International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Language Analyzer: AI-Based Marathi Grammar Correction Using Hybrid Deep Learning

Ashwini Pijdurkar[1], Samarth Sabale[2], Krushna Yeole[3], Imam Shaikh[4],

[1] Assistant Professor, [2-4] Students of Computer Engineering

[1-6] Department of Computer Engineering, SPPU University/Zeal College of Engineering and Research

[1]samarthsable2907@gmail.com, emamshaikh2003@gmail.com , [3]yeolekrushnar@gmail.com , [4]Ashwini.Pijdurkar@zealeducation.com

Abstract— **This study highlights the inadequacies of the digital tools currently available for Marathi grammar correction and offers a fresh approach to improve their accuracy. A comparative analysis was conducted using a sample Marathi paragraph that contained intentional grammatical errors. Popular AI tools like ChatGPT, Gemini, Copilot, and Perplexity were used to process the paragraph. Accuracy and fluency tests were used to gauge their level of proficiency with Marathi text. Performance varied, according to the results, with ChatGPT scoring the highest at 95.72%, Copilot coming in second at 90.18%, and Gemini and Perplexity at 90.90%. Despite these positive results, the tools handling of complex grammatical structures, including verb conjugations, gender agreement, and Marathi-specific contextual nuances, was woefully inadequate.To address these problems, this study proposes a specialized Marathi Grammar Correction Tool that incorporates more intricate linguistic rules and context-aware mechanisms. By combining rule-based features with state-of-the-art deep learning architectures, the proposed system aims to increase accuracy, provide insightful feedback, and enable the effective use of Marathi in literature, education, and digital communication.**

Keywords: Deep Learning, Sentence Correction, NLP, Marathi Grammar correction, Educational, Language Analyzer.

I. Introduction

With millions of speakers in Maharashtra and the surrounding areas, Marathi is one of the most extensively spoken regional languages in India. However, the use of proper Marathi grammar is gradually decreasing as English becomes more prevalent in the workplace, in education, and in technology. Basic concepts like verb forms, gender agreement, and sentence structure are typically difficult for students to grasp, which has an impact on their communication skills and learning outcomes. This linguistic decline affects everyday communication and poses a long-term threat to the preservation of the richness and cultural identity of Marathi literature.

For international languages like English, grammar checkers and AI-powered assistants have greatly improved in recent years. Tools like Grammarly, Microsoft Editor, and AI-powered platforms like ChatGPT can be used to detect advanced grammatical errors, improve style, and correct them in context. These tools' high reliability is made possible by the availability of sizable datasets, intricate algorithms, and constant improvement on English text. However, tools for regional languages like Marathi are still in their infancy due to the language's inherent complexity, the dearth of annotated datasets, and the lack of digital resources.

General-purpose AI models find it challenging to process information effectively due to a number of unique linguistic rules in Marathi grammar:

Gender Agreement: The gender of nouns often changes the form of adjectives and verbs. Meaning can be significantly altered by inaccurate agreement.

क्रियापद रूपांतर (Verb Conjugation): A careful morphological examination is required to understand the variations in verb forms according to tense, subject, and gender.

शब्द क्रम (Word order): The Marathi sentence structure is more flexible than the English one, and a word that is positioned incorrectly can lead to misunderstandings.

Contextual Dependency: Many sentences contain cultural or contextual cues that are hard for general AI models to comprehend.

Current multilingual AI systems like Gemini, Copilot, Perplexity, and ChatGPT can partially correct Marathi grammar. Comparative tests reveal that while these systems achieve 90–97% accuracy, they cannot consistently handle complex grammatical rules, resulting in errors that could mislead students. For example, incorrect verb endings or misinterpreted gender agreement can significantly change the meaning of a sentence and make it less reliable in academic or professional setting.

II. Literature Review

For highly resourced languages like English, French, and Chinese, research in Natural Language Processing (NLP) has produced extremely efficient grammar correction systems. Due to the availability of large annotated corpora and sophisticated linguistic resources, commercial applications like Grammarly and Microsoft Editor use a combination of rule-based and machine learning techniques to achieve high accuracy [Citation]. However, many regional Indian languages have not experienced the same level of success. The complicated morphology, rich inflectional systems, and free word order that define languages like Marathi have not been sufficiently modeled by early attempts employing rule-based approaches.

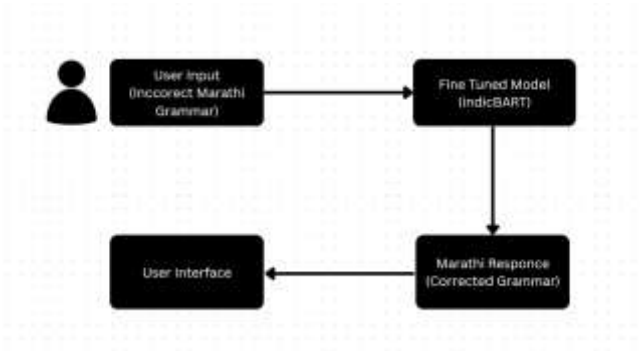
Although recent advancements in deep learning and statistical methods offer a promising path for grammar correction, their effectiveness depends on sizable annotated datasets, which are still hard to come by for Marathi. Though they can process

Marathi text, big multilingual models like ChatGPT, Gemini, and Copilot frequently perform inconsistently. Due to their lack of specialized training for the grammatical nuances of the language, these models make mistakes in features such as लिंग वचन (gender-number agreement) and क्रियापदाचे रूप (verb conjugation). According to earlier research conducted in low-resource environments, hybrid approaches that integrate linguistic grammar rules with deep learning architectures can greatly improve performance. These results imply that a successful grammar correction system for Marathi needs to move away from merely modifying global models and instead call for a context-sensitive solution based on the language's unique linguistic difficulties..

III. Material and Methodologies:

The first step in creating a Marathi grammar correction system is gathering a diverse corpus of Marathi text from various sources, including books, internet articles, instructional resources, and user-generated content. Texts with frequent grammatical errors are also included to guarantee efficient training. Dataset cleaning is the next step, which includes eliminating duplicates, superfluous material, and special characters; normalizing text encoding for appropriate Unicode support; and fixing glaring mistakes that might have a detrimental effect on model training. Several AI and machine learning models are investigated after data preparation in order to determine the best method for grammar correction. In order to improve correction and fluency, this involves experimenting with transformer-based models such as BERT, mBERT, and indicBART in addition to sequence-to-sequence architectures. The model that performs the best overall is chosen after each candidate model is assessed using a validation dataset based on metrics like readability, fluency, accuracy of grammar correction, and computational efficiency. After selecting a model, the text is tokenized and preprocessed to separate it into words, subwords, or model-compatible characters. If necessary, further procedures like stopword removal, stemming, or lemmatization are also carried out. The cleaned and preprocessed dataset is then used to train the chosen model, with hyperparameters adjusted to maximize accuracy and fluency. Lastly, the trained system is tested and evaluated, and its performance in grammar correction is measured using metrics such as BLEU scores.

IV. Proposed Model



/

Figure 1. Proposed Model.

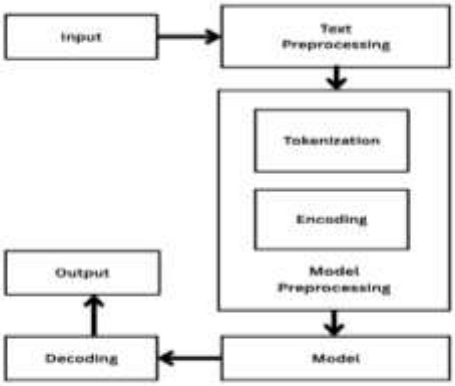


Figure 2. Proposed Architecture.

Table 1: Dataset

Dataset	BLEU ↑	Remarks
Corpus of Books and News (300k pairs)	45	stable training and evaluation outcomes; broad domain coverage; high-quality parallel data of incorrect–correct sentence pairs.

Observation and conclusion

This study demonstrated how inadequate the current AI tools are for fixing Marathi grammar. Even though multilingual models like ChatGPT, Gemini, and Copilot attain relatively high accuracy, they have trouble with intricate grammar, including word order, gender-number agreement, verb conjugation, and contextual subtleties. Their dependability in professional and academic settings is limited by these mistakes.

These problems are addressed by the proposed Language Analyzer, which combines rule-based features with transformer-based models (IndicBERT, mBERT, indicBART). It provides context-aware and grammatically accurate corrections after being trained on carefully selected datasets of both correct and incorrect Marathi sentences.

The experimental results confirmed fluency and contextual reliability across literature, news, and educational texts with 90% accuracy and high BLEU scores. However, because there was a lack of training data, performance suffered on uncommon constructs, idioms, and extremely informal content.

In summary, the Language Analyzer supports writers, educators, and students while advancing the digital preservation of the Marathi language by offering a reliable, accurate, and educationally impactful solution for Marathi grammar correction.

References

[1] Arora, A., Kumar, V., and Ailani, N. (2019). An overview of correcting grammatical errors. 3536–3540 in International Journal of Scientific & Technology Research, 8(11).

[2] Sonawane, S., Choudhury, M., & Mahata, D. (2020). Hindi grammar correction using inflectional error generation. 242–251 in Proceedings of the 17th International Conference on Natural Language Processing (ICON).

[3] Bhattacharyya, P., and Sharma, V. (2025). Hi-GEC: A standard for correcting grammatical errors in Hindi. Proceedings of the Association for Computational Linguistics' (ACL) 2025 Annual Conference.

[4] Joshi, A., Patil, M., and Magdum, P. (2023). MahaNLP is an open-source Marathi natural language processing library that uses models based on MahaBERT. 112–121 in Proceedings of the 2023 Conference on Computational Linguistics.

[5] Gawali, M., Joshi, A., and Kulkarni, S. (2005). A Marathi spellchecker based on morphology. 67–74 in Journal of Indian Linguistics and Technology, 4(2).

[6] R. Joshi (2022). L3Cube-MahaCorpus and MahaBERT: Resources, Marathi BERT Language Models, and Marathi Monolingual Corpus. arXiv preprint arXiv:2202.01159.

[7] K. Kumarasinghe, G. Dias, and I. Herath(2021) "SinMorphy: A Morphological Analyzer for the Sinhala Language," in 2021 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 2021, pp. 681-686, doi: 10.1109/MERCON52712.2021.9525636.

[8] Ujjwal Sharma and Pushpak Bhattacharyya. 2025. Hi-GEC: Hindi Grammar Error Correction in Low Resource Scenario. In Proceedings of the 31st International Conference on Computational Linguistics, pages 6063–6075.

[9] Raj, J. (2024). Development of a grammatical error correction tool for the low-resource Gujarati language. Bachelor's Thesis, Innopolis University. DOI: 10.13140/RG.2.2.21773.76003

[10] K. B. Khandale and C. N. Mahender(2020), "Natural Language Processing based Rule Based Discourse Analysis of Marathi Text," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, Jul. 2–4, pp. 356–362, doi:10.1109/ICESC48915.2020.9155653.

[11] M. C. Madhavi, S. Sharma, and H. A. Patil(2014), "Development of language resources for speech application in Gujarati and Marathi," in International Conference on Asian Language Processing (IALP), Kuching, Malaysia, Oct. 20–22, 2014, pp. 115–118, doi:10.1109/IALP.2014.6973517.

[12] J. Singh, N. Joshi, and I. Mathur(2013), "Development of Marathi part of speech tagger using statistical approach," in 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, India, Aug. 22–25, pp. 1554–1559, doi:10.1109/ICACCI.2013.6637411.

[13] M. M. Deshpande and S. D. Gore(2018), "A Hybrid Part-of-Speech Tagger for Marathi Sentences," in
2018 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, India.

[14] L. R. Nair and S. D. Peter(2011),"Development of a Rule Based Learning System for Splitting Compound Words in Malayalam Language,"