

The logo for RADemics, featuring the text "RADemics" in white on a blue arrow-shaped background pointing to the right. The arrow is part of a larger blue horizontal bar that is attached to a dark blue vertical bar on the left side of the page.

RADemics

# AI-Driven Automated English Speaking Skill Assessment Using Deep Learning Techniques

A decorative graphic consisting of several thin, curved lines in shades of blue and grey, originating from the bottom left and extending upwards and to the right, resembling stylized grass or reeds.

A. S. Gousia Banu, Bhupendra Kumar  
Patel

ST PETERS ENGINEERING AUTONOMOUS COLLEGE,  
NAVEEN GOVERNMENT COLLEGE

# AI-Driven Automated English Speaking Skill Assessment Using Deep Learning Techniques

<sup>1</sup>A. S. Gousia Banu, Professor, Department of CSE, St Peters Engineering Autonomous College, Hyderabad, Telangana, India. [gbanuzia@gmail.com](mailto:gbanuzia@gmail.com)

<sup>2</sup>Bhupendra Kumar Patel, HoD, English Principal, Naveen Government College, Nawagarh, Janjgir Champa, Chhattisgarh State, India. [bk1962.patel@gmail.com](mailto:bk1962.patel@gmail.com)

## Abstract

Rapid advancement in artificial intelligence technologies has significantly influenced modern educational systems, particularly in the field of language learning and assessment. Accurate evaluation of English-speaking proficiency remains a challenging task due to the multidimensional characteristics of spoken communication involving pronunciation, fluency, grammatical accuracy, vocabulary usage, and semantic coherence. Conventional speaking assessment methods depend largely on human evaluators, leading to subjectivity, scoring inconsistencies, and limited scalability in large learning environments. AI-driven automated speaking assessment systems offer an effective solution by enabling objective, consistent, and scalable evaluation of spoken language performance. Integration of speech processing techniques, automatic speech recognition, and advanced feature extraction methods enables detailed analysis of acoustic and linguistic characteristics embedded within speech signals. Deep learning models trained on large speech datasets learn complex hierarchical patterns related to articulation, speech rhythm, and sentence structure, enabling accurate prediction of language proficiency levels. Hybrid deep learning architectures combining convolutional neural networks, recurrent neural networks, and attention-based mechanisms enhance the capability of automated systems to analyze both temporal and contextual aspects of spoken responses. Extraction of acoustic features such as Mel Frequency Cepstral Coefficients, pitch variation, and energy patterns along with semantic and syntactic representations derived from speech transcripts supports comprehensive evaluation of communication competence. Automated speaking assessment frameworks provide real-time evaluation and personalized feedback, facilitating continuous improvement of speaking abilities within digital learning environments. Such intelligent systems contribute significantly to online education platforms, language training programs, and large-scale language proficiency testing by ensuring efficient and unbiased assessment processes. Development of AI-driven evaluation frameworks therefore represents a major advancement in educational technology, offering reliable tools for objective speaking skill assessment and supporting adaptive language learning in modern technology-enabled educational ecosystems.

Keywords: Artificial Intelligence, Deep Learning, Automated Speech Assessment, Speech Processing, English Speaking Skill Evaluation, Natural Language Processing.

## Introduction

English language proficiency has become a fundamental requirement in global communication, academic development, and professional advancement [1]. Among the core language skills,

speaking ability represents a critical component that reflects communicative competence and linguistic proficiency. Effective speaking involves accurate pronunciation, appropriate vocabulary usage, grammatical correctness, fluency, and the ability to convey meaningful ideas in a coherent manner [2]. Evaluation of these characteristics plays a crucial role in language learning environments, standardized language testing, and professional skill assessment [3]. Traditional speaking evaluation practices generally rely on human examiners who assess spoken responses based on predefined rubrics and subjective judgment [4]. Such approaches introduce challenges associated with evaluator bias, scoring inconsistencies, and limitations in scalability when large numbers of learners require assessment. Increasing demand for reliable and efficient language evaluation systems has created the need for automated solutions capable of performing consistent and objective speaking assessments across diverse educational contexts [5].

Rapid progress in artificial intelligence technologies has opened new opportunities for transforming language education and evaluation practices [6]. AI-driven systems enable computational analysis of speech signals and linguistic patterns generated during spoken communication [7]. Speech processing techniques allow conversion of raw audio signals into structured data representations that capture phonetic articulation, rhythm patterns, and speech energy distribution [8]. Machine learning algorithms trained on annotated speech datasets can identify correlations between speech characteristics and language proficiency levels [9]. Deep learning models provide enhanced capability for recognizing complex speech patterns through hierarchical feature learning, enabling accurate interpretation of pronunciation, fluency, and linguistic organization within spoken responses. Integration of such technologies has encouraged development of automated systems designed to evaluate speaking performance with improved accuracy and efficiency [10].