] O. Mamyrbayev, A. Toleu, G. Tolegen, N. Mekebayev and D. Pham, Neural architectures for gender detection and speaker identification, *Cogent Engineering* **7**(1) (2020), Article: 1727168, DOI: 10.1080/23311916.2020.1727168.

[11] B. Medhi, Analysis of formant frequency F1, F2, and F3 in Assamese vowel phonemes using LPC Model, *International Journal of Engineering Research & Technology* **6**(5) (2017), 616 – 618.

[12] A. Raahul, R. Sapthagiri, K. Pankaj and V. Vijayarajan, Voice based gender classification using machine learning, *IOP Conference Series: Materials Science and Engineering* **263**(4), 042083, DOI: 10.1088/1757-899X/263/4/042083.

[13] L. R. Rabiner and R. W. Schafer, Introduction to digital speech processing, *Foundations and Trends® in Signal Processing* **1**(1-2) (2007), 1 – 194, DOI: 10.1561/2000000001.

[14] E. Ramdinmawii and V. K. Mittal, Gender identification from speech signal by examining the speech production characteristics, in: *Proceedings of the 2016 International Conference on Signal Processing and Communication* (ICSC), Noida, India (2016), pp. 244 – 249, DOI: 10.1109/ICSPCom.2016.7980584.

[15] P. B. Ramteke, A. A. Dixit, S. Supanekar, N. V. Dharwadkar and S. G. Koolagudi, Gender identification from children's speech, in: *Proceedings of the 2018 Eleventh International Conference on Contemporary Computing* (IC3), Noida, India (2018), pp. 1 – 6, DOI: 10.1109/IC3.2018.8530666.

[16] H. A. Sánchez-Hevia, R. Gil-Pita, M. Utrilla-Manso and M. Rosa-Zurera, Age group classification and gender recognition from speech with temporal convolutional neural networks, *Multimedia Tools and Applications* **81** (2022), 3535 – 3552, DOI: 10.1007/s11042-021-11614-4, 2022.

[17] 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), Durban, South Africa (2021), pp. 1 – 6, DOI: 10.1109/icABCD51485.2021.9519308.

[18] S. Sharma, A. Shukla and P. Mishra, Speaker and gender identification on Indian languages using multilingual speech, *International Journal of Innovative Science, Engineering & Technology* **1**(4) (2014), 522 – 525.

[19] 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* **6**(1) (2022), 27 – 42, DOI: 10.1080/24751839.2021.1983318.

[20] B. Zhong, Y. Liang, J. Wu, B. Quan, C. Li, W. Wang, J. Zhang and Z. Li, Gender recognition of speech based on decision tree model, in: *Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology* (ICCIA 2019), Advances in Computer Science Research series, Vol. 90, 2019, Atlantis Press, DOI: 10.2991/iccia-19.2019.91.