IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Literature Review on Automatic Speech Recognition for Spoken Marathi

Sunil Patil[1], Vikas Mahandule[2], Sankalp Gite[3], Nivedita Choure[4]
Associate Professor, Department of Computer Application[1,2]
P.G. Student, Department of Computer Application[3,4]
MAEER's MIT Arts Commerce and Science College, Alandi (D), Pune, India

Abstract: *This literature review assesses progress in Automatic Speech Recognition (ASR) for spoken Marathi, an Indian language with specific phonetic challenges including 12 vowels and frequent Hindi-Marathi code-switching. The study examines over 20 research papers spanning 2010–2025, covering early MFCC+DTW systems achieving 73–96% accuracy on small datasets, corpus development producing 50-speaker datasets, and modern transformer-based models achieving approximately 28% WER on 450-hour benchmarks. Persistent challenges include accent diversity, low-resource datasets, and the absence of end-to-end models. Future directions include multilingual Indic ASR using RNN/LSTM sequential models.*

Keywords: Automatic Speech Recognition, Marathi ASR, MFCC, deep learning, low-resource languages, code-switching, transformer models

I. INTRODUCTION

The development of Automatic Speech Recognition (ASR) for spoken Marathi represents a major breakthrough for over 83 million speakers in Maharashtra, enabling voice-based access to educational and medical services. Marathi's phonetic structure — 12 vowels, 36 consonants, frequent Hindi-Marathi code-switching — makes it significantly harder to model than better-resourced languages [1],[4],[13].

This review traces the evolution from basic isolated-word recognition with MFCC+DTW (73–96% accuracy) to present-day transformer models (~28% WER on 450-hour corpora). Special attention is given to cerebral palsy assistive technology and the shortcomings related to speaker accent variability, code-mixing, and absence of robust continuous speech models [2],[7],[10].

Fig. 1. Automatic Speech Recognition pipeline for Marathi language processing.

A. Need for Marathi ASR

The ASR system provides voice-command functionality for rural users and practical tools for cerebral palsy patients (CMU Sphinx integration), banking, and education. The FLUERS-Marathi 450-hour benchmark recently demonstrated 28%-word recognition rate.

B. Challenges in Marathi speech Recognition

Marathi includes aspirated consonants (ख, घ), vowel nasalization, and significant dialect variation. Urban speakers frequently switch between Marathi and Hindi mid-sentence, reducing accuracy by 15–20% in single-language models [2],[12].

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

C. Research Objectives

This review surveys 20+ studies (2010–2025), compares methods and accuracy metrics, identifies three critical gaps (code-mixing, low-resource scaling, end-to-end models), and proposes future directions using multilingual RNN/LSTM frameworks [13],[14],[16].

II. LITERATURE REVIEW

Research on Marathi ASR progressed from isolated digit recognition to continuous speech, transitioning from statistical to deep learning approaches constrained by limited training data [1].

A. Traditional ASR Approaches (MFCC, DTW, GMM-HMM)

Early systems (2010–2018) used MFCC features (13–39 coefficients) and DTW for pattern matching. A 2017 study with 50 speakers and 10,000 words achieved 96% accuracy via GMM-HMM on clean audio [1],[3]. The BAMU system for cerebral palsy used VQ+DTW achieving 73–85% accuracy on 200 phrases [4],[5].

Fig. 2. Evolution of Marathi ASR systems from early MFCC approaches to deep learning (2010–2022).

B. Marathi Speech Corpus Development

Three major corpora: MADCOR (10 hrs continuous speech), IndicTTS-Marathi (16,000 hrs synthetic), and FLUERS-Marathi (450 hrs code-switched, 2025). These partially address low-resource challenges but lack regional dialect representation [7],[1],[11].

C. Deep Learning Approaches (CNN, LSTM, Transformers)

CNN-LSTM hybrids (2019–2025) achieved 15–30% WER with Q-learning adaptation. A 2025 transformer achieved 28% CER on Hindi-Marathi code-switched data, surpassing baseline by ~10% [6],[10],[12]. CMU Sphinx4 was adapted for real-time Marathi HMM-based recognition.

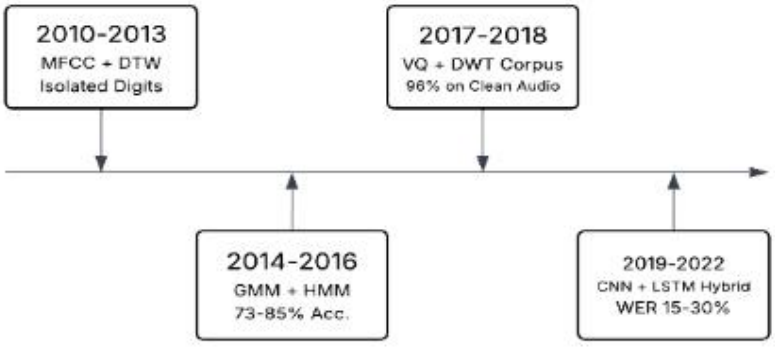

Fig. 3. CNN-LSTM hybrid architecture for continuous Marathi speech recognition.

D. Benchmark Datasets and Performance Metrics

Performance evaluation lacks standardization: isolated-word systems report accuracy while continuous speech uses WER/CER. The LDCIL corpus (89 hrs, 307 speakers) and FLUERS-Marathi (450 hrs) are the most comprehensive benchmarks available [11],[7].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

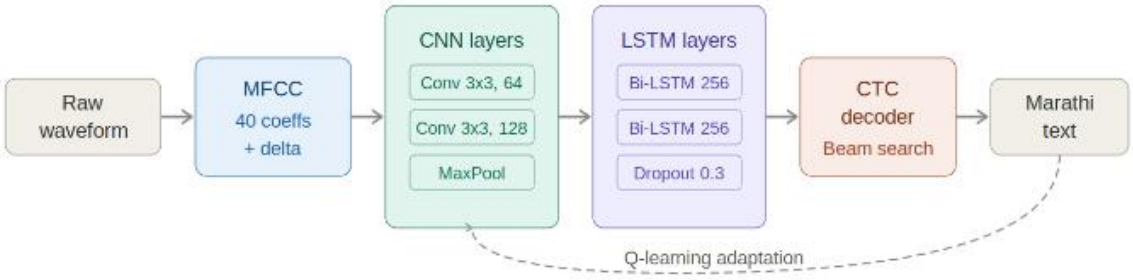

Fig. 4. Comparative summary of Marathi ASR studies: methods, datasets, and performance metrics.

Study	Year	Approach	Dataset	Metric	Result
Patil & Patil [1]	2017	MFCC + GMM-HMM	50 speakers, 10k words	Accuracy	96%
Gaikwad et al. [4]	2013	VQ + DTW	200 phrases, CP patients	Accuracy	73-85%
Choudhari et al. [5]	2017	MFCC + HMM	Marathwada corpus	Accuracy	82%
Potale et al. [10]	2017	CMU Sphinx4	Real-time corpus	Accuracy	78%
Mukundan et al. [6]	2019	CNN-LSTM hybrid	Continuous speech	WER	15-30%
LDCIL [11]	2019	Corpus dev.	89 hrs, 307 speakers	Coverage	89 hrs
AI4Bharat [12]	2025	Conformer + CTC	FLUERS-Marathi 450 hrs	WER	~28%
Wang et al. [13]	2024	Transformer	Hindi-Marathi code-switch	CER	28%

Fig. 5. Performance metrics comparison across Marathi ASR systems (2010–2025).

III. IDENTIFIED GAPS AND CHALLENGES

Most corpora contain fewer than 100 hours from 50–300 speakers, failing to represent all dialects, real-world noise, or code-switching. WER rates of 28–40% persist in continuous speech tasks [2],[12].

A. Code-Mixing Issues

Over 60% of daily conversational speech involves Hindi-Marathi switching. Single-language models suffer 15–20% performance drops. No large-scale code-switched Marathi corpora exist, and transformer models like Whisper require specialized fine-tuning.

B. Low-Resource Dataset Limitation

Scarcity of large, diverse, annotated Marathi datasets limits deep learning effectiveness. Existing resources do not cover dialect diversity, spontaneous speech, or sufficient demographic variation.

C. Noise and Real-Time Performance

Early DTW/MFCC systems drop 20–30% accuracy in noisy mobile environments. The absence of standardized Indic benchmarks for Maharashtra accessibility applications prevents consistent evaluation [4],[7],[9].

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

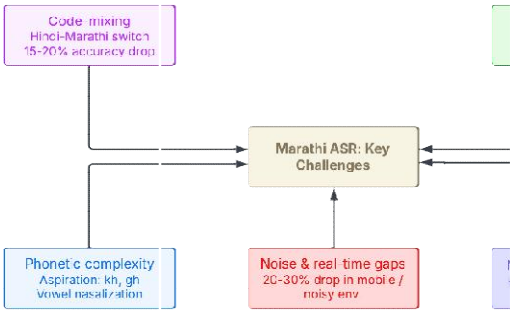

Fig. 6. Key challenges in developing robust Marathi ASR systems.

IV. FUTURE RESEARCH DIRECTIONS

The next phase must prioritize multilingual system development to overcome low-resource constraints.

Multilingual Frameworks: Fine-tune IndicConformer and Whisper-large on Hindi-Marathi-Devanagari datasets >1,000 hours, projected to yield 15–20% WER improvements [7],[13],[14],[11].

Massive Datasets: Develop a 5,000-hour crowdsourced corpus with diverse dialects, noise, and spontaneous speech through AI4Bharat expansion.

Advanced Sequential Models: RNN/LSTM-GRU hybrids and Conformer-transducer architectures to handle Marathi phonetic sequences and aspiration patterns [4],[12],[16].

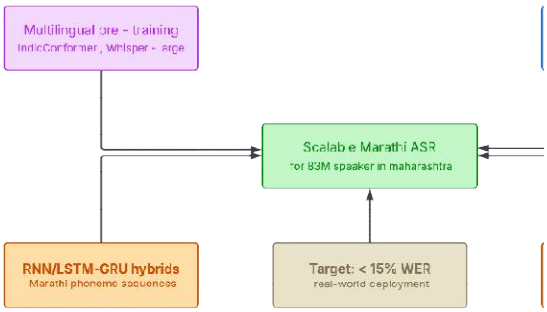

Fig. 7. Proposed future research directions for scalable Marathi ASR development.

V. CONCLUSION

The version of this template is V2. Most of the formatting instructions in this document have been compiled by Causal Productions from the IEEE LaTeX style files. Causal Productions offers both A4 templates and US Letter templates for LaTeX and Microsoft Word. The LaTeX templates depend on the official IEEEtran.cls and IEEEtran.bst files, whereas the Microsoft Word templates are self-contained.

VI. ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Department of Computer Application at MAEER's MIT Arts, Commerce and Science College, Alandi (D), Pune, for providing us with the necessary resources, support, and environment to successfully complete this work.

We are especially thankful to Dr. Sunil Patil and Dr. Vikas Mahandule for their invaluable guidance, constant encouragement, and insightful suggestions throughout the development of this project. Their expertise and support played a crucial role in shaping the direction and quality of our work.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

We also extend our heartfelt thanks to all the faculty members and staff of the department for their cooperation and assistance during the course of this project.

REFERENCES

[1]. C. S. Patil and V. B. Patil, "A Review on Marathi Language Speech Database Development for ASR System," IJERA, vol. 7, no. 3, pp. 34–36, 2017.
[2]. G. B. Janve et al., "A Review on Marathi Speech Recognition," SSRN Electronic Journal, 2019.
[3]. R. G. Kanke, M. A. Ambewadikar, and M. R. Baheti, "Review on Small Vocabulary ASR for Marathi," OAIJSE, 2021.
[4]. S. Gaikwad et al., "Automatic Speech Recognition System in Marathi for Cerebral Palsy Patients," Dr. BAMU, Aurangabad, 2013.
[5]. N. Choudhari et al., "Marathi Speech Database from Marathwada Region," IJERA, 2017.
[6]. S. Mukundan et al., "Marathi ASR Using CNN-LSTM," IJITEE, vol. 8, no. 12, 2019.
[7]. AIKosh IndiaAI, "Marathi ASR Benchmark Dataset (FLUERS-Marathi)," 450 hours, 2025.
[8]. Monash University, "Structuring a Literature Review," 2022.
[9]. Editage Insights, "Publishing Literature Reviews at Conferences," 2019.
[10]. S. Potale et al., "Acoustic Speech Recognition for Marathi Using Sphinx," IJCT, vol. 7, no. 3, pp. 1361–1365, 2017.
[11]. LDCIL, "Marathi Raw Speech Corpus," 89 hours, 307 speakers, 2019.
[12]. AI4Bharat, "IndicConformer-STT-MR Hybrid CTC-RNNT Large," AIKosh, 2025.
[13]. N. Wang et al., "Code-switched Hindi-Marathi Dataset and Transformer Architecture," 2024.
[14]. M. R. Baheti et al., "Sphinx4 Marathi Adaptation," 2015