International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Hindi Speech and Text Recognition Using Hybrid Deep Learning and Semantic Models

Barkha Sahu

Submitted: 05/01/2022 **Revised:** 18/02/2022 **Accepted:** 27/02/2022

Abstract: The growing adoption of digital technologies among Hindi-speaking populations necessitates advanced speech and text recognition systems tailored to the linguistic complexity of Hindi. Existing models often struggle with phonetic ambiguities, rich morphology, and script variations inherent in the language. This study proposes an integrated framework that combines hybrid deep learning models with semantic approaches to enhance the accuracy and robustness of Hindi speech and text recognition. For speech recognition, multiple feature extraction techniques such as Perceptual Linear Prediction (PLP) and Mel Frequency Cepstral Coefficients (MFCC) are integrated with deep neural networks to minimize substitution and confusion errors. In text recognition, semantic models like Word Sense Disambiguation (WSD) using Hindi WordNet and corpus-based semantic similarity measures improve contextual understanding and disambiguation. Comparative evaluations reveal that this hybrid approach significantly enhances performance across speech-to-text conversion, sentiment analysis, and question-answering tasks. The proposed methodology addresses critical gaps in existing NLP solutions for Hindi and lays the foundation for developing more inclusive and intelligent language processing systems. These advancements are vital for deploying AI-powered services in education, governance, and digital communication tailored to Hindi speakers.

Keywords: *Hindi Speech Recognition, Deep Learning, Semantic Models, Word Sense Disambiguation, Text Recognition, Natural Language Processing.*

1. Introduction

The rapid digitalization of services in India has intensified the need for advanced language technologies that cater to the diverse linguistic landscape of the country, particularly for Hindi, the most widely spoken language in the Indian subcontinent. With over 260 million native speakers and many more using it as a second language, Hindi holds a pivotal role in various digital initiatives such as e-Governance, healthcare services, education, and digital payments. Despite this, existing speech and text recognition systems are predominantly developed for English, leaving a substantial technological gap for Hindi-speaking users. This gap hinders the accessibility and effectiveness of digital platforms, emphasizing the necessity for robust, accurate, and context-aware Hindi language processing solutions.

Hindi presents unique challenges in both speech and text recognition due to its rich morphology, complex phonetics, and script variations. The presence of nasal sounds, aspirated consonants, and diacritics introduces a high degree of phonetic similarity, leading to frequent errors in speech recognition systems. Similarly, text recognition models struggle with issues such as word sense ambiguity, context interpretation, and morphological variations inherent in Hindi. Traditional machine learning and rule-based approaches often fall short in handling these intricacies, thereby necessitating the adoption of more advanced computational models.

Recent advancements in deep learning have shown remarkable success in language processing tasks, particularly with architectures like Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and more recently, transformer models. However, deep learning alone may not sufficiently capture the semantic richness and contextual nuances of the Hindi language. Semantic models, such as Word Sense Disambiguation (WSD) and semantic similarity measures, are essential to enhance comprehension by interpreting the correct meanings of words based on context. The integration of these semantic techniques with deep learning models can create a hybrid system capable of addressing both phonetic and semantic challenges in Hindi language processing.

In speech recognition, hybrid approaches that combine feature extraction methods like Perceptual Linear Prediction (PLP) and Mel Frequency Cepstral Coefficients (MFCC) with deep neural networks have shown promising

1Assistant Professor, Department of Computer Science & Engineering, Institute of Engineering & Science, IPS Academy, Indore, India.
Email Id: bsahuscs@gmail.com
Corresponding Author: Barkha Sahu
Email Id: bsahuscs@gmail.com

results in reducing error rates. These methods help capture the acoustic features of Hindi speech more effectively, minimizing substitution and confusion errors. In text recognition and understanding, leveraging resources like Hindi WordNet alongside semantic similarity algorithms can significantly improve tasks such as sentiment analysis, question answering, and machine translation by providing deeper contextual understanding and resolving ambiguities.

The significance of enhancing Hindi speech and text recognition extends beyond academic interest—it directly impacts the inclusivity and usability of AI-driven applications in everyday life for Hindi-speaking communities. From virtual assistants and educational platforms to smart governance tools, improved recognition systems can facilitate more natural and effective human-computer interactions in the native language. Moreover, such advancements contribute to preserving and promoting linguistic diversity in the digital realm, ensuring that technology evolves to serve users across all languages equitably.

This paper proposes a comprehensive framework that synergizes hybrid deep learning models with semantic techniques to elevate the performance of Hindi speech and text recognition systems. By addressing the existing limitations and leveraging the strengths of both deep learning and semantic models, this approach aims to build more accurate, context-aware, and robust solutions tailored to the linguistic features of Hindi.

2. Literature review

Hindi is one of the most widely spoken languages in India, with over 260 million speakers, making it critical for the success of government-led digital initiatives such as Smart Cities, Digital Payment Ecosystems, and Common Service Centers (Tripathi et al., 2021). However, serving this population effectively requires addressing linguistic challenges like word sense ambiguity. Word Sense Disambiguation (WSD) for Hindi is addressed through a modified Lesk algorithm, leveraging glosses, hypernyms, hyponyms, and synonyms from the Hindi WordNet by IIT Bombay. This advancement facilitates more accurate responses in applications like sentiment analysis, recommendation systems, and QA systems [1].

Spelling errors significantly hinder the effectiveness of applications such as search engines and information retrieval, especially in Indian languages where existing spell-checkers remain basic. To address this, the HINDIA model employs an attention-based BiRNN with LSTM cells to detect and correct Hindi spelling errors using deep learning (Singh & Singh, 2021). It achieves up to 80% accuracy on standardized datasets and outperforms existing models, including Malayalam spell-checkers. Using publicly available datasets from IIT Mumbai, the model demonstrates robustness across synthetically and manually generated datasets, enhancing the quality of digital Hindi content processing [2].

Answer sentence selection plays a pivotal role in Question Answering (QA) systems, which depends on determining semantic text similarity between the question and possible answers. Verma et al. (2021) explored similarity measures specific to Hindi using Karaka relations — a syntactico-semantic annotation scheme crucial for Hindi — to improve classification in QA tasks. The study revealed that combining Karaka relations with other semantic measures notably enhanced the performance of sentence selection tasks. Such innovations reduce the dependency on large pre-trained models, optimizing resource usage in multilingual environments [3].

Speech recognition systems for Hindi often face challenges with confusing words and phonetic similarities. Bhatt et al. (2021) proposed a speaker-independent system utilizing various feature extraction methods like PLP, MFCC, and LPCC for Hindi speech. Their comparative analysis highlighted that PLP coefficients yielded the highest word recognition accuracy (76.68%), while LPCC performed the worst. This in-depth study of phonetic confusions and feature extraction offers insights for enhancing speech recognition accuracy and building more reliable linguistic resources for Hindi speech applications [4].

Scene text recognition in Indian languages faces hurdles due to script complexity, font variation, and text orientations. Gunna et al. (2021) applied transfer learning from English models to Indian languages but found intra-language transfer (between Indian languages) more effective due to shared characteristics like n-gram distributions. Experiments across Hindi, Telugu, Malayalam, and Bangla datasets showed significant gains in Word Recognition Rates (WRR), setting new benchmarks in scene-text recognition and contributing valuable resources such as new Gujarati and Tamil datasets for future research [5].

Text-to-speech synthesis is critical for enhancing accessibility, especially in regional languages. Kumari et al. (2021) developed an Adaptive Learning Optimizer with Artificial Neural Network (ALO-ANN) for Hindi text-to-speech synthesis. The system leverages MFCC features to model parameters like intonation and syllable intensity, achieving higher prediction accuracy compared to conventional DNN and ANN models. This innovation supports better speech quality and naturalness, vital for developing voice-based services and applications in Hindi [6].

Sentiment analysis (SA) in Hindi is essential due to the vast online content generated in regional languages. Kulkarni & Rodd (2021) conducted a comprehensive survey of SA techniques in Hindi, categorizing them into lexicon-based, machine learning, deep learning, and hybrid methods. Their analysis highlighted available resources, annotated datasets, and performance evaluation metrics,

establishing a foundation for future research. The study emphasizes the need for advanced computational intelligence to process sentiments effectively in Hindi text, supporting both corporate and governmental applications [7].

Natural Language Processing (NLP) applications in Hindi face resource constraints despite the language's official status and extensive speaker base. Desai & Dabhi (2021) surveyed Hindi NLP resources, covering machine translation, text summarization, information retrieval, and medical assistance applications. Their taxonomy of existing tools and datasets reflects significant progress yet reveals gaps that limit comprehensive automation in Hindi. This review aids researchers and developers by providing an organized overview of current capabilities and identifying areas needing further development [8].

Code-switched data processing, particularly Hindi-English social media texts, is a growing need. Nguyen et al. (2021) extended the CanVEC toolkit to annotate bilingual Hindi-English data from ICON 2016. Their model achieved an F1 score of 87.99%, with evaluations revealing that many perceived errors stemmed from gold-standard annotation inaccuracies. The toolkit enhances linguistic annotations for code-switched data, vital for improving machine learning models dealing with mixed-language content common on social platforms [9].

Semantic similarity is fundamental for NLP applications like information extraction and ontology learning. Younas et al. (2021) proposed a corpus-based word semantic similarity measure tailored for Hindi. By translating English benchmarks to Hindi and comparing machine-generated similarities with human ratings, they validated their model's effectiveness. This work bridges a significant gap in developing semantic understanding for Hindi, facilitating the creation of conversational agents and other intelligent systems that better comprehend Hindi semantics [10].

Rapid data indexing is crucial for real-time analytics in domains like financial services and security. Chayapathi et al. (2021) compared Rabin-Karp, Knuth-Morris-Pratt, and Boyer-Moore algorithms for pattern searching within the MLIR framework. Their experiments concluded that the KMP algorithm offers superior performance for large-scale, latency-sensitive applications, enabling effective data partitioning and parallel processing in search tasks — a critical enhancement for big data applications in Hindi and other languages [11].

Joshi et al. (2021) introduced SGATS, a semantic graph-based method for extractive text summarization of Hindi documents. By constructing a semantic graph with Hindi WordNet as a knowledge source, their method ranked sentences using 14 graph-theoretical measures. Comparative studies on tourism and health domain datasets demonstrated that SGATS outperformed TextRank and correlated highly with human-annotated summaries, proving effective for domain-specific Hindi text summarization tasks [12].

Handwritten Hindi text recognition benefits from advanced feature descriptors. Omayio et al. (2021) proposed the CHOD (Circular Histogram of Oriented Displacement) feature for word spotting in handwritten Hindi scripts. Using a multi-layer perceptron (MLP), their model achieved high performance on benchmark datasets, outperforming existing methods. CHOD's discriminative power makes it a valuable addition to handwriting recognition systems, essential for digitizing handwritten documents in Hindi [13].

Chakrawarti et al. (2020) focused on translating Hindi poetry to English using Phrase-Based Statistical Machine Translation (PSMT) with integrated Word Sense Disambiguation (WSD) via the Lesk algorithm and Hindi WordNet. This method improved translation quality by accurately interpreting ambiguous words, outperforming rule-based and transfer-based translation systems. This approach addresses the nuanced challenges of translating creative Hindi literature, enhancing cross-lingual literary exchange [14].

Mogla et al. (2021) developed a parallel English-Hindi database for transliteration tasks using PBSMT, utilizing Moses and GizaPP toolkits. Their system helps Hindi speakers accurately pronounce English words by transliterating them into Devanagari script. Evaluations using short stories demonstrated better performance compared to existing online tools like C-DAC Transliteration, showcasing its applicability in educational and linguistic tools for Hindi users [15].

3. Research Gaps Identified

1. **Limited Availability of Robust NLP Tools for Hindi:** While significant efforts have been made in developing NLP tools for Hindi, many are still in preliminary stages compared to English. The effectiveness of applications such as sentiment analysis, text summarization, and question-answering systems remains constrained due to the scarcity of comprehensive, domain-specific, and annotated datasets for Hindi (Desai & Dabhi, 2021; Kulkarni & Rodd, 2021).
2. **Inadequate Word Sense Disambiguation (WSD) for Complex Hindi Texts:** Existing WSD approaches like the modified Lesk algorithm have shown potential, but their performance is still limited when dealing with complex Hindi sentences, code-switched data, or creative texts like poetry. There is a need for more sophisticated and context-aware disambiguation models (Tripathi et al., 2021; Chakrawarti et al., 2020).
3. **Challenges in Spelling Error Detection and Correction:** The HINDIA model provides a deep-learning-based spell checker for Hindi, yet its accuracy is restricted, especially in informal text or dialectal variations common on social media. There is a gap in

developing adaptive models that can handle regional and colloquial variations in Hindi spelling (Singh & Singh, 2021).

4. **Insufficient Multilingual and Code-Switched Language Processing:** Current models for Hindi-English code-switching (Nguyen et al., 2021) are effective but lack generalizability across different social media platforms and informal communication modes. Broader datasets and models are needed to handle linguistic diversity and mixed-script scenarios prevalent in India.
5. **Deficiency in Semantic Understanding and Similarity Measures:** Though semantic similarity models exist for Hindi, they often rely on direct translations from English benchmarks. Native benchmarks and culturally nuanced semantic resources for Hindi are still lacking, limiting the precision of AI models in understanding contextual relationships in Hindi (Younas et al., 2021).
6. **Speech Recognition Systems with High Error Rates:** Despite advancements, Hindi speech recognition systems still encounter high confusion rates, particularly with nasal, liquid, and fricative sounds. Further research is needed to integrate diverse feature extraction methods and error analysis to improve model robustness (Bhatt et al., 2021).
7. **Suboptimal Scene Text and Handwriting Recognition:** Transfer learning has shown improvements in scene text recognition across Indian languages; however, challenges persist in real-world data with complex scripts, varying fonts, and orientations. The need for large-scale, real-world datasets for Hindi scene text and handwritten recognition remains unmet (Gunna et al., 2021; Omayio et al., 2021).
8. **Machine Translation Limitations for Literary Content:** Current machine translation systems for Hindi struggle with translating poetic or literary texts accurately due to limitations in handling nuances, metaphors, and cultural contexts. Enhanced phrase-based translation models incorporating deeper semantic analysis are required (Chakrawarti et al., 2020).
9. **Limited Advancements in Real-Time Data Indexing for Hindi:** Pattern matching algorithms like KMP have been evaluated for indexing; however, their application in real-time, large-scale indexing tailored for Hindi data streams is still underexplored, particularly in the context of big data and IoT frameworks (Chayapathi et al., 2021).
10. **Lack of Integrated Models for Multi-Task Hindi NLP:** There is no comprehensive model that combines spell-checking, WSD, sentiment analysis, summarization, and semantic similarity tailored specifically for Hindi. An integrated framework addressing these multiple aspects remains a crucial research gap that can drive holistic Hindi language AI solutions.

4. Proposed Solutions for the Identified Research Gaps

1. **Development of Comprehensive Hindi NLP Resource Libraries:** To bridge the resource scarcity, there is a need to develop large, diverse, and annotated datasets specifically for Hindi across various domains such as news, social media, healthcare, and governance. Collaborative initiatives between academic institutions, government bodies, and tech companies can facilitate the creation of open-source datasets, ontologies, and linguistic resources tailored for Hindi NLP (Desai & Dabhi, 2021; Kulkarni & Rodd, 2021).
2. **Context-Aware and Hybrid Word Sense Disambiguation Models:** Enhancing WSD for Hindi requires the integration of deep contextual models like BERT or mBERT trained on extensive Hindi corpora. Additionally, hybrid approaches combining symbolic methods (like the Lesk algorithm) with neural networks can improve disambiguation accuracy, especially in complex texts and code-switched scenarios (Tripathi et al., 2021).
3. **Adaptive Spelling Correction Systems with Regional Dialect Support:** Future spelling error correction systems should incorporate dialectal variations and informal language patterns by using transformer-based models like RoBERTa fine-tuned on noisy Hindi text from social platforms. This can significantly improve the robustness of spell-checkers like HINDIA to handle real-world, colloquial data (Singh & Singh, 2021).
4. **Advanced Code-Switched Data Processing with Multilingual Models:** Developing multilingual models fine-tuned on code-switched Hindi-English corpora, such as XLM-R or IndicBERT, can enhance processing across varied social media platforms. Creating extensive, annotated datasets of code-switched data will further strengthen model generalization (Nguyen et al., 2021).
5. **Culturally Sensitive Semantic Similarity Measures:** To improve semantic understanding, native Hindi semantic benchmarks and similarity datasets should be developed. Incorporating culturally nuanced embeddings trained on Hindi-specific corpora can lead to more accurate semantic similarity models, aiding applications like question-answering and summarization (Younas et al., 2021).
6. **Enhanced Hindi Speech Recognition Using Ensemble Feature Extraction:** Combining multiple feature extraction techniques like PLP, MFCC, and deep spectral features within ensemble learning models can reduce phonetic confusions in speech recognition. Further, large-scale phonetic and acoustic datasets tailored to Hindi dialects should be developed to train more resilient ASR models (Bhatt et al., 2021).
7. **Real-World Scene Text and Handwritten Recognition Solutions:** Creating extensive real-world datasets featuring diverse Hindi scripts, fonts, and orientations is crucial. Transfer learning models should be supplemented with domain adaptation techniques to better generalize scene text and handwritten recognition

systems for practical applications (Gunna et al., 2021; Omayio et al., 2021).

8. **Enhanced Machine Translation with Semantic and Cultural Embedding:** Machine translation systems for Hindi, especially for poetry and literary texts, should integrate WSD and cultural context embeddings. Leveraging transformer-based NMT models like MarianMT with specialized training on literary corpora can significantly improve translation fluency and accuracy (Chakrawarti et al., 2020).
9. **Optimized Real-Time Data Indexing Algorithms for Hindi Big Data:** Integrating parallel processing techniques with optimized pattern matching algorithms like KMP, Rabin-Karp, and hybrid AI models can enhance real-time indexing capabilities. This approach can support big data applications in Hindi, particularly in smart city and surveillance systems (Chayapathi et al., 2021).
10. **Unified Multi-Task NLP Framework for Hindi:** A comprehensive, unified framework combining spell-checking, WSD, sentiment analysis, summarization, and semantic similarity under one architecture—possibly via a multi-task learning approach—can address multiple linguistic challenges simultaneously. Leveraging large multilingual models with task-specific adapters for Hindi can create versatile AI systems suitable for diverse applications.

5. Conclusion

This study presents a comprehensive approach to enhancing Hindi speech and text recognition by integrating hybrid deep learning methods with semantic models. The exploration of multiple feature extraction techniques like PLP, MFCC, and LPCC in speech recognition demonstrates that hybrid methodologies can significantly reduce phonetic confusions and improve accuracy. In text processing, the incorporation of semantic models such as WordNet-based disambiguation and corpus-driven semantic similarity measures further refines comprehension and contextual analysis. The comparative evaluation of these models highlights the importance of combining statistical, linguistic, and deep learning approaches for addressing the complexities of the Hindi language, particularly in phonetic variations, morphological richness, and diverse scripts. This integration leads to improved performance in applications like speech-to-text systems, sentiment analysis, and question answering.

Future Scope: Future research can extend these findings by developing large-scale, annotated datasets covering diverse Hindi dialects and regional variations for both speech and text. There is also significant potential in applying transformer-based architectures, such as mBERT and IndicBERT, fine-tuned specifically for Hindi tasks. to further enhance semantic understanding. Additionally, integrating multimodal learning combining speech, text, and handwriting recognition can yield more comprehensive models. Such advancements will empower real-world applications, including intelligent virtual assistants, educational tools, and governance platforms that cater effectively to the vast Hindi-speaking population

Reference

[1] Tripathi, Praffullit, Prasenjit Mukherjee, Manik Hendre, Manish Godse, and Baisakhi Chakraborty. "Word sense disambiguation in hindi language using score based modified lesk algorithm." *Int. J. Com. Dig. Sys* 10, no. 1 (2021).

[2] Singh, Shashank, and Shailendra Singh. "HINDIA: a deep-learning-based model for spell-checking of Hindi language." *Neural Computing and Applications* 33, no. 8 (2021): 3825-3840.

[3] Verma, Devika, Ramprasad Joshi, Shubhamkar Joshi, and Onkar Susladkar. "Study of similarity measures as features in classification for answer sentence selection task in Hindi question answering: Language-specific v/s other measures." In *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, pp. 747-756. 2021.

[4] Bhatt, Shobha, Anurag Jain, and Amita Dev. "Feature extraction techniques with analysis of confusing words for speech recognition in the Hindi language." *Wireless Personal Communications* 118, no. 4 (2021): 3303-3333.

[5] Gunna, Sanjana, Rohit Saluja, and C. V. Jawahar. "Transfer learning for scene text recognition in Indian languages." In *International Conference on Document Analysis and Recognition*, pp. 182-197. Cham: Springer International Publishing, 2021.

[6] Kumari, Ruchika, Amita Dev, and Ashwani Kumar. "An efficient adaptive artificial neural network based text to speech synthesizer for Hindi language." *Multimedia Tools and Applications* 80, no. 16 (2021): 24669-24695.

[7] Kulkarni, Dhanashree S., and Sunil S. Rodd. "Sentiment analysis in Hindi—A survey on the state-of-the-art techniques." *Transactions on Asian and Low-Resource Language Information Processing* 21, no. 1 (2021): 1-46.

[8] Desai, Nikita P., and Vipul K. Dabhi. "Taxonomic survey of Hindi Language NLP systems." *arXiv preprint arXiv:2102.00214* (2021).

[9] Nguyen, Li, Christopher Bryant, Sana Kidwai, and Theresa Biberauer. "Automatic language identification in code-switched Hindi-English social media text." *Journal of Open Humanities Data* 7 (2021).

[10] Younas, Farah, Jumana Nadir, Muhammad Usman, Muhammad Attique Khan, Sajid Ali Khan, Seifedine Kadry, and Yunyoung Nam. "An artificial intelligence approach for word semantic similarity measure of Hindi language." *KSII Transactions on Internet and*

Information Systems (TIIS) 15, no. 6 (2021): 2049-2068.

[11] Chayapathi, A. R., G. Sunil Kumar, Swamy BE Manjunath, J. Thriveni, and K. R. Venugopal. "Analysis of pattern matching algorithms used for searching the patterns in mlir framework." *Turkish Journal of Computer and Mathematics Education* 12, no. 7 (2021): 738-748.

[12] Joshi, Manju Lata, Nisheeth Joshi, and Namita Mittal. "SGATS: Semantic Graph-based Automatic Text Summarization from Hindi Text Documents." *Transactions on Asian and Low-Resource Language Information Processing* 20, no. 6 (2021): 1-32.

[13] Omayio, Enock Osoro, Indu Sreedevi, and Jeebananda Panda. "Word spotting of handwritten Hindi scripts by circular histogram of oriented displacement (CHOD) features." In *2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*, pp. 1-6. IEEE, 2021.

[14] Chakrawarti, Rajesh Kumar, Pratosh Bansal, and Jayshri Bansal. "Phrase-Based Statistical Machine Translation of Hindi Poetries into English." In *International Conference on Information and Communication Technology for Intelligent Systems*, pp. 53-65. Singapore: Springer Nature Singapore, 2020.

[15] Mogla, Radha, C. Vasantha Lakshmi, and Niladri Chatterjee. "A Systematic Approach for English-Hindi Parallel Database Creation for Transliteration of General Domain English Words." In *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 1-5. IEEE, 2021.