

A Review of Language Identification Technology

Xinyu Hu^{1*}

¹ Luoyang Campus, Former Information Engineering University, Luoyang, China

Email Address

huxinyu1002015@163.com (Xinyu Hu)

*Correspondence: huxinyu1002015@163.com

Received: 11 July 2023; **Accepted:** 26 August 2023; **Published:** 13 September 2023

Abstract:

Language Identification (LID) is the technology that aims to identify the language category to which a given audio sample belongs. As a critical aspect of both fundamental research and practical applications in the field of speech intelligence, LID plays a vital role in improving the accuracy and reliability of speech intelligence technology, attracting increasing attention. This paper first introduces the background and basic principles of language identification research and then elaborates on two major categories of methods: traditional statistical models for language identification and deep learning-based approaches, discussing the development process of each method. Finally, the advantages and limitations of these methods are analyzed, and the prospects of language identification technology are discussed.

Keywords:

Language Identification, Acoustic Features, Phoneme Features, Deep Neural Network

1. Background and Basic Principles

Language, as one of the most important and direct means of human information interaction, holds an irreplaceable position in the development of human civilization. With over seven thousand known dialects worldwide, Mandarin Chinese has the most speakers, while English is the most widely used language. With the rapid development of globalization, international communication has become increasingly common, creating a growing demand for communication and interaction among people from different language backgrounds. Relying solely on human language learning to meet the escalating demands of multilingual communication is impractical. Therefore, multilingual intelligent speech technology is urgently needed to provide support. With the rapid development of artificial intelligence and big data analysis, speech intelligence technology has become an integral part of social development, leading to an increased demand and wide-ranging applications.

Language Identification (LID) is the technology that aims to identify the language category to which a given audio sample belongs. As a crucial aspect of both fundamental research and practical applications in the field of speech intelligence, LID plays an important role in improving the accuracy and reliability of speech intelligence technology. It is commonly used as a preprocessing technique in various

speech recognition tasks [1]. For example, in Automatic Speech Recognition (ASR) systems, automatic language detection is conducted as the initial step. Without accurate language detection, speech corpora cannot be properly processed, and grammar rules cannot be applied, leading to the failure of subsequent speech recognition steps. Additionally, language identification technology finds extensive applications in machine translation, speech search engines, information services, public safety, and smart homes, demonstrating significant practical value.

Language identification is an important task in the field of natural language processing, aiming to classify given audio segments or speech signals into their respective language categories. Language identification involves three main processes: feature extraction from speech signals, construction of language models, and language decision-making for test speech. During the training process, various language-related feature models are constructed based on the extracted features of different language datasets. In the testing phase, the features of the speech to be identified are processed using specific recognition strategies and similarity metrics based on the previously trained language-specific models. Figure 1 illustrates this process. In the training phase, the speech signals are preprocessed to facilitate feature extraction from various language datasets. Then, the extracted features are used to construct a feature vector sequence, which is used to train one or more language models containing language information. In the recognition phase, the speech to be identified is subjected to preprocessing, feature extraction, and feature vector construction. Classification models or algorithms are then employed to measure the similarity between the test features and the corresponding language models trained in the training phase, thus determining the language of the speech.

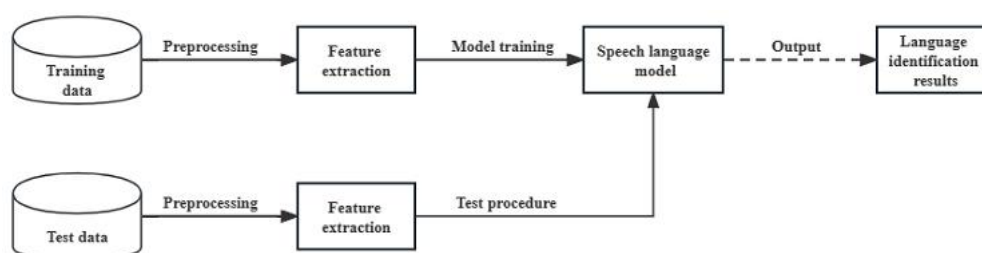


Figure 1. Basic principles of language recognition.

With the advancement of acoustic research, language identification technology has evolved through several stages. Initially, in the 1960s, Hidden Markov Models (HMM) emerged and were widely used in various statistical models, giving birth to HMM-based language identification systems. Subsequently, acoustic research led to the development of language identification methods based on phoneme recognition. With further research on speech signals and auditory and perceptual features, approaches based on low-level acoustic features such as Mel-Frequency Cepstral Coefficients (MFCC) gained popularity. In recent years, as breakthroughs have been made in the training process of Deep Neural Networks, deep learning-based language identification methods have gained prominence. The following chapters will elaborate on the development process of various language identification methods, followed by a summary and analysis of the advantages and limitations in Chapter 4.

2. Statistical Model-based Language Identification Methods

Statistical model-based methods consist of two steps: feature extraction and statistical modeling, as shown in Figure 1. The purpose of feature extraction is to obtain discriminative feature vectors for language identification, while the objective of statistical modeling is to describe the distribution characteristics of the feature vectors. As mentioned earlier, obtaining high-level linguistic features is challenging, so most language identification systems are based on short-term acoustic features or phoneme sequences. Although these features may not be highly discriminative, they are computationally simple and scalable.

2.1. Language Identification Based on Acoustic Features

In the field of language identification, methods based on low-level acoustic features have been extensively researched and applied. Low-level acoustic features refer to the fundamental attributes of sound signals, such as spectral shape and resonance peak positions. These features describe the differences in statistical distribution characteristics among acoustic units, and they can be extracted using signal processing techniques to represent the characteristics of the corresponding speech and perform language identification tasks.

In 1980, Buzo proposed a technique for language identification using resonance peaks to represent acoustic features, modeling their frequencies using vector quantization distortion [2]. In 1982, Poritz et al. used Hidden Markov Model (HMM) techniques for language identification, extracting rhythmic features from pitch and energy contours and employing classical classifiers for decision-making [3]. In 1989, Goodman et al. significantly reduced the error rate of language identification by modeling the resonance peak features based on Linear Prediction Coding (LPC) [4] [5]. In 1991, Savic et al. proposed an automatic language identification method based on multiple language-specific features, including HMM and language-specific pitch contour features [6].

In 2002, acoustic features obtained by modeling speech signals using Gaussian Mixture Models (GMM) [7] achieved an average error rate of 16.32% in language identification tasks for English, German, Spanish, and French [8]. Since then, the GMM approach has become the mainstream model for acoustic modeling and has been widely applied to language identification tasks. In the same year, Torres et al. proposed a language identification method combining Shifted Delta Cepstral (SDC) features with GMM-Universal Background Model (UBM) [9], which achieved comparable results to the phoneme recognizer-based methods. However, GMM-UBM modeling requires a large amount of training data to estimate the covariance matrix of the UBM model. In relation to the large number of parameters, training data is often insufficient [10]. Subsequent studies have continuously improved upon this method. In 2003, Qu et al. proposed an improved GMM-UBM model [11]. In 2004, Campbell et al. used Support Vector Machines (SVM) to classify GMM mean super vectors (GSV-SVM), resulting in improved language identification system performance [12]. In 2006, Kenny et al. employed Factor Analysis (FA) to reduce noise influence when processing low-level acoustic features [13]. In 2007, Joint Factor Analysis (JFA) from speaker recognition algorithms began to be gradually applied to language identification [14]. In 2011, this led to the development of methods based on Total Variability (TV) modeling and the use of Identity Vectors (i-vectors) for identification [15]. Subsequently, variation compensation methods based on i-vector space, such as

Linear Discriminative Analysis (LDA), were used to improve system performance [16].

2.2. Language Identification Based on Phoneme Features

In the field of language identification, methods based on phoneme features are a common research direction. Phonemes are the smallest phonetic units in a language, and by analyzing and comparing phonemes across different languages, language identification can be achieved.

Firstly, methods based on phoneme features for language identification involve constructing phoneme models for different languages and using the differences between these models to determine the language of the input speech. The basic idea is that the phoneme distribution and frequency vary among different languages, enabling the classification of languages by modeling and statistical analysis of phonemes. Secondly, phoneme features are extracted based on speech signals, capturing spectral features related to phonemes, thereby providing information about the language. In addition, building large-scale corpora is necessary to train phoneme models for language identification. These corpora contain phonetically annotated data for different languages, which is used to train models and learn the language-specific features. By testing the trained models, the input speech can be matched against different languages, resulting in language identification.

In 1994, Tucker et al. conducted language identification research using phoneme features based on the TIMIT database [17]. They investigated three features: the acoustic differences between phonemes in each language, the relative frequencies of phonemes, and the combination of the two, achieving the highest accuracy of 97%. In the same year, Lamel et al. proposed a language identification method based on phoneme acoustic likelihood, calculating the likelihood of current speech with phonemes of specific languages and selecting the most similar language as the output result [18].

Also in 1994, Zissman et al. used a phoneme-level speech recognizer (PR) to decode speech signals, converting them into phoneme sequences, and established corresponding phoneme-level N-gram language models for language identification. This method is known as Phone Recognizer followed by Language Modeling (PRLM) [19]. The trained phoneme recognizer can extract monophone or triphone states of speech signals. In the same year, Berkling et al. used a combination of phoneme features and linguistic knowledge for language identification [20].

Since then, the PRLM method has been continuously developed. In 2004, William proposed the Phoneme Recognizer combined with Support Vector Machine (PR SVM) method for language identification. It maps phoneme sequences to frequency vectors on all N-grams and constructs an SVM kernel function based on these frequency vectors to establish a phoneme sequence classification model instead of the original N-gram scoring model [21]. In 2005, Dong Zhu et al. used syllables and words as modeling units to utilize higher-level language information for identification based on higher-level phonetic units [22]. In 2006, Matejka proposed using Factor Analysis (FA) to process low-level acoustic features and reduce noise influence [13]. In 2007, Joint Factor Analysis (JFA) from speaker recognition algorithms began to be gradually applied to language identification [14]. In 2011, this led to the development of methods based on Total Variability (TV) modeling and the use of Identity Vectors (i-vectors) for identification [15]. Afterward, variation compensation methods based

on i-vector space, such as Linear Discriminative Analysis (LDA), were used to improve system performance [16].

3. Deep Learning-based Language Identification Methods

Deep learning, as a powerful machine learning technique, has made significant breakthroughs in the field of speech recognition in recent years, particularly in language identification tasks, showing high potential for application. By constructing multi-layer neural networks, deep learning can automatically learn to extract high-level speech features, enabling effective discrimination between different languages.

3.1. Statistical Model-based Methods using Deep Neural Networks

Deep Neural Networks (DNNs) possess powerful feature extraction capabilities. For a DNN trained to classify phonemes, the features closer to the classification layer are more representative of phonetic content, while being less susceptible to noise or channel interference. In particular, DNNs can input long-duration speech segments, better utilizing contextual information. Therefore, abstract features extracted by DNNs have stronger robustness and phonetic representation compared to low-level acoustic features. Leveraging this characteristic, Song et al. [25] proposed using the last layer of the phoneme discriminative network as feature construction for the i-vector language identification model, achieving promising results. Another method utilizing DNNs for statistical modeling involves using statistical measurements of speech features extracted by DNNs to construct DNN i-vector systems [26]. Traditional i-vector methods utilize a Universal Background Model (UBM) to generate several Gaussian components and calculate the posterior probabilities of each frame's speech feature on these Gaussian components using Bayesian formula. Based on these posterior probabilities, the zeroth- and first-order statistical measurements of a speech feature segment can be computed and used to train i-vector models. In DNN i-vector methods, DNNs replace UBMs, and phonemes (more accurately, context-dependent phonemes after clustering, i.e., Senones) replace the Gaussian components in UBM. Because the forward process of a DNN can directly calculate the posterior probabilities of each frame's feature given each phoneme, obtaining the corresponding statistical measurements of a speech segment and training i-vector models becomes straightforward. Compared to UBM i-vector methods, DNN i-vector methods utilize phonemes as language units and can improve robustness in posterior probability calculation through discriminative training. Ferrer et al.'s experiments showed that when the DNN training data is consistent with the i-vector training/recognition data, the DNN i-vector system surpasses the UBM i-vector system in language identification tasks; however, it performs poorly in scenarios with data mismatch.

3.2. End-to-End Language Identification based on Deep Neural Networks

In deep learning-based language identification methods, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are the most commonly used architectures. CNNs effectively extract local features and possess translation invariance, making them capable of capturing basic language characteristics. RNNs capture contextual dependencies in speech signals, distinguishing different languages by learning temporal information from speech sequences.

In 1989, Cole et al. discussed the possibility of combining speech recognition and artificial neural networks for language identification [27]. In 1992, Morgan utilized

Recurrent Context Expander (RCE) neural networks for automatic language identification [28]. In 1998, considering RNN's ability to capture contextual information from language models [29], Braun was the first to apply RNNs to language identification [30].

Since 2011, Deep Neural Networks have achieved significant success in the field of speech recognition, prompting researchers to introduce deep learning methods into language identification tasks. In 2014, researchers first applied DNN models to large-scale short-duration language tasks, performing feature extraction, transformation, and classification using neural networks [31]. In 2016, an end-to-end language identification framework based on the Attention Mechanism was proposed [32]. In 2018, David et al. introduced the x-vector method, mapping variable-length speech segments to fixed-dimensional embeddings using Time-Delayed Neural Networks (TDNNs), which significantly improved short-duration language identification [33]. TDNNs extract feature vectors containing rich contextual information, effectively enhancing system performance. In 2022, Zhao et al. proposed an ECAPA-TDNN+ model with Contrastive Predictive Coding (CPC) for multi-task learning in language identification [34].

4. Analysis of Strengths and Limitations of Language Identification Methods

Methods based on low-level acoustic features primarily utilize features such as spectral shape, pitch, and resonance peaks to classify and identify languages. They are an early approach that has been extensively researched and applied. Traditional acoustic feature extraction methods have demonstrated some effectiveness in language identification tasks. However, due to the limitations of hand-crafted feature design, they often fail to capture higher-level differences between languages and are sensitive to factors such as sound quality and environmental noise, resulting in poorer performance in complex language identification tasks. Their advantages lie in their simplicity of implementation and lower computational resource requirements.

Methods based on phoneme features utilize phonetic information in speech signals for classification and identification. Firstly, phoneme feature-based language identification methods segment speech signals into phoneme sequences using phoneme recognizers. Then, they extract features based on these phoneme sequences, such as phoneme duration and frequency. Finally, classifiers are used to train and classify these features to achieve language identification. Methods based on phoneme features have shown good performance in complex language identification tasks. Researchers have improved phoneme recognition accuracy and used it as the basis for language identification by introducing more contextual information and speech models. This approach captures higher-level differences between languages, resulting in improved language identification performance. However, the accuracy of phoneme recognition significantly affects the performance of language identification, making high requirements for speech transcription and phoneme annotation, posing a significant challenge.

Methods based on Deep Neural Networks for language identification utilize deep learning algorithms to model and learn features from speech signals. Typically, multi-layer neural network structures such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) are employed to extract and abstract features from the input speech signal, followed by training and classification using classifiers

to achieve language identification. Deep learning methods have shown good performance in language identification tasks due to their strong nonlinear modeling capabilities, which allow them to automatically learn feature representations from input data. Researchers have further improved language identification accuracy by introducing more complex network structures, optimizing training algorithms, and increasing training data, resulting in language identification models with better generalization abilities. However, deep neural network training requires a large amount of training data, and model training and optimization can be complex, demanding high computational resources.

5. Conclusions

This paper initially introduced the background and basic principles of language identification research, then discussed two major categories of language identification methods: traditional statistical model-based methods and deep learning-based methods, providing an overview of the development process of various methods. Traditional language identification methods based on low-level acoustic features and phoneme features have been extensively researched and applied. Deep learning-based methods have emerged as a promising approach, leveraging the strong representation learning capabilities of Deep Neural Networks. Future research can focus on improving models, expanding datasets, and optimizing feature representations to further enhance the performance of deep learning-based language identification. By comparing and analyzing the three methods, the paper concludes that methods based on low-level acoustic features are simple to implement but have limited performance. Methods based on phoneme features capture higher-level differences between languages but face challenges with phoneme recognition accuracy and data mismatch. Deep learning-based methods have advantages in accuracy but demand high requirements for data and computational resources. Future research can explore combining different methods to further enhance language identification performance.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] K. Li.; T. Edwards, Statistical models for automatic language identification, ICASSP '80. IEEE International Conference on Acoustics, Speech, and Signal Processing, Denver, CO, USA, 1980, pp. 884-887, DOI: 10.1109/ICASSP.1980.1170832.
- [2] A. Buzo, A. Gray, R. Gray and J. Markel, Speech coding based upon vector quantization, ICASSP '80. IEEE International Conference on Acoustics, Speech, and Signal Processing, Denver, CO, USA, 1980, pp. 15-18, DOI: 10.1109/ICASSP.1980.1170996.

- [3] Poritz, A. Linear Predictive Hidden Markov Models and The Speech Signal. ICASSP'82. IEEE International Conference on Acoustics, Speech, And Signal Processing. IEEE, 1982, 7: 1291-1294. DOI: 10.1109/ICASSP.1982.1171633.
- [4] Lyon, R. Acoustics, Speech, And Signal Processing, *IEEE International Conference On ICASSP'82*. 1982.
- [5] Goodman, F.J.; Martin A.F.; and Wohlford, R.E.; Improved automatic language identification in noisy speech, *International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, UK, 1989, pp. 528-531 vol.1, DOI: 10.1109/ICASSP.1989.266480.
- [6] Savic, M.; Gupta, S.; Acosta, E. An automatic language identification system, in *Acoustics, Speech, and Signal Processing*, IEEE International Conference on, Toronto, Ont., Canada, 1991 pp. 817-820. DOI: 10.1109/ICASSP.1991.150462
- [7] C. M. del Alamo, F. J. Caminero Gil, C. dela Torre Munilla and L. Hernandez Gomez, Discriminative training of GMM for speaker identification, 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Atlanta, GA, USA, 1996, pp. 89-92 vol. 1, DOI: 10.1109/ICASSP.1996.540297.
- [8] Backfried, G.; Rainoldi, R.; Riedler, J. Automatic language identification in broadcast news, Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290), Honolulu, HI, USA, 2002, pp. 1406-1410 vol.2, DOI: 10.1109/IJCNN.2002.1007722.
- [9] Kohler, M.A.; Kennedy, M. Language identification using shifted delta cepstral, The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002., Tulsa, OK, USA, 2002, pp. III-69, DOI: 10.1109/MWSCAS.2002.1186972.
- [10] Han, Y.R. Research on speech language recognition technology based on deep learning network, Northwest University, Xian, 2021. DOI: 10.27405/d.cnki.gxbdu.2021.001561.
- [11] Dan, Q.; Wang, B.X. Automatic language identification based on GMBM-UBBM, International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003, Beijing, China, 2003, pp. 722-727, DOI: 10.1109/NLPKE.2003.1276000.
- [12] Torres-Carrasquillo, P.A. Language identification using Gaussian mixture models. *Michigan State University*, East Lansing, 2002.
- [13] Kenny, P. Joint factor analysis of speaker and session variability: Theory and algorithms. *CRIM, Montreal*, 2005, 14(28-29), 2.
- [14] Kenny, P.; XBoulianne, P.; Ouellet, P.; Dumouchel, P. Joint Factor Analysis Versus Eigenchannels in Speaker Recognition, in *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, 15(4), 1435-1447, DOI: 10.1109/TASL.2006.881693.
- [15] Dehak, N.; Kenny, P.J.; Dehak R, et al. Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 2010, 19(4), 788-798. DOI: 10.1109/TASL.2010.2064307.

- [16]Jiang, Y. Speaker recognition based on short speech and channel change. Nanjing University of Science and Technology, Nan Jing, 2013.
- [17]Tucker, R.C.F.; Carey, M.J.; Parris, E.S. Automatic language identification using sub-word models, Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing, Adelaide, SA, Australia, 1994, pp. I/301-I/304 vol.1, DOI: 10.1109/ICASSP.1994.389295.
- [18]Lamel, L.F.; Gauvain, J.L. Language identification using phone-based acoustic likelihoods, Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing, Adelaide, SA, Australia, 1994, pp. I/293-I/296 vol.1, DOI: 10.1109/ICASSP.1994.389297.
- [19]Zissman, M.A.; Singer, E. Automatic language identification of telephone speech messages using phoneme recognition and N-gram modeling, Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing, Adelaide, SA, Australia, 1994, pp. I/305-I/308 vol.1, DOI: 10.1109/ICASSP.1994.389377.
- [20]Berkling, K.M.; Arai T.; Barnard, E. Analysis of phoneme-based features for language identification, Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing, Adelaide, SA, Australia, 1994, pp. I/289-I/292 vol.1, DOI: 10.1109/ICASSP.1994.389298.
- [21]Campbell, W.M.; Campbell, J.R.; Reynolds, D.A.; Jones, D.A.; Leek, T.R. High-level speaker verification with support vector machines, 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, 2004, pp. I-73, DOI: 10.1109/ICASSP.2004.1325925.
- [22]Zhu, D.; Adda-Decker, M.; Antoine, F. Different size multilingual phone inventories and context-dependent acoustic models for language identification. Ninth European Conference on Speech Communication and Technology. 2005.
- [23]Matejka, P.; Burget, L.; Schwarz, P.; Cernocky, J. Brno University of Technology System for NIST 2005 Language Recognition Evaluation, 2006 IEEE Odyssey - The Speaker and Language Recognition Workshop, San Juan, PR, USA, 2006, pp. 1-7, DOI: 10.1109/ODYSSEY.2006.248093.
- [24]Boil, H.; Qian, Z.; Ziaei, A, et al. Automatic Assessment of Language Background in Toddlers Through Phonotactic and Pitch Pattern Modeling of Short Vocalizations. Workshop on Child Computer Interaction (WOCCI 2014). 2014.
- [25]Ferrer, L.; Lei, Y.; McLaren, M.; Scheffer, N. Study of Senone-Based Deep Neural Network Approaches for Spoken Language Recognition, in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 1, pp. 105-116, Jan. 2016, DOI: 10.1109/TASLP.2015.2496226.
- [26]Song,, Y.; Jiang, B.; Bao, Y.B, et al. I-vector representation based on bottleneck features for language identification. *Electronics Letters*, 2013, 49(24), 1569-1570. <https://doi.org/10.1049/el.2013.1721>.
- [27]Leung, H.C.; Zue, V.W. Some phonetic recognition experiments using artificial neural nets, ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing, New York, NY, USA, 1988, pp. 422-425 vol.1, DOI: 10.1109/ICASSP.1988.196608.

- [28]Morgan, D.P.; Riek, L.; Mistretta, W.J.; Scofield, C.L.; Grouin, P.; Hull, F. Experiments in language identification with neural networks, [Proceedings 1992] IJCNN International Joint Conference on Neural Networks, Baltimore, MD, USA, 1992, pp. 320-325 vol.2, DOI: 10.1109/IJCNN.1992.226968.
- [29]Wigström, H. A model of a neural network with recurrent inhibition. *Kybernetik*, 1974, 16, 103-112, DOI: <https://doi.org/10.1007/BF00271633>.
- [30]Braun, J.; Levkowitz, H. Automatic language identification with recurrent neural networks, 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227), Anchorage, AK, USA, 1998, pp. 2184-2189 vol.3, DOI: 10.1109/IJCNN.1998.687199.
- [31]Lopez-Moreno, I.; Gonzalez-Dominguez, J.; Plhot, O.; Martinez, D.; Gonzalez-Rodriguez, J.; Moreno, P. Automatic language identification using Deep Neural Networks, 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 5337-5341, DOI: 10.1109/ICASSP.2014.6854622.
- [32]Mounika, K.V.; Achanta, S.; Lakshmi, R.; Vuppala, A.; Gangashetty, S. An Investigation of Deep Neural Network Architectures for Language Recognition in Indian Languages. Interspeech, San Francisco, California, USA, 2016.
- [33]Snyder, D.; Garcia-Romero, D.; McCree A, et al. Spoken language recognition using x-vectors. *Odyssey*, 2018, 2018, 105-111.
- [34]Zhao, J.C. Multitask learning language recognition method based on contrast predictive coding model. *Data Acquisition and Processing*, 2022, 37(02), 288-297. DOI: 10.16337/j.1004-9037.2022.02.003.



© 2023 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)