IEEE ACCEPTED VERSION 2023

Bengali News Abstractive Summarization: T5 Transformer and Hybrid Approach

Khan Md Hasib[1], Md. Atiqur Rahman[2], Mustavi Ibne Masum[3], Friso De Boer[4], Sami Azam[5], Asif Karim[6]

[1]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh

[2,3]Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh

[4,5,6]Faculty of Science and Technology, Charles Darwin University, NT, Australia

Corresponding Author: asif.karim@cdu.edu.au, sami.azam@cdu.edu.au

***Abstract*—In today's fast-paced world, everyone wants things to happen quickly. Thanks to the internet, news spreads super fast. But not all news is important. News summarization helps by giving a short version of each news story, so readers can easily figure out what type of news they want to read. There are two main types of summarization: Abstractive Text Summarization and Extractive Text Summarization. The process of abstractive text summarization is much more complex than that of extractive text summarization. This study proposes a model for generating extractive summaries, which are then utilized as input to generate abstractive summaries. The model uses the Bengali Text Summarization (BenSumm) model for extractive summarization and the Bangla Text-to-Text Transfer Transformer (BanlaT5) for abstractive summarization. The research also compares summarization acquired straight from the BanglaT5 model with summarization obtained via the proposed model. Abstractive summarization in the Bengali language has been accomplished using the Text-to-Text Transfer Transformer(T5) in this research. Although abstractive summarization of the Bengali language has been accomplished over the years using a variety of techniques, the field of using T5 in this field has only recently been discovered, and there is still a wide range of opportunities to be explored. The study has achieved promising results.**

***Index Terms*—T5 Transformer, Textrank, unsupervised learning**

I. Introduction

Text summarizing is the process of extracting the core idea of the source corpus from a big body of text and condensing it into a small number of sentences. In this era of the internet, many new articles are available here and there. Reading the whole news article is time-consuming. A summary of the whole article can save both time and energy. Furthermore, news summaries play a crucial role in offering users a concise understanding of essential elements within the content, allowing them to grasp the fundamental concepts of an article before deciding to read the complete text. There are two categories of text summarizing, extractive and abstractive. Abstractive summarization does not simply copy important phrases from the source text but also potentially comes up with relevant new phrases. Abstractive summarization generates new, pertinent terms in addition to just copying significant passages from the source text [1]. The summarizer creates new words, phrases, and sentences to create the summary. On the other hand, the process of extractive summarizing involves selecting the main phrases from a text to condense their meaning instead of starting from scratch while generating new text. Bhattacharjee et al. [2] used the neural attention approach to create abstract Bengali summaries of the news. They used LSTM as both encoder and decoder. From the Quantitative analysis of their proposed model, they took 100 generated summaries and their corresponding actual summaries from which they achieved a score of 0.3 ROUGE-1, 0.31 ROUGE-L, and 0.3 BLEU score. Chowdhury et al. [3] introduced an unsupervised method of text summarization of Bengali text documents. Their model BenSumm can be used for both abstractive and extractive text summarization. They used the clustering method to cluster the words of sentences from which the summaries are generated from sentence selection produced from different clusters. From the BenSumm model for abstractive summarization, they got 12.17 ROUGE-1, 1.92 ROUGE-2, and 11.35 ROUGE-L scores, and for extractive summarization 61.62 ROUGE-1, 55.97 ROUGE-2, and 61.09 ROUGE-L score.

Sethi et al. [4] generated summarized text of news articles. They used BART and T5 transformer models for generating summaries and have done a comparative analysis between both models. The analysis showed that BART outperforms T5 but the difference is not much. The BART model has an average F1 score of 33% and the T5 model has an average F1 score of 26%. Zolotareva et al. [5] used transfer learning for abstractive summarization. Transfer learning means using the T5 approach to summarize the text. Their model achieved F1 score of 0.473 ROUGE-1, 0.265 ROUGE-2, and 0.361 ROUGE-L. Fendji [6] used the T5 model for text summarization for SMS. The work had been done on the French Wikipedia. The summaries generated by the model are also in the French language. They achieved a ROUGE-L score of 77.0 for SMS in the French language having a length of more than 500 characters. Bohra et al. [7] did a comparative analysis of the T5 model on different datasets. The datasets they used are CNN/Daily-Mail(CNNDM), MSMO, and XSUM. The work has generated abstractive summaries from the datasets. From their evaluation, the highest scores they achieved are a ROUGE-1 score of 40.791, a ROUGE-2 of 18.551, and a ROUGE-L score of 34.80 on the CNNDM dataset. They gained a BLEU-1 score of 43.9 on the MSMO dataset, a BLEU-2 score of 21.3 on the MSMO dataset, a BLEU-3 score of 11.7 on the MSMO dataset, a BLEU-4 score of 8.0 on CNNDM dataset and a BLEU score of 14.58 on CNNDM dataset. Etemad et al. [8] also used T5 for abstractive text summarization. They fine-tuned the T5 model for a better result. Their model achieved a ROUGE-1 score of 43.02, ROUGE-2 score of 14.50, ROUGE-L score of 37.43, and ROUGE-LSUM score of 37.49.

Abstractive summary generation is a more difficult task than extractive. Trained language models have played a pivotal role in enhancing various natural language processing (NLP) applications, including text summarization. These models have significantly contributed to the advancement and effectiveness of automated summarization techniques. The field of NLP research predominantly utilizes two state-of-the-art pre-trained language models: Text-to-Text Transfer Transformer (T5) [9] and Bidirectional Encoder Representations from Transformers (BERT) [10]. These models have gained significant popularity and are widely employed due to their impressive performance and versatility in various NLP tasks. Text summarization using the T5 transformer model in the Bengali language is untouched. It has been demonstrated that the T5 transformer is excellent at producing summaries from texts [11]. The following are the main goals of this paper:

- Compile a large Bengali news dataset for benchmarking exercise.
- Propose a hybrid model that combines the strengths of the BenSumm model and the T5 model for text summarization.
- Compare the performance of the proposed hybrid model and the T5 model and conduct a comprehensive evaluation of the results to determine the effectiveness of the hybrid approach in generating summaries for Bengali text.

II. Dataset

The dataset used in this experiment is collected from Kaggle [1]. The dataset contains 80,148 news articles of three categories. The categorical article distribution is given in table I.

The dataset contains three columns, category, summary, and text. The summary column contains one line short summary of the whole text. An example from the dataset is given below.

[1]https://www.kaggle.com/datasets/hasanmoni/bengali-text-summarizationm

TABLE I: Category-wise article distribution

Categories	Amount of data (%)
Bangladesh	71
International	8
Others	21

Text: নগরের সিইপিজেড এলাকার চানখালি সড়কে তিন পার্বত্য জেলার বৌদ্ধ সম্প্রদায়ের লোকজন নিয়ে গঠিত ফরা রং খ্রে-দ ক্যংয়ে (বৌদ্ধবিহার) কঠিনচীবর দান ও জ্ঞাতিসম্মেলন কাল শুক্রবার অনুষ্ঠিত হবে। এ উপলক্ষে নেওয়া কর্মসূচির মধ্যে রয়েছে ধর্মীয় পতাকা উত্তোলন, শোভাযাত্রা, ভিক্ষু সংঘের পিণ্ডদান, চীবরদান ও সদ্ধর্মদেশনা। এতে খাগড়াছড়ি জেলা পরিষদের চেয়ারম্যান চাইথোঅং মারমা, বান্দরবান জেলা পরিষদের সাবেক চেয়ারম্যান মংক্যচিং চৌধুরী, রাঙামাটি জেলা পরিষদ সদস্য অংসুইপ্রু চৌধুরীসহ ধর্মীয় নেতারা উপস্থিত থাকবেন। বিজ্ঞপ্তি।

Reference summary: ফরা রং খ্রে-দ ক্যংয়ে কঠিন চীবরদান ও জ্ঞাতি সম্মেলন কাল

III. Background Study

In this section, we will provide a brief description of the used models. The T5 model is based on the transformer architecture proposed by Vaswani et al. [12]. The transformer model is only an encoder-decoder architecture [13] attached to a multi-head attention mechanism and is depicted in the fig. 1. The high-level overview of the encoder and decoder is described below:

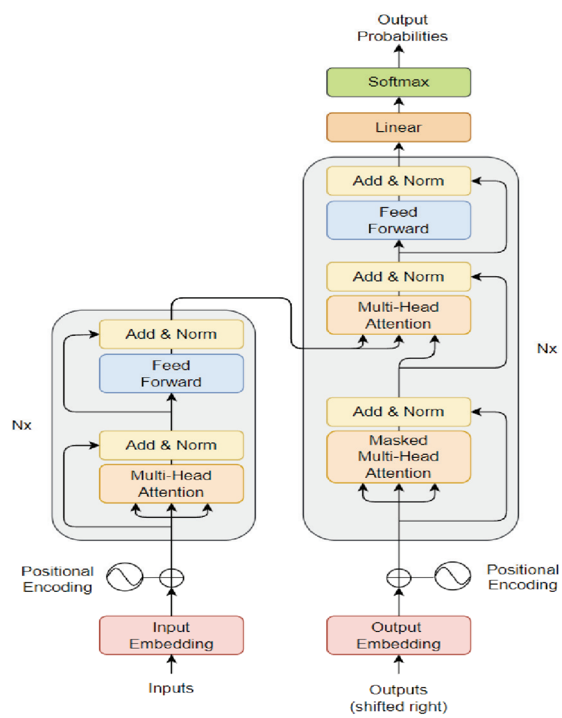


Fig. 1: The Transformer - model architecture

A. *Encoder*

This layer has two sub-layers. Self-attention mechanism and feed-forward neural networks. The transformer encoder first embeds the input text into a vector using embedding methods before applying positional encoding to preserve the token order. The input is then passed through to the self-attention layer. In this layer, the key, query, and value vector dot products of randomly initialized vectors are used to determine each word's attention. The final attention vector is calculated by adding together all of the attention vectors after each word's attention has been calculated. Lastly, The output of the attention layer is passed through the feed-forward layer.

B. *Decoder*

Encoder-decoder attention, feed-forward, and multi-head attention layers are the three sub-layers contained in the decoder module. The decoder applies encoder-decoder attention to the input from the previous last encoder and generates an output [14]. Additionally, the decoder employs multi-head attention to concentrate on the necessary portions of the sentences. The cycle repeats until the special token [EOF] is reached. Then, the decoder feeds each step to its bottommost section to produce the final result.

T5: The T5 model applies a simplified version of layer normalization with rescaled activations and residual skip connections.

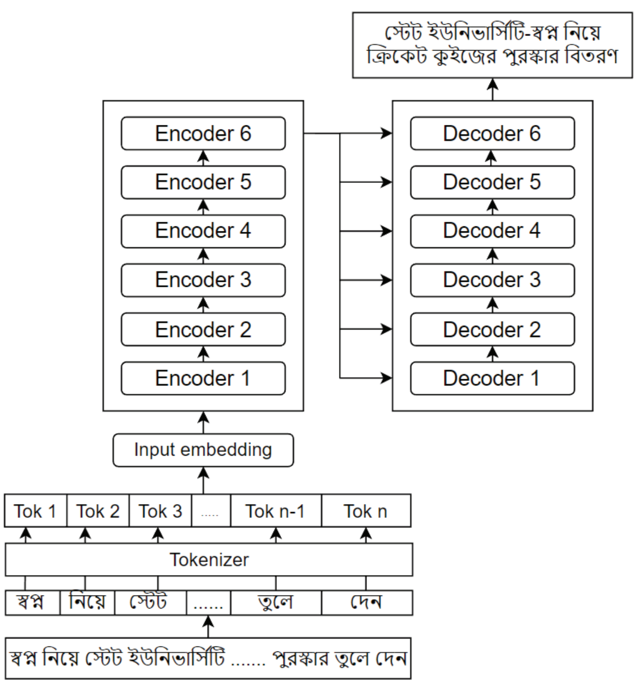


Fig. 2: Summary generation of T5 transformer

Dropout [15] is implemented on various segments of the model, and it uses a form of autoregressive self-attention for the encoder-decoder attention. Relative position embedding is used instead of sinusoidal embedding and position embedding parameters are shared across all layers for efficiency. The overall process of generating the summary is depicted in the fig. 2.

BenSumm model:

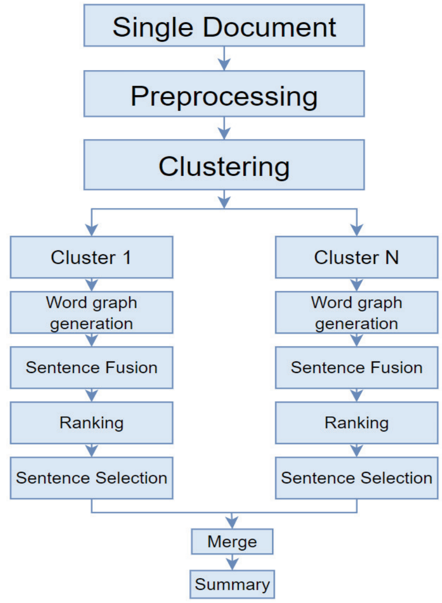


Fig. 3: Overview of our BenSumm model

This model involved several steps. The first step is preprocessing. This step includes tokenization, removal of stopwords, Part-Of-Speech(POS) tagging, and filtering of punctuation. The next step is sentence clustering. The clustering step allows the grouping of similar sentences from a given document. Hierarchical agglomerative [16] clustering with the ward's method is used. Then the number of clusters for a given document is measured using the silhouette value. The formula is given below:

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

where x is the mean distance to the instances of the next closest cluster and y denotes the mean distance to the other instances of the intra-cluster. Using the cluster of related sentences, the word graph is constructed. This is an unsupervised method and only needs POS tagger. The parts-of-speech (POS) [17] tags and the words are both represented as vertices. Connecting the neighboring words from the phrases creates directed edges. To extract abstractive fusions from these connected phrases, word graphs are produced for each cluster. Using the ranking approach, the clusters are made up of numerous weighted sentences. Each cluster's top-ranked sentence is used to output the summary. The top-ranked sentences are all combined to create the final summary. Fig. 3 shows the overall BenSumm model architecture.

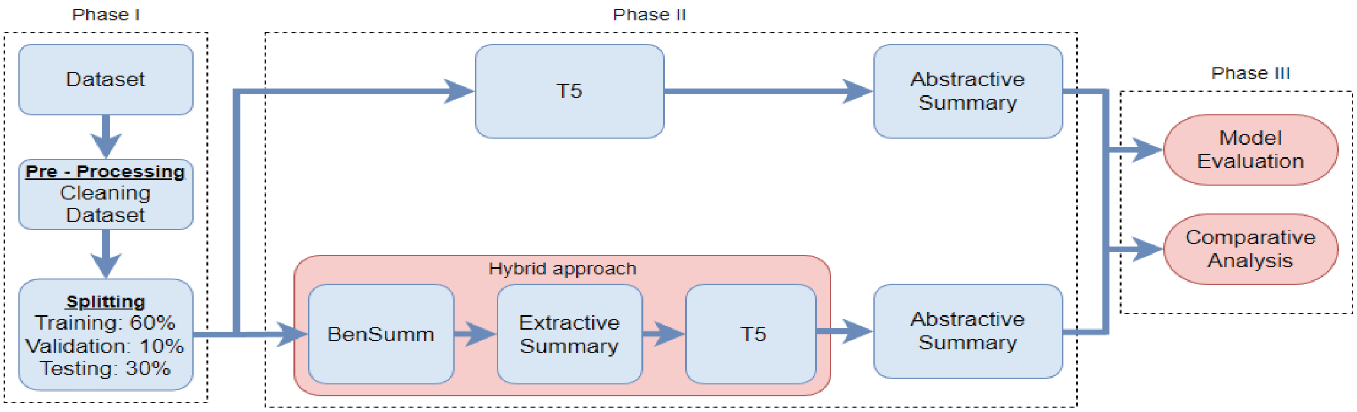


Fig. 4: Proposed methodology

C. Evaluation Metrics

We used Bilingual Evaluation Understudy (BLEU) [18] and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [19] since they have been shown to have a strong correlation with human judgments, the BLEU and ROUGE similarity metrics are fairly reliable.

BLEU: BLEU was introduced in 2002 and is widely used to assess the quality of machine-generated texts. BLEU scores take into consideration the similarity between predicted unigrams or higher n-grams and a set of reference sentences of one or more candidates. The output of BLEU ranging 0 to 1. If the generated text exactly matches the reference text, the output is 1. For the opposite case, the output is 0. The equation of calculation BLEU is shown in the equation 2:

$$logBLEU = min(1 - \frac{l_r}{l_c}, 0) + \sum_{n=1}^{N} \omega_n \log_{p_n} \tag{2}$$

In the above equation, $\frac{l_r}{l_c}$ is the ratio of the length of the similar reference corpus, and the candidate description, ω_n are positive weights and p_n represent the geometric average of the modified n-gram precision.

ROUGE: ROUGE was introduced in 2004. ROUGE is a set of evaluation metrics used in NLP and text summarization to measure the quality and similarity of generated summaries to reference summaries [20]. ROUGE-1 measures the overlap of unigrams, ROUGE-2 measures the overlap of bigrams, and ROUGE-L considers the longest common subsequence. These metrics provide precision, recall, and F1-score, offering a quantitative assessment of summary quality. The equation of calculation ROUGE is shown in the equation 3:

$$ROUGE - N = \frac{\sum_{s \in R_{sum}} \sum_{g_n \in S} C_m(g_n)}{\sum_{s \in R_{sum}} \sum_{g_n \in S} C(g_n)} \tag{3}$$

In the above equation, n denotes the length of an n-gram, g_n, and $C_m(g_n)$ denotes the largest number of n-grams found in the candidate, along with ground truth summaries, and R_{sum} denotes reference summaries.

IV. Proposed Methodology

In this paper, we propose a hybrid model using a T5 transformer and the BenSumm [3] model. Using BenSumm we generate the extractive summary and feed the output to the T5 model. T5 model outputs the abstractive summary which is our ultimate target.

The first phase contains the dataset preprocess, and the second phase is fine-tuning the models and generating the summary. The last phase includes a comparison of the results. Fig. 4 depicts the proposed method of our work.

A. Phase I

As mentioned before, the dataset is using this experiment is collected from Kaggle. The dataset contains more than 80 thousand texts and summaries. There are several texts containing English words. We cut out the texts that contain the English alphabet or words. Then after dropping the duplicate texts, the number of texts is 76, 487. Then we divided the dataset into the train, validation, and test sets with a ratio of $60 : 10 : 30$. There are 48186, 5354, and 22947 sets of train, validation, and test data, respectively.

B. Phase II

This phase contains fine-tuning the T5 model. To make the hybrid model, we combined BenSumm and T5 models. We generated the extractive summaries using the BenSumm model. As the BenSumm model is unsupervised, we did not need to train this model. We only fine-tuned the T5 model. The algorithm for fine-tuning the T5 model is presented below:

C. Phase III

The last phase is about result comparison. To compare the result we employed ROUGH [19] and BLEU [18] metrics. We also evaluated the generated summary manually to check if the summaries generated by the models are correct or wrong.

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 ← t5model(text, summary)
        loss.backward()
        optimizer.step()
        running_loss ← loss.item() * len(summary)
    end
    epoch_loss ← running_loss / len(dataloader)
end
```

V. Experiment and Results

For programming language, Python was used. Ben-Summ model uses the Scikit-learn library [21]. We used the PyTorch framework to train and test the T5 model.

A. Hyperparameter settings

We selected the best hyperparameter [22] manually. For the train T5 model, we split the dataset into a 60:30:10 ratio where 60% data is for training, 30% data for testing, and 10% data for validation. The best hyperparameters for the T5 model are the following:

- Batch Size = 8
- Epoch = 4
- Learning Rate = 0.0001
- Seed = 42
- Optimizer = AdamW

We only mention the hyperparameters that give us the best performance.

B. Fine tuning Result

The table II shows the loss of each epoch for both models.

TABLE II: Epoch loss table

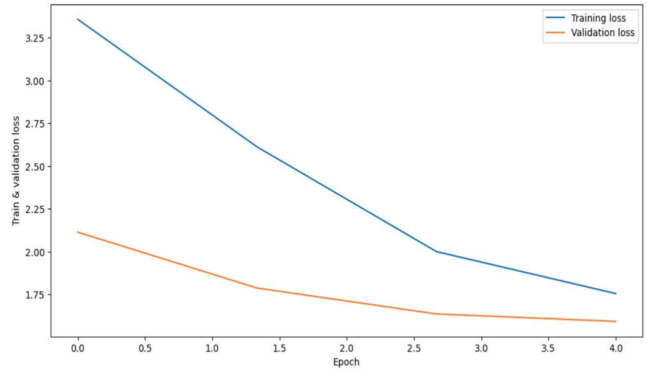
Epoch	Train Loss (%) Hybid Model	Train Loss (%) T5 model	Validation Loss (%) Hybid Model	Validation Loss (%) T5 Model
1	4.7696	2.1145	2.8988	2.1145
2	3.2207	2.6124	2.4654	1.7878
3	2.7650	2.0007	2.3006	1.6365
4	2.4916	1.7554	2.2302	1.5930

The fig. 5 shows the Fine-tuning and validation loss vs epoch curve for both models.

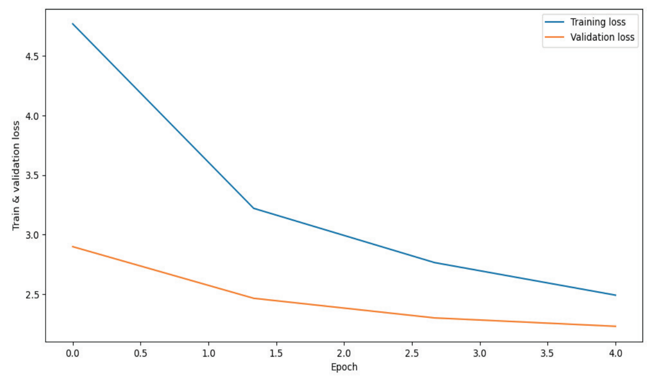
C. Result

Table III shows the ROUGE score of the Hybrid (Ben-Summ + T5) model and table IV shows the ROUGE score of the T5 model.

The Rough score clearly shows that the T5 model performs better than the hybrid model. Each rough score has a precision, recall, and F1 score which is denoted by 'P', 'R', and 'F1' in the fig. 6.



(a) Loss curve of T5 model



(b) Loss curve of hybrid model

Fig. 5: Fine-tune and validation loss curves for both of the models

TABLE III: Rouge score of hybrid model

	ROUGE-1	ROUGE-2	ROUGE-L
Precision	0.352	0.1711	0.3332
Recall	0.3531	0.1727	0.3343
F1	0.35	0.1704	0.3313

TABLE IV: Rouge score of T5 model

	ROUGE-1	ROUGE-2	ROUGE-L
Precision	0.4461	0.2465	0.4243
Recall	0.4558	0.2531	0.4335
F1	0.4478	0.2477	0.4259

The table V represents the BLEU score of the hybrid and T5 model.

TABLE V: Obtained BLEU score

	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Hybrid Model	0.2728	0.1636	0.0940	0.0526
T5	0.3617	0.2414	0.1514	0.0897

Similar to the ROUGH scores, the BLEU score of the T5 model is higher than the hybrid model. The comparison of the BLEU score is shown in the fig. 7.

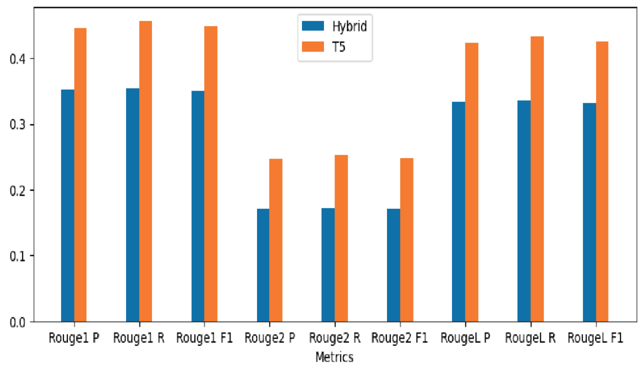


Fig. 6: Comparison of Rough score

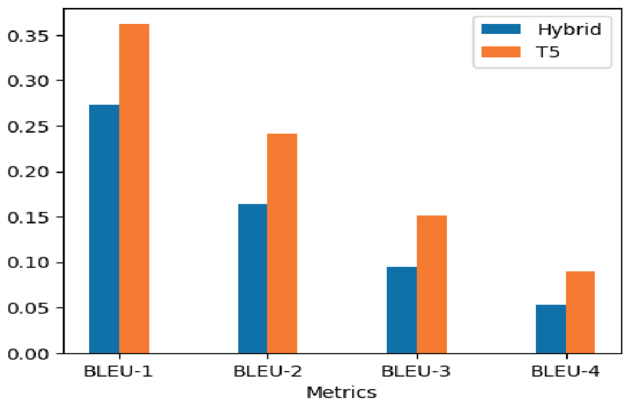


Fig. 7: Comparison of BLEU score

The generated summary of T5 and the hybrid model is slightly different. Some of them are almost the same. The sample of generated summary of both models is given below.

Example 1:

Text: এই সিদ্ধান্তের পর ইতিমধ্যে যেসব শিক্ষাপ্রতিষ্ঠান বেশি বেতন নিয়েছে, তাদের সরকারি সিদ্ধান্ত অনুযায়ী সমন্বয় করতে হবে বা ফেরত দিতে হবে। মাধ্যমিক ও উচ্চমাধ্যমিক শিক্ষা বোর্ডের প্রবিধানমালা অনুযায়ী, সরকারের নির্দেশনা অনুযায়ী সংশ্লিষ্ট শিক্ষাপ্রতিষ্ঠানের পরিচালনা কমিটি শিক্ষার্থীদের কাছ থেকে আদায়যোগ্য বেতন ও ফির হার নির্ধারণ করতে পারে। এ নিয়ে অভিভাবকেরা আন্দোলনে নেমেছেন। উইলস লিটল ফ্লাওয়ার স্কুল অ্যান্ড কলেজ নোটিশ দিয়ে বলেছে, শিক্ষা মন্ত্রণালয়ের পরবর্তী নির্দেশ না আসা পর্যন্ত নতুন ও পুরোনো সব ভর্তি কার্যক্রমের অংশ হিসেবে ব্যাংকে টাকা জমা স্থগিত থাকবে। ইতিমধ্যে যারা ভর্তি হয়েছে এবং ভর্তিযোগ্য হয়েও ভর্তি হতে পারেনি, তারাও ক্লাসে অংশ নিতে পারবে। গত রোববার শিক্ষা মন্ত্রণালয় এক আদেশে পরবর্তী নির্দেশ না দেওয়া পর্যন্ত বেসরকারি শিক্ষাপ্রতিষ্ঠানে শিক্ষার্থীদের বর্ধিত মাসিক বেতন ও অন্যান্য ফি বন্ধের নির্দেশ দেয়।

Reference summary: বেসরকারি স্কুলে বেতন ঠিক করে দেবে সরকার

Hybrid models' summary: বর্ধিত বেতন ও ফি স্থগিত, ক্লাসে অংশ নিতে পারবে শিক্ষার্থীরা

T5 models' summary: বেসরকারি শিক্ষাপ্রতিষ্ঠানে বেতন বাড়বে ২৫ শতাংশ

Example 2:

Text: ময়মনসিংহের ত্রিশাল উপজেলায় পুলিশের সঙ্গে সংঘর্ষে আওয়ামী লীগের দুজন নেতা-কর্মী গুলিবিদ্ধ হয়েছেন। তাঁদের মধ্যে ইস্রাফিল হোসেন (২২) নামের একজনকে ময়মনসিংহ মেডিকেল কলেজ হাসপাতালে ভর্তি করা হয়েছে। একপর্যায়ে পুরো ত্রিশাল বাসস্ট্যান্ড এলাকা রণক্ষেত্রে পরিণত হয়। প্রায় চার ঘণ্টা এ সংঘর্ষ চলে। এ সময় সাংসদ রেজা আলীর অফিসে আগুন ধরিয়ে দেয় দুর্বৃত্তরা। দফায় দফায় ককটেল বিস্ফোরণ ঘটানো হয়। পরিস্থিতি নিয়ন্ত্রণে আনতে পুলিশ ব্যাপক লাঠিপেটা, কাঁদানে গ্যাসের শেল ও গুলি ছোড়ে। এর আগে আজ বিকেল চারটার দিকে আওয়ামী লীগের মনোনয়নবঞ্চিত সাবেক সাংসদ রুহুল আমীন মাদানী ও উপজেলা আওয়ামী যুবলীগের সভাপতি পৌর মেয়র এ বি এম আনিছুজ্জামানের অনুসারী আওয়ামী লীগের কয়েক হাজার নেতা-কর্মী ত্রিশাল পৌর শহরের কয়েকটি পয়েন্টে অবস্থান নেন। সেখানে নারীরাও আহত হন। উপস্থিত আনিস ও মাদানীর অনুসারীরা জানান, বহিরাগত রেজা আলীর মনোনয়ন বাতিল না করা পর্যন্ত আন্দোলন চলবে। স্থানীয় সাতজন মনোনয়নপ্রত্যাশীর মধ্য থেকে যেকোনো একজনকে মনোনয়ন দিলে কারও কোনো আপত্তি থাকবে না বলেও তাঁরা জানান।

Reference summary: ত্রিশালে মনোনয়নবঞ্চিত আ.লীগ নেতার কর্মীদের সঙ্গে পুলিশের সংঘর্ষ

Hybrid models' summary: পুলিশের সঙ্গে সংঘর্ষে আ.লীগের দুই নেতা-কর্মী গুলিবিদ্ধ

T5 models' summary: পুলিশের সঙ্গে সংঘর্ষে আ.লীগের দুই নেতা-কর্মী গুলিবিদ্ধ

We have tested the models using news from newspapers that are outside our used dataset. The summaries generated by both models are entirely identical. The results are shown below:

Example 3:

Text: ছিনতাইয়ে বাধা দেওয়ায় রাজধানীর মোহাম্মদপুর এলাকায় ছিনতাইকারীর ছুরিকাঘাতে মো. রিয়াজ নামের এক কিশোর আহত হয়েছে। বুধবার রাতে চাঁদ উদ্যানে এ ঘটনা ঘটে। গুরুতর আহত অবস্থায় তাকে ঢাকা মেডিকেল কলেজ হাসপাতালে ভর্তি করা হয়েছে। সে মোহাম্মদপুরের একটি বেসরকারি স্কুলের সপ্তম শ্রেণির শিক্ষার্থী। পুলিশ জানায়, বুধবার রাত ৯টার দিকে বন্ধুর সঙ্গে ব্যাটারিচালিত অটোরিকশায় ঘুরতে বের হয় রিয়াজ। অটোরিকশাটি চাঁদ উদ্যানের সামনে পৌঁছামাত্র চার-পাঁচ জন ছিনতাইকারী ছুরির মুখে তাদের গতিরোধ করে। রিয়াজের বড় ভাই মো. মিরাজ সাংবাদিকদের বলেন, ছিনতাইকারীরা রিয়াজের মুঠোফোন ছিনিয়ে নিতে গেলে সে বাধা দেয়। এতে ছিনতাইকারীরা তাঁর পিঠে ছুরি মেরে পালিয়ে যায়। একপর্যায়ে ছিনতাইকারীরা রিয়াজের মুঠোফোন ছিনতাইয়ে ব্যর্থ হয়ে তার পিঠে ছুরি মেরে পালিয়ে যায়। পরে স্বজনেরা প্রথমে তাকে শহীদ সোহ্‌রাওয়ার্দী মেডিকেল কলেজ হাসপাতালে নিয়ে যায়। অবস্থার অবনতি হওয়ায় পরে ঢাকা মেডিকেল কলেজ হাসপাতালে স্থানান্তর করা হয় তাকে। ঢাকা মেডিকেল কলেজ হাসপাতাল পুলিশ ক্যাম্পের পরিদর্শক মো. বাচ্চু মিয়া ঘটনার সত্যতা নিশ্চিত করেছেন।

Reference summary: ছিনতাইয়ে বাধা দেওয়ায় স্কুলছাত্রকে ছুরিকাঘাত, হাসপাতালে ভর্তি

Hybrid models' summary: ছিনতাইয়ে বাধা দেওয়ায় ছুরিকাঘাতে কিশোর আহত

T5 models' summary: ছিনতাইয়ে বাধা দেওয়ায় ছুরিকাঘাতে কিশোর আহত

Although the hybrid model's ROUGH and BLEU scores are lower, the generated summaries are not wrong and irrelevant. The reason for this kind of result is that the extractive summaries that are the input of the T5 model are shorter than the original texts and do not include all sentences of the original texts. Due to the absence of some of the sentences, generating the summary is different from the reference summary. This hybrid is useful when the input text is extensive for the T5 model that we use in this experiment. By generating the extractive summary, the length of the text is reduced and only contains valuable sentences that are the input of the T5 transformer model.

VI. Conclusion and Future Work

In this paper, we compared the result of the T5 model and our proposed hybrid model on Bengali news articles. From the last example in section V-C, we can see that the T5-generated summary and hybrid model-generated summary exhibit complete similarity. By evaluating the results, we see that T5 performs better in generating summaries close to the reference. The f1 score of ROUGE-L is 28% higher than the hybrid model. The T5 outshines the hybrid model in the case of the BLEU-1 score also. T5 achieved 32% higher than the hybrid model. Though the T5 model is dominating the hybrid model, the hybrid model is useful when the input text is extremely large for the T5 model as the number of input tokens is limited for the T5 model. From the result, we can see that the generated summaries of the hybrid model are not wrong and contains the gist of the article.

There are some scopes for improvement. Currently, the texts are only taken from news articles. In the future, we will use more data from different online portals that are from different domains, such as medical texts, textbook paragraphs, etc. Instead of the BenSumm model for extractive summarization, we will use other techniques to shorten the extremely large texts that may help boost the performance.

References

[1] A. Karim, S. Azam, B. Shanmugam, K. Kannoorpatti, and M. Alazab, "A comprehensive survey for intelligent spam email detection," *IEEE Access*, vol. 7, pp. 168 261–168 295, 2019.

[2] P. Bhattacharjee, A. Mallick, and M. Saiful Islam, "Bengali abstractive news summarization (bans): a neural attention approach," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020*. Springer, 2021, pp. 41–51.

[3] R. R. Chowdhury, M. T. Nayeem, T. T. Mim, M. Chowdhury, S. Rahman, and T. Jannat, "Unsupervised abstractive summarization of bengali text documents," *arXiv preprint arXiv:2102.04490*, 2021.

[4] P. Sethi, S. Sonawane, S. Khanwalker, and R. B. Keskar, "Automatic text summarization of news articles," in *2017 International Conference on Big Data, IoT and Data Science (BID)*. IEEE, 2017, pp. 23–29.

[5] E. Zolotareva, T. M. Tashu, and T. Horváth, "Abstractive text summarization using transfer learning." in *ITAT*, 2020, pp. 75–80.

[6] J. L. E. K. Fendji, D. M. Taira, M. Atemkeng, and A. M. Ali, "Wats-sms: A t5-based french wikipedia abstractive text summarizer for sms," *Future Internet*, vol. 13, no. 9, p. 238, 2021.

[7] M. Bohra, P. Dadure, P. Pakray *et al.*, "Comparative analysis of t5 model for abstractive text summarization on different datasets," 2022.

[8] A. G. Etemad, A. I. Abidi, and M. Chhabra, "Fine-tuned t5 for abstractive summarization." *International Journal of Performability Engineering*, vol. 17, no. 10, 2021.

[9] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[11] A. Garg, S. Adusumilli, S. Yenneti, T. Badal, D. Garg, V. Pandey, A. Nigam, Y. K. Gupta, G. Mittal, and R. Agarwal, "News article summarization with pretrained transformer," in *Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India, December 5–6, 2020, Revised Selected Papers, Part I 10*. Springer, 2021, pp. 203–211.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[13] A. Kathed, S. Azam, B. Shanmugam, A. Karim, K. C. Yeo, F. De Boer, and M. Jonkman, "An enhanced 3-tier multimodal biometric authentication," in *2019 international conference on computer communication and informatics (ICCCI)*. IEEE, 2019, pp. 1–6.

[14] A. Al Shiam, S. M. Redwan, M. H. Kabir, and J. Shin, "A neural attention-based encoder-decoder approach for english to bangla translation," *Computer Science*, vol. 31, no. 1, p. 91, 2023.

[15] S. Azam, S. Montaha, A. R. H. Rafid, A. Karim, M. Jonkman, F. De Boer, G. McCallum, I. B. Masters, and A. Chang, "An automated broncho-arterial (ba) pair segmentation process and assessment of ba ratios in children with bronchiectasis using lung hrct scans: A pilot study," *Biomedicines*, vol. 11, no. 7, p. 1874, 2023.

[16] L. Q. Hu, A. Yadav, H. Liu, S. Azam, A. Karim, B. Shanmugam, A. H. Siddique, and M. Hasan, "Analysis of lemon company's cross-border e-commerce logistics distribution mode selection," in *LISS 2020: Proceedings of the 10th International Conference on Logistics, Informatics and Service Sciences*. Springer, 2021, pp. 601–615.

[17] M. Tareq, M. F. Islam, S. Deb, S. Rahman, and A. Al Mahmud, "Data-augmentation for bangla-english code-mixed sentiment analysis: Enhancing cross linguistic contextual understanding," *IEEE Access*, 2023.

[18] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.

[19] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text summarization branches out*, 2004, pp. 74–81.

[20] A. Khan, S. A. Ishita, F. Zaman, A. I. Ashik, and M. M. Hoque, "Intelligent combination of approaches towards improved bangla text summarization," in *2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)*. IEEE, 2023, pp. 1–6.

[21] K. M. Hasib, N. A. Towhid, K. O. Faruk, J. Al Mahmud, and M. Mridha, "Strategies for enhancing the performance of news article classification in bangla: Handling imbalance and interpretation," *Engineering Applications of Artificial Intelligence*, vol. 125, p. 106688, 2023.

[22] A. Ghadimi and H. Beigy, "Sgcsumm: An extractive multi-document summarization method based on pre-trained language model, submodularity, and graph convolutional neural networks," *Expert Systems with Applications*, vol. 215, p. 119308, 2023.