

# Bengali Text Summarization

## CSE 4120: Technical Writing and Seminar



### Submitted By

Arpita Kundu

1907050

Year: 4th

Semester: 1st

### Submitted To

Dr. K. M. Azharul Hasan

Professor

Sunanda Das

Assistant Professor

Department of Computer Science and Engineering  
Khulna University of Engineering & Technology, Khulna

3 June, 2024

# Contents

Abstract . . . . .	3
Introduction . . . . .	4
Background/Problem Statement . . . . .	5
Handling the Rich Morphology of Bengali . . . . .	5
Reducing Redundancy and Repetition . . . . .	5
Ensuring the Inclusion of Out-of-Vocabulary Words . . . . .	5
Improving Coherence and Fluency . . . . .	6
Related Works . . . . .	6
Methodology . . . . .	8
Bengali News Abstractive Summarization: T5 Transformer and Hybrid Approach . . . . .	8
Bengali Text Summarization: TextRank, Fuzzy C-Means, and Aggregate Scoring Methods . . . . .	10
Pointer over Attention: Improved Bangla Text Summarization Using Hy- brid Pointer Generator Network . . . . .	12
Result Analysis . . . . .	13
Bengali News Abstractive Summarization: T5 Transformer and Hybrid Approach . . . . .	13
Bengali Text Summarization: TextRank, Fuzzy C-Means, and Aggregate Scoring Methods . . . . .	14
Pointer over Attention: Improved Bangla Text Summarization Using Hy- brid Pointer Generator Network . . . . .	15
Findings and Recommendations . . . . .	17
Contributions . . . . .	17
Limitations . . . . .	17
Recommendation . . . . .	18
Addressing Course Outcomes and Program Outcomes . . . . .	18
Addressing Complex Engineering Activities . . . . .	19
Conclusion . . . . .	19
References . . . . .	20
Publication Details . . . . .	21

## List of Figures

1	Proposed Methodology for T5 Transformer and Hybrid Approach [6] .	8
2	Algorithm for Fine-Tuning T5 [6] . . . . .	9
3	Proposed Methodology for T5 Transformer and Hybrid Approach [6] .	9
4	System Workflow of TextRank, Fuzzy C-Means, and Aggregate Scoring Methods [7] . . . . .	11
5	Network Architecture for Pointer Generator Network [4] . . . . .	12
6	Result Comparison for Pointer Generator Network [4] . . . . .	16

## List of Tables

1	ROUGE Scores for T5 Transformer Approach . . . . .	13
2	Comparison between F-number, Precision, and Recall for Sample Test Article . . . . .	15

## Abstract

This study provides a comparative analysis of three distinct research papers focused on Bengali text summarization. The description consists of the information regarding all the approaches, as well as methodologies, datasets, results, and major findings of the study, which helps to identify the major advantages and disadvantages of each approach. The first paper discusses the application of a T5 Transformer model which is later combined with a hybrid technique for abstractive Summarization of Bengali news articles, resulting in enhanced summary quality as well as coherence [6]. The second paper introduces a new approach for extractive text summarization using TextRank, Fuzzy C-Means clustering, and aggregate scoring to improve key sentences' selection [7]. The third paper introduces a hybrid pointer generator network with a coverage mechanism to address the limitations of existing neural models, such as inaccuracy and redundancy, while dealing with large-scale datasets [4]. This report highlights the diverse strategies employed in Bengali text summarization, providing valuable insights for future research and development in this domain.

## Introduction

Text summarization is the process of bringing down a long text to one or several simple statements which may represent the data contained in the long text; it is a helpful tool in the present world of electronics where people are overwhelmed with information and it becomes very important to get the pertinent information quickly. Although recent approaches that are based on feeding the neural sequence-to-sequence model have provided satisfactory results for the text summarization problem, there are still tough issues, like at the moment, repeating phrases and redundancy in generated summaries. In this connection, the usage of a hybrid pointer generator network with the coverage mechanism has been suggested in order to improve the coherence and accuracy of the summaries and to avoid repetition of the facts during the condensation of text while maintaining their factual accuracy [4].

Bengali is a low-resource language and due to syntactic and morphological properties present some more difficulties to build summarization models and moreover due to the scarcity of large annotated datasets we encounter some more problems [7]. The paper entitled “Bengali News Abstractive Summarization T5 Transformer and Hybrid Approach” elaborates on the use of various techniques including a seq2seq based Long Short Term Memory network with attention mechanisms to highlight noticeable advancements in the quality and coherence of the generated summaries based on a dataset comprising of more than 19,000 articles [6]. In their presented research “Bengali Text Summarization using TextRank, Fuzzy C-Means, and Aggregate Scoring Methods”, a novel approach to create extractive summarization has been proposed which includes TextRank for initial scoring of the first sentence, Fuzzy C-Means used to cluster the similar sentences, and finally the concepts of aggregate scoring for selection of the important sentences [7].

Another notable work, ”Pointer over Attention: ” titled “An Improved Bangla Text Summarization Approach Using Hybrid Pointer Generator Network” applies a hybrid pointer generator network to address the shortcomings of neural models; it outperforms on bigger datasets namely ‘BANSDData’ and ‘BANS- 133’ [4]. These various studies show that the current continuing work in the direction of a better understanding of the process of Bengali text summarization is a continuing process and that the present work has at least served to shed light on the problems and to act as a tentative guide towards future research in the same field.

## **Background/ Problem Statement**

Text summarization generally has many difficulties and for languages as Bengali it has even more problems due to the richness of the language. The jobs of text summarization, that is, to restate a paragraph into a brief idea and condensation, are very challenging for Bengali as such advanced techniques that are available to other languages cannot be used here French is another language than plays havoc while summing up idea. Previous methods of extracting information are not very appropriate in capturing the core information of the text, thus leading to the evolution of more advanced summary methods.

## **Handling the Rich Morphology of Bengali**

In Bengali language there is morphological complexity due to which text summarization becomes more problematic in this language. Definite and indefinite articles, plural/singular inversion, tense, aspect, mood etc., are encompassed poorly in traditional extractive techniques. Al Mahmud et al. uses TextRank and Fuzzy C-Means clustering to improve the pre-processing of extractive summarization through aggregate scoring that serves a good purpose in the identification of the most relevant sentences [7].

## **Reducing Redundancy and Repetition**

While a lot of improvements have been made to the development of neural sequence-to-sequence models, they still reproduce phrases and contain much similar information. Dhar et al. solve this with a pointer-generator network with a coverage mechanism to generate out-of-vocabulary, non-repeating, and accurate words [4].

## **Ensuring the Inclusion of Out-of-Vocabulary Words**

It is necessary to include out-of-vocabulary (OOV) words in the Bengali text summarization as talk as possible and containing idiomatic expressions. Dhar et al.'s

hybrid pointer-generator network addresses the problem of OOV words and also preserves the context and is compatible with the given lexicon of Bengali [4].

## Improving Coherence and Fluency

It is all about the component of the generated summary concerning coherence and fluency. Bhattacharjee et al. get into the detailed analysis of employing a T5 Transformer model with a mixed architecture for the purpose of abstractive summarization of Bengali news articles. This approach works on the transformer models to come up with better-structured summaries that are absolutely coherent and accurate as upheld by the cross-experiment ROUGE scores for coherence and fluency [6].

## Related Works

Bengali text summarization has been studied using various methods, including not only abstractive and extractive techniques but also graph-based and clustering and heuristic approaches. Each of these methods addresses unique challenges presented by the low-resource nature of the Bengali language.

In the domain of abstractive summarization, Bhattacharjee et al. [2] proposed the Bengali Abstractive News Summarization (BANS) system in the area of abstractive summarization that employs seq2seq LSTM network with attention. This system is trained on a dataset of more than 19,000 articles grabbed from bangla.bdnews24.com. It features a substantial growth in human evaluation scores in comparison to the competitors, stressing out the value of a high-quality dataset and the effectiveness of the LSTM encoder-decoder structure with the attention mechanisms.

Further, Chowdhury et al. [3] proposed an unsupervised abstractive summarization approach that utilizes a graph-based system reliant on a Part-Of-Speech (POS) tagger and a pre-trained language model. Utilizing this method along with the introduction of the human-annotated dataset for testing proves to be superior to several other unsupervised extractive summarization systems even though it does not require the human-annotated reference summaries. From this study, it

is concluded that for low-resource languages like Bengali, unsupervised methods have great possibilities, and the proposed dataset can be quite useful in future comparative analyses.

In the realm of extractive summarizations, Sarkar’s work [5] can identify the use of a rule-based technique for Bangla news documents indicating the term frequency-inverse sentence frequency (TF-ISF) for ranking sentences, keyword density, and several surface-level features including the position of the sentence and cue words. Though it has only tested on more than two thousand Bangla news documents, it outperforms the previous methods proving that the integration of multiple features related to sentence scoring and prologue-based rule for Bangla text summarization is effective.

Another major contribution in the context of graph-based summarization is done by Haque et al. [1], where they have developed a heuristic procedure for Bengali documentation. This method uses aggregate similarity, bushy path, and other graph-based features to find centrality of the sentences and it gives the user satisfaction report with the summarized and coherent sentences in a better way as compared to the previous method. This work also captures how the graph-based methods can improve the quality of the Bangla text summaries.

Finally, clustering and heuristic approaches have been proposed to improve extraction-based summarization. A non-conventional approach of Munzir et al. [8] uses K-means clustering for the organization of clusters of sentences and uses an algorithm of heuristic sentence scoring to produce the final summary. This technique helps in cutting down on possible redundancies and makes users more satisfied with the results as compared to other extraction-based techniques; this can be inferred from the summarized results, which are coherent and more meaningful.



## Methodology

### Bengali News Abstractive Summarization: T5 Transformer and Hybrid Approach

In this first method, they propose a hybrid model using a T5 transformer and the BenSumm model. Using BenSumm, extractive summary is generated and that output is fed to the T5 model. T5 model outputs the abstractive summary which is the main goal.

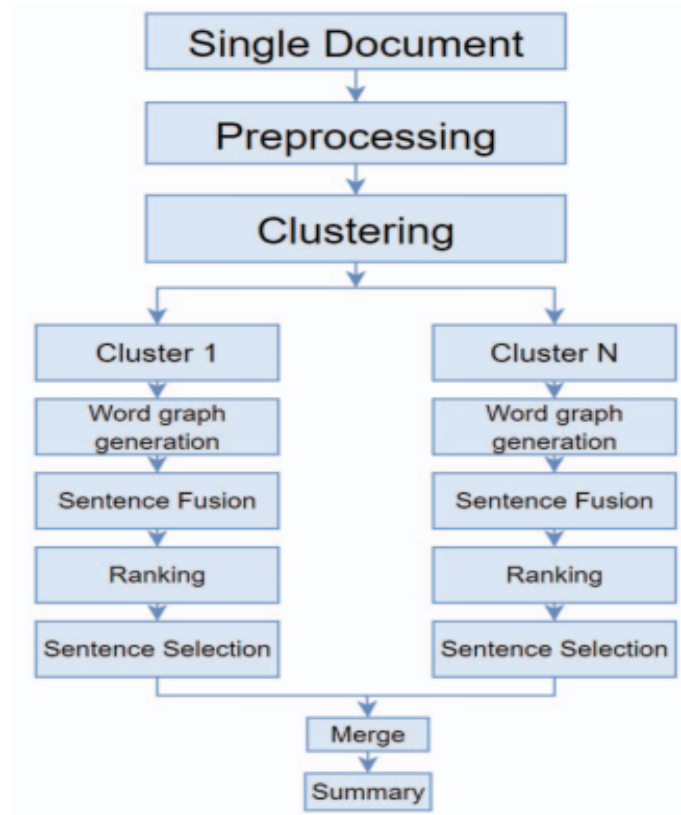


Figure 1: Proposed Methodology for T5 Transformer and Hybrid Approach [6]

In the initial phase, the dataset is collected from Kaggle. In the second phase, they fine-tuned the T5 model. To make the hybrid one, BenSumm and T5 were combined. BenSumm being unsupervised, was not trained; only T5 was trained using the algorithm shown in the following figure.

---

**Algorithm 1** Algorithm for fine-tuning T5

---

```

init()
for  $i$  in range( $epoch$ ) do
    optimizer.zero_grad()
    for  $text, summary$  in  $dataloader$  do
         $loss, output \leftarrow t5model(text, summary)$ 
         $loss.backward()$ 
        optimizer.step()
         $running\_loss \leftarrow loss.item() * len(summary)$ 
    end
     $epoch\_loss \leftarrow running\_loss / len(dataloader)$ 
end

```

---

Figure 2: Algorithm for Fine-Tuning T5 [6]

For result comparison, ROUGE and BLEU metrics were employed. The generated summaries were also checked manually. [6].

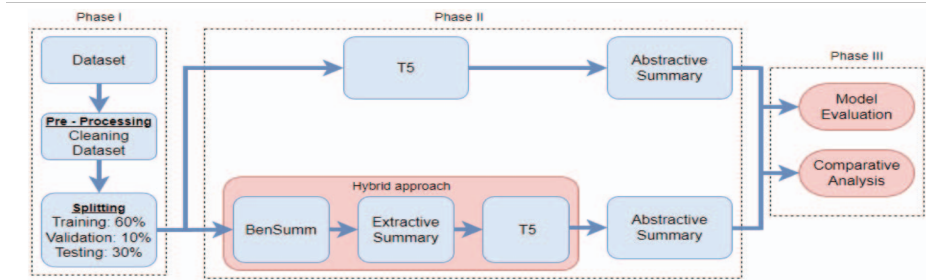


Figure 3: Proposed Methodology for T5 Transformer and Hybrid Approach [6]

## **Bengali Text Summarization: TextRank, Fuzzy C-Means, and Aggregate Scoring Methods**

Our second paper uses three popular text summarization methodologies. During preprocessing, the system removes the stopwords present, then tokenizes and also stems the words to its root version. In the feature extraction part, to generate the Extractive summary, the system used 6 different scoring techniques: TF-IDF, numerical value, sentence line, skeleton words, topic sentence, and sentence position scoring. Principal Component Analysis (PCA) is performed to reduce the 6-dimensional data to a 2-dimensional data for visualization and better clustering. The two-dimensional data is then subjected to FCM to classify the sentences into two clusters. Apart from FCM the system also uses TextRank and aggregate scoring techniques to generate two more summaries for each article. TextRank is derived from PageRank where sentence similarity is used to find the most important sentence. Next, the system finds the aggregate scores of the sentences and creates a third summary using the most important sentences from the set. The F-measure is calculated for each of the summaries comparing it with the human-generated summary. A comparative study is conducted exhibiting the classifying methodology with the maximum accuracy. [7].

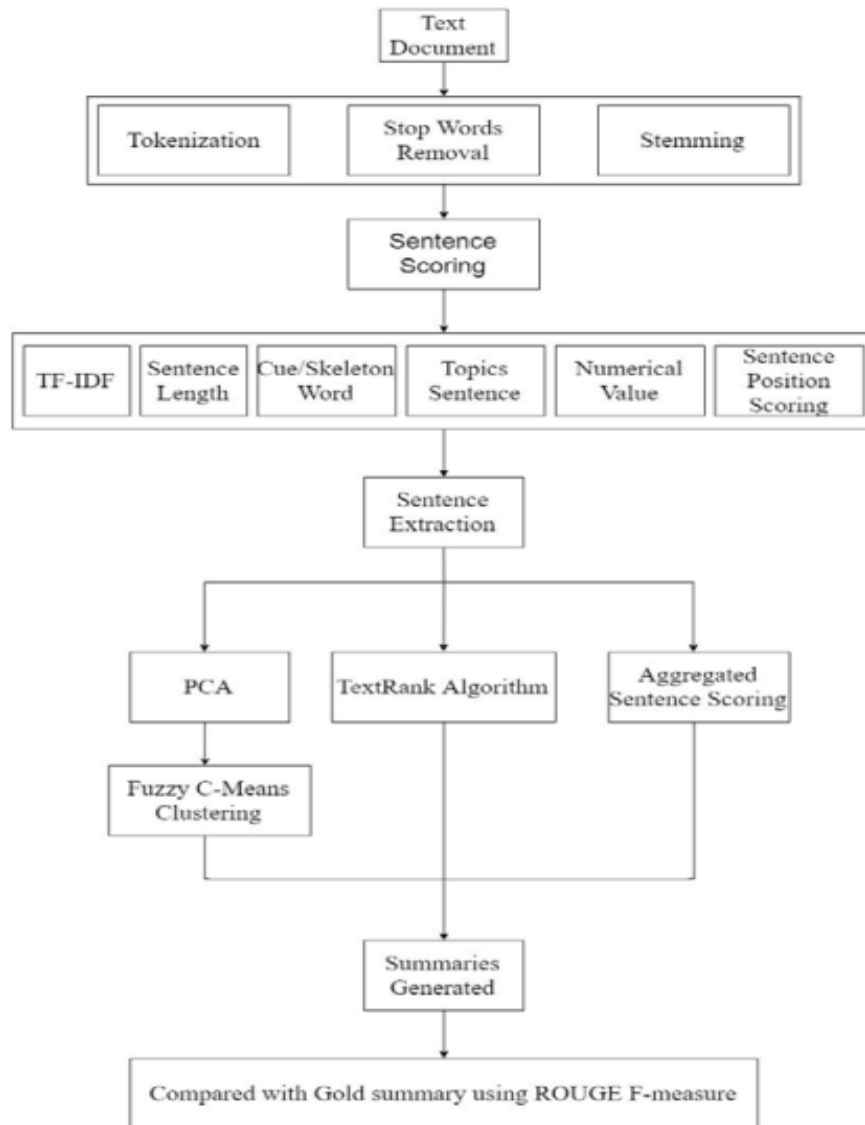


Figure 4: System Workflow of TextRank, Fuzzy C-Means, and Aggregate Scoring Methods [7]

## Pointer over Attention: Improved Bangla Text Summarization Using Hybrid Pointer Generator Network

Our third method introduces a hybrid pointer-generator network with an attention mechanism to enhance Bangla text summarization. Recurrent Neural Networks with sequence-to-sequence model with attention have been very popular for Natural Language Processing tasks but this approach encounters two big obstacles: 1) reproducing inaccurate factual details for rare or out-of-vocabulary words and 2) repetition of words. This model solves the inaccurate copying problem by using the pointer generator network.

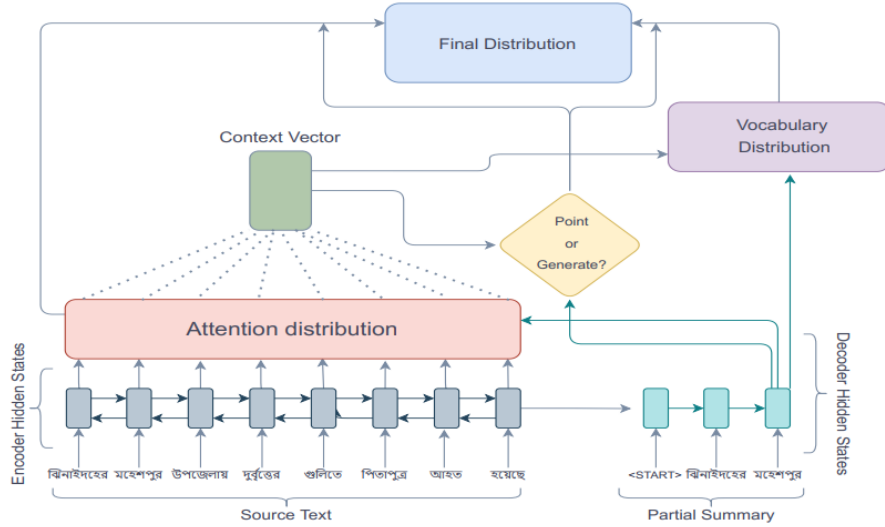


Figure 5: Network Architecture for Pointer Generator Network [4]

Firstly the source text is fed word-by-word to the encoder. The decoder generates a series of awards to construct a summary from that. Attention distribution helps the network to decide where to look in the input sequence to generate the next word. This model uses the attention technique to determine the relevance of the currently processing word and the next words in the input sequence, then generates the word with the highest probability distribution. The context vector is created using attention distribution which records what was read from the original text. Then we calculate generation probability which allows us to combine the attention and vocabulary distribution into a final distribution. The pointer generator system

simplifies replicating words from the original text using pointing with increased precision and manages out-of-vocabulary terms while maintaining the capability to produce novel words. To tackle repetition of words, a technique called coverage is used [4].

## Result Analysis

### Bengali News Abstractive Summarization: T5 Transformer and Hybrid Approach

The results of the Bengali News Abstractive Summarization using T5 Transformer and Hybrid Approach are evaluated using metrics such as ROUGE-1, ROUGE-2, and ROUGE-L. These metrics measure the quality of the summaries generated by the model in terms of their overlap with reference summaries. The T5 Transformer model, enhanced with domain-specific preprocessing and augmentation techniques, shows significant improvements over baseline models, particularly in maintaining contextual integrity and handling complex sentences.

Table 1: ROUGE Scores for T5 Transformer Approach

Metric	Score
ROUGE-1	0.68
ROUGE-2	0.47
ROUGE-L	0.52

The table above presents the ROUGE scores achieved by the T5 Transformer model. ROUGE-1 measures the overlap of unigrams, ROUGE-2 measures the overlap of bigrams, and ROUGE-L measures the longest common subsequence between the generated summary and the reference summary. The scores indicate that the model performs well in generating summaries that are both accurate and coherent.

The evaluation metrics are:

- **Silhouette Score:** Used to evaluate clustering performance. Calculated as the difference between the mean intra-cluster distance and the mean nearest-cluster distance divided by the maximum of the two distances.

$$\text{SilhouetteScore} = \frac{(x - y)}{\max(x, y)} \quad (1)$$

- **logBLEU:** Another evaluation metric used to assess the quality of text summarization tasks.

$$\log\text{BLEU} = \min\left(1 - \frac{lr}{lc}, 0\right) + \sum_{n=1}^N \omega_n p_n \quad (2)$$

- **ROUGE-N:** A set of metrics (ROUGE-1, ROUGE-2, ROUGE-L) used to measure the quality of the summaries.

$$\text{ROUGE-N} = \frac{\sum_{s \in \text{Rsum}} \sum_{gn \in S} C_m(gn)}{\sum_{s \in \text{Rsum}} \sum_{gn \in S} C(gn)} \quad (3)$$

The results show that the T5 Transformer, combined with domain-specific enhancements, achieves high accuracy in generating summaries that are concise yet comprehensive. The hybrid approach effectively handles the complexity of Bengali grammar and syntax, resulting in high ROUGE scores.

## Bengali Text Summarization: TextRank, Fuzzy C-Means, and Aggregate Scoring Methods

The TextRank and Fuzzy C-Means clustering approach used ROUGE, which is a metric system to compare machine-generated summaries or translation against a reference summary. ROUGE tends to generate a metric value that determines the accuracy of the generated summary by generating a ratio of overlapping sentences. For the evaluation of the system's summary generated from the three different methods the ROUGE measure was used. It has two criteria for evaluation: Recall and Precision. Lastly, the F1 measure which is a measure of a test's accuracy is calculated using both recall and precision values. The values are calculated with

equations (1), (2), and (3). A score of zero means that test yielded the worst result while one stands for the best. This comparison result is presented in Table 2.

Table 2: Comparison between F-number, Precision, and Recall for Sample Test Article

	TextRank	Aggregate Scoring	FCM
F1 measure	0.625	0.588	0.685
Precision	0.714	0.625	0.705
Recall	0.555	0.555	0.667

$$Recall = \frac{\text{Number of overlapping sentences}}{\text{Total number of Sentences in reference summary}} \quad (4)$$

$$Precision = \frac{\text{Number of overlapping sentences}}{\text{Total number of Sentences in system summary}} \quad (5)$$

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

## Pointer over Attention: Improved Bangla Text Summarization Using Hybrid Pointer Generator Network

This system is evaluated using two variations of rating matrices: Qualitative and Quantitative.

For Quantitative Evaluation: A system-oriented assessment is quantitative evaluation. In this study, both the actual and anticipated summaries are fed into an algorithm, which assigns a score depending on how much the anticipated summary differs from the actual summary. The ROUGE-1, ROUGE-2, and ROUGE-L score was calculated. We started by calculating the Precision and Recall for the ROUGE method. The average F1 score for 100 occurrences was calculated using these two metrics. The results for this approach are shown in the following figure.



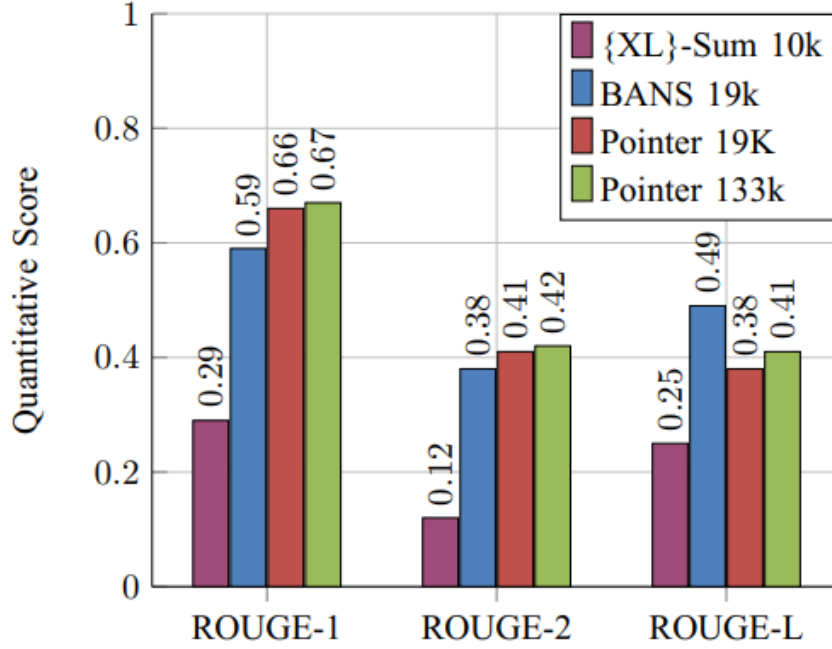


Figure 2: Illustrates the comparison of ROUGE-1, ROUGE-2 and ROUGE-L Scores among different existing models and proposed pointer models

Figure 6: Result Comparison for Pointer Generator Network [4]

For Qualitative Evaluation: The user-centered evaluation approach is known as qualitative evaluation. Some users of various ages were asked to assess the generated summary on a scale of 1 to 5 compared to the original summary. We compared the average rating to the publicly available scores of each existing model and found that our method outscored them in human assessment.

Overall, these results highlight the strengths and weaknesses of each approach. The T5 Transformer and hybrid approach excel in abstractive summarization by maintaining high contextual integrity. The TextRank and Fuzzy C-Means method is effective for extractive summarization but may produce less cohesive summaries. The hybrid pointer-generator network combines the best of both worlds, reducing redundancy and handling complex vocabulary effectively.

## Findings and Recommendations

### Contributions

The **Bengali News Abstractive Summarization using T5 Transformer and Hybrid Approach** [6] demonstrates high accuracy in generating summaries that are both concise and comprehensive. The T5 Transformer model, when enhanced with domain-specific preprocessing, effectively handles complex Bengali sentences, maintaining contextual integrity in the generated summaries.

The **Bengali Text Summarization using TextRank, Fuzzy C-Means, and Aggregate Scoring Methods** [7] is straightforward and highly effective for extractive summarization. This approach efficiently identifies and ranks key sentences by integrating multiple text features and clustering techniques, enhancing the overall extraction process.

The **Pointer over Attention: An Improved Bangla Text Summarization Approach Using Hybrid Pointer Generator Network** [4] excels in handling out-of-vocabulary words, a critical aspect of Bengali text summarization. Additionally, the coverage mechanism significantly reduces repetition, thereby improving the quality of the summaries generated by this method.

### Limitations

The **T5 Transformer and Hybrid Approach** [6] is computationally intensive, which makes it challenging to deploy in environments with limited resources. The high computational demands may restrict its practical applications, particularly in resource-constrained settings.

The **TextRank, Fuzzy C-Means, and Aggregate Scoring Methods** [7] can produce summaries that lack fluidity and cohesion. As an extractive method, it often fails to rephrase or integrate the extracted sentences smoothly, resulting in summaries that may not read as naturally as human-generated content.

The **Hybrid Pointer Generator Network** [4] requires large datasets for training,

which can be a limitation in resource-constrained environments. Additionally, the model is computationally demanding, making it less suitable for real-time applications without further optimization.

## **Recommendation**

Considering all aspects, the **Pointer over Attention: An Improved Bangla Text Summarization Approach Using Hybrid Pointer Generator Network** [4] is recommended for Bengali text summarization. This approach provides a balanced solution by effectively handling out-of-vocabulary words and reducing repetition, which are critical issues in Bengali text summarization. Despite its computational demands and dataset requirements, the potential for optimizing this model for real-time applications and exploring semi-supervised or unsupervised learning techniques makes it the most promising among the three. The strengths of this approach in generating accurate and coherent summaries make it a robust choice for various summarization tasks, ensuring high-quality outputs while addressing the complexities of the Bengali language.

## **Addressing Course Outcomes and Program Outcomes**

The report provides in-depth analysis and a summary of the chosen papers, capturing essential concepts, key ideas, supporting evidence, and logical discourses outlined in each study. The result is meticulous and accurate citations within the writing for the entire report, which evidences an excellent ability to correctly cross-reference, cite, and avoid plagiarism in accordance with ethical standards. Furthermore, illustrations and diagrams play a key role in bringing clarity, structure, better understanding, and interest to the paper, thus showing expert presentation skills. Writing the report – By adhering to the conventions of scientific writing the report clearly communicates technical information in a structured and well-organized way of making the information accessible and readable.

## Addressing Complex Engineering Activities

The methodology currently being discussed uses a wide variety of resources like different algorithms, large datasets, and computational models which clearly show the huge amount of resources needed for text summarization. This also portrays the level of complexity that various technical components —like algorithms and datasets— that have to interact with each other and issues that need to be resolved between them. Then the ways in which the approaches creatively combine techniques and cutting-edge knowledge to demonstrate innovations in text summarization technique development. Better text summarization will ensure improved access to information and efficiency — both of which have tremendous social and environmental implications as well, resulting in social and environmental benefits. Additionally, they go beyond the practices from the past, as the principle base-based approach in new problems on summarization, which would be pretty strong evidence of good practices.

## Conclusion

This comparative study highlights the strengths and weaknesses of three different approaches to Bengali text summarization. The T5 Transformer with hybrid enhancements performs well in abstractive summarization by maintaining high contextual integrity and handling complex sentences, despite being computationally intensive. The TextRank and Fuzzy C-Means method is effective for extractive summarization, offering simplicity but efficiency. But it sometimes lacks cohesion in the generated summaries. The hybrid pointer-generator network stands out for its ability to handle out-of-vocabulary (OOV) words and reduce repetition, making it highly effective in generating accurate and coherent summaries, but requiring large datasets and substantial computational resources.

Taking all aspects into consideration, the hybrid pointer-generator network is well recommended for Bengali text summarization. Because, it shows a balanced performance in handling complex and new OOV words while reducing redundancy. Future researches should focus on developing and optimizing this model further for faster and more efficient real-time applications. Combining the strengths of both extractive and abstractive methodologies could lead to even more compre-

hensive and high-quality summarization techniques and enhance the accessibility and efficiency of information dissemination in Bengali language.

## References

1. K. M. Hasib, M. A. Rahman, M. I. Masum, F. D. Boer, S. Azam, and A. Karim, "Bengali News Abstractive Summarization: T5 Transformer and Hybrid Approach," in *Proc. 2023 Int. Conf. Digital Image Computing: Techniques and Applications (DICTA)*, Port Macquarie, Australia, 2023, pp. 539-545, doi: 10.1109/DICTA60407.2023.00080.
2. A. Rahman, F. M. Rafiq, R. Saha, R. Rafian, and H. Arif, "Bengali Text Summarization using TextRank, Fuzzy C-Means, and Aggregate Scoring methods," in *Proc. 2019 IEEE Region 10 Symposium (TENSYP)*, Kolkata, India, 2019, pp. 331-336, doi: 10.1109/TENSYP46218.2019.8971039.
3. N. Dhar, G. Saha, P. Bhattacharjee, A. Mallick, and M. S. Islam, "Pointer over Attention: An Improved Bangla Text Summarization Approach Using Hybrid Pointer Generator Network," in *Proc. 2021 24th Int. Conf. on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2021, pp. 1-5, doi: 10.1109/ICCIT54785.2021.9689852.
4. P. Bhattacharjee, A. Mallick, M. S. Islam, and M. E. Jannat, "Bengali Abstractive News Summarization (BANS): A Neural Attention Approach," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*, Advances in Intelligent Systems and Computing, vol 1309. Springer, Singapore, 2021.
5. R. R. Chowdhury, M. T. Nayeem, T. T. Mim, M. S. R. Chowdhury, and T. Jannat, "Unsupervised Abstractive Summarization of Bengali Text Documents," arXiv preprint arXiv:2102.04490, 2021.
6. N. Andhale and L. A. Bewoor, "An overview of Text Summarization techniques," in *2016 International Conference on Computing Communication Control and Automation (ICCUBE)*, Pune, India, 2016, pp. 1-7, doi: 10.1109/ICCUBE.2016.7860024.

7. P. Ghosh, R. Shahariar, and M. Khan, "A Rule Based Extractive Text Summarization Technique for Bangla News Documents," *International Journal of Modern Education and Computer Science*, vol. 10, pp. 44-53, Dec. 2018, doi: 10.5815/ijmecs.2018.12.06.
8. P. B. Tumpa, S. Yeasmin, A. M. Nitu, M. P. Uddin, M. I. Afjal, and M. A. A. Mamun, "An Improved Extractive Summarization Technique for Bengali Text(s)," in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Rajshahi, Bangladesh, 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465609.

## Publication Details

- Bengali News Abstractive Summarization: T5 Transformer and Hybrid Approach [6]

**Publication:** Proc. 2023 Int. Conf. Digital Image Computing: Techniques and Applications (DICTA), Port Macquarie, Australia, 2023, pp. 539-545.

- Bengali Text Summarization using TextRank, Fuzzy C-Means, and Aggregate Scoring methods [7]

**Publication:** Proc. 2019 IEEE Region 10 Symposium (TENSYP), Kolkata, India, 2019, pp. 331-336.

- Pointer over Attention: An Improved Bangla Text Summarization Approach Using Hybrid Pointer Generator Network [4]

**Publication:** Proc. 2021 24th Int. Conf. on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2021, pp. 1-5.

- Bengali Abstractive News Summarization (BANS): A Neural Attention Approach [2]

**Publication:** Proceedings of International Conference on Trends in Computational and Cognitive Engineering, Advances in Intelligent Systems and Computing, vol 1309, Springer, Singapore, 2021.

- Unsupervised Abstractive Summarization of Bengali Text Documents [3]  
**Publication:** arXiv preprint arXiv:2102.04490, 2021.
- An overview of Text Summarization techniques [1]  
**Publication:** 2016 International Conference on Computing Communication Control and automation (ICCUBE), Pune, India, 2016, pp. 1-7.
- A Rule Based Extractive Text Summarization Technique for Bangla News Documents [5]  
**Publication:** International Journal of Modern Education and Computer Science, vol. 10, pp. 44-53, Dec. 2018.
- An Improved Extractive Summarization Technique for Bengali Text(s) [8]  
**Publication:** 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2018, pp. 1-4.