An Efficient Transformer Based Framework for Marathi Abstractive Summarization

Ms. Sonali Waje[1], Dr. C. M. Raut[2]
[1]*M.E. (Computer Engg.) Datta Meghe College of Engineering, Navi Mumbai.*
[2]*Professor, Datta Meghe College of Engineering, Navi Mumbai.*

***Abstract*—Automatic Text Summarization (ATS) is an important task in the field of Natural Language Processing (NLP) that focuses on generating concise and meaningful summaries from large volumes of textual data. With the rapid growth of digital content available in multiple languages, ATS has gained significant attention from researchers in recent years. The ability to automatically produce accurate summaries has numerous practical applications, including news summarization, research article summarization, financial report analysis, and information retrieval. Although substantial progress has been made in English text summarization due to the availability of large-scale datasets and pretrained language models, Marathi and other Indian regional languages have received comparatively less research attention. The Marathi language presents unique challenges for NLP because of its rich morphological structure and complex syntactic characteristics. This study presents a comparative analysis of different transformer-based models for Automatic Text Summarization in Marathi, aiming to evaluate their effectiveness in generating coherent and contextually relevant summaries.**

***Index Terms*—Abstractive Text summarization (ATS), Transformer Models, IndicBART, mBART, mT5, Marathi language.**

I. INTRODUCTION

The rapid growth of digital information has created a strong need for efficient techniques to process and understand large volumes of textual data. Automated text summarization has therefore become an important research area in the field of Natural Language Processing (NLP). The main objective of text summarization is to generate a concise and meaningful summary from lengthy documents while preserving the important information. Such systems help users save time and improve information retrieval efficiency in domains such as news analysis, education, healthcare, and digital media.

Text summarization techniques are generally categorized into extractive, abstractive, and hybrid approaches. Extractive summarization identifies and selects the most important sentences directly from the original document without modifying their structure. In contrast, abstractive summarization generates summaries in a human-like manner by understanding the context of the text and producing new sentences with improved coherence and readability. Hybrid methods combine the advantages of both extractive and abstractive techniques to produce informative and fluent summaries. In Marathi language processing, earlier research focused mainly on extractive summarization methods. Traditional approaches such as TextRank and Latent Semantic Analysis (LSA) were widely used to identify important sentences from Marathi documents. Although these methods produced acceptable results, they often lacked semantic understanding and generated summaries that were less natural and less coherent. Since extractive methods only select existing sentences from the source text, they are unable to effectively paraphrase or restructure information.

Marathi is one of the major Indo-Aryan languages spoken by millions of people in India. Despite its widespread usage, Marathi remains relatively underrepresented in NLP research compared to resource-rich languages such as English. Marathi possesses complex linguistic characteristics including rich morphology, flexible word order, and diverse grammatical structures, which make automatic summarization a challenging task. Conventional rule-based and statistical techniques are often insufficient to capture the semantic relationships and contextual meaning required for high-quality abstractive

summarization. Recent advancements in deep learning and transformer-based architectures have significantly improved the performance of various NLP applications. Pretrained models such as BERT, mBART, T5, and IndicBART have demonstrated remarkable capabilities in text generation and language understanding through transfer learning. Among these models, IndicBART is specifically designed for Indian languages and supports multilingual sequence-to-sequence text generation. Its architecture enables better contextual understanding and generation for Indic languages, making it suitable for Marathi abstractive text summarization.

This research focuses on improving Marathi abstractive text summarization using transformer-based models, particularly mBART and IndicBART. The proposed work utilizes Marathi datasets obtained from the L3Cube-MahaSum and XL-Sum corpora for training, validation, and testing purposes. The study fine-tunes the pretrained IndicBART model to generate concise, coherent, and informative summaries in Marathi. The performance of the generated summaries is evaluated using standard metrics such as ROUGE-1, ROUGE-2, ROUGE-L, and BLEU score. The research aims to analyze the effectiveness of transformer-based approaches in enhancing summary quality in terms of fluency, informativeness, and semantic accuracy.

The major contribution of this work is the comparative improvement and evaluation of mBART and IndicBART models for Marathi abstractive text summarization. By exploring advanced transformer architectures for low-resource Indian languages, this study contributes toward the development of more effective NLP solutions for Marathi and other Indic languages.

II. LITERATURE REVIEW

2.1 Review of concept and theories

Text summarization is an important task in Natural Language Processing (NLP) that aims to generate a concise and meaningful summary from a larger document while preserving its key information. With the increasing volume of digital textual information such as news articles, research papers, and government reports, automatic text summarization has become essential for efficient information management and knowledge extraction.

Text summarization techniques are broadly categorized into extractive summarization and abstractive summarization. Extractive summarization selects important sentences directly from the original document based on statistical or linguistic features. Early methods relied on techniques such as TF–IDF weighting, sentence position scoring, and similarity measures to identify relevant sentences [1]. These methods often used clustering techniques such as K-means to remove redundancy and improve summary diversity.

Later research introduced machine learning approaches for summarization. Algorithms such as Support Vector Machines, Naïve Bayes, and neural networks were applied to determine sentence importance using various linguistic and statistical features. However, these methods required extensive feature engineering and struggled to capture contextual relationships between sentences.

With the advancement of deep learning, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) were introduced for abstractive summarization tasks [2]. These models use encoder–decoder architectures with attention mechanisms to generate summaries that better capture the semantic meaning of the original text. Studies conducted on languages such as Urdu and Bengali demonstrated that attention-based neural architectures improve summarization quality compared to traditional methods [3].

In recent years, Transformer-based architectures have significantly improved the performance of text summarization systems. Transformer models rely on self-attention mechanisms that allow the model to understand relationships between words across long text sequences [12]. Popular transformer models include BERT, BART, T5, PEGASUS, mBART, and mT5, which have achieved state-of-the-art performance in various NLP tasks including summarization [18].

For low-resource languages such as Marathi, specialized transformer models have been developed to address linguistic challenges such as complex morphology and limited training datasets [5]. Research indicates that Indic-specific transformer models such as IndicBART perform better than general multilingual models because they are trained on large corpora of Indian languages [10].

Another important concept in summarization research is the evaluation of generated summaries. Traditional evaluation metrics such as ROUGE and BLEU measure lexical overlap between the generated summary and reference summary [15]. However, these metrics sometimes fail to capture semantic similarity. Therefore, newer evaluation methods such as BERTScore and semantic similarity metrics have been proposed to better evaluate summary quality and contextual relevance [22].

The availability of large datasets is also critical for training summarization models. Recent research introduced datasets such as MahaSUM and XL-Sum, which provide large collections of Marathi news articles paired with summaries for training and evaluation of abstractive summarization systems [4][11].

Thus, the evolution of summarization techniques from statistical methods to deep learning and transformer-based architectures has significantly improved summarization performance, particularly in multilingual and low-resource language scenarios.

III. PROPOSED SYSTEM

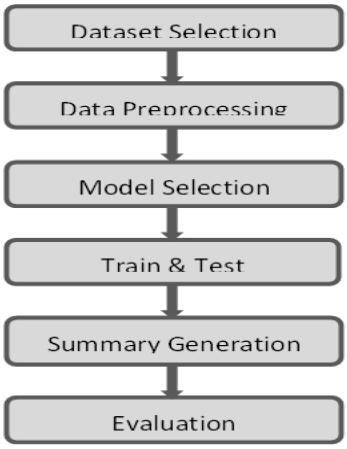

Fig 1: Proposed Methodology

Fig.1 depicts the proposed methodology used for developing an ATS system for Marathi language using different transformer models. It consists of steps like dataset selection, data preprocessing, model selection & fine tuning, summary generation and evaluation.

A. Dataset selection

BBC Research developed large-scale multilingual summarization dataset XL-Sum that represents information across 44 languages with Marathi among them. The system provides human-written abstract text summaries that stay connected to the original context. This study utilized the Marathi subset from the XL-Sum dataset for benchmark performance evaluation.

The BBC Marathi news article collection provides 5,855 records for Marathi language which contains the fields like ID (A unique identifier for each news article), URL (The link to the original BBC Marathi news article), Title (Headline of the article), Article (Full text of the news article), Summary (Manually generated abstractive summary). The dataset divides its contents into 3 sets as Training Set: 4,684 records, Validation Set: 732 records and Test Set: 439 records.

B. Data preprocessing

The method of converting raw data into clean and structured arrangement before feeding it into a machine learning model. It involves the techniques to improve model accuracy and efficiency like Cleaning, stop word Removal, Stemming & Lemmatization, Handling Encoding Issues, Text Normalization.

C. Model selection

The proposed system uses mT5, mBART, and IndicBART models for Marathi abstractive text summarization. These transformer-based models are selected because of their strong capability in multilingual text generation and contextual understanding. Among them, IndicBART is specifically designed for Indian languages, making it highly suitable for Marathi summarization tasks. The models are fine-tuned using XL-Sum dataset, and their performance is evaluated using ROUGE and BLEU scores.

D. Evaluation

Evaluation: The performance of the text summarization models is evaluated using the ROUGE metric, which is one of the most widely adopted evaluation methods for automatic text summarization. ROUGE measures the similarity between the generated summary and the reference summary by calculating the overlap of n-grams and the longest common subsequence. In this study, the summarization models are assessed using ROUGE-1, ROUGE-2, ROUGE-L, and BLEU scores to evaluate the quality and accuracy of the generated summaries.

1. ROUGE-1: This metric measures the degree of unigram (single-word) overlap between the generated summary and the reference summary. It provides an effective way to assess how well the generated summary captures the important words present in the reference summary. A higher ROUGE-1 score indicates greater content coverage and better summarization performance.
Equation (1) shows ROUGE-1 calculation formula.
The formula for ROUGE-1 Recall is:

$$\text{ROUGE-1} = \frac{\sum \text{Number of overlapping unigrams}}{\sum \text{Total unigrams in reference summary}} \quad (1)$$

Where:
Overlapping unigrams = common words between generated and reference summaries
Reference summary = actual human-written summary

2. ROUGE-2: This metric measures the overlap of bigrams (two-word sequences) between the generated summary and the reference summary. It evaluates how effectively the generated summary preserves the sequence of words found in the reference summary, providing insights into its coherence and fluency. A higher ROUGE-2 score indicates better contextual consistency and improved summary quality.
Equation (2) shows ROUGE-2 calculation formula.
The formula for ROUGE-2 Recall is:

$$\text{ROUGE-2} = \frac{\sum \text{Number of overlapping bigrams}}{\sum \text{Total bigrams in reference summary}} \quad (2)$$

Where:
Overlapping bigrams = matching two-word sequences between generated and reference summaries
Reference summary = original human-written summary

3. ROUGE-L: This metric measures the longest common subsequence (LCS) shared between the generated summary and the reference summary. It evaluates the ability of the generated summary to preserve the content order and overall structure of the reference summary. A higher ROUGE-L score indicates that the generated summary more closely follows the sequence and organization of the reference summary.
Equation (3) shows ROUGE-L calculation formula.
The formula for ROUGE-L Recall is:

$$\text{ROUGE-L} = \frac{\text{LCS(Generated Summary,Reference Summary)}}{\text{Length of Reference Summary}} \quad (3)$$

Where:
LCS = Longest Common Subsequence between generated and reference summaries
Reference Summary Length = total number of words in the reference summary

4. BLEU Score: The BLEU (Bilingual Evaluation Understudy) score evaluates the quality of the generated summary by measuring the precision of matching n-grams between the generated summary and the reference summary. A higher BLEU score indicates greater similarity and better agreement with the reference summary.
The BLEU score formula is:

$$\text{BLEU} = \text{BP} \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

Where:

- BP = Brevity Penalty
- p_n= Precision of n-grams
- w_n= Weight assigned to each n-gram
- N = Maximum n-gram length

BLEU score evaluates the fluency and accuracy of generated summary by comparing it with the reference summary. Higher BLEU values indicate better quality summaries.

IV. RESULTS

Table 1 presents the evaluation scores for all three models on the Marathi test set (500 records). mT5 and IndicBART are shown in both base and fine-tuned configurations. mBART-50 is included as a base-only reference.

Table 1

Model	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	METEOR	BERTScore F1
mT5 (Base)	37.42%	16.85%	33.17%	10.23%	28.64%	72.38%
mT5 (Finetuned)	42.14%	27.36%	38.72%	18.75%	44.70%	78.51%
IndicBART (Base)	22.76%	9.43%	20.18%	5.67%	18.92%	65.84%
IndicBART (Finetuned)	27.46%	15.26%	25.46%	11.77%	26.74%	74.23%
mBART-50 (Base only)	6.38%	0.75%	5.34%	0.18%	6.80%	62.18%

Base scores for mT5 and IndicBART are sourced from pre-fine-tuning evaluations. Finetuned scores are from the best model checkpoints selected by highest validation ROUGE-L. mBART-50 was evaluated on the base model only. Improvement After Fine-tuning Following table represents the difference between base model and fine-tuned models with different parameters.

Table 2

Metric	mT5 Base	mT5 Finetuned	Δ mT5	IndicBART Base	IndicBART Finetuned	Δ IndicBART
ROUGE-1	37.42%	42.14%	+4.72%	22.76%	27.46%	+4.70%
ROUGE-2	16.85%	27.36%	+10.51%	9.43%	15.26%	+5.83%
ROUGE-L	33.17%	38.72%	+5.55%	20.18%	25.46%	+5.28%
BLEU	10.23	18.75	+8.52	5.67	11.77	+6.10
METEOR	28.64%	44.70%	+16.06%	18.92%	26.74%	+7.82%
BERTScore F1	72.38%	78.51%	+6.13%	65.84%	74.23%	+8.39%

V. ANALYSIS

- mBART-50 base confirms the translation-model limitation: with ROUGE-1 of only 6.38% and BLEU of 0.18, it performs well below both mT5 and IndicBART base, validating the decision to exclude it from fine-tuning.
- mT5 outperforms IndicBART across all metrics in both base and fine-tuned configurations. The largest gains after fine-tuning are observed in METEOR (+16.06% for mT5, +7.82% for IndicBART) and ROUGE-2 (+10.51% for mT5, +5.83% for IndicBART).
- BERTScore is the closest metric between the two fine-tuned models (78.51% vs 74.23%), indicating that both models produce semantically coherent summaries despite differences in n-gram overlap scores.
- Fine-tuning yields significant improvements for both models: mT5 improves ROUGE-1 by +4.72% and IndicBART by +4.70% over their respective base versions. The improvement is more pronounced on ROUGE-2 and METEOR, metrics sensitive to fluency and word choice.
- IndicBART base performance is notably lower than mT5 base (BLEU: 5.67 vs 10.23), suggesting that mT5's XLSum pre-training on Marathi data provides a stronger initialization for the summarization task.

Why mBART-50 Was Not Fine-tuned?

mBART-50 (facebook/mbart-large-50) was evaluated only in its base form and excluded from fine-tuning for the following reasons:

- Designed for translation, not summarization. mBART-50 was pre-trained on multilingual machine translation. Using it with src_lang = tgt_lang = mr_IN (same language on both sides) conflicts with its pre-training objective.
- Very low base performance. ROUGE-1 of 6.38% and BLEU of 0.18 indicate the model generates minimal meaningful Marathi content — far below mT5 base (37.42%) and IndicBART base (22.76%). Fine-tuning a model with such a weak inductive bias for the task would require substantially more data and epochs.
- Hardware constraints. At ~2.4 GB, fine-tuning mBART-50 on an 8 GB VRAM GPU is feasible but leaves little headroom. Given the low baseline, the expected return on compute did not justify the cost.
- Severe output repetition. The base model produces highly repetitive Marathi text when used for summarization a structural limitation of applying a translation model to an abstractive generation task.

Figure 2 shows the GUI prepared for proposed methodology to generate summary for entered text.

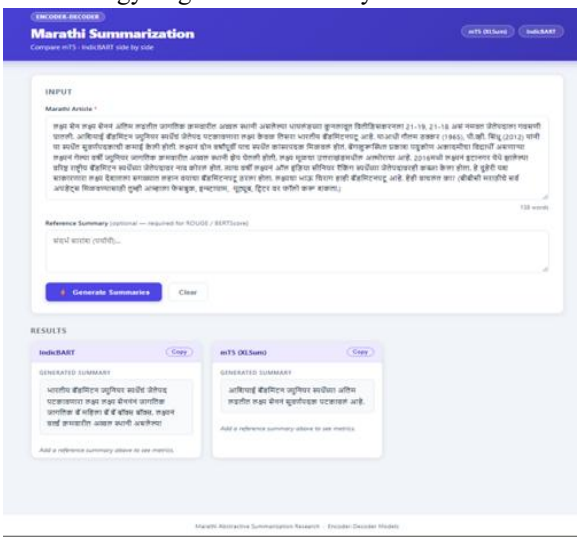

Figure 2. GUI for proposed methodology

VI. CONCLUSION

The comparative analysis of the transformer-based models provided valuable insights into the performance of different multilingual architectures for Marathi abstractive text summarization. The fine-tuned models demonstrated improved performance by generating summaries that were more coherent, contextually relevant, and semantically meaningful. The results highlight the effectiveness of transformer-based deep learning models in addressing the challenges of natural language processing for low-resource regional languages such as Marathi.

Although the proposed approach achieved promising results, there is scope for further enhancement. Future work can focus on utilizing larger and more diverse datasets, exploring advanced transformer architectures, and adopting more comprehensive evaluation metrics to improve summarization quality. In addition, multilingual summarization techniques and domain-specific applications can be investigated to develop more robust and versatile automated text summarization systems.

In conclusion, this research successfully demonstrates the applicability of modern transformer-based models for Marathi abstractive text summarization. The study establishes a strong foundation for future research in low-resource language processing and contributes to the advancement of automated summarization techniques for Indian regional languages.

REFERENCES

[1] Yang, Y., Y. Tan, J. Min, and Z. Huang, "Automatic text summarization for government news reports based on multiple features," *The Journal of Supercomputing*, vol. 80, no. 3, pp. 3212–3228, 2024.

[2] Awais, M., and R. M. A. Nawab, "Abstractive text summarization for the Urdu language: Data and methods," *IEEE Access*, vol. 12, pp. 61198–61210, 2024.

[3] Rifat, S. S., M. A. Rahman, and M. R. Islam, "Abstractive text summarization for Bangla language using NLP and machine learning approaches," in *Proc. Int. Conf. Electrical, Computer and Communication Engineering (ECCE)*, Feb. 13–15, 2025.

[4] Joshi, K., A. Kunchukuttan, P. Bhattacharyya, and S. Varma, "L3Cube-MahaSum: A comprehensive dataset and BART models for abstractive text summarization in Marathi," *arXiv preprint*, arXiv:2410.09184, Oct. 2024.

[5] Patil, S., A. Kulkarni, and P. Deshpande, "Enhancing Marathi abstractive text summarization with IndicBART," in *Proc. Int. Conf. Inventive Computation Technologies (ICICT)*, IEEE, 2025.

[6] Kulkarni, S., P. Patil, and R. Joshi, "Multilingual and cross-lingual text summarization of Marathi and English using transformer-based models and their systematic evaluation," *International Journal of Computer Applications*, vol. 186, no. 26, pp. 15–22, Jun. 2024.

[7] Deshmukh, S., and R. Kulkarni, "Experimental evaluation and enhancement of automatic unsupervised extractive text summarization of Marathi text using machine learning algorithm," *Journal of Machine and Computing*, vol. 2, no. 1, pp. 45–52, 2022.

[8] Hosseini, M., A. Shakery, and H. Zamani, "Optimizing question-answering framework through integration of text summarization model and third-generation generative pre-trained transformer," in *Proc. 14th Int. Conf. Computer and Knowledge Engineering (ICCKE)*, 2024.

[9] Kumar, A., S. Sharma, and P. Gupta, "Hybrid extractive-abstractive model for Indic news summarization," in *Proc. Int. Conf. Natural Language Processing (ICON)*, 2024.

[10] Gupta, A., R. Sharma, and S. Verma, "Transformer-based low-resource summarization for Indian languages," in *Proc. IEEE Int. Conf. Big Data*, 2024.

[11] Kunchukuttan, A., P. Bhattacharyya, and S. Varma, "Marathi XL-Sum evaluation study for abstractive text summarization," *Computer Science Series*, Springer, 2024.

[12] Zhang, J., X. Liu, and Y. Wang, "Fine-tuning PEGASUS for Indic language abstractive summarization," in *Proc. IEEE Int. Conf. Computing and Communication*, 2024.

[13] Liu, T., H. Chen, and Y. Zhang, "Comparative analysis of TextRank and BERT-based extractive models for news summarization," in *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2024.

[14] Fan, A., S. Bhosale, H. Schwenk, Z. Ma, A. El-Kishky, S. Goyal, *et al.*, "Cross-lingual abstractive text summarization using M2M100 transformer," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[15] Zhang, Y., T. Liu, and H. Chen, "Semantic-aware evaluation metrics for abstractive text summarization," *Information Processing & Management*, vol. 61, no. 3, 2024.

[16] Kumar, A., S. Sharma, and R. Patel, "Deep learning techniques for regional language natural language processing: A comprehensive review," *Neural Networks*, vol. 168, pp. 45–62, 2024.

[17] Deshmukh, P., A. Kulkarni, and S. Patil, "Comparative analysis of transformer architectures for Marathi natural language processing," *IEEE Access*, vol. 13, pp. 11245–11260, 2025.

[18] Sharma, R., A. Singh, and K. Gupta, "Benchmark evaluation of transformer models for Indic natural language processing tasks," *IEEE Access*, vol. 13, pp. 15420–15435, 2025.

[19] Verma, S., R. Iyer, and P. Kulkarni, "Low-resource natural language processing for Indian languages: Challenges and transformer-based solutions," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 2, pp. 1–18, 2024.

[20] Gupta, A., S. Mehta, and R. Sharma, "A comparative study of extractive and abstractive deep learning models for text summarization," *Expert Systems with Applications*, vol. 240, pp. 121–134, 2024.

[21] Joshi, M., P. Sharma, and A. Kulkarni, "A comparative study of ROUGE and BERTScore for low-resource abstractive summarization evaluation," *Information Processing and Management*, vol. 62, no. 1, pp. 102–115, 2025.

[22] Zhang, L., H. Liu, and J. Chen, "Improving factual consistency in abstractive text summarization," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1450–1462, 2024.

[23] Wang, Y., X. Li, and M. Chen, "Reinforcement learning-based optimization strategies for abstractive text summarization," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1785–1796, 2025.

[24] Patil, S., A. Deshpande, and R. Kulkarni, "Pointer-generator networks for abstractive summarization in Marathi," in *Springer Computer Science Series*, Springer, pp. 210–220, 2024.