International Journal of Interdisciplinary Telecommunications and Networking
Volume 18 • Issue 1 • January-December 2026

Phonetic Feature Detection and Comparative Analysis of Chinese Dialects Using Computer-Based Methods

Yanan Wang
https://orcid.org/0009-0007-1140-3891
Academic Affairs and Scientific Research Department, Hebi Vocational College of Energy and Chemistry, China
Received: July 17th, 2025 | **Accepted:** January 27th, 2026

ABSTRACT

Computer-aided dialect speech recognition is an important interdisciplinary field combining linguistics and computer science. The phonological complexity of Chinese dialects poses major challenges for traditional analysis methods. While computational approaches have gained traction, issues like data scarcity and poor model generalization remain. This paper reviews current techniques and proposes a new deep learning approach enhanced with transfer learning to address these limitations. Experimental results show improved accuracy and robustness over conventional methods. A comparative analysis highlights key advancements and future research directions. This work contributes to both linguistic and information sciences and supports efforts in documenting endangered dialects.

KEYWORDS

Computer Aided, Chinese Dialect, Speech Recognition, Deep Learning, Transfer Learning

INTRODUCTION

Language functions not only as a communication medium but also as a carrier of sociocultural and historical information (Y. Kumar et al., 2023; Xu et al., 2021). In linguistically diverse regions, such as China, the rich tapestry of dialects reflects a complex sociolinguistic landscape. According to surveys conducted by the Chinese Academy of Social Sciences, over 80 distinct dialects are spoken across the country, each characterized by unique phonetic, lexical, and syntactic features that encapsulate local traditions, folklore, and cultural values (Shi et al., 2022). However, the daily use and intergenerational transmission of many dialects are obviously decreasing (L. Chen, 2023). Many dialects are now endangered, with some even classified as critically endangered by United Nations Educational, Scientific and Cultural Organization, highlighting a broader trend of linguistic erosion.

The loss of dialects is not merely a linguistic issue; it signifies a profound erosion of cultural identity and heritage (Kasiyarno & Apriyanto, 2025). As a result, the preservation of dialectal diversity has emerged as an urgent and pressing concern. Traditional methods for documenting dialects have predominantly relied on manual phonetic transcription and auditory comparison (Sertolli et al., 2021). While these methods have been valuable in their own right, they are labor-intensive, subjective, and often inefficient when dealing with large datasets. In the context of growing digital archives and the

DOI: 10.4018/IJITN.400786

increasing prevalence of speech-driven applications, there is a clear and urgent need for automated solutions that can scale up dialect preservation efforts in a more efficient and sustainable manner.

Recent advancements in machine learning, particularly deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), offer promising and innovative solutions to this challenge (Zhang et al., 2021). CNNs have demonstrated exceptional capabilities in extracting spectral features from speech signals, while RNNs are adept at capturing the temporal dependencies inherent in speech patterns (Kheddar et al., 2024; Wei et al., 2022). Additionally, transfer learning has emerged as a powerful approach, enabling the reuse of knowledge from high-resource languages, such as Mandarin, to enhance recognition performance in under-resourced dialects. Despite some progress in multilingual speech recognition, the application of these techniques to Chinese dialects remains relatively underexplored, leaving a significant research gap that needs to be addressed.

This study aims to bridge this gap by developing an advanced hybrid deep learning model that integrates CNNs for spectral feature extraction, RNNs for temporal modeling, and transfer learning to improve recognition accuracy. The proposed model will be rigorously tested on a diverse range of dialectal speech corpora and compared against traditional approaches, such as hidden Markov models (HMMs) and support vector machines (SVMs). By leveraging the strengths of these cutting-edge techniques, the hybrid model offers a more scalable and efficient approach to dialect recognition, with potential applications in digital preservation and revitalization efforts.

Under this background, the core scientific problem that this paper focuses on is how to realize high-precision and extensible speech recognition for multi-Chinese dialects under the condition of limited labeled data. In order to solve this problem, this paper proposes a hybrid model that combines CNNs, circular neural network and transfer learning. By jointly modeling the spectrum characteristics, time series dependence, and cross-dialect knowledge transfer, the recognition performance in low-resource dialect scenes is improved. This study provides a systematic and scalable solution for multi-dialect speech recognition in both method and application.

LITERATURE REVIEW

The field of Chinese dialect speech recognition has undergone notable advancements, particularly with the integration of deep learning and transfer learning techniques. These approaches have significantly contributed to overcoming challenges, such as data scarcity, tonal complexity, and phonetic variability across dialects.

Li et al. (2024) provided a comprehensive survey that chronicled the evolution of dialect recognition methodologies. Their review marked the transition from traditional rule-based approaches to modern data-driven techniques, highlighting the importance of acoustic-phonetic modeling in dialect classification. Specifically, the study emphasized the role of phonetic feature extraction, including acoustic, phonemic, and prosodic elements, as foundational components in distinguishing between dialects. However, the study also acknowledged that rule-based systems often exhibit limited adaptability when confronted with irregular or unseen linguistic patterns.

Wang et al. (2021) introduced an innovative end-to-end dialect identification system based on multilingual automatic speech recognition. By leveraging a transfer learning strategy, their system transferred knowledge from high-resource languages (such as Mandarin and English) to low-resource Chinese dialects, thereby improving recognition accuracy in data-constrained environments. Their experimental results demonstrated that multilingual pretraining can significantly enhance model generalizability, particularly in dialects with sparse training corpora.

Expanding on the transfer learning paradigm, Taghizadeh and Faili (2022) proposed a model that employs shared hidden layers across related language tasks to facilitate cross-linguistic knowledge transfer. The survey provides an overview of cross-lingual dependency parsing approaches, particularly focusing on the model transfer method. It discusses how these approaches address challenges, such as