Speech Identification for Remote Assessments: Age and Gender Recognition in Distance Learning

Liang Wenlong

Faculty of Electrical Engineering and Informatics (VIK), Budapest University of Technology and Economics, Budapest, Hungary, wenlong.ethan@gmail.com

Wu Yue

Bánki Donát Faculty of Mechanical and Safety Engineering; Doctoral School on Safety and Security Science, Obuda University, Budapest, Hungary, wuyue.budapest@gmail.com

Abstract: Ensuring the authenticity of student identity in online education remains a topic of research to date. With the rapid development of Internet-based distance learning system, the demand for learner identity verification methods is increasing day by day. This study explores how voice-based age and gender recognition technologies can be applied to improve the authority of online learning and examination platforms. We use deep speaker embedding technology to extract and analyze speaker identity attributes using advanced speaker verification models such as x-vector, ECAPA-TDNN and ResNet. We employ deep speaker embedding techniques to extract and analyse speaker identity attributes using state-of-the-art speaker verification models (e.g., x-vector, ECAPA-TDNN and ResNet). Using the collected speaker dataset, we evaluate the effectiveness of these models for age and gender classification, demonstrating their potential to reduce the risk of impersonation and improve the security of exam proctoring. Our findings highlight that integrating automatic speech recognition can enhance identity verification in digital educational environments while maintaining the student learning experience. This study contributes to improving biometric security in distance learning by evaluating the feasibility of voice-based authentication in distance learning.

Keywords: deep learning, distance learning, speaker identification, biometric security.。

1 Introduction

With the rapid development of digital education platforms and the continuous evolution of the global distance learning environment, distance teaching and online examination systems have gradually become popular. However, there are still severe challenges in ensuring the authenticity of students' identities during the remote assessment process. Traditional identity verification methods (such as account and password login) are difficult to effectively prevent cheating behaviors such as proxy examination and disguise. Therefore, an increasing number of studies have begun to focus on biometric identity-based solutions to enhance the security and credibility of online evaluation systems [1].

Among numerous biometric technologies, voice is regarded as an identity recognition method with broad application prospects due to its advantages of convenient collection and low hardware threshold. Speech signals naturally carry the individual characteristics of the speaker, such as gender and age, making them of great value in the tasks of speaker verification and classification. In recent years, the development of deep learning technology, especially the introduction of embedded modeling methods such as x-vector and ECAPA-TDNN, has significantly improved the accuracy and robustness of speech identity recognition systems.

This study aims to explore the utilization of gender and age recognition results in speech to assist the student identity verification mechanism in remote examination scenarios. Specifically, this paper attempts to evaluate the mainstream speaker embedding extraction model on the Hungarian speech dataset, and uses classifiers such as MLP, Logistic Regression and Random Forest to automatically identify the gender and age of speakers. So as to analyze its feasibility and application value in identity-assisted verification.

Although methods such as face recognition [2] have been used for remote identity verification, the identity attributes contained in the voice (such as gender and age) have not been fully utilized in the online education scenario. Most of the existing studies focus on speaker verification [3], while attempts to integrate speech recognition into examination systems are still relatively limited.

Therefore, this paper focuses on the automatic classification ability of age and gender in speech recognition, and attempts to introduce an auxiliary verification system in the remote examination and assessment mechanism, thereby enhancing the security and discrimination of identity verification.

2 Literature Review

Speaker Verification is a technology that uses the characteristics of an individual's voice for identity confirmation and is widely applied in fields such as security and distance learning. In recent years, deep learning technology has greatly promoted the development of this field. Traditional manual features such as MFCC (Mel-Frequency Cepstral Coefficients) have gradually been replaced by more representational deep embedding models, such as x-vector and ECAPA-TDNN [4]. The x-vector model proposed by Snyder et al. converts variable-length speech sequences into fixed-dimensional vectors through time-averaged pooling, becoming an important baseline in the speaker verification task [5].

In the speech recognition system, age and gender recognition, as auxiliary identity information, can significantly improve the accuracy and robustness of the system. Existing studies have shown that there are statistical differences in acoustic characteristics among different gender and age groups, which provides distinguishable features for classifiers [6]. Furthermore, embedding gender and age as prior features into the recognition process also helps to reduce the search space, improve the efficiency of speaker clustering and matching, and is particularly suitable for the fields of remote evaluation and forensic medicine [1].

The current mainstream speech embedding models include x-vector, ECAPA-TDNN and ResNet based on residual structure. Most of these models are trained with the goal of speaker verification, but they are also widely used to be transferred to age estimation and gender recognition tasks, forming the so-called Multi-task Learning framework. Some of the works adopted the joint loss function and trained the three subtasks of gender, age and identity simultaneously, achieving better performance [7]. However, most of the existing studies rely on controlled corpora or English speech, and seldom explore the generalization ability on multilingual and real-world data.

Although existing studies have achieved certain results in the task of speech gender and age classification, in real application scenarios such as distance learning, they still face challenges such as noise interference, changes in speech rate, and short speech segments. Therefore, this paper takes speaker recognition in remote evaluation as the background, uses the pre-trained x-vector model to extract embeddings, and builds gender and age classification models on Hungarian speech data to further compare the performance differences of methods such as MLP, Logistic Regression and Random Forest. To explore the relationship between embedding stability and recognition accuracy and provide a reference for actual deployment.

In recent years, identity verification systems in distance learning scenarios have mostly relied on face recognition. However, face recognition still has deficiencies in aspects such as privacy protection, lighting adaptability and device requirements [2]. Therefore, this study attempts to introduce a voice-based identity recognition method, using the gender and age characteristics of the speaker as supplementary information to enhance the robustness, interpretability and user privacy security of the identity verification system.

3 Methodology

This research aims to explore and evaluate age- and gender-based speech recognition methods for authentication in the contexts of distance learning and online examinations. In order to achieve these objectives, five specific research processes are outlined in this paper: the preparation of a dataset, the extraction of speech embedding, Speaker Identification System in Distance Learning , the training and evaluation of a model.

3.1 DataSets

At the current stage of this study, the speech data set provide by the Dr. Abed. The data comes from real distance learning or experimental scenes, and has certain speaker diversity and real speech environment characteristics, which is suitable for the verification set of the preliminary experiment. The dataset is manually segmented, processed, and labelled by speaker. Further public data sets such as VoxCeleb are planned for subsequent experimental phases [8]. Public data sets have wider adaptability, which helps to improve the generalization ability of the model.A comparison of these public datasets is shown in Table 1.

Name	Env.	Free	#of Speaker	#of Utter.
AESDD	Acted Emotion	✓	>5	~500
ANAD	Acted Emotion	✓	Multiple	1384
ANDOSL	Clean speech	-	204	33,900
Forensic Comparison	Telephony	✓	552	1264
SITW	Multi-media	✓	299	2800
Common Voice	Multi-media	✓	40	>2500hs
NIST SRE	Clean speech	-	>2000	*
TIMIT	Clean speech	-	630	6300
VoxCeleb1	Multi-media	✓	1251	>100,000
VoxCeleb2	Multi-media	✓	6112	>1100,000

Table 1

Comparison of existing speaker identification datasets. (Env.: Recording Environment; Of Utter.: Approximate number of utterances. *Varies by year.)

Source: based on literature [8]

3.2 Speech pre-processing for Speaker Identification

In the process of speech identification, the quality of speech data has a very important impact on the performance of the model. Therefore, it is necessary to preprocess the collected voice data. In this study, the original audio is systematically preprocessed to ensure the stability and validity of the model input. Preprocessing includes the following steps:

Convert all speech samples to a single-channel.wav file with a sampling rate of 16kHz, and complete the format conversion using the Librosa library.

Using the TextGrid annotation file matching with the voice file (including the start and end time and content label of each voice), accurate extraction of effective voice fragments [6].

By parsing the TextGrid file, the time periods marked as empty strings ("") or silent marks ("sil") are eliminated to preserve valid speech segments.

3.3 Embedding and Extracting

Speaker Embedding is extracted from the pre-trained depth model.It mainly uses X-Vector and ECAPA-TDNN models based on the SpeechBrain framework to extract the speaker 512-dimensional vector for each sample.These vectors contain key information such as the speaker's identity, physical characteric [9]. By analogy to the classification method in face recognition, if the original voice file is compared to a face photo. Then the speaker embedding is equivalent to face embedding. Comparing with each other after extraction of the embedding is like comparing whether two faces are similar with FaceNet. Classifying with the embedding is similar to identifying age and gender with FaceNet [10].

The extracting of the Speaker embedding is to encode the processed audio through the pre-trained depth model and obtain a high-dimensional vector that can represent the identity of the speaker. This process is the basis for speech recognition, gender and age classification.

Compared with traditional handcrafted features (such as Mel-frequency cepstral coefficients, MFCC), **x-vector**-based speaker embeddings demonstrate stronger robustness and discriminability under complex acoustic conditions. In this study, a pretrained x-vector model is employed to extract speaker embeddings, which serve as the input features for subsequent **gender and age classification** tasks.

As illustrated in **Figure 1**, the embedding extraction pipeline consists of four main stages: loading the pretrained model, audio loading and preprocessing, embedding extraction, and feature saving. During the preprocessing phase, raw audio files are first standardized to a uniform format (e.g., 16kHz mono) and undergo basic processing such as **denoising and amplitude normalization**. Then, the 512 feature embeddings corresponding to each audio segment are extracted through the x-vector model.

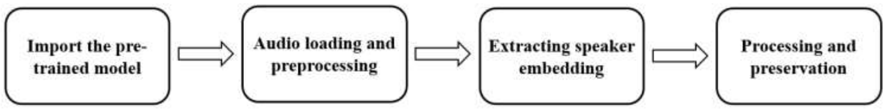

Figure 1
Speaker Embedding Extraction Workflow
Source: author' own construction

All the extracted embedding vectors and their corresponding labels (i.e., gender and age group) are uniformly saved the.csv file, and these vectors are directly used as input features for the subsequent training of the classification model. These embeddings effectively capture the speaker's acoustic and behavioral traits, and are used as the core input features for downstream classification models such as MLP, Logistic Regression, and Random Forest. The entire embedding extraction process provides fundamental support for constructing gender and age classification models in our system.

3.4 Age and Gender Classification Method

The extracted speech embeddings are input into two independent classification models, which are respectively used for gender recognition (male/female) and age group prediction (young/middle-aged/old). Each classifier consists of two layers of fully connected neural networks (MLP) and is trained using the cross-entropy loss function. In the task of gender recognition, we further introduce multiple traditional machine learning models (such as MLP, Logistic Regression, and Random Forest) for comparative experiments, all using the extracted x-vector embeddings as input features. On real distance learning datasets, the MLP model demonstrated the best performance, verifying the feasibility of integrating gender and age classification as auxiliary signals in the identity recognition task.

3.5 System frame workflow

This system aims at the identity verification problem of online examinations in the remote learning environment and designs a speaker verification system based on the speech embedding and auxiliary classification module. The system mainly consists of the following five modules. The overall structure is illustrated in Figure 2.

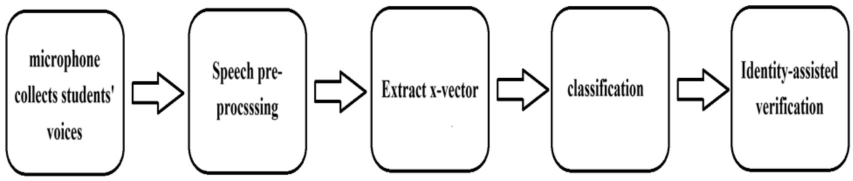

Figure 2
System structure
Source: authors’ own construction

1. Voice Collection

The system is deployed at the back end of the online examination platform. Users upload voice clips (2-3 seconds of daily voice) through a computer or device microphone as the basic input for identity verification..

2. Speech Pre-processing

The sampling rate conversion (uniformly to 16kHz single channel), denoising and format standardization operations of the original audio were carried out through the Librosa tool. Further, effective speech segment extraction was conducted in combination with the TextGrid file to remove silent and non-target speech segments, ensuring the accuracy and robustness of subsequent feature extraction.

3. Speaker Embedding

Using the pre-trained x-vector model provided by SpeechBrain, each processed audio segment is transformed into a 512-dimensional speaker embedding vector, capturing its acoustic features and using them as a unified classification input.

4. Classification

The embedded vectors are input into various classification models such as MLP, Logistic Regression, and Random Forest for gender (male/female) and age group (young/middle-aged/old) identification respectively. The model demonstrated a relatively high accuracy rate (up to 99.6% at most) in the experiment and had good generalization ability.

5. Identification

This system is deployed at the back end of the online examination platform. During the examination process, it receives the candidates' voices in real time and completes the recognition of their identity features. When there are discrepancies between the classification results of gender or age and the registration information, the invigilators can be reminded to conduct further verification, thereby providing a secure and convenient auxiliary identity verification mechanism for the distance education platform.

3.6 Experiment

This study adopted the same experimental setup in the two subtasks of gender recognition and age recognition. All experiments were based on the pre-extracted x-vector embedding vectors as input features, and comparative experiments were conducted respectively using three classification models: multi-layer perceptron (MLP), Logistic Regression, and Random Forest.

The training set and the test set are divided in a ratio of 8:2. The model training and evaluation were accomplished using the Scikit-learn framework, and the input features were uniformly standardized (StandardScaler). During the training process, end-to-end neural network training is not involved, and only shallow classifiers are constructed based on the embedding layer.

The model evaluation metrics include Accuracy, Precision, Recall, F1-score and Confusion Matrix.

4 Results

4.1 Analysis of Silent Elimination Effect

Figure 3 shows a comparison of the original language and the waveform after silent clipping. The green highlighted waveform area represents the valid speech segment retained after the silence is removed. This preprocessing step indicates that the silent detection algorithm can effectively eliminate non-speech intervals and provide clearer speech input for subsequent model training.

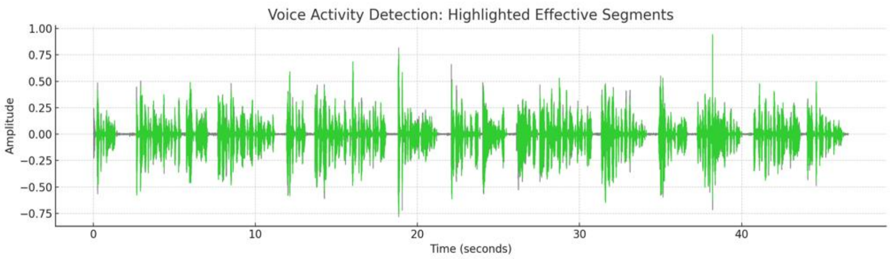

Figure 3
Data visualization of silent segment removal in speech preprocessing
Source: authors’ own construction

4.2 Statistics of audio length distribution

As shown in Figure 4, we have plotted the histogram of the length distribution of the segmented speech segments from three speech datasets (PD, Dep, Pato). The results show that the speech segments in the two datasets of Dep and Pato are mainly concentrated between 2.0 and 2.5 seconds, which meets the preset input duration requirements of this study. While the PD dataset has a wider duration distribution and provides more diverse training samples.

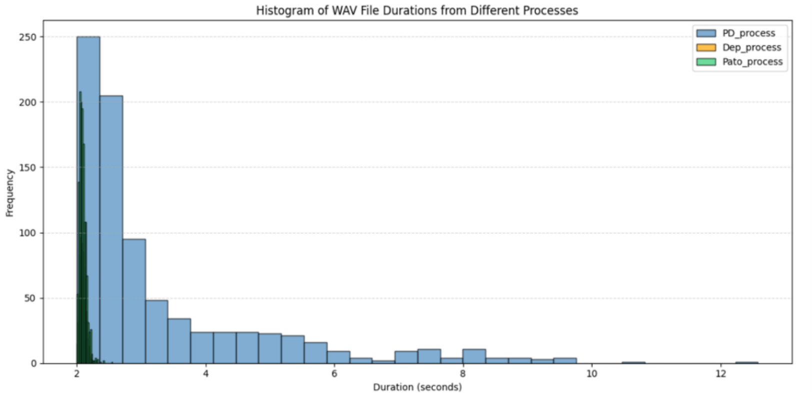

Figure 4

Histogram of Processed Speech Segment Durations

Source: authors' own construction

4.3 Embedded extraction process display

Figure 5 shows the process of extracting speaker embeddings from the pre-trained model (ECAPA-TDNN or x-vector), including audio loading, preprocessing, embedding generation and saving. The extracted 512-dimensional feature vector is used for the subsequent gender and age classification tasks and is the core basis of the system construction.

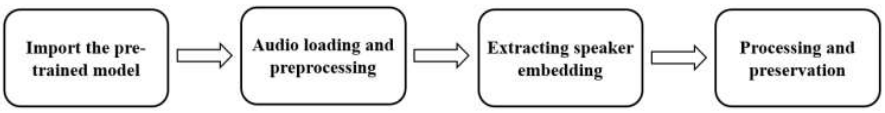

Figure 5

Speaker Embedding Extraction Workflow

Source: authors' own construction

4.4 Gender Classification Performance Analysis

In the gender classification experiment, multiple classification models were constructed based on the extracted x-vector speaker embedding. Among them, the MLP model shows extremely high classification performance. As shown in Figure 6, among the 230 female samples in the test set, 229 were correctly identified, and only 1 was misjudged as male. Among the 216 male samples, 213 were correctly classified, and only 3 were wrongly identified as female. The overall accuracy rate of the model reached 99.0%, and both Precision and Recall reached 0.99, indicating that the model has a high discriminative ability on samples of different genders. It is notable that the recall of the MLP model for females (1.00) is slightly higher than that for males (0.99), which may be related to the distribution differences of gender characteristics in the speech data. Overall, this model shows good generalization ability in distinguishing the gender of speakers in the assisted distance learning scenario, providing reliable support for the subsequent multimodal identity recognition system.

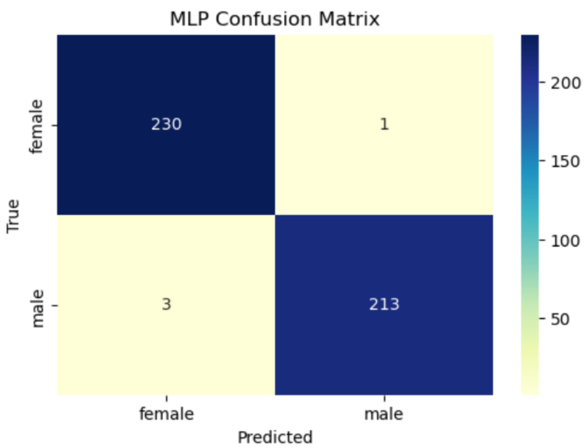

Figure 6
Result of MLP Confusion Matrix
Source: authors’ own construction

To evaluate the effectiveness of deep speaker embeddings in supporting speaker identity verification, we conducted a series of gender classification experiments using x-vector features extracted from the speech dataset. Three different classifiers were implemented and compared: Multi-Layer Perceptron (MLP), Logistic Regression, and Random Forest.

The experimental results are summarized in Table X. The MLP model achieved the best overall performance, with an accuracy, precision, recall, and F1-score all reaching 99.6%. Logistic Regression followed closely, attaining a performance of 99.5% across all metrics. Although Random Forest showed slightly lower performance compared to the other two models, it still yielded competitive results with a 98.6% accuracy.

These results suggest that x-vector embeddings are highly effective in capturing speaker-specific information for binary gender classification tasks. The consistently high accuracy across models demonstrates the reliability of these embeddings in distinguishing between male and female speakers, which is crucial for downstream tasks such as speaker verification in remote learning scenarios. As shown in Table 2 and Figure 6, all three classifiers exhibit strong performance in distinguishing speaker gender. The MLP model achieves the highest F1-score of 0.996. These findings confirm that gender recognition can be reliably integrated into remote assessment systems as an auxiliary signal for speaker identity verification, thus improving the robustness and security of distance learning environments.

Moel	Accuracy	Precision	Recall	F1-score
MLP	99.6%	99.6%	99.6%	99.6%
Logistic Regression	99.5%	99.5%	99.5%	99.5%
Random Forest	98.6%	98.6%	98.6%	98.6%

Table 2
Performance Comparison of Gender Classification Models
Source: authors' own construction

4.5 Age Classification Performance Analysis

In the age identification task, we divided the samples into five groups by age: teenagers (≤18 years old), young adults (19-29 years old), middle-aged adults (30-44 years old), middle-aged and elderly adults (45-59 years old), and elderly adults (≥60 years old). Based on the extracted x-vector embedding features, we conducted multi-classification recognition experiments for age groups using three mainstream classification models, namely Logistic Regression, MLP and Random Forest respectively. The experimental results show that the MLP model has the best comprehensive performance among the five age groups, with an accuracy rate of 76%. It has high Precision (0.81, 0.53) and Recall (0.94, 0.88) in the young and middle-aged groups, demonstrating a strong ability to distinguish age segments. In contrast, the overall performance of Logistic Regression was slightly weaker, with an accuracy rate of 71%. Although it had certain advantages in the adolescent group (Precision 0.74) and the middle-aged group (Recall 0.87), there were certain deviations in the identification of the middle-aged and elderly and the elderly segments. The overall performance of the Random Forest model was the worst, with an accuracy rate of only 57%. Especially on the samples of middle-aged and elderly people and the elderly, there was a lot of confusion, indicating that its ability to distinguish age groups with blurred boundaries is weak. It can be seen from the model comparison that deep neural networks (such as MLP) have a stronger nonlinear modeling ability when dealing with high-dimensional embedding features, and can better depict the potential relationship between speech and age(Table 3). Therefore, they are more suitable for age recognition tasks. This experiment verified the feasibility and practicability of the speech-based age estimation method in assisting identity verification in the distance learning scenario. The detailed confusion matrix is presented in Figure 7.

Model	Accuracy	Macro F1	Weighted F1
LogisticRegression	0.71	0.71	0.71
MLP	0.76	0.80	0.77
Random Forest	0.57	0.64	0.58

Table 3

Performance Comparison of Age Classification Models

Source: authors' own construction

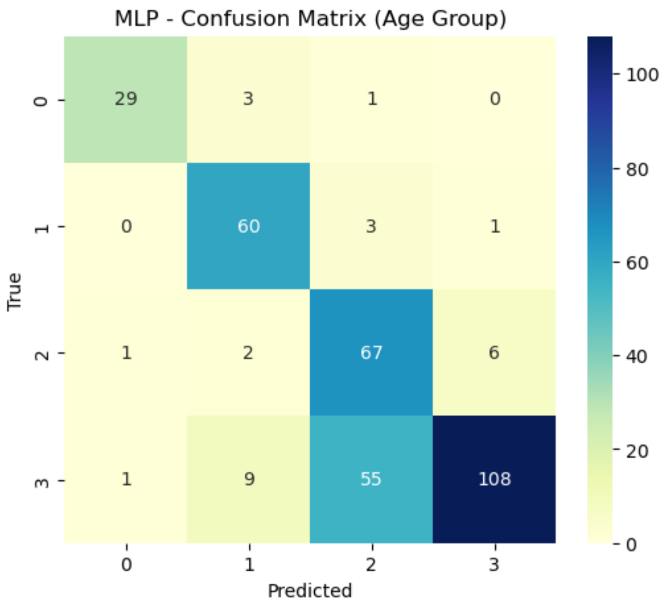

Figure 7

Result of MLP Confusion matrix

Source: authors' own construction

5 Discussion

5.1 Concluding Remarks

In recent years, with the application of deep learning in the field of speaker recognition, methods based on speaker embedding have gradually become mainstream. The x-vector method proposed by Snyder et al. (2018) [11] has demonstrated excellent performance on various tasks, proving its effectiveness in modeling speaker features.

In this study, the pre-trained x-vector model was adopted as the embedding feature extraction method, achieving good results in both gender recognition and age classification tasks, which is consistent with the related research trends mentioned in the Literature Review.

In the gender recognition task, the MLP model in this study achieves an accuracy of 99.6%, which is higher than the accuracy of about 95% in the literature [12] based on the traditional i-vector+PLDA method, showing that the deep learning model has a stronger ability to fit the speaker's gender features.

In the age recognition task, this study divided age into five intervals for classification, and MLP achieved an accuracy rate of approximately 76%, which is similar to the 74.2% accuracy rate result obtained based on CNN feature extraction and multi-category SVM classifier in reference [13].

It can be seen from this that the system in this study can also maintain high performance in a small-sample distance learning environment, further verifying the feasibility and practicability of the authentication scheme based on speaker embedding features.

5.2 Limitation and Suggestions

This study specifically compared three models, namely MLP, Logistic Regression and Random Forest:

MLP achieved the optimal performance in both tasks (gender and age), verifying the advantages of deep neural networks when dealing with high-dimensional and complex embedding vectors [3].

Although Logistic Regression and Random Forest have also reached a relatively high level in gender recognition, their performance has significantly declined in age classification, indicating that there are limitations in the ability of linear models and tree models to capture continuity and minor differences.

5.3 Implications for the Application of Identity Verification in Distance learning

In the Literature Review, a study [2] proposed the use of facial recognition and keystroke behavior as remote examination authentication methods, while the utilization of voice biometric features is still in its infancy.

This study verified the feasibility of assisted identity verification through voice embedding features from the perspectives of gender and age, providing a new technical option for future online education platforms.

Especially during the remote examination process, the system can infer the gender and age of the examinees in real time and verify them with the registration information. When a discrepancy in gender or age is detected, an alarm can be triggered, further reducing the risk of cheating.

By integrating gender classification into speaker verification frameworks, we propose a lightweight, privacy-friendly auxiliary check for distance learning authentication. For example, if a student's registered gender is "female" but the voice classification consistently predicts "male," the system could trigger an alert for potential impersonation.

Conclusion

This study proposes an identity verification framework based on speech recognition, aiming to enhance the authenticity and security of identity verification in the distance learning environment. This framework relies on the individual biological attribute characteristics contained in the speech, combined with the current mainstream deep embedding technology and classification algorithms, providing a lightweight and scalable auxiliary authentication method for remote examination and online education scenarios.

In terms of system construction, this paper has completed the complete process of speech data collection, preprocessing and embedding extraction, and adopted the pre-trained model based on x-vector for speaker feature extraction. Subsequently, by inputting the embedding vectors into various classifiers (such as MLP, Logistic Regression, and Random Forest), the experimental verification of the two core tasks of gender and age recognition was achieved.

The experimental results show that the MLP model has an accuracy rate as high as 99.6% in the gender recognition task and achieves an overall accuracy rate of 76% in the age recognition task. Especially in the young and middle-aged groups, the recognition effect is remarkable. Compared with the existing research results, this paper achieves a relatively high performance in speaker attribute recognition, verifying the feasibility and effectiveness of introducing gender and age as auxiliary features into the remote identity verification process.

The future work can consider the following directions: (1) Introduce more individual attribute labels (such as emotional state, language ability, etc.) to enhance the discrimination dimension of the model. (2) Introduce more languages to enhance the model's generalization ability. (3) Attempt more complex model structures to optimize the performance of age group recognition. (2) Attempt a multimodal joint modeling approach that integrates facial, voice and behavioral features

References

[1] J. Renzella, A. Cain, and J.-G. Schneider, "Verifying student identity in oral assessments with deep speaker," Comput. Educ. Artif. Intell., 3, p. 100044, Jan. 2022, doi: 10.1016/j.caeai.2021.100044.

[2] J. L. Bailie and M. A. Jortberg, "Online Learner Authentication: Verifying the Identity of Online Users," vol. 5, no. 2, 2009.

[3] Z. Bai and X.-L. Zhang, "Speaker recognition based on deep learning: An overview," Neural Netw., 140, pp. 65–99, Aug. 2021, doi: 10.1016/j.neunet.2021.03.004.

[4] B. Desplanques, J. Thienpondt, and K. Demuynck, "ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification," in Interspeech 2020, Oct. 2020, pp. 3830–3834. doi: 10.21437/Interspeech.2020-2650.

[5] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-Vectors: Robust DNN Embeddings for Speaker Recognition," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2018, pp. 5329–5333. doi: 10.1109/ICASSP.2018.8461375.

[6] L. Fischbach, C. Kleen, L. Flek, and A. Lameli, "Does Preprocessing Matter? An Analysis of Acoustic Feature Importance in Deep Learning for Dialect Classification," in Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025), R. Johansson and S. Stymne, Eds., Tallinn, Estonia: University of Tartu Library, Mar. 2025, pp. 159–169. Accessed: Apr. 24, 2025. [Online]. Available: https://aclanthology.org/2025.nodalida-1.16/

[7] D. Kwasny and D. Hemmerling, "Joint gender and age estimation based on speech signals using x-vectors and transfer learning," Dec. 02, 2020, arXiv: arXiv:2012.01551. doi: 10.48550/arXiv.2012.01551.

[8] A. Nagrani, J. S. Chung, W. Xie, and A. Zisserman, "Voxceleb: Large-scale speaker verification in the wild," Comput. Speech Lang., 60, p. 101027, Mar. 2020, doi: 10.1016/j.csl.2019.101027.

[9] M. Ravanelli et al., "SpeechBrain: A General-Purpose Speech Toolkit," Jun. 08, 2021, arXiv: arXiv:2106.04624. doi: 10.48550/arXiv.2106.04624.

[10] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823. Accessed: Apr. 29, 2025. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html

[11] F. T. Pinki, N. Khatun, and S. M. M. Islam, "Content based paddy leaf disease recognition and remedy prediction using support vector machine," in 2017 20th International Conference of Computer and Information Technology (ICCIT), Dec. 2017, pp. 1–5. doi: 10.1109/ICCITECHN.2017.8281764.

[12] "Teaching a Project-Based Course at a Safe Distance: An Experience Report | IEEE Conference Publication | IEEE Xplore." Accessed: Apr. 29, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/9206192