Natural Language Processing Journal 11 (2025) 100138

Contents lists available at ScienceDirect

Natural Language Processing Journal

journal homepage: www.elsevier.com/locate/nlp

BeliN: A novel corpus for Bengali religious news headline generation using contextual feature fusion

Md Osama [a,b], Ashim Dey [a], Kawsar Ahmed [a], Muhammad Ashad Kabir [c,*]

[a] Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Raozan, Chittagong, 4349, Bangladesh
[b] Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology (BAUST), Saidpur, Nilphamari, 5310, Bangladesh
[c] School of Computing, Mathematics and Engineering, Charles Sturt University, Bathurst, NSW, 2795, Australia

ARTICLE INFO

Dataset link: https://github.com/akabircs/BeliN

Keywords:
Bengali
Headline generation
Religious
News article
Feature fusion
Aspect
Sentiment
Transformer

ABSTRACT

Automatic text summarization, particularly headline generation, remains a critical yet under-explored area for Bengali religious news. Existing approaches to headline generation typically rely solely on the article content, overlooking crucial contextual features such as sentiment, category, and aspect. This limitation significantly hinders their effectiveness and overall performance. This study addresses this limitation by introducing a novel corpus, BeliN (Bengali Religious News) – comprising religious news articles from prominent Bangladeshi online newspapers, and *MultiGen* – a contextual multi-input feature fusion headline generation approach. Leveraging transformer-based pre-trained language models such as BanglaT5, mBART, mT5, and mT0, *MultiGen* integrates additional contextual features – including category, aspect, and sentiment – with the news content. This fusion enables the model to capture critical contextual information often overlooked by traditional methods. Experimental results demonstrate the superiority of *MultiGen* over the baseline approach that uses only news content, achieving a BLEU score of 18.61 and ROUGE-L score of 24.19, compared to baseline approach scores of 16.08 and 23.08, respectively. These findings underscore the importance of incorporating contextual features in headline generation for low-resource languages. By bridging linguistic and cultural gaps, this research advances natural language processing for Bengali and other under-represented languages. To promote reproducibility and further exploration, the dataset and implementation code are publicly accessible at https://github.com/akabircs/BeliN.

1. Introduction

A newspaper title serves as an essential element, often shaping the reader's first impression by being both representative of the content and attention-grabbing (Cai et al., 2023). Representative headlines play a crucial role in information retrieval systems, which prioritize keywords in headlines to enhance searchability and relevance (De Francisci Morales et al., 2012). Consequently, researchers have extensively explored automated techniques, such as text summarization, to generate compelling and accurate headlines from articles (Koh et al., 2022).

Text summarization systems can be categorized into two main approaches: extractive and abstractive techniques (Rao et al., 2024). Early text summarization techniques predominantly relied on extractive processes, which identify and select significant portions of text, such as key phrases or sentences, directly from the document. These methods generate summaries by reproducing the most critical points verbatim, ensuring fidelity to the original content (Banerjee et al., 2023; Giarelis et al., 2023). Abstractive text summarization, a more recent development in the field of Automatic Text Summarization (ATS), generates concise summaries by reformulating key ideas from the original text. Unlike extractive methods, which replicate exact phrases, abstractive summarization produces a condensed script that captures the essence of the document in a clear and coherent manner (Alomari et al., 2022; Rao et al., 2024). ATS aims to identify and appropriately prioritize the informative components of the source article, making it particularly useful for summarizing blogs, newspapers, and other text-based media (El-Kassas et al., 2021).

Headline generation, a specialized form of text summarization, can also employ both extractive and abstractive approaches. However, compared to the extractive technique, the abstractive method more accurately generates real-world headlines because it captures the underlying meaning and context of the content, allowing for greater flexibility and creativity in rephrasing, while the extractive approach

* Correspondence to: Charles Sturt University, Panorama Ave, Bathurst, NSW 2795, Australia.
E-mail addresses: osama_cse@baust.edu.bd (Md Osama), ashim@cuet.ac.bd (A. Dey), u1804017@student.cuet.ac.bd (K. Ahmed), akabir@csu.edu.au (M.A. Kabir).

https://doi.org/10.1016/j.nlp.2025.100138
Received 31 December 2024; Received in revised form 22 February 2025; Accepted 8 March 2025

tends to rely on directly selecting portions of the input text (Ahuir et al., 2024). Traditionally, headlines are generated solely based on the content of the article (Ayana et al., 2017), a method that, while effective in capturing key information, limits the potential for engaging and compelling headline creation (Hagar and Diakopoulos, 2019). This approach tends to focus purely on summarizing the main points, often neglecting the broader context or emotional resonance required to capture a reader's attention in today's fast-paced media environment (Banerjee and Urminsky, 2024). Furthermore, while there has been extensive research in headline generation for high-resource languages such as English (El-Kassas et al., 2021), relatively few studies have focused on Bengali, particularly in the context of Bengali religious news (Akash et al., 2023; Saad et al., 2024). One notable study, Shironaam (Akash et al., 2023), focused on Bengali and incorporated additional contextual information, such as category, in the headline generation process. However, their sample of religious news was very small, and they did not consider other significant contextual elements, such as sentiment and aspect.

As of 2024, Bengali is spoken by over 237 million native speakers and an additional 41 million second-language speakers, making it the fifth most spoken native language and the seventh most spoken language worldwide (Eberhard et al., 2024). This widespread usage highlights the importance of Bengali in global linguistic diversity. Moreover, the scarcity of sufficient religious content and the sensitivity associated with religious discourse motivate us to explore this domain. Religious content is particularly delicate, as even a minor change in a headline can affect emotions and religious beliefs. Therefore, ensuring accurate and context-aware summarization of Bengali religious news is crucial.

In this paper, we introduce a novel Bengali religious news corpus, named *BeliN*, with a multi-input approach that incorporates additional features alongside the news content to generate more accurate and contextually relevant headlines. Specifically, we include the article's category, content aspect, and sentiment as additional input features, as illustrated in Fig. 1. This approach aims to narrow the domain and enhance the quality of the generated headlines. The rationale behind this method is that a headline should not only capture the essence of the article but also reflect its main context concisely. By integrating multiple inputs, we seek to provide a more comprehensive understanding of the article's content, thereby producing more precise and relevant headlines. The inclusion of these features in the headline generation process is expected to improve the model's performance and better reflect the underlying article's themes (Akash et al., 2023). This multi-input approach not only narrows the domain but also enriches the generated headlines with additional contextual information, making them more informative and representative of the article.

By leveraging these additional inputs, we aim to create a more robust and accurate headline-generation system for Bengali news articles, particularly within the domain of religious news. The proposed Bengali news headline generation system, named *MultiGen*, employs and evaluates state-of-the-art pretrained transformer models such as mT5, mT0, mBART, and BanglaT5. Our experimental evaluations show that BanglaT5 outperforms others, offering significant improvements in headline accuracy and contextual relevance. The key Contributions of this work are as follows:

- We have developed a novel dataset, *BeliN*, focused on the underexplored domain of religious news in the low-resource Bengali language. The dataset comprises 2,520 news articles and their corresponding headlines, making it valuable for various NLP tasks, including headline generation, text summarization, news categorization, sentiment analysis, and aspect classification.
- We have proposed the *MultiGen* approach, which incorporates additional features such as category, aspect, and sentiment as auxiliary information to enhance the headline generation process. Our approach demonstrates the importance of integrating multiple inputs, achieving significant improvements over the baseline approach.
- We have employed and evaluated the performance of state-of-the-art pre-trained models for generating compelling and attention-grabbing news headlines.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of the related work that forms the foundation of this research. Section 3 introduces the BeliN corpus, presenting its key characteristics and statistical summary. The methodology and approaches adopted in this study are detailed in Section 4. Section 5 discusses the experimental setup, metrics, and models and presents the evaluation results. A detailed analysis, including a discussion of the findings and limitations, is presented in Section 6. Finally, Section 7 concludes the paper and outlines potential future directions.

2. Related work

Generating news headlines has been a prominent area of research within the field of natural language processing (NLP) (Shaibani and Elnagar, 2024; Zeyad and Biradar, 2024). Although significant advancements have been made in text summarization (Yadav et al., 2023; Bharathi Mohan et al., 2023; Cajueiro et al., 2023) and headline generation (Ayana et al., 2017; Liu et al., 2018), progress in low-resource languages, such as Bengali, remains limited (Akash et al., 2023; Salehin et al., 2019; Hayat et al., 2023; Kabir et al., 2024). The task of generating headlines, particularly for religious news in Bengali, poses unique challenges due to the scarcity of annotated datasets (Akash et al., 2023).

In high-resource languages like English and Chinese, a variety of datasets have supported significant advancements in summarization and headline generation research summarized in Table 1. Numerous datasets, including Newsroom (Grusky et al., 2018), CNN Daily Mail (Nallapati et al., 2016), CNN Corpus (Lins et al., 2019), and CCSum (Jiang and Dreyer, 2024), have been developed for summarization in English. While some of these datasets are publicly accessible, others are privately listed, as listed in Table 1. Headline generation, as a specific subset of abstractive summarization, has also benefited from datasets, such as XSum (Narayan et al., 2018), NewSHead (Gu et al., 2020), PENS (Ao et al., 2021), New York Times (Sandhaus, 2008), DUC (National Institute of Standard and Technology, 2014), and Gigaword (Graff and Cieri, 2003; Napoles et al., 2012), all designed for English. Most of these datasets utilize textual content to produce summaries or headlines, with two datasets, NewSHead (Gu et al., 2020) and PENS (Ao et al., 2021), explicitly incorporating category as a feature. While the majority of these works emphasize the abstractive approach, some also explore the extractive methodology.

Beyond English, several datasets have been developed for abstractive headline generation in other languages. These include SuDer (Sen and Yanikoglu, 2018) in Turkish, AFRIHG (Ogunremi et al., 2024) in African, RIA (Gavrilov et al., 2019) and Lenta (Yutkin, 2019) in Russian, and Mukhyansh (Madasu et al., 2023) and Varta (Aralikatte et al., 2023) in Indic languages. Additionally, LCSTS (Hu et al., 2015) has been created in Chinese for abstractive summarization.

While most existing studies focus primarily on utilizing article content alone for headline generation, few studies have explored integrating additional features to enhance the quality of generated headlines. These additional features include leveraging metadata, social media interactions, user engagement data, and personalization techniques. Iwama and Kano (2019) used page metadata – such as headline position, font size, and article page number – to generate multiple headlines for a single article, and generated headlines align with the article's visual presentation. CLH3G (Liu et al., 2022) enhances headline generation by integrating the author's historical headlines to capture and maintain their writing style. By leveraging contrastive learning, their approach effectively learns stylistic features, ensuring that generated headlines align not only with the article content but also with the author's unique style. MEBART (Fang et al., 2024) designed to generate

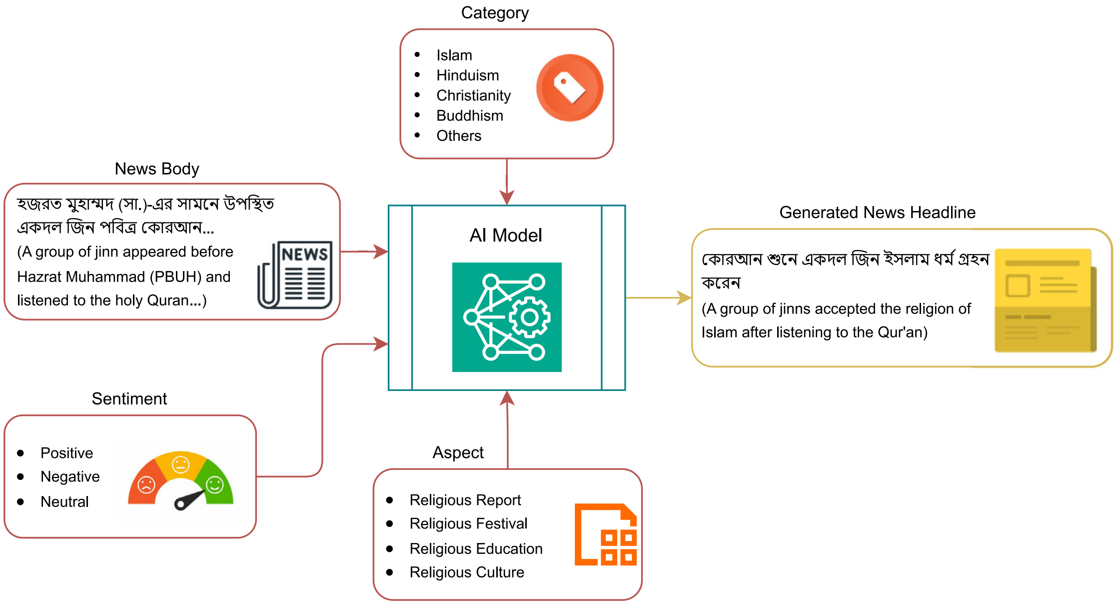


Fig. 1. An overview of multi-input headline generation.

Table 1
Summary of the related work.

Study	Dataset/Corpus				Feature				Task	Approach
	Name	Language	Religious	Availability	Content	Category	Aspect	Sentiment		
Grusky et al. (2018)	Newsroom	English	×	Public[a]	✓	×	×	×	S	A & E
Nallapati et al. (2016)	CNN Daily Mail	English	×	Public[b]	✓	×	×	×	S	A
Lins et al. (2019)	CNN Corpus	English	×	Private	✓	×	×	×	S	E
Jiang and Dreyer (2024)	CCSum	English	×	Public[c]	✓	×	×	×	S	A
Narayan et al. (2018)	XSum	English	×	Public[l]	✓	×	×	×	H	A
Gu et al. (2020)	NewSHead	English	×	Public[d]	✓	✓	×	×	H	A
Ao et al. (2021, 2023)	PENS	English	×	Public[e]	✓	✓	×	×	H	A
Jin et al. (2020)	CNN, New York Times (Sandhaus, 2008)	English	×	Private	✓	×	×	×	H	A
Takase et al. (2016)	DUC (National Institute of Standard and Technology, 2014), Gigaword (Graff and Cieri, 2003; Napoles et al., 2012)	English	×	Public[f,g]	✓	×	×	×	H	A
Cai et al. (2023)	Newsroom (Grusky et al., 2018), Gigaword (Graff and Cieri, 2003; Napoles et al., 2012)	English	×	Public[a,g]	✓	×	×	×	S	A & E
Singh et al. (2021)	Newsroom (Grusky et al., 2018), Gigaword (Graff and Cieri, 2003; Napoles et al., 2012), CNN Daily Mail (Nallapati et al., 2016)	English	×	Public[a,g,b]	✓	×	×	×	S	A
Sen and Yanikoglu (2018)	SuDer	Turkish	×	Private	✓	×	×	×	H	A
Ogunremi et al. (2024)	AFRIHG	African	×	Private	✓	×	×	×	H	A
Bukhtiyarov and Gusev (2020)	RIA (Gavrilov et al., 2019), Lenta (Yutkin, 2019)	Russian	×	Public[h,i]	✓	×	×	×	H	A
Hu et al. (2015)	LCSTS	Chinese	×	Public[j]	✓	×	×	×	S	A
Madasu et al. (2023)	Mukhyansh	Indic languages	×	Private	✓	×	×	×	H	A
Aralikatte et al. (2023)	Varta	Indic, English	×	Public[k]	✓	×	×	×	H	A
Li et al. (2021)	LCSTS (Hu et al., 2015), XSum (Narayan et al., 2018)	Chinese, English	×	Public[j,l]	✓	×	×	×	H	A
Gavrilov et al. (2019)	RIA, New York Times (Sandhaus, 2008)	Russian, English	×	Private (NYT)	✓	×	×	×	H	A
Salehin et al. (2019)	Own dataset	Bengali	×	Private	✓	×	×	×	H	A

(continued on next page)

Table 1 (*continued*).

Study	Dataset/Corpus				Feature				Task	Approach
	Name	Language	Religious	Availability	Content	Category	Aspect	Sentiment		
Hasan et al. (2021)	XL-Sum	Bengali+43 others	×	Public[m]	✓	×	×	×	S	A
Ahmad et al. (2022)	Potrika	Bengali	×	Public[n]	✓	✓	×	×	H	A
Saad et al. (2024)	BNAD	Bengali	✓	Public[o]	✓	✓	×	×	H	A
Akash et al. (2023)	Shironaam	Bengali	✓	Public[p]	✓	✓	×	×	H	A
Ours	BeliN	Bengali	✓	Public[q]	✓	✓	✓	✓	H	A

S = Summarization, H = Headline, A = Abstractive, E = Extractive.

[a] https://paperswithcode.com/dataset/newsroom.

[b] https://www.kaggle.com/datasets/gowrishankarp/newspaper-text-summarization-cnn-dailymail.

[c] https://github.com/amazon-science/ccsum.

[d] https://github.com/google-research-datasets/NewSHead.

[e] https://msnews.github.io/pens.html.

[f] https://www-nlpir.nist.gov/projects/duc/data.html.

[g] https://www.kaggle.com/datasets/arngowda/gigaword-corpus.

[h] https://github.com/RossiyaSegodnya/ria_news_dataset.

[i] https://github.com/yutkin/Lenta.Ru-News-Dataset.

[j] https://huggingface.co/datasets/hugcyp/LCSTS.

[k] https://github.com/rahular/varta.

[l] https://github.com/EdinburghNLP/XSum/tree/master/XSum-Dataset.

[m] https://github.com/csebuetnlp/xl-sum.

[n] https://doi.org/10.17632/v362rp78dc.4.

[o] https://doi.org/10.5281/zenodo.11069882.

[p] https://github.com/dialect-ai/BenHeadGen.

[q] https://github.com/akabircs/BeliN.

headlines optimized for social media by capturing prevalent trends and user preferences, producing headlines that are more engaging for online audiences. The FPG framework (Yang et al., 2023) enhances personalized news headline generation by balancing user-specific preferences with factual consistency. It leverages similarity-based attention to key facts and employs contrastive learning to improve headline accuracy.

In the Bengali language, several datasets have been developed for summarization and headline generation tasks. XL-Sum (Hasan et al., 2021), a multilingual dataset covering 44 languages, includes Bengali and focuses on abstractive summarization using news content. Potrika (Ahmad et al., 2022) stands out as a substantial dataset offering a large collection of Bengali news samples with different categories, excluding religious news. BNAD (Saad et al., 2024) is another notable dataset in Bengali, which integrates category information with news articles for headline generation, though it includes only 1,275 religious news samples across its domains. Shironaam (Akash et al., 2023) is a large-scale dataset encompassing 13 news domains, including religious news, albeit with limited samples in this domain.

Among the previously mentioned research efforts, none of the non-Bengali datasets include the religious news domain. While some Bengali datasets, such as BNAD (Saad et al., 2024) and Shironaam (Akash et al., 2023), incorporate religious news along with categories, they lack subdivisions within the religious content, aspect, and sentiment of a news article. In our proposed dataset, *BeliN*, we address this gap by focusing on the less explored religious news domain. We have compiled religious news samples from various online newspapers, integrating additional features such as category, aspect, and sentiment. The dataset includes five religious categories: Islam, Hinduism, Christianity, Buddhism, and others. Furthermore, it captures four distinct aspects: religious reports, festivals, education, and culture. Additionally, we annotate the sentiment polarity of the news articles as positive, negative, or neutral to further enhance the dataset's utility.

Numerous headline-generation systems have been developed utilizing various datasets, with the majority relying solely on content as input for generating headlines (Ao et al., 2023; Karaca and Aydın, 2023; Ogunremi et al., 2024; Bukhtiyarov and Gusev, 2020; Jin et al., 2020; Gu et al., 2020; Napoles et al., 2012; Cai et al., 2023; Singh et al., 2021; Li et al., 2021; Sandhaus, 2008). This content-only approach has also been widely adopted in Bengali, as seen in research on text summarization (Hasan et al., 2021), and headline generation (Salehin et al., 2019) that utilize custom datasets. In the news content-only approach, the absence of linguistic context (*i.e.*, the surrounding language or text that provides clarification and deeper meaning beyond the literal interpretation of words (Theledi and Pule, 2024)) along with the lack of additional guidance, often leads to limited headline diversity and challenges in evaluation due to reliance on a single ground truth (Liu et al., 2020a). To address this limitation, incorporating additional contextual features is crucial for generating more nuanced and linguistically informed headlines. While Shironaam (Akash et al., 2023) introduced a multi-input framework by incorporating category and image captions alongside content, our proposed *MultiGen* approach goes a step further by integrating contextual features such as aspect and sentiment, in addition to content and category. By leveraging sentiment to reflect the natural tone of the news, our method aims to generate more robust and contextually accurate headlines. The *MultiGen* approach demonstrates superior performance compared to the baseline content-only systems, highlighting its effectiveness over existing methods like Shironaam (Akash et al., 2023).

In summary, while headline generation has seen substantial progress in high-resource languages, research in low-resource languages like Bengali remains underexplored. This study bridges the gap by introducing *BeliN*, a specialized dataset for Bengali religious news, and *MultiGen*, a state-of-the-art approach tailored for this domain. Together, they lay the groundwork for broader advancements in low-resource language processing and headline generation.

3. The BeliN corpus

This section provides a comprehensive overview of the *BeliN* corpus,[1] a meticulously curated dataset designed to advance the task of Bengali news headline generation. The corpus development process, as illustrated in Fig. 2, follows a structured approach that integrates raw data collection, labeling, and statistical analysis to ensure a high-quality and contextually enriched dataset. The process begins with sourcing raw data from diverse Bengali news websites and religious news portals to achieve a representative dataset. This is followed by a detailed labeling methodology, where additional metadata such as categories,

[1] https://github.com/akabircs/BeliN

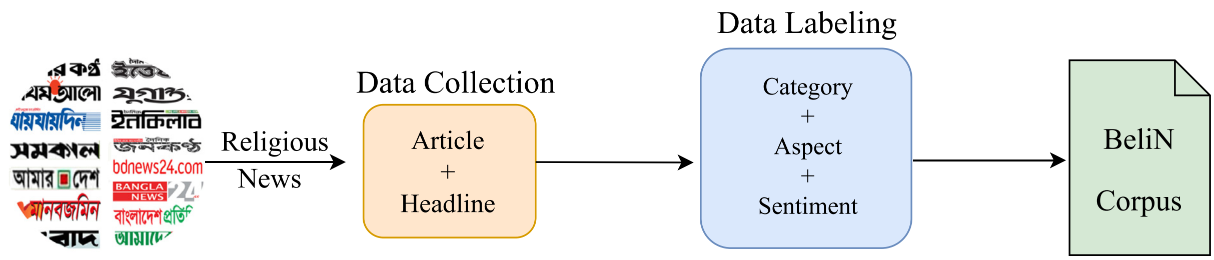


Fig. 2. An overview of *BeliN* corpus development process.

Table 2
List of newspapers.

Newspaper	URL
Prothom alo	https://www.prothomalo.com/religion
Kaler kantho	https://www.kalerkantho.com/online/Islamic-lifestylie
Bangladesh pratidin	https://www.bd-pratidin.com/islam
NayaDiganta	https://www.dailynayadiganta.com/diganta-islami-jobon/133
Jugantor	https://www.jugantor.com/all-news/islam-life
Daily Ittefaq	https://www.ittefaq.com.bd/religion
Samakal	https://samakal.com/search?q=religion
Dhaka Tribune	https://www.dhakatribune.com/topic/religion
Bhorer Kagoj	https://www.bhorerkagoj.com/religion
Jai Jai Din	https://www.jaijaidinbd.com/islam-and-religion
Alokito Bnagladesh	https://www.alokitoBengalidesh.com/islam
Daily Inqilab	https://dailyinqilab.com/islamic-world
Daily Vorer Pata	https://www.dailyvorerpata.com/cat.php?cd=293
Daily Khabar Patra	https://khoborpatrabd.com/?s=religion

sentiments, and aspects are assigned to enrich the data. These auxiliary features enhance the context for generating precise and meaningful headlines. The final stage involves a thorough statistical analysis to evaluate the dataset's composition and suitability for training and evaluating generative models. Each of these steps is systematically discussed in the following subsections

3.1. Raw data collection

The raw data collection for the *BeliN* corpus was conducted with the objective of creating a robust dataset for Bengali news headline generation, specifically targeting religious news. Articles and corresponding headlines were manually gathered from a diverse set of Bengali news websites and religious news portals, ensuring the inclusion of high-quality and contextually relevant data. The sources for these articles are listed in Table 2, reflecting a wide coverage of topics and perspectives within the religious domain.

The manual collection approach was crucial in ensuring the integrity of the data, particularly in capturing nuanced contexts and maintaining linguistic authenticity. Unlike automated scraping techniques, this method allowed for the careful selection of articles and headlines that aligned with the focus of the BeliN corpus. This step laid the groundwork for the subsequent labeling process, where additional metadata was assigned to further enrich the dataset's contextual depth.

3.2. Dataset labeling

Each article in the BeliN corpus was meticulously labeled to ensure accuracy and consistency. The dataset includes category, aspect, and sentiment information along with the article text to aid in headline generation. While aspect and sentiment were manually annotated by independent annotators, the category of religious affiliation was directly sourced from the respective news portals' classification. To maintain the reliability of the extracted information, we implemented a two-step annotation process. Initially, two independent annotators labeled the collected articles based on predefined criteria outlined below for aspect and sentiment polarity, ensuring consistency and reducing subjective bias. A third annotator carefully reviewed their annotations and resolved any discrepancies through consensus. This rigorous approach ensured the dataset's accuracy, consistency, and structural integrity. The final corpus comprises a diverse collection of religious news articles from various Bengali sources, systematically organized into five key columns, as detailed below.

i. Article: The full text of the news article.
ii. Headline: The original headline of the news article.
iii. Category: The religious affiliation of the news article — Islam, Hinduism, Christianity, Buddhism, Others.
iv. Aspect: This identifies the specific focus or theme of the article's content, which can include categories such as religious reports, festivals, education, culture, or other related topics, which can be one of the following:

   (a) Religious Report: Religious reports typically encompass a range of religious discussions, including sacred events and mythological and religious tales. These reports highlight significant news about religious communities, broader spiritual perspectives, and important religious philosophies. For example, "In Pakistan, a church was vandalized and set on fire. Two members of the community were arrested for blasphemy".[2]
   (b) Religious Festival: This aspect includes news about religious festivals, ceremonies, and rituals, emphasizing major events across different religious communities. It covers key aspects such as festivals, worship, and other religious practices. During festive periods, updates on events from various religious communities are also featured. For example, "Durga Puja is celebrated with grandeur in Abu Dhabi".[3]
   (c) Religious Education: This aspect focuses on religious education and spiritual growth. It highlights updates from various educational institutions, religious schools, and

[2] https://www.prothomalo.com/world/pakistan/i3riizd976
[3] https://www.bd-pratidin.com/probash-potro/2023/10/22/932700

policies on religious education. For example, "Those deeds by which one can attain paradise with the Prophet Muhammad (peace be upon him)".[4]

(d) Religious Culture: Under this news aspect, reports typically focus on religious culture and significant religious figures. This category highlights news related to religious personalities, mythologies, and cultural events of religious significance. For example, "A 10-day Islamic book fair begins in Mymensingh".[5]

v. Sentiment: It represents the sentiment of the article (Sharma et al., 2024), which can be classified as:

(a) Positive: This sentiment label signifies that the article's content expresses favorable, supportive, or optimistic views toward the subject matter. For example, "Faith grows through contemplation and research".[6]

(b) Negative: This sentiment label denotes that the article's content conveys unfavorable, critical, or pessimistic views toward the subject matter. For example, "The national wealth is at risk of self-destruction",[7]

(c) Neutral: This sentiment label indicates that the article's content maintains an impartial, balanced, or indifferent stance without showing strong positive or negative opinions. For example, "The rare coin of the Islamic era in Saudi Arabia".[8]

A sample of the dataset has been given in Table 3. The BeliN corpus captures the complexity of religious news, reflecting diverse aspects and sentiments within the articles.

3.3. Dataset statistics

This subsection provides a detailed statistical analysis of the BeliN corpus, highlighting its composition and diversity. The dataset, specifically curated for religious news, spans multiple categories, aspects, and sentiment polarities. Such granularity ensures the dataset's utility for training and evaluating generative models, offering a rich context for generating headlines. Table 4 presents the statistics of the BeliN corpus, which encompasses religious news across different categories. It includes counts for five major categories. Each category is further analyzed based on four aspects, and sentiment counts (positive, negative, and neutral) are provided for each aspect. The table concludes with a total count of 2520 entries in the dataset, with individual category counts distributed accordingly.

Table 5 compares the features of Shironaam and BeliN in the religious domain. BeliN includes additional features like aspect and sentiment, while Shironaam includes topic words and image captions. BeliN also has a larger number of samples (2520 news) compared to Shironaam (294 news).

Table 6 shows the quantitative statistics of Shironaam and BeliN based on the average number of words, sentences, and vocabulary size. The BeliN dataset demonstrates strong novelty in its n-grams, with 4.42% novel unigrams, 21.48% novel bigrams, 42.10% novel trigrams, and 56.47% novel 4-grams, as shown in Table 7. These results highlight the distinctiveness of headlines in comparison to the articles. While Shironaam shows slightly higher percentages of novel n-grams, BeliN still provides valuable insights into Bengali religious news, showcasing significant diversity in language use. This makes BeliN a valuable resource for research in Bengali language processing. The figures presented illustrate the distribution of article and headline lengths in the dataset. Fig. 3(a) shows the frequency of article lengths, measured in words, revealing the common word counts for articles. Fig. 3(b) depicts the frequency of headline lengths, also measured in words, providing insight into the typical brevity or elaboration of headlines compared to the full articles. Together, these figures offer a visual representation of the structure and variation in article and headline lengths within the dataset.

The rich contextual information embedded in the BeliN corpus has significant potential for various natural language processing (NLP) tasks beyond headline generation. These include text generation, topic modeling, news categorization, and news headline sentiment analysis within the realm of religious news. By leveraging the detailed annotations and multi-aspect nature of the dataset, researchers and developers can create more sophisticated and human-like AI systems capable of understanding and generating content with high contextual awareness.

4. The *MultiGen* approach

The traditional news content-only approach to headline generation faces several challenges. Relying solely on the news content often results in a lack of linguistic context and guidance, limiting headline diversity and creating evaluation difficulties due to dependence on a single ground-truth reference (Liu et al., 2020a). In this approach, the model is designed to take solely news content as input to generate headlines, which may not capture the full spectrum of possible interpretations or nuances of the article. Additionally, without the inclusion of contextual features, generated headlines may fail to align with the emotional tone or thematic aspects of the content, diminishing their relevance and overall quality.

To overcome these limitations, the *MultiGen* approach introduces a multi-input framework for Bengali news headline generation. Unlike the conventional approach that relies exclusively on news content, *MultiGen* integrates additional contextual features such as aspect, category, and sentiment. These features provide a more comprehensive understanding of the article, allowing the model to generate headlines that are contextually relevant and linguistically informed. Aspect and category help the model focus on specific narrative elements, while sentiment captures the emotional tone, ensuring that the generated headlines are better aligned with the article's mood and message. This enriched approach improves both the diversity and quality of the generated headlines. Incorporating contextual features has also proven effective in other similar NLP tasks, including text classification (Kiefer, 2022), information retrieval (Chen, 2019), and sentiment analysis (Zhu et al., 2023; Aziz et al., 2025).

By incorporating these additional features, *MultiGen* enhances the model's ability to understand the nuances of the article, producing headlines that are both informative and contextually aligned with the content's emotional tone. The overall framework of *MultiGen*, as illustrated in Fig. 4, showcases the seamless integration of these diverse inputs within an encoder–decoder architecture, enabling more accurate and contextually aware headline generation. The remainder of this section provides a detailed description of the *MultiGen* approach, starting with the preprocessing steps, followed by the fusion of multiple inputs, and concluding with the encoder–decoder architecture used to achieve enhanced headline generation performance.

4.1. Preprocessing

Preprocessing is a critical phase that prepares raw text data for input into generative models, ensuring the data is clean, consistent, and ready for effective processing. It involves two main tasks:

[4] https://www.kalerkantho.com/online/Islamic-lifestylie/2023/10/25/1330109

[5] https://www.kalerkantho.com/online/Islamic-lifestylie/2023/10/05/1324183

[6] https://www.kalerkantho.com/online/Islamic-lifestylie/2023/10/06/1324389

[7] https://www.kalerkantho.com/online/Islamic-lifestylie/2023/09/02/1314294

[8] https://www.kalerkantho.com/online/Islamic-lifestylie/2023/09/10/1316757

Table 3
Samples of the BeliN corpus.

Article	Headline	Category	Aspect	Sentiment
দীর্ঘ একটি মাস পবিত্র মাহে রমজানে ইবাদত- বন্দেগি করে প্রত্যেক মুমিন হৃদয় প্রশান্তিতে ভরপুর। আল্লাহতায়ালা মানব জাতিকে সৃষ্টি করেছেন একমাত্র তাঁরই ইবাদতের জন্য।..........(After a month of worship and devotion during the holy month of Ramadan, every believer's heart is filled with peace. Allah, the Almighty, created humankind solely for His worship....)	শাওয়াল মাসের ছয় রোজার মর্যাদা (The dignity of the six fasts of the month of Shawwal)	Islam	Education	Positive
হালাল জীবিকা উপার্জন মানুষের জীবনের অপরিহার্য অংশ। ব্যবসা-বাণিজ্য জীবিকা উপার্জনের অন্যতম একটি মাধ্যম। বর্ণিত আছে, ‘দশ ভাগের নয় অংশ জীবিকা ব্যবসা থেকে ....(Earning a halal livelihood is an essential part of a person's life. Business and trade are among the primary means of earning a livelihood. It is mentioned, ``Nine out of ten parts of sustenance come from business .....")	ব্যবসা মুসলিমের মর্যাদার প্রতীক (Business is a symbol of Muslim status)	Islam	Culture	Positive
বিদ্যার দেবীর আরাধনায় বুধবার (১৪ ফেব্রুয়ারি) উৎসবমুখর পরিবেশে দেশে উদযাপিত হচ্ছে সরস্বতী পূজা। হিন্দু সম্প্রদায়ের অন্যতম ধর্মীয় এ উৎসবে অগণিত ভক্ত বিদ্যা ও জ্ঞানের অধিষ্ঠাত্রী দেবী সরস্বতীর চরণে...........(The worship of the goddess of knowledge, Saraswati Puja, is being celebrated in a festive atmosphere across the country on Wednesday (February 14). In this significant religious festival of the Hindu community, countless devotees are offering prayers at the feet of Goddess Saraswati, the deity of wisdom and knowledge...)	বিদ্যার দেবীর আরাধনায় সরস্বতী পূজা উদযাপন (Celebrating Saraswati Puja in worship of Goddess Vidya)	Hinduism	Festival	Positive
বাংলাদেশ বিশ্বের এশিয়া মহাদেশ ও ভারতীয় উপমহাদেশভুক্ত একটি স্বাধীন ও সার্বভৌম মুসলিম রাষ্ট্র। খ্রিস্টীয় ষষ্ঠ শতকে মরুময় আরবে ইতিহাসের `আইয়ামে জাহেলিয়া' তথা জেহালত .......(Bangladesh is an independent and sovereign Muslim state located in the continent of Asia and part of the Indian subcontinent. In the 6th century AD, in the desert lands of Arabia, there was a historical period known as the `Age of Ignorance' or `Ayam-e-Jahiliyyah'...)	বঙ্গবন্ধুর অসাম্প্রদায়িক ও আন্তঃধর্মীয় চেতনা (Bangabandhu's non-communal and inter-religious spirit)	Others	Report	Neutral

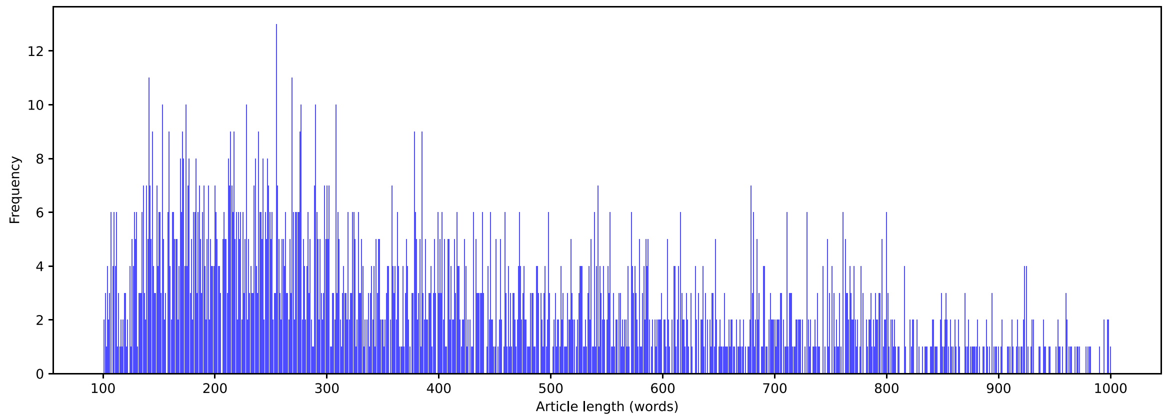
Table 4
Descriptive statistics of the *BeliN* corpus.

Category	Aspect				Sentiment			Total
	Report	Festival	Education	Culture	Positive	Negative	Neutral	
Islam	860	68	890	183	1457	299	245	2001
Hinduism	135	67	16	24	128	58	56	242
Christianity	7	12	7	2	19	5	4	28
Buddhism	12	13	1	3	25	3	1	29
Others	190	1	16	13	88	90	42	220
Total	1204	161	930	225	1717	455	348	2520

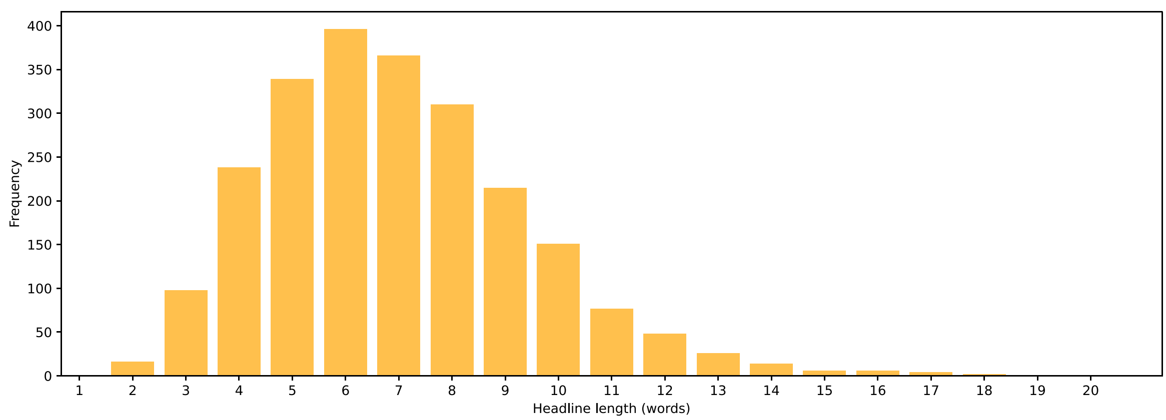
- Text Normalization: This step uses the BUET normalizer (Hasan et al., 2020) to standardize characters with Unicode NFKC. Non-textual elements like URLs and emojis are removed, excessive whitespace is managed, and redundant punctuation characters are reduced.
- Input Processing: Text is formatted for model training by adding task-specific prefixes, such as “Summarize the Article as Headlines”, to guide the model. Appropriate tokenizers, like AutoTokenizer for BanglaT5 and mBART, are used to process articles. The text is truncated to 512 tokens for input and 64 tokens for headlines, ensuring computational efficiency. Finally, tokenized inputs and labels are combined into a dictionary for training, enhancing coherence and output quality.

4.2. Fusing article with category, aspect, and sentiment

The proposed multi-input approach enhances Bengali news headline generation by integrating multiple contextual signals—article (**A**), category (**C**), aspect (**P**), and sentiment (**S**)—into a unified input sequence. By leveraging this enriched input, the model benefits from a broader



(a)



(b)

Fig. 3. Distribution of article and headline lengths in the dataset. (a) Frequency of article lengths (in words), illustrating the common word counts for articles. (b) Frequency of headline lengths (in words), highlighting the typical brevity or elaboration of headlines compared to the full articles.

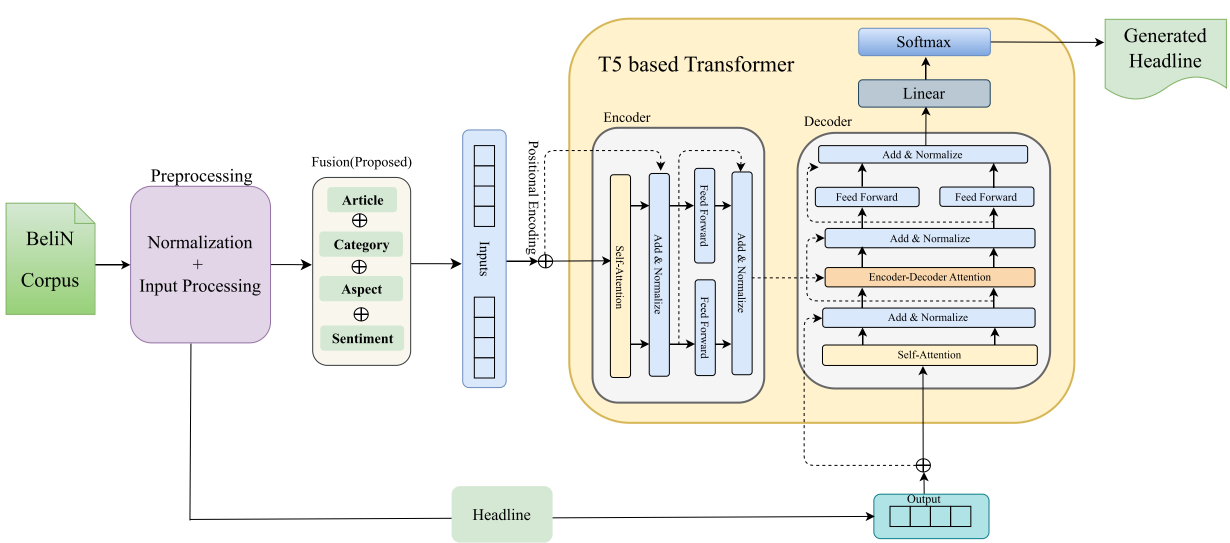


Fig. 4. *MultiGen* architecture for headline generation.

contextual understanding, enabling it to generate more precise and contextually aligned headlines. This approach constructs the input sequence by fusing these components with a [SEP] token, ensuring a clear distinction between different elements while preserving their individual contributions. In contrast to the baseline approach, which relies solely on article content, the proposed fusion approach incorporates additional contextual elements, significantly enhancing the model's capacity to produce contextually nuanced headlines. The fused input sequence

Table 5
Feature comparison of the *BeliN* dataset and Shironaam.

Features	Shironaam (Akash et al., 2023)	BeliN (this study)
Article	✓	✓
Headline	✓	✓
Category	✓	✓
Aspect	×	✓
Sentiment	×	✓
Topic words	✓	×
Image caption	✓	×
Total Samples	294[a]	2520

[a] Religious news.

F is defined as:

$$\mathbf{F} = [A_1, \dots, A_n, [\text{SEP}], C_1, \dots, C_k, [\text{SEP}], P_1, \dots, P_l, [\text{SEP}], S_1, \dots, S_m] \quad (1)$$

where $A_1, \dots, A_n$ represents the sequence of tokens from the news article, forming the primary textual content; [SEP] is a special separator token used to distinguish different segments of the input; $C_1, \dots, C_k$ represents the sequence of tokens corresponding to the category of the news article (*i.e.*, "Islam"); $P_1, \dots, P_l$ represents the sequence of tokens for the aspect of the news (*i.e.*, "Education" in a business-related article); and $S_1, \dots, S_m$ represents the sequence of tokens denoting the sentiment of the news article (*i.e.*, "Positive").

This fusion strategy, referred to as *MultiGen*, is designed to enhance the quality of headline generation by tailoring the outputs to specific categories, aspects, and sentiment polarities. By capturing richer contextual information, this approach generates headlines that resonate more closely with the article's intent and emotional tone, ultimately delivering a more engaging and context-aware reader experience.

4.3. Encoder–decoder architecture

The T5 (Text-to-Text Transfer Transformer) model employs an encoder–decoder architecture specifically designed for sequence-to-sequence tasks. This architecture consists of two primary components: the encoder and the decoder, both built using transformer layers. The encoder processes the input sequence and converts it into fixed-length continuous hidden representations (**Z**) that encapsulate both semantic and syntactic features. The decoder then utilizes these hidden representations to generate the output sequence by predicting tokens iteratively, conditioned on the encoder's output and previously generated tokens.

Encoder. The encoder comprises a stack of transformer layers, each including multi-head self-attention mechanisms and feedforward neural networks. Given an input sequence $\mathbf{I} = (I_1, I_2, \dots, I_n)$, where each I_t represents a token in the input text, the encoder first converts these tokens into continuous vector representations through an embedding layer. Additionally, positional encodings are applied to retain information about word order. The sequence then passes through multiple transformer layers, capturing both local and global dependencies through self-attention mechanisms. Mathematically, the encoder maps the input sequence **I** into a set of hidden states $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$:

$$\mathbf{Z} = TransformerEncoder(\mathbf{I}) \quad (2)$$

where **Z** represents the contextualized representations of the input tokens after being processed by the encoder layers.

Decoder. The decoder is responsible for generating the output sequence $\mathbf{H} = (H_1, H_2, \dots, H_m)$, where each H_t corresponds to a predicted token in the generated text (*i.e.*, a headline). It consists of transformer layers that incorporate self-attention over previously generated tokens and cross-attention over the encoder's hidden states **Z**. The decoder generates tokens sequentially, with each predicted token conditioned on the previously generated tokens and the encoder output:

$$P(\mathbf{H} \mid \mathbf{Z}) = \prod_{t=1}^{m} P(H_t \mid H_{<t}, \mathbf{Z}) \quad (3)$$

where $P(H_t \mid H_{<t}, \mathbf{Z})$ represents the probability of predicting the next token H_t given all previously generated tokens $H_{<t}$ and the encoded representation **Z**.

Baseline approach. In the baseline approach, the encoder processes only the news article $\mathbf{A} = (A_1, A_2, \dots, A_n)$ as the input sequence, generating hidden states that the decoder uses to produce the headline:

$$\mathbf{Z} = TransformerEncoder(\mathbf{A}) \quad (4)$$

$$P(\mathbf{H} \mid \mathbf{Z}) = \prod_{t=1}^{m} P(H_t \mid H_{<t}, \mathbf{Z}) \quad (5)$$

It means that only the content of the article is used to generate the headline without incorporating any auxiliary information.

Proposed approach. In our proposed approach, we introduce additional contextual features such as category (**C**), aspect (**P**), and sentiment (**S**) to enrich the input sequence. These features are concatenated with the news content using special separator tokens [SEP], forming a fused input sequence:

$$\mathbf{F} = [A_1, \dots, A_n, [\text{SEP}], C_1, \dots, C_k, [\text{SEP}], P_1, \dots, P_l, [\text{SEP}], S_1, \dots, S_m] \quad (6)$$

The fused input **F** is then processed by the encoder to generate hidden states:

$$\mathbf{Z} = TransformerEncoder(\mathbf{F}) \quad (7)$$

Finally, the decoder generates the output sequence using this enriched contextual representation:

$$P(\mathbf{H} \mid \mathbf{Z}) = \prod_{t=1}^{m} P(H_t \mid H_{<t}, \mathbf{Z}) \quad (8)$$

By incorporating auxiliary information into the input sequence, our approach enhances the quality and contextual relevance of the generated headlines. The additional context helps the model capture deeper semantic nuances, leading to more informative and accurate headline generation.

5. Experimental evaluation

5.1. Experimental settings

For our Bengali religious news headline generation task, we employed a combination of hardware and software resources to ensure efficient model training and evaluation. Hardware resources included Google Colab Pro with an NVIDIA A100 GPU, Kaggle's environment with NVIDIA T4 × 2 GPUs, and a local machine equipped with an NVIDIA T4 GPU to maximize computational efficiency. On the software side, the project was developed on a Windows 11 machine with a 1TB HDD and 512 GB SSD. We utilized TensorFlow and PyTorch for deep learning, NLTK for text preprocessing, and Hugging Face's Transformers library for implementing encoder–decoder Transformer architectures. The dataset was split into training (1870 samples, 74%), validation (150 samples, 6%), and testing (500 samples, 20%) subsets, allowing for thorough model training, hyperparameter tuning, and performance evaluation. This setup provided a solid foundation for the successful development and assessment of the headline generation models.

Table 6
Comparison of the *BeliN* and Shironaam datasets based on average words, average sentences, and vocabulary size.

Dataset	Article			Headline		
	Avg. words	Avg. sentences	Vocabulary	Avg. words	Avg. sentences	Vocabulary
Shironaam[a] (Akash et al., 2023)	943.43	7.55	3,497	13.03	1.02	416
BeliN (this study)	1001.18	32.75	9,750	17.13	1.06	1,410

[a] for religious news only.

Table 7
Percentage of n-grams in the *BeliN* and Shironaam datasets.

Dataset	Unigram	Bigram	Trigram	4-gram
Shironaam[a] (Akash et al., 2023)	4.50%	22.48%	45.19%	60.22%
BeliN (this study)	4.42%	21.48%	42.10%	56.47%

[a] for religious news only.

5.2. Evaluation metrics

To evaluate the performance of our developed system, we utilized several evaluation metrics. These metrics include BLEU, ROUGE-1, ROUGE-2, ROUGE-L, BERTScore and METEOR. A larger metric value generally indicates better performance; however, the interpretation varies across metrics. While BLEU, ROUGE, and METEOR scores typically reflect higher quality with larger values, their evaluation focuses on different aspects, such as lexical overlap (*i.e.*, BLEU, ROUGE) or semantic similarity (*i.e.*, BERTScore).

ROUGE (Recall-Oriented Understudy for Gisting Evaluation). This metric is a widely used family of metrics for evaluating natural language processing tasks, particularly text summarization (Lin, 2004). It measures the similarity between model-generated summaries and reference summaries based on n-gram overlap and the longest common subsequence (LCS). Three key ROUGE metrics are commonly employed: ROUGE-1 evaluates the overlap of unigrams (individual words) between the generated and reference summaries, calculating Precision and Recall. ROUGE-2 extends this to bi-grams (pairs of consecutive words), measuring the degree of structural similarity between the summaries. ROUGE-L, on the other hand, focuses on the longest common subsequence (LCS), which captures the longest sequence of words that appears in both summaries, regardless of order, providing a measure of structural alignment. Together, these ROUGE metrics offer a comprehensive view of content and structural similarity between generated and reference summaries, helping assess the effectiveness of the summarization model.

BLEU (Bilingual Evaluation Understudy). BLEU (Papineni et al., 2002) is a widely used metric for evaluating the quality of machine-translated text. It measures the similarity between model-generated and human-generated reference translations based on n-gram precision. BLEU is calculated by comparing the n-grams (sequences of n words) generated by the model to the n-grams in the reference translations. The BLEU score ranges from 0 to 1, with higher scores indicating better translation quality. The BLEU score can be defined using the following equation.

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^{N} \frac{1}{N} \log\left(precision_n\right)\right) \tag{9}$$

where BP is the brevity penalty to account for shorter translations, and $precision_n$ is the modified precision for n-grams of size n. The brevity penalty BP is calculated using the following formula:

$$BP = \begin{cases} 1 & c > r \\ \exp(1 - \frac{r}{c}) & c \le r \end{cases} \tag{10}$$

where c is the length of the model output, and r is the effective reference length, which is the length of the reference translation closest to the length of the model output. BLEU is a valuable metric for evaluating the overall quality of machine-generated translations, but it has some limitations, particularly in capturing semantic similarity and fluency. It is best used in conjunction with other evaluation metrics for a comprehensive assessment of translation quality.

METEOR (Metric for Evaluation of Translation with Explicit ORdering). This metric (Banerjee and Lavie, 2005) is designed to evaluate machine translation by considering synonyms, stemming, and paraphrasing. It calculates precision (P) and recall (R) based on alignments between the generated and reference texts. The METEOR score is the harmonic mean of precision and recall, adjusted by a penalty for fragmentation. The METEOR score can be defined using the following equation.

$$\text{METEOR} = F_{\text{mean}} \cdot (1 - BP) \tag{11}$$

where BP is the brevity penalty and F_{mean} is a harmonic mean of precision (P) and recall (R). They can be calculated as:

$$BP = \gamma \cdot \left(\frac{Chunks}{Matchedunigrams}\right)^{\beta} \tag{12}$$

where γ and β are parameters that control the severity of the penalty (typically $\gamma = 0.5$ and $\beta = 3$), $Chunks$ refers to contiguous matched segments in the candidate sentence, and *Matched unigrams* is the number of unigram matches between the candidate and reference sentences.

$$P = \frac{\text{Number of matched unigrams}}{\text{Total unigrams in the candidate sentence}} \tag{13}$$

$$R = \frac{\text{Number of matched unigrams}}{\text{Total unigrams in the reference sentence}} \tag{14}$$

$$F_{\text{mean}} = \frac{P \cdot R}{\alpha \cdot P + (1 - \alpha) \cdot R} \tag{15}$$

where α is a parameter (typically set to 0.9) that gives higher weight to recall.

BERTScore. It evaluates the quality of text generation using contextual embeddings derived from pre-trained BERT models (Zhang et al., 2020). In this study, we used the BanglaBERT pre-trained BERT model from the Hugging Face Transformers library to compute BERTScore. The metric is typically represented by the F1-score and calculates the cosine similarity between token embeddings of the generated and reference texts by incorporating precision (P), recall (R). Precision measures how well the generated text tokens align with the reference text. Recall measures how well the reference tokens are represented in the generated text. They are calculated as:

$$\text{BERTScore} = \text{F1-score} = 2 \cdot \frac{P \cdot R}{P + R} \tag{16}$$

$$P = \frac{1}{|X|} \sum_{x \in X} \max_{y \in Y} cosine(x, y) \tag{17}$$

$$R = \frac{1}{|Y|} \sum_{y \in Y} \max_{x \in X} cosine(y, x) \tag{18}$$

where X and Y represent the sets of token embeddings for the generated (model output) and reference texts, respectively.

5.3. Pre-trained language models

The pre-trained language models (Li et al., 2024) employed for this task include both T5-based and BART-based models. These models are fine-tuned to transform Bengali news articles into concise, informative

Table 8
Details of the pre-trained language models used.

Model	Hugging Face link	Parameters	Pretrained on
Bangala-T5 (Bhattacharjee et al., 2023)	https://huggingface.co/csebuetnlp/BanglaT5	247M	Bengali2B+
mT0-base (Muennighoff et al., 2023)	https://huggingface.co/bigscience/mt0-base	582M	mC4
mT5-Base (Xue et al., 2021)	https://huggingface.co/google/mt5-base	582M	mC4
mBART-50 (Tang et al., 2021)	https://huggingface.co/facebook/mbart-large-50	610M	CC25

Table 9
The hyperparameter search space used in tuning and the selected optimal hyperparameters for each model.

Hyper-parameters	Hyper-parameter Space	Bangla-T5	mBART	mT5	mT0
Learning Rate	2×10^{-5}, 1×10^{-4}, 1×10^{-3}	1×10^{-4}	1×10^{-3}	1×10^{-4}	2×10^{-5}
Epochs	3–10	5	5	5	5
Batch Size	4, 8	8	8	8	8
Input Token Length	512, 1024	512	512	512	512
Target Token Length	16, 32, 64, 128	64	64	64	64

headlines. The T5-based models include BanglaT5, mT5, and mT0, while mBART represents the BART-based model. Detailed information about these models is provided in Table 8.

- T5-based models: BanglaT5, mT5, and mT0 are based on the T5 architecture (Raffel et al., 2020), which frames all NLP tasks as text-to-text problems. The T5 model uses a sequence-to-sequence framework where the encoder processes the input text and the decoder generates the output text. This approach allows for flexibility in handling various NLP tasks, such as translation, summarization, and text generation, by converting them into a text-to-text format.
- BART-based model: mBART (Liu et al., 2020b) is based on the BART architecture (Lewis et al., 2020), which utilizes a denoising autoencoder approach for pre-training. This model follows a sequence-to-sequence framework similar to T5 but incorporates a denoising objective during pre-training, where parts of the input are corrupted and the model learns to reconstruct the original text. This pre-training strategy helps the model become robust to noise and enhances its ability to generate coherent and contextually relevant text.

5.4. Hyper-parameter tuning

Hyperparameter tuning is a critical step in optimizing the performance of generative models for headline generation. In this study, we experimented with various hyperparameter configurations for Bangla-T5, mBART, mT5, and mT0 to achieve the best results. Table 9 summarizes the hyperparameter search space and the selected configurations for each model.

The selection of appropriate hyperparameters directly impacts the performance and efficiency of the models. The learning rate, for instance, was found to be a key factor in stabilizing training. While Bangla-T5 and mT5 performed optimally with a learning rate of 1×10^{-4}, mBART required a higher rate of 1×10^{-3}, and mT0 performed best with 2×10^{-5}. A batch size of 8 was chosen to balance computational requirements and model convergence, ensuring stable training. Input and target token lengths were also crucial for effectively handling the variability in article lengths and headline requirements. Setting the input token length to 512 ensured that the models could process sufficiently detailed content, while a target token length of 64 allowed the generation of concise and precise headlines without truncation. The number of epochs, fixed at 5, provided a balance between overfitting and undertraining for all models. These carefully chosen hyperparameters enabled the models to achieve strong performance in generating contextually relevant and coherent headlines, underscoring the importance of systematic hyperparameter tuning.

5.5. Results

This part evaluates and analyzes the performance of the headline generation models developed in this research. We have provided a comprehensive overview of the evaluation metrics employed, including ROUGE, BLEU, METEOR, and BERTScore. These metrics are widely used in natural language processing tasks to assess the quality of generated text. State of the Art (SOTA) analysis is a critical component of model evaluation, as it establishes a benchmark against which the performance of proposed models can be compared. By analyzing improvements in performance metrics relative to existing methods, researchers can demonstrate the effectiveness and advancements offered by their proposed approaches. The performance of the developed models was evaluated using a combination of these metrics. Table 10 reports the results of our models, comparing the proposed approaches against their respective baseline models and providing insights into the SOTA improvements.

The proposed approach consistently outperforms the baseline across all models in terms of BLEU, ROUGE, BERTScore, and METEOR scores. The BLEU scores indicate a notable enhancement in fluency and coherence of the generated headlines, with the proposed models achieving significant improvements over their baselines. For instance, the mT5 model exhibits a BLEU score increase of 13.1%, while BanglaT5 shows a remarkable improvement of 15.7%, emphasizing the effectiveness of the proposed methods in producing high-quality outputs. The ROUGE scores further demonstrate the superiority of the proposed approach, revealing higher precision and recall compared to the baseline models. This reflects an improved relevance and informativeness in the generated headlines. For example, mT0's proposed method achieves a ROUGE-1 score that is 21.2% better than its baseline, indicating that it captures more relevant content. Additionally, BERTScore and METEOR metrics underscore the semantic accuracy and contextual relevance of the generated headlines. The improvements in these scores highlight the effectiveness of the proposed models in understanding and generating text that aligns well with human expectations. Among all the models, BanglaT5 stands out as the top performer in the proposed approach, achieving the highest scores across all metrics. The substantial improvements in BLEU, ROUGE, BERTScore, and METEOR suggest that BanglaT5 effectively leverages additional contextual information, such as aspect categories and sentiment analysis, to generate headlines that are not only accurate and informative but also contextually relevant. Overall, the findings from this research indicate that the proposed models significantly enhance the quality of Bengali news headline generation, offering promising directions for future research and development in this area.

6. Discussion

6.1. Analyzing generated headlines

Table 11 illustrates sample-generated headlines, showcasing the impact of our *MultiGen* approach on generating high-quality, contextually

Table 10
Performance comparison of the proposed *MultiGen* approach and the baseline across various transformer-based pre-trained models.

Model	Approach	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore	METEOR
mT5	Baseline	10.31	13.47	4.22	13.03	69.34	9.80
	Proposed	11.66	17.54	5.68	16.85	71.74	10.86
	Δ SOTA	+13.1%	+30.4%	+34.6%	+29.1%	+3.5%	+10.8%
mT0	Baseline	12.08	18.84	7.10	17.95	70.34	13.90
	Proposed	13.13	22.94	7.94	21.48	72.62	14.40
	Δ SOTA	+8.7%	+21.2%	+11.8%	+19.0%	+3.2%	+3.6%
mBART	Baseline	15.23	23.01	7.90	21.88	73.21	13.12
	Proposed	16.58	24.36	7.78	22.63	74.63	14.60
	Δ SOTA	+8.8%	+5.9%	−1.5%	+3.4%	+1.9%	+11.3%
BanglaT5	Baseline	16.08	22.84	7.97	23.08	73.57	15.40
	Proposed	18.61	26.70	10.60	24.19	75.12	16.65
	Δ SOTA	+15.7%	+17.0%	+33.0%	+4.8%	+2.1%	+8.1%

relevant headlines for Bengali news articles. The proposed approach demonstrates significant improvements over the baseline, particularly in preserving the essence of the articles while ensuring linguistic fluency and contextual accuracy. The generated headlines from both the baseline and the proposed *MultiGen* approach were compared against the reference headlines to evaluate their quality and contextual relevance. The analysis reveals notable differences in the performance of the two models, particularly in their ability to align with the reference headlines.

The baseline model, while capable of generating coherent headlines, often fails to capture the nuanced meaning or context of the input article. For instance, in sample #1, the reference headline focuses on the etiquette of reciting the Qur'an. The baseline model, however, generates a different headline highlighting the virtues of the Qur'an, which does not address the context intended in the reference. In contrast, the proposed *MultiGen* approach successfully generates a headline emphasizing the manners of recitation, aligning more closely with the reference.

Similarly, in sample #2, the baseline approach generated headline incorrectly attributes lying as a characteristic of Islam, reflecting a lack of semantic understanding. Conversely, the proposed approach accurately conveys the importance of truthfulness, capturing the core message of the article and adhering more closely to the reference.

For sample #3, the reference headline highlights poverty alleviation through Zakat. While the baseline approach generates a headline related to wealth distribution in Islam, it does not focus on the specific theme of poverty alleviation. The *MultiGen* approach, however, generates a headline that aligns with the reference and integrates the article's broader social and economic implications.

Sample #4 further illustrates the shortcomings of the baseline approach, which generates a headline that fails to capture the critical detail that today is Christmas Day. In contrast, the proposed approach successfully incorporates this temporal context, producing a headline that accurately conveys the date-related information and aligns closely with the reference.

Beyond improving headline generation, automatic text summarization for religious news has several practical applications. Summarization enhances information accessibility by condensing lengthy religious articles into concise, digestible formats, making it easier for a broader audience to engage with important discussions. It also plays a crucial role in supporting multilingual communities by facilitating cross-cultural understanding through translated summaries. Additionally, religious news outlets can leverage summarization for efficient media coverage, ensuring accurate and timely reporting. The ability to integrate sentiment and contextual understanding further improves fairness in religious news reporting, mitigating potential biases. Moreover, automatic summarization can assist in content moderation on digital platforms, helping to filter and highlight relevant discussions. These applications underscore the broader significance of our approach in enhancing the accessibility, discoverability, and neutrality of religious news articles.

Overall, the *MultiGen* approach consistently outperforms the baseline model by generating headlines that are contextually and temporally accurate, semantically rich, and closely aligned with the reference. This demonstrates the efficacy of incorporating additional contextual information such as category, aspect, and sentiment, enabling the proposed model to better understand and reflect the underlying themes of the input articles.

6.2. Findings and observations

This section highlights key insights from the results presented in Table 10, with a focus on error analysis to identify areas where the headline generation models may have performed sub-optimally. Below are the primary observations:

Low BLEU scores. The mT5 model, under the baseline approach, demonstrated relatively low BLEU scores compared to other models. This suggests that the generated headlines often lacked fluency or coherence, resulting in lower n-gram overlaps with the reference headlines.

Variability in ROUGE scores. While the proposed approach generally outperformed the baseline across all models, variability in ROUGE scores was observed. For instance, the ROUGE-2 scores for mT5 and mBART under the proposed approach were slightly lower than other models, indicating difficulties in capturing bi-gram similarities effectively.

Performance discrepancies. BanglaT5 consistently exhibited superior performance, particularly in terms of ROUGE scores, highlighting its ability to generate headlines that closely align with reference headlines. Conversely, mT0's relatively lower BLEU and ROUGE scores suggest room for improvement in fluency and relevance.

Impact of additional context. The improved performance of the proposed approach, which leverages additional contextual features such as aspect categories and sentiment, underscores the significance of incorporating auxiliary inputs for headline generation. This approach facilitates a deeper understanding of article contexts, leading to more coherent and informative outputs.

Room for improvement. Despite the promising results, there is potential for further improvement in headline generation. Refining model architectures, optimizing hyperparameters, and enhancing preprocessing techniques could mitigate observed shortcomings and elevate the quality of generated headlines.

Table 11
Qualitative analysis of headlines generated by the proposed *MultiGen* approach.

Sample #	News content	Headline		
		Reference	Baseline	*MultiGen*
1	কোরআনুল কারিম মানব জাতির হেদায়েতের মাধ্যম। এর মাধ্যমে মুক্তি অনুসন্ধান করা হয়। এতে রয়েছে আরোগ্য.....(The Qur'an is the medium of guidance for mankind. Through this liberation is sought. It contains healing.....)	কোরআন তিলাওয়াতের আদব (Etiquette of reciting the Qur'an)	কোরআনে রয়েছে অনেক গুন (There are many virtues in the Qur'an)	কোরআন তিলাওয়াতের কয়েকটি আদব (Some etiquette of reciting the Qur'an)
2	ইসলামের এক গুরুত্বপূর্ণ শিক্ষা, মানবের এক গুরুত্বপূর্ণ বৈশিষ্ট্য, সত্যবাদিতা। এছাড়া একজন মুমিন কামিল মুমিন হতে পারে না। একজন মানব......(An important teaching of Islam, an important characteristic of man, is truthfulness. Besides, a believer cannot be a perfect believer. A human....)	মিথ্যাবাদিতা মুনাফিকদের বৈশিষ্ট্য (Lying is the characteristic of hypocrites)	মিথ্যাবাদিতা ইসলামের বৈশিষ্ঠ্য (Lying is characteristic of Islam)	সত্যবাদিতার গুরুত্ব (The importance of truthfulness)
3	ইসলামে যাকাত একটি আর্থিক ইবাদত যা পঞ্চস্তম্ভের একটি গুরুত্বপূর্ণ স্তম্ভ। এর ধর্মীয় গুরুত্বের পাশাপাশি রয়েছে সামাজিক, নৈতিক ও অর্থনৈতিক.....(Zakat is a financial act of worship in Islam which is an important pillar of the Panchastambha. Besides its religious importance, it has social, moral and economic importance...)	দারিদ্র বিমোচনে জাকাত (Zakat for poverty alleviation)	ইসলামে ধনসম্পদ বন্টন ব্যবস্থার প্রবর্তন (Introduction of wealth distribution system in Islam)	জাকাতের মাধ্যমে দারিদ্র বিমোচনের বিধান (Provision of poverty alleviation through Zakat)
4	ভেতরে-বাইরে রঙিন কাগজে ঢেকেছে ঢাকার তেজগাঁওয়ের পবিত্র জপমালা রানীর গির্জা। এই গির্জার চারপাশে বর্ণিল আলোকসজ্জা.....(The Holy Rosary Rani Church in Tejgaon, Dhaka is covered with colored paper inside and outside. Colorful illumination around this church...)	শুভ বড়দিন আজ (Merry Christmas today)	খ্রীষ্টান্দের শুভ বড়দিন (Merry Christmas to Christians)	খ্রীষ্ট ধর্মবলম্বীদের সবচেয়ে বড় ধর্মীয় উৎসব বড়দিন আজ (Christmas is the biggest religious festival of Christians today)

6.3. Limitations and future work

Despite the encouraging results, the study encountered several limitations. A significant challenge was the scarcity of high-quality annotated datasets for Bengali, which constrained model training and evaluation, potentially limiting the generalizability of results. Hardware limitations also restricted fine-tuning large-scale generative models, impacting training efficiency and performance. Furthermore, the complexity of Bengali morphology and syntax introduced additional challenges, occasionally resulting in inaccuracies in the generated headlines (Rahman and Mamun, 2024).

Future work could address these limitations through innovative data augmentation techniques, exploration of advanced model architectures, and domain-specific customization. Expanding the dataset to include a wider variety of domains and incorporating user feedback in training loops could further enhance the models. Additionally, leveraging LLMs for real-time and multilingual headline generation could broaden the applicability and effectiveness of the approach (Kabir et al., 2024). This study provides a solid foundation for advancing headline generation in Bengali, offering valuable insights for future research in natural language processing and text generation.

7. Conclusion

This research work has explored the potential of a contextual multi-input feature fusion approach, using various generative models for religious news headline generation, with a particular focus on the Bengali language. Central to this work is the introduction of the novel *BeliN* corpus, a curated dataset of Bengali religious news articles and corresponding headlines. This dataset addresses the scarcity of resources for Bengali and serves as a foundational contribution to advancing natural language processing for low-resource languages. We have implemented and evaluated state-of-the-art pre-trained models, including mT5, mT0, mBART, and BanglaT5, within the proposed *MultiGen* approach, incorporating additional contextual information such as aspect, category, and sentiment analysis. Rigorous experimentation and detailed analysis demonstrate that the proposed approach significantly outperforms traditional baseline methods, achieving higher accuracy, coherence, and contextual relevance in generated headlines.

The findings underscore the importance of integrating contextual features in headline generation and highlight the efficacy of the *BeliN* corpus in enabling this advancement. This research contributes to natural language processing and offers practical insights for developing sophisticated text summarization systems in underrepresented languages, thereby promoting linguistic inclusivity and cross-cultural communication.

CRediT authorship contribution statement

Md Osama: Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Ashim Dey:** Writing – review & editing, Supervision, Methodology. **Kawsar Ahmed:** Writing – original draft, Software, Data curation, Conceptualization. **Muhammad Ashad Kabir:** Writing – review & editing, Visualization, Validation, Supervision, Conceptualization, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We want to thank Sadia Tasnim (Begum Rokeya University, Rangpur) for their assistance with collecting and annotating the dataset.

Data availability

Data and code used in this study are publicly available at https://github.com/akabircs/BeliN.

References

Ahmad, I., Alqurashi, F., Mehmood, R., 2022. Potrika: Raw and balanced newspaper datasets in the bangla language with eight topics and five attributes. arXiv:2210.09389.

Ahuir, V., González, J.-Á., Hurtado, L.-F., Segarra, E., 2024. Abstractive summarizers become emotional on news summarization. Appl. Sci. 14 (2), 713. http://dx.doi.org/10.3390/app14020713.

Akash, A.U., Nayeem, M.T., Shohan, F.T., Islam, T., 2023. Shironaam: Bengali news headline generation using auxiliary information. In: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Dubrovnik, Croatia, pp. 52–67. http://dx.doi.org/10.18653/v1/2023.eacl-main.4.

Alomari, A., Idris, N., Sabri, A.Q.M., Alsmadi, I., 2022. Deep reinforcement and transfer learning for abstractive text summarization: A review. Comput. Speech & Lang. 71, 101276. http://dx.doi.org/10.1016/j.csl.2021.101276.

Ao, X., Luo, L., Wang, X., Yang, Z., Chen, J.-H., Qiao, Y., He, Q., Xie, X., 2023. Put your voice on stage: Personalized headline generation for news articles. ACM Trans. Knowl. Discov. from Data 18 (3), 20. http://dx.doi.org/10.1145/3629168.

Ao, X., Wang, X., Luo, L., Qiao, Y., He, Q., Xie, X., 2021. PENS: A dataset and generic framework for personalized news headline generation. In: Zong, C., Xia, F., Li, W., Navigli, R. (Eds.), Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Online, pp. 82–92. http://dx.doi.org/10.18653/v1/2021.acl-long.7.

Aralikatte, R., Cheng, Z., Doddapaneni, S., Cheung, J.C.K., 2023. Varta: A large-scale headline-generation dataset for indic languages. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (Eds.), Findings of the Association for Computational Linguistics: ACL 2023. Association for Computational Linguistics, Toronto, Canada, pp. 3468–3492. http://dx.doi.org/10.18653/v1/2023.findings-acl.215.

Ayana, Shen, S.-Q., Lin, Y.-K., Tu, C.-C., Zhao, Y., Liu, Z.-Y., Sun, M.-S., 2017. Recent advances on neural headline generation. J. Comput. Sci. Tech. 32 (4), 768–784. http://dx.doi.org/10.1007/s11390-017-1758-3.

Aziz, A., Chowdhury, N.K., Kabir, M.A., Chy, A.N., Siddique, M.J., 2025. MMTF-DES: A fusion of multimodal transformer models for desire, emotion, and sentiment analysis of social media data. Neurocomputing 623, 129376. http://dx.doi.org/10.1016/j.neucom.2025.129376.

Banerjee, S., Lavie, A., 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: Goldstein, J., Lavie, A., Lin, C.-Y., Voss, C. (Eds.), Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/Or Summarization. Association for Computational Linguistics, Ann Arbor, Michigan, pp. 65–72, URL https://aclanthology.org/W05-0909.

Banerjee, S., Mukherjee, S., Bandyopadhyay, S., Pakray, P., 2023. An extract-then-abstract based method to generate disaster-news headlines using a DNN extractor followed by a transformer abstractor. Inf. Process. Manage. 60 (3), 103291. http://dx.doi.org/10.1016/j.ipm.2023.103291.

Banerjee, A., Urminsky, O., 2024. The language that drives engagement: A systematic large-scale analysis of headline experiments. Mark. Sci. http://dx.doi.org/10.1287/mksc.2021.0018.

Bharathi Mohan, G., Prasanna Kumar, R., Parathasarathy, S., Aravind, S., Hanish, K.B., Pavithria, G., 2023. Text summarization for big data analytics: A comprehensive review of GPT 2 and BERT approaches. In: Data Analytics for Internet of Things Infrastructure. Springer Nature Switzerland, Cham, pp. 247–264. http://dx.doi.org/10.1007/978-3-031-33808-3_14.

Bhattacharjee, A., Hasan, T., Ahmad, W.U., Shahriyar, R., 2023. BanglaNLG and BanglaT5: Benchmarks and resources for evaluating low-resource natural language generation in bangla. In: Vlachos, A., Augenstein, I. (Eds.), Findings of the Association for Computational Linguistics. EACL, Association for Computational Linguistics, Dubrovnik, Croatia, pp. 726–735. http://dx.doi.org/10.18653/v1/2023.findings-eacl.54.

Bukhtiyarov, A., Gusev, I., 2020. Advances of transformer-based models for news headline generation. In: Proceedings of the 9th Artificial Intelligence and Natural Language Conference. AINL, Springer, pp. 54–61. http://dx.doi.org/10.1007/978-3-030-59082-6_4.

Cai, P., Song, K., Cho, S., Wang, H., Wang, X., Yu, H., Liu, F., Yu, D., 2023. Generating user-engaging news headlines. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. pp. 3265–3280. http://dx.doi.org/10.18653/v1/2023.acl-long.183.

Cajueiro, D.O., Nery, A.G., Tavares, I., K. De Melo, M., dos Reis, S.A., Weigang, L., Celestino, V.R.R., 2023. A comprehensive review of automatic text summarization techniques: method, data, evaluation and coding. arXiv:2301.03403.

Chen, N., 2019. Ci-snf: Exploiting contextual information to improve snf based information retrieval. Inf. Fusion 52, 175–186. http://dx.doi.org/10.1016/j.inffus.2018.08.004.

De Francisci Morales, G., Gionis, A., Lucchese, C., 2012. From chatter to headlines: harnessing the real-time web for personalized news recommendation. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. pp. 153–162. http://dx.doi.org/10.1145/2124295.2124315.

Eberhard, D.M., Simons, G.F., Fennig, C.D., 2024. Ethnologue: Languages of the world. Twenty-seventh edition.. https://www.ethnologue.com/. (last Accessed 30 December 2024).

El-Kassas, W.S., Salama, C.R., Rafea, A.A., Mohamed, H.K., 2021. Automatic text summarization: A comprehensive survey. Expert Syst. Appl. 165, 113679. http://dx.doi.org/10.1016/j.eswa.2020.113679.

Fang, Z., Yu, M., Fu, Z., Zhang, B., Huang, X., Tang, X., Yang, Y., 2024. How to generate popular post headlines on social media? AI Open 5, 1–9. http://dx.doi.org/10.1016/j.aiopen.2023.12.002.

Gavrilov, D., Kalaidin, P., Malykh, V., 2019. Self-attentive model for headline generation. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (Eds.), Advances in Information Retrieval, 41st European Conference on IR Research. ECIR, Springer International Publishing, Cham, pp. 87–93. http://dx.doi.org/10.1007/978-3-030-15719-7_11.

Giarelis, N., Mastrokostas, C., Karacapilidis, N., 2023. Abstractive vs. Extractive summarization: An experimental review. Appl. Sci. 13, 13. http://dx.doi.org/10.3390/app13137620.

Graff, D., Cieri, C., 2003. English gigaword. http://dx.doi.org/10.35111/0z6y-q265, https://catalog.ldc.upenn.edu/LDC2003T05.

Grusky, M., Naaman, M., Artzi, Y., 2018. Newsroom: A dataset of 1.3 million summaries with diverse extractive strategies. In: Walker, M., Ji, H., Stent, A. (Eds.), Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Association for Computational Linguistics, New Orleans, Louisiana, pp. 708–719. http://dx.doi.org/10.18653/v1/N18-1065.

Gu, X., Mao, Y., Han, J., Liu, J., Wu, Y., Yu, C., Finnie, D., Yu, H., Zhai, J., Zukoski, N., 2020. Generating representative headlines for news stories. In: Proceedings of the Web Conference 2020. WWW '20, Association for Computing Machinery, New York, NY, USA, pp. 1773–1784. http://dx.doi.org/10.1145/3366423.3380247.

Hagar, N., Diakopoulos, N., 2019. Optimizing content with a/b headline testing: Changing newsroom practices. Media Commun. 7 (1), 117–127. http://dx.doi.org/10.17645/mac.v7i1.1801.

Hasan, T., Bhattacharjee, A., Islam, M.S., Mubasshir, K., Li, Y.-F., Kang, Y.-B., Rahman, M.S., Shahriyar, R., 2021. XL-Sum: Large-scale multilingual abstractive summarization for 44 languages. In: The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. ACL-IJCNLP, Association for Computational Linguistics, pp. 4693–4703. http://dx.doi.org/10.18653/v1/2021.findings-acl.413.

Hasan, T., Bhattacharjee, A., Samin, K., Hasan, M., Basak, M., Rahman, M.S., Shahriyar, R., 2020. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for bengali-english machine translation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. EMNLP, Association for Computational Linguistics, pp. 2612–2623. http://dx.doi.org/10.18653/v1/2020.emnlp-main.207.

Hayat, S.M.A.I., Das, A., Hoque, M., 2023. Abstractive bengali text summarization using transformer-based learning. In: 6th International Conference on Electrical Information and Communication Technology. EICT, pp. 1–6. http://dx.doi.org/10.1109/EICT61409.2023.10427906.

Hu, B., Chen, Q., Zhu, F., 2015. LCSTS: A large scale Chinese short text summarization dataset. In: Màrquez, L., Callison-Burch, C., Su, J. (Eds.), Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Lisbon, Portugal, pp. 1967–1972. http://dx.doi.org/10.18653/v1/D15-1229.

Iwama, K., Kano, Y., 2019. Multiple news headlines generation using page metadata. In: van Deemter, K., Lin, C., Takamura, H. (Eds.), Proceedings of the 12th International Conference on Natural Language Generation. Association for Computational Linguistics, Tokyo, Japan, pp. 101–105. http://dx.doi.org/10.18653/v1/W19-8612.

Jiang, X., Dreyer, M., 2024. CCSum: A large-scale and high-quality dataset for abstractive news summarization. In: Duh, K., Gomez, H., Bethard, S. (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). Association for Computational Linguistics, Mexico City, Mexico, pp. 7306–7336. http://dx.doi.org/10.18653/v1/2024.naacl-long.406.

Jin, D., Jin, Z., Zhou, J.T., Orii, L., Szolovits, P., 2020. Hooks in the headline: Learning to generate headlines with controlled styles. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp. 5082–5093. http://dx.doi.org/10.18653/v1/2020.acl-main.456.

Kabir, M., Islam, M.S., Laskar, M.T.R., Nayeem, M.T., Bari, M.S., Hoque, E., 2024. BenLLM-eval: A comprehensive evaluation into the potentials and pitfalls of large language models on bengali NLP. In: Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., Xue, N. (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). ELRA and ICCL, Torino, Italia, pp. 2238–2252, URL https://aclanthology.org/2024.lrec-main.201.

Karaca, A., Aydın, Ö., 2023. Generating headlines for turkish news texts with transformer architecture based deep learning method. Gazi Üniversitesi Mühendislik-MimarlıK Fakültesi Derg. 39, 485–495. http://dx.doi.org/10.17341/gazimmfd.963240.

Kiefer, S., 2022. CaSE: Explaining text classifications by fusion of local surrogate explanation models with contextual and semantic knowledge. Inf. Fusion 77, 184–195. http://dx.doi.org/10.1016/j.inffus.2021.07.014.

Koh, H.Y., Ju, J., Liu, M., Pan, S., 2022. An empirical survey on long document summarization: Datasets, models, and metrics. ACM Comput. Surv. 55 (8), 1–35. http://dx.doi.org/10.1145/3545176.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L., 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp. 7871–7880. http://dx.doi.org/10.18653/v1/2020.acl-main.703.

Li, J., Tang, T., Zhao, W.X., Nie, J.-Y., Wen, J.-R., 2024. Pre-trained language models for text generation: A survey. ACM Comput. Surv. 56 (9), http://dx.doi.org/10.1145/3649449.

Li, P., Yu, J., Chen, J., Guo, B., 2021. HG-news: News headline generation based on a generative pre-training model. IEEE Access 9, 110039–110046. http://dx.doi.org/10.1109/ACCESS.2021.3102741.

Lin, C.-Y., 2004. ROUGE: A package for automatic evaluation of summaries. In: Text Summarization Branches Out. Association for Computational Linguistics, Barcelona, Spain, pp. 74–81, URL https://aclanthology.org/W04-1013.

Lins, R.D., Oliveira, H., Cabral, L., Batista, J., Tenorio, B., Ferreira, R., Lima, R., de França Pereira e Silva, G., Simske, S.J., 2019. The CNN-corpus: A large textual corpus for single-document extractive summarization. In: Proceedings of the ACM Symposium on Document Engineering 2019. DocEng '19, Association for Computing Machinery, New York, NY, USA, pp. 1–10. http://dx.doi.org/10.1145/3342558.3345388.

Liu, D., Gong, Y., Yan, Y., Fu, J., Shao, B., Jiang, D., Lv, J., Duan, N., 2020a. Diverse, controllable, and keyphrase-aware: A corpus and method for news multi-headline generation. In: Webber, B., Cohn, T., He, Y., Liu, Y. (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. EMNLP, Association for Computational Linguistics, Online, pp. 6241–6250. http://dx.doi.org/10.18653/v1/2020.emnlp-main.505.

Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., Zettlemoyer, L., 2020b. Multilingual denoising pre-training for neural machine translation. Trans. Assoc. Comput. Linguist. 8, 726–742. http://dx.doi.org/10.1162/tacl_a_00343.

Liu, H., Guo, W., Chen, Y., Li, X., 2022. Contrastive learning enhanced author-style headline generation. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, pp. 5063–5072. http://dx.doi.org/10.18653/v1/2022.emnlp-main.338.

Liu, T., Li, H., Zhu, J., Zhang, J., Zong, C., 2018. Review headline generation with user embedding. In: Sun, M., Liu, T., Wang, X., Liu, Z., Liu, Y. (Eds.), Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer International Publishing, Cham, pp. 324–334. http://dx.doi.org/10.1007/978-3-030-01716-3_27.

Madasu, L., Kanumolu, G., Surange, N., Shrivastava, M., 2023. Mukhyansh: A headline generation dataset for indic languages. In: Huang, C.-R., Harada, Y., Kim, J.-B., Chen, S., Hsu, Y.-Y., Chersoni, E., Pranav, A., Zeng, W.H., Peng, B., Li, Y., Li, J. (Eds.), Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation. Association for Computational Linguistics, Hong Kong, China, pp. 620–634, URL https://aclanthology.org/2023.paclic-1.62.

Muennighoff, N., Wang, T., Sutawika, L., Roberts, A., Biderman, S., Le Scao, T., Bari, M.S., Shen, S., Yong, Z.X., Schoelkopf, H., Tang, X., Radev, D., Aji, A.F., Almubarak, K., Albanie, S., Alyafeai, Z., Webson, A., Raff, E., Raffel, C., 2023. Crosslingual generalization through multitask finetuning. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Toronto, Canada, pp. 15991–16111. http://dx.doi.org/10.18653/v1/2023.acl-long.891.

Nallapati, R., Zhou, B., dos Santos, C., Gulçehre, Ç., Xiang, B., 2016. Abstractive text summarization using sequence-to-sequence RNNs and beyond. In: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning. pp. 280–290. http://dx.doi.org/10.18653/v1/K16-1028.

Napoles, C., Gormley, M., Van Durme, B., 2012. Annotated gigaword. In: Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-Scale Knowledge Extraction. In: AKBC-WEKEX '12, Association for Computational Linguistics, pp. 95–100, URL https://aclanthology.org/W12-3018.pdf.

Narayan, S., Cohen, S.B., Lapata, M., 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (Eds.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, pp. 1797–1807. http://dx.doi.org/10.18653/v1/D18-1206.

National Institute of Standard and Technology, 2014. Document understanding conferences. https://www-nlpir.nist.gov/projects/duc/data.html. (Accessed: 30 December 2024).

Ogunremi, T., sessi Akojenu, S., Soronnadi, A., Adekanmbi, O., Adelani, D.I., 2024. AfriHG: News headline generation for african languages. In: 5th Workshop on African Natural Language Processing. p. 4, URL https://openreview.net/forum?id=fw7g7pNUDl.

Papineni, K., Roukos, S., Ward, T., Zhu, W.-J., 2002. Bleu: a method for automatic evaluation of machine translation. In: Isabelle, P., Charniak, E., Lin, D. (Eds.), Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, pp. 311–318. http://dx.doi.org/10.3115/1073083.1073135.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. 21 (140), 1–67, URL http://jmlr.org/papers/v21/20-074.html.

Rahman, A., Mamun, A., 2024. The rise of clickbait headlines: A study on media platforms from Bangladesh. Athens J. Mass Media Commun. 10, 109–130. http://dx.doi.org/10.30958/ajmmc.10-2-3.

Rao, A., Aithal, S., Singh, S., 2024. Single-document abstractive text summarization: A systematic literature review. ACM Comput. Surv. 57 (3), http://dx.doi.org/10.1145/3700639.

Saad, A.M., Mahi, U.N., Salim, M.S., Hossain, S.I., 2024. Bangla news article dataset. Data Brief 57, 110874. http://dx.doi.org/10.1016/j.dib.2024.110874.

Salehin, M., Rafat, A., Khan, F., Abujar, S., 2019. Generating bengali news headlines: An attentive approach with sequence-to-sequence networks. In: Proceedings of the 8th International Conference System Modeling and Advancement in Research Trends. SMART, pp. 256–261. http://dx.doi.org/10.1109/SMART46866.2019.9117554.

Sandhaus, E., 2008. The new york times annotated corpus. http://dx.doi.org/11272.1/AB2/GZC6PL, URL https://hdl.handle.net/11272.1/AB2/GZC6PL.

Sen, M., Yanikoglu, B., 2018. Document classification of SuDer turkish news corpora. In: Proceedings of the 2018 26th Signal Processing and Communications Applications Conference. SIU, pp. 1–4. http://dx.doi.org/10.1109/SIU.2018.8404790.

Shaibani, A.Y., Elnagar, A.M., 2024. A survey of text summarization and headline generation methods in arabic: A survey of text summarization and headline generation methods in arabic. In: Proceedings of the 2024 9th International Conference on Machine Learning Technologies. ICMLT '24, Association for Computing Machinery, New York, NY, USA, pp. 317–323. http://dx.doi.org/10.1145/3674029.3674078.

Sharma, N.A., Ali, A.S., Kabir, M.A., 2024. A review of sentiment analysis: tasks, applications, and deep learning techniques. Int. J. Data Sci. Anal. 1–38. http://dx.doi.org/10.1007/s41060-024-00594-x.

Singh, R.K., Khetarpaul, S., Gorantla, R., Allada, S.G., 2021. SHEG: summarization and headline generation of news articles using deep learning. Neural Comput. Appl. 33, 3251–3265. http://dx.doi.org/10.1007/s00521-020-05188-9.

Takase, S., Suzuki, J., Okazaki, N., Hirao, T., Nagata, M., 2016. Neural headline generation on abstract meaning representation. In: Su, J., Duh, K., Carreras, X. (Eds.), Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Austin, Texas, pp. 1054–1059. http://dx.doi.org/10.18653/v1/D16-1112.

Tang, Y., Tran, C., Li, X., Chen, P.-J., Goyal, N., Chaudhary, V., Gu, J., Fan, A., 2021. Multilingual translation from denoising pre-training. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP. Association for Computational Linguistics, pp. 3450–3466. http://dx.doi.org/10.18653/v1/2021.findings-acl.304.

Theledi, K., Pule, V.M., 2024. President's speech and terminology used during the COVID-19 pandemic: The interpretation of linguistic meaning in context and situational context. In: Public Health Communication Challenges To Minority and Indigenous Communities. IGI Global, pp. 92–107. http://dx.doi.org/10.4018/979-8-3693-0624-6.ch007.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., Raffel, C., 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In: Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., Zhou, Y. (Eds.), Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Online, pp. 483–498. http://dx.doi.org/10.18653/v1/2021.naacl-main.41.

Yadav, A.K., Ranvijay, Yadav, R.S., Maurya, A.K., 2023. State-of-the-art approach to extractive text summarization: a comprehensive review. Multimedia Tools Appl. 82 (19), 29135–29197. http://dx.doi.org/10.1007/s11042-023-14613-9.

Yang, Z., Lian, J., Ao, X., 2023. Fact-preserved personalized news headline generation. In: 2023 IEEE International Conference on Data Mining. ICDM, pp. 1493–1498. http://dx.doi.org/10.1109/ICDM58522.2023.00197.

Yutkin, D., 2019. Lenta. https://github.com/yutkin/Lenta.Ru-News-Dataset. (Accessed: 30 December 2024).

Zeyad, A.M.A., Biradar, A., 2024. Advancements in the efficacy of flan-T5 for abstractive text summarization: A multi-dataset evaluation using ROUGE and bertscore. In: 2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI). pp. 1–5, URL https://api.semanticscholar.org/CorpusID:271748003.

Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., Artzi, Y., 2020. BERTScore: Evaluating text generation with BERT. In: International Conference on Learning Representations. p. 43, URL https://openreview.net/forum?id=SkeHuCVFDr.

Zhu, L., Zhu, Z., Zhang, C., Xu, Y., Kong, X., 2023. Multimodal sentiment analysis based on fusion methods: A survey. Inf. Fusion 95, 306–325. http://dx.doi.org/10.1016/j.inffus.2023.02.028.