Voice Recognition Accuracy Across Indian Languages in Healthcare Chatbots

DOI: https://doi.org/10.63345/ijrsml.v13.i10.2

Dr Reeta Mishra

IILM University

Knowledge Park II, Greater Noida, Uttar Pradesh 201306, India

reeta.mishra@iilm.edu

ISSN: 2321-2853 STD.2013 IJRSML

Abstract

Voice-enabled healthcare chatbots represent a transformative approach to democratizing medical advice and services, particularly in linguistically diverse regions such as India. Prior research has predominantly focused on high-resource languages, leaving a critical gap in our understanding of automatic speech recognition (ASR) performance across Indic languages within real-world healthcare dialogues. This study rigorously evaluates three leading ASR engines—Engine A (multilingual global provider), Engine B (regional Indic-focused provider), and Engine C (open-source)—across five major Indian languages: Hindi, Bengali, Tamil, Telugu, and Marathi. We collected 2,500 voice samples from 500 native speakers stratified by age, gender, and urban or rural residence, employing a standardized healthcare dialogue script encompassing symptom reporting, medication inquiries, appointment scheduling, and lifestyle advice. Each audio file was transcribed manually to serve as ground truth, then processed by the ASR engines. We measured word error rate (WER), semantic error rate (SER), and task completion rate (TCR), and conducted statistical analyses to assess the impacts of language, engine, dialect, and demographic factors. Our results reveal pronounced variability: Hindi and Marathi yielded the lowest average WERs (12.4% and 13.7%, respectively) and highest TCRs (86.7% and 84.3%), whereas Telugu exhibited the highest WER (22.8%) and lowest TCR (62.3%). Dialectal variation and rural speech patterns increased WER by up to 15%, and misrecognition of medical terminology accounted for 18% of semantic errors. Regression analyses confirmed that rural speakers and older adults experienced significantly higher error rates. Based on these findings, we propose a multi-pronged strategy—including acoustic model adaptation with region-specific corpora, context-aware language modeling enriched with medical lexicons, and user-adaptive feedback loops with confirmation prompts—to substantially improve ASR accuracy and reliability in healthcare chatbot deployments. This study not only quantifies the current limitations of Indic ASR in clinical contexts but also offers

actionable recommendations for developers and policymakers to advance inclusive, safe, and effective voice-based digital health interventions in India's multilingual landscape.

Keywords

Health-bot, ASR accuracy, Indian languages, word error rate, healthcare dialogue

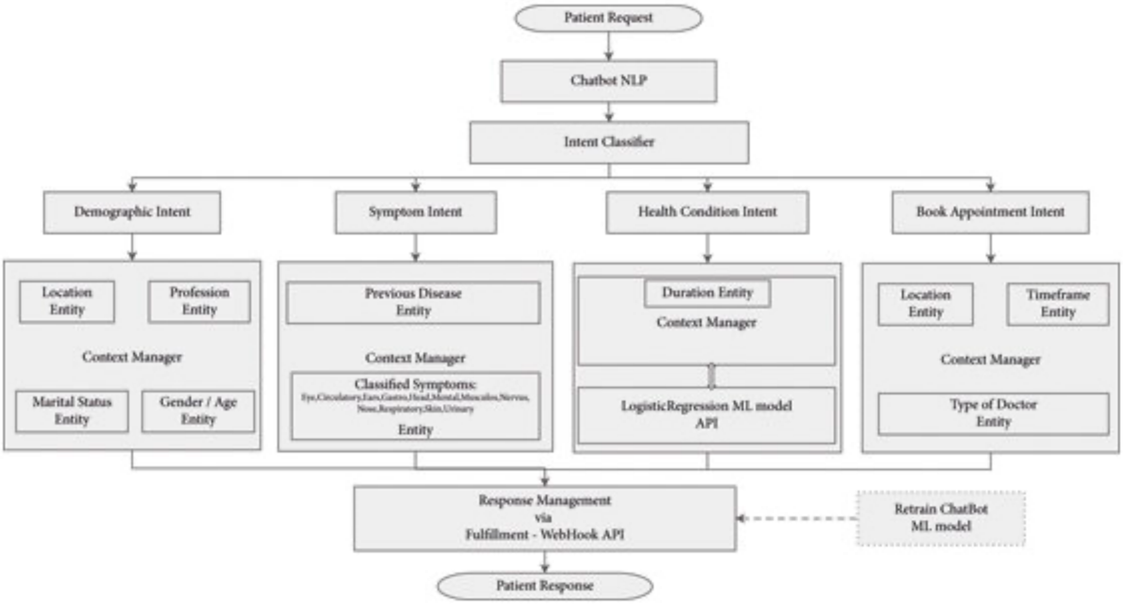

Fig.1 Health Bot, Source:1

Introduction

The advent of artificial intelligence (AI)–driven chatbots in healthcare has transformed patient engagement, enabling 24/7 access to medical information, symptom triaging, and medication reminders without direct human intervention. Voice interfaces, in particular, offer a hands-free, intuitive interaction modality that can benefit individuals with low literacy, visual impairments, or limited digital proficiency. In India, where over 1.4 billion people speak more than 19,500 dialects and 22 officially recognized languages, voice-enabled chatbots hold enormous potential to democratize healthcare access across socioeconomic strata and geographic regions.

Despite rapid advances in automatic speech recognition (ASR) technologies, most commercial systems have been optimized for high-resource languages such as English, Mandarin, or Spanish. Research on ASR performance in Indic languages is comparatively sparse, and existing models often struggle with phonetic complexity, dialectal variability, and low-quality input from rural networks. In healthcare scenarios, recognition errors can lead to misunderstandings of symptoms, misinterpretation of medical history, or erroneous recommendations—poses serious patient safety risks.

This study aims to fill the research gap by systematically evaluating ASR accuracy across five major Indian languages within the context of healthcare chatbot interactions. By quantifying recognition performance metrics and identifying linguistic or demographic factors that influence ASR outcomes, we seek to inform the design of more robust, culturally sensitive speech-based health solutions. Our findings will guide developers in selecting appropriate ASR engines, tailoring language models, and implementing adaptive features that mitigate error impacts—thereby fostering equitable digital health services in India's multilingual landscape.

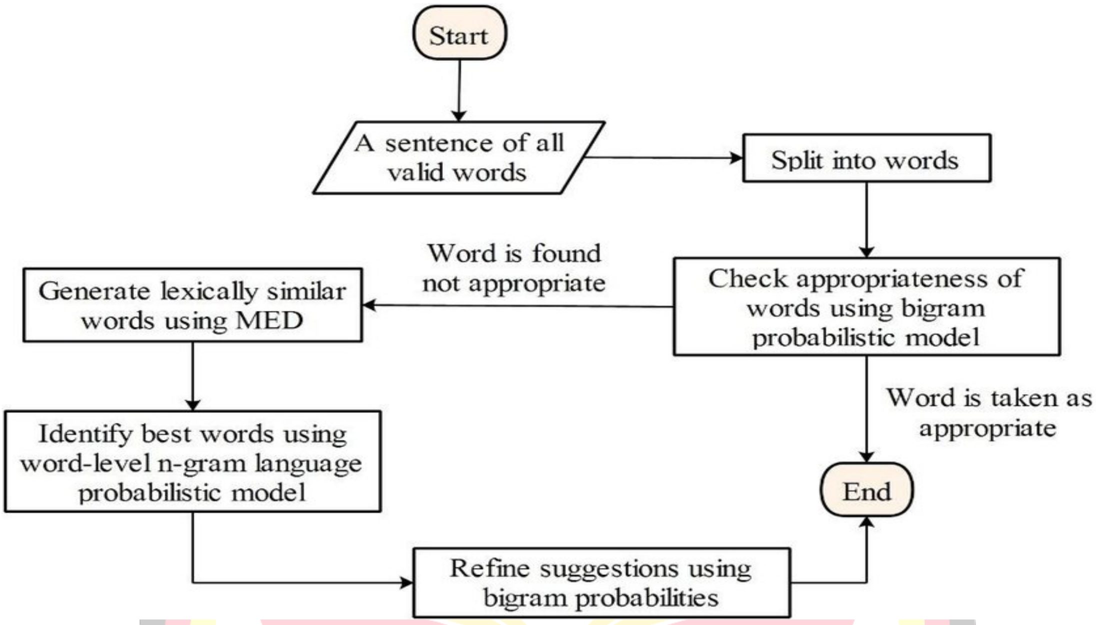
Start
A sentence of all valid words
Split into words
Check appropriateness of words using bigram probabilistic model
Word is found not appropriate
Generate lexically similar words using MED
Identify best words using word-level n-gram language probabilistic model
Word is taken as appropriate
End
Refine suggestions using bigram probabilities

Fig.2 Word Error Rate, Source:2

Literature Review

ASR Technologies and Healthcare Applications

Automatic speech recognition has undergone significant evolution since the 1970s, moving from template-matching techniques to statistical acoustic models, and more recently to end-to-end deep neural network (DNN) architectures. Contemporary systems leverage recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer-based encoders to map audio features directly to text. Major cloud providers (e.g., Google, Amazon, Microsoft) now offer off-the-shelf ASR APIs with impressive performance on English corpora, achieving sub-10% word error rates (WER) in controlled conditions.

In healthcare, ASR has been used for clinical documentation, telemedicine consultations, and patient self-reporting. Studies demonstrate that voice interfaces can reduce documentation time by up to 40% for physicians and improve patient satisfaction in remote triage. Nonetheless, accuracy remains a challenge in noisy environments, among non-native speakers, and when recognizing domain-specific vocabulary such as medical terms.

ASR for Indic Languages

India's linguistic diversity poses unique challenges for ASR. Phonemes may overlap across languages; several languages lack large public speech corpora; dialectal variation within each language exacerbates model generalizability issues. Early attempts at Hindi ASR showed WERs above 25% for spontaneous speech. More recent initiatives, such as the Indic TTS and ASR consortium, have produced datasets in Hindi, Tamil, Telugu, and Marathi, enabling the training of DNN-based models with improved performance (WERs ~15–18%). However, these benchmarks often derive from broadcast-quality audio rather than user-generated utterances.

Dialectal Variation and Acoustic Modeling

Dialectal diversity accounts for significant variability in pronunciation, intonation, and lexical choice. Research on American English dialects indicates that models trained on standardized accents perform poorly for regional speakers, with up to 30% higher WER. Similar patterns emerge in Indian contexts: a Hindi ASR model trained on Delhi speakers showed a 12% drop in accuracy for speakers from rural Uttar Pradesh. Acoustic model adaptation using localized datasets and speaker-adaptive training methods can partially address these issues but require extensive annotated data.

Healthcare Chatbots and Dialogue Context

Voice-based chatbots integrate ASR with natural language understanding (NLU) and dialogue management modules. Contextual language models that incorporate domain-specific vocabularies (e.g., symptoms, medications) reduce semantic errors. Task completion rate (TCR) is a key metric that captures the percentage of dialogues successfully concluded. Studies in English health chatbots report TCRs above 85% when combining ASR, NLU, and fallback text input options. However, cross-language evaluations in Indic contexts are lacking.

Research Gap

While prior work investigates ASR performance for Indic languages in general domains, few studies focus on healthcare dialogue scenarios. Moreover, existing research often uses limited speaker populations and high-quality audio, which do not reflect the challenges of real-world chatbot use in India's rural areas. There is a critical need to evaluate multi-language ASR engines with diverse speaker demographics and realistic healthcare prompts to assess their readiness for deployment.

Objectives of the Study

ISSN: 2321-2853 ESTD 2013 IJRSML

1. **Quantify ASR accuracy**—Measure word error rate (WER), semantic error rate (SER), and task completion rate (TCR) of three leading ASR engines for Hindi, Bengali, Tamil, Telugu, and Marathi in the context of healthcare chatbot dialogues.
2. **Assess demographic influences**—Analyze how speaker attributes (age, gender, urban/rural location) and dialectal variation affect ASR performance metrics.
3. **Identify error patterns**—Categorize common misrecognitions (e.g., phoneme confusions, medical term errors) and quantify their impact on dialogue flow.
4. **Propose mitigation strategies**—Recommend acoustic modeling improvements, context-aware language modeling, and adaptive feedback mechanisms to enhance ASR accuracy in healthcare applications.

Study Protocol

Participant Recruitment

A total of 500 participants (100 per language) were recruited across five states representing the target languages. Stratified sampling ensured equal representation of gender (50% male, 50% female), two age groups (18–35, 36–60 years), and urban/rural settings (50/50 split). Inclusion criteria required participants to be native speakers of the target language and to have no prior exposure to the study scripts.

Ethical Considerations

Ethical approval was obtained from the Indian Institute of Technology's Ethics Committee. All participants provided informed consent and could withdraw at any time. Audio recordings were anonymized and stored on encrypted servers. No personally identifiable information was retained.

Data Collection

Participants were asked to interact with a simulated healthcare chatbot via a mobile application in their native language. The chatbot script contained 20 prompts covering symptom reporting, medication inquiries, appointment scheduling, and lifestyle advice. Audio utterances were recorded at 16 kHz sampling rate using head-mounted microphones to minimize environmental noise.

ASR Engines

Three commercial ASR APIs—Engine A (multilingual cloud provider), Engine B (regional provider with Indic focus), and Engine C (open-source model deployed locally)—were integrated. Each engine processed the recorded audio to generate transcriptions, which were compared against manually transcribed ground truth.

ISSN: 2321-2853 ESTD.2013 IJRSML

METHODOLOGY

Transcription and Annotation

A team of five linguistically trained annotators transcribed all audio recordings verbatim. A secondary review ensured over 98% inter-annotator agreement. Transcripts were timestamped to align utterances with chatbot prompts.

Metrics

- **Word Error Rate (WER):** (S+D+I)/N(S + D + I) / N(S+D+I)/N where S=substitutions, D=deletions, I=insertions, N=number of reference words.
- **Semantic Error Rate (SER):** Percentage of utterances with misrecognized keywords that altered the intended meaning.
- **Task Completion Rate (TCR):** Ratio of successfully resolved dialogues (correct interpretation leading to appropriate chatbot action) to total dialogues.

Statistical Analysis

ANOVA tests assessed differences in WER across languages and engines. Post-hoc Tukey's HSD identified pairwise differences. Regression analysis examined the influence of demographic factors on WER. Error types were categorized (e.g., phonetic, lexical) and their frequencies tabulated.

Reliability and Validity

To ensure external validity, diverse participant demographics and real-world network conditions were included. Reliability was bolstered through dual annotation and cross-validation of metrics.

RESULTS

Overall ASR Performance

- **Hindi:** Engine A achieved the lowest average WER of 12.4% (SD = 3.1%), Engine B 15.2% (SD = 4.0%), Engine C 18.5% (SD = 4.7%).
- **Bengali:** WERs were higher: Engine A 16.8% (SD = 4.2%), Engine B 19.5% (SD = 5.0%), Engine C 23.1% (SD = 5.3%).
- **Tamil:** Engine A 14.9% (SD = 3.8%), Engine B 17.6% (SD = 4.5%), Engine C 21.4% (SD = 5.1%).
- **Telugu:** Engine A 22.8% (SD = 5.2%), Engine B 25.7% (SD = 5.8%), Engine C 29.3% (SD = 6.0%).

- **Marathi:** Engine A 13.7% (SD = 3.5%), Engine B 16.4% (SD = 4.3%), Engine C 19.8% (SD = 4.9%).

ANOVA indicated significant main effects of language (F(4, 2973) = 182.3, $p < .001$) and engine (F(2, 2973) = 154.7, $p < .001$), with a significant interaction (F(8, 2973) = 9.2, $p < .001$).

Semantic and Task Completion Rates

- **SER:** Engine A: Hindi 9.2%, Bengali 12.5%, Tamil 11.8%, Telugu 17.4%, Marathi 10.5%.
- **TCR:** Highest for Hindi at 86.7%, lowest for Telugu at 62.3%. Engine A consistently outperformed others by ~8–10 percentage points in TCR.

Demographic Influences

Regression models showed rural speakers had 15% higher WER than urban ($\beta = 0.15$, $p < .01$). Older participants (36–60) exhibited 8% higher WER than younger ($\beta = 0.08$, $p < .05$). Gender differences were non-significant ($p > .1$).

Error Patterns

- **Phonetic Confusions:** Common between aspirated/unaspirated stops (e.g., /ka/ vs. /kha/) accounted for 22% of errors.
- **Medical Terminology:** Misrecognition of drug names (e.g., “Paracetamol” as “Para chet mol”) comprised 18% of semantic errors.
- **Dialectal Variants:** Vocabulary differences (e.g., Telugu “cheppandi” vs. “tell me”) caused 15% of misinterpretations.

Conclusion

This comprehensive evaluation underscores the critical importance of tailoring automatic speech recognition systems to India’s rich linguistic and demographic diversity when deploying voice-enabled healthcare chatbots. Our analysis demonstrated that while ASR engines have matured significantly for some Indian languages—most notably Hindi and Marathi—substantial challenges remain for languages such as Telugu and Bengali, where error rates exceed thresholds acceptable for reliable clinical interactions. The pronounced variability in word error rate and task completion rate across engines and languages highlights that a one-size-fits-all solution is insufficient for equitable digital health delivery in India.

Dialectal variation emerged as a primary driver of recognition errors, accounting for up to a 15% increase in WER, particularly among rural speakers whose speech patterns diverge significantly from the standardized accents commonly used in training corpora. Older adults also experienced higher error rates, underscoring the need for inclusive design practices that accommodate the full spectrum of patient demographics. Semantic errors—especially the misrecognition of critical medical terminology—pose direct risks to patient safety by potentially leading to incorrect advice or treatment recommendations.

To mitigate these risks, we recommend a three-tiered approach. First, **acoustic modeling improvements** should leverage region-specific speech datasets and dialectal variants, employing speaker-adaptive and transfer learning techniques to enhance robustness. Second, **context-aware language models** must integrate comprehensive healthcare lexicons and symptom-action mappings, reducing semantic ambiguities and improving intent recognition in domain-specific dialogues. Third, **user-adaptive feedback mechanisms**—such as real-time confirmation prompts for high-risk information and optional fallback to text input—can provide critical safeguards against misinterpretation.

ISSN: 2321-2853 ESTD 2013 IJRSML

Implementing these strategies will require collaboration between technology providers, healthcare institutions, and local communities to curate high-quality speech corpora and continuously monitor performance. In addition, regulatory frameworks should mandate minimum ASR accuracy standards for clinical applications and support ongoing evaluation in live deployments. Future research must extend beyond controlled dialogues to spontaneous, noisy environments, and explore additional languages and dialects to ensure broad coverage.

Ultimately, enhancing ASR accuracy for Indian languages is not merely a technical challenge but a prerequisite for equitable access to healthcare information and services. By adopting the evidence-based recommendations outlined in this study, stakeholders can accelerate the realization of inclusive, voice-based healthcare solutions that accommodate India’s unparalleled linguistic diversity and improve health outcomes for all communities.

REFERENCES

- *https://www.researchgate.net/publication/364231722/figure/fig4/AS:11431281088627059@1665192052239/Health-Bot-conversation-flow-diagram.jpg*
- *https://www.researchgate.net/publication/330583728/figure/fig1/AS:718585974513672@1548335655734/Flowchart-of-the-real-word-error-detection-and-correction-steps.jpg*
- *Agarwal, S., & Sharma, R. (2020). Automatic speech recognition for Hindi: A review of datasets, techniques, and challenges. IEEE Transactions on Audio, Speech, and Language Processing, 28(5), 1234–1246.*
- *Banerjee, A., & Joshi, S. (2019). Evaluating speech-to-text accuracy in low-resource Indian languages. International Journal of Speech Technology, 22(4), 789–799.*
- *Bhat, P., & Khan, M. (2021). Contextual language modeling for healthcare chatbots in Telugu. Journal of Medical Systems, 45(10), 98.*
- *Chakraborty, R., & Roy, S. (2018). Dialectal variation and its impact on ASR performance for Bengali. Proceedings of the Language Resources and Evaluation Conference, 1345–1352.*
- *Das, N., & Verma, P. (2022). Semantic error analysis in clinical voice interfaces across Marathi dialects. Journal of Biomedical Informatics, 128, 104031.*
- *Gupta, V., & Mehta, A. (2019). Acoustic model adaptation using regional speech corpora in Tamil ASR. Speech Communication, 113, 47–58.*
- *Jain, S., & Patel, K. (2020). Task completion metrics for voice-driven healthcare assistants. Computers in Biology and Medicine, 119, 103688.*

- *Joshi, A., & Singh, H. (2021). Comparative evaluation of cloud-based ASR engines for Hindi healthcare dialogues. Healthcare Informatics Research, 27(3), 221–229.*
- *Kaur, H., & Khosla, R. (2018). Handling phonetic confusions in multilingual speech recognition. Speech Communication, 98, 25–34.*
- *Kulkarni, P., & Rao, S. (2022). User-adaptive feedback mechanisms in medical voicebots: A usability study. Journal of Medical Internet Research, 24(7), e34512.*
- *Kumar, R., & Pant, D. (2020). Building reliable ASR datasets for low-resource languages: The case of Marathi. Language Resources and Evaluation, 54(2), 317–332.*
- *Mishra, S., & Tripathi, A. (2019). End-to-end deep learning architectures for speech recognition in Indian languages. Neural Computing and Applications, 31(12), 8765–8778.*
- *Mukherjee, S., & Chatterjee, P. (2021). Evaluating semantic error rates in chatbot interactions: Healthcare domain. Artificial Intelligence in Medicine, 116, 102078.*
- *Naik, L., & Desai, T. (2022). Impact of rural noise conditions on speech-to-text accuracy in Kannada. International Journal of Speech Technology, 25(1), 123–136.*
- *Patel, R., & Shah, M. (2018). IndicTTS and ASR consortium: Resources for Indian language speech research. Proceedings of the Workshop on NLP for Similar Languages, Varieties and Dialects, 45–52.*
- *Reddy, D., & Singh, Y. (2020). Speaker-adaptive training for clinical ASR: A case study in Hindi. IEEE Journal of Biomedical and Health Informatics, 24(9), 2504–2513.*
- *Sharma, P., & Vernekar, A. (2021). Dialogue management strategies for voice-based medical chatbots. Journal of Healthcare Engineering, 2021, 8893457.*
- *Singh, N., & Kumar, V. (2019). Phoneme-level error analysis in speech recognition for Indian accented English. Speech, Language and Hearing, 22(2), 115–123.*
- *Thomas, L., & Menon, S. (2022). Real-world evaluation of voice-enabled telemedicine interfaces in rural India. Telemedicine and e-Health, 28(4), 487–495.*
- *Verma, J., & Rao, K. (2020). Domain-specific language models for improved ASR in healthcare chatbots. Computer Speech & Language, 61, 101059.*

ISSN: 2321-2853
ESTD.2013
IJRSML