Large Scale Multi-Lingual Multi-Modal Summarization Dataset

Yash Verma*
IISER Kolkata
yashv7523@gmail.com

Anubhav Jangra*
Indian Institute of Technology Patna
anubhav0603@gmail.com

Raghvendra Kumar
Indian Institute of Technology Patna
raghvendra.kumar1004@gmail.com

Sriparna Saha
Indian Institute of Technology Patna
sriparna@iitp.ac.in

Abstract

Significant developments in techniques such as encoder-decoder models have enabled us to represent information comprising multiple modalities. This information can further enhance many downstream tasks in the field of information retrieval and natural language processing; however, improvements in multi-modal techniques and their performance evaluation require large-scale multi-modal data which offers sufficient diversity. Multi-lingual modeling for a variety of tasks like multi-modal summarization, text generation, and translation leverages information derived from high-quality multi-lingual annotated data. In this work, we present the current largest multi-lingual multi-modal summarization dataset (M3LS), and it consists of over a million instances of document-image pairs along with a professionally annotated multi-modal summary for each pair. It is derived from news articles published by British Broadcasting Corporation(BBC) over a decade and spans 20 languages, targeting diversity across five language roots, it is also the largest summarization dataset for 13 languages and consists of cross-lingual summarization data for 2 languages. We formally define the multi-lingual multi-modal summarization task utilizing our dataset and report baseline scores from various state-of-the-art summarization techniques in a multi-lingual setting. We also compare it with many similar datasets to analyze the uniqueness and difficulty of M3LS. [1]

1 Introduction

The world we live in today is very diverse, with over 7,000+ languages spoken across the globe[2]. These languages have varying traits and are spoken by communities of various sizes depending upon the popularity of the language. For example, Mandarin consists of over 50,000 *hanzi* (characters) and is spoken by over 1.117 billion people[3], while there exist languages like Rotokas, which is an indigenous language spoken by about 4,320 people on the island of Bougainville, Papua New Guinea, which consists of only 12 letters[4].

Table 1: Comparison of proposed dataset with existing large-scale multi-lingual and multi-modal datasets.

Multi-lingual Summarization Datasets			
Dataset Name	**Dataset Size**	**#Languages**	**Domain**
XL-Sum (Hasan et al., 2021)	1M	44	News
MLSUM (Scialom et al., 2020)	1.5M	5	News
WikiLingua (Ladhak et al., 2020)	770K	18	Tutorials
MLGSum (Wang et al., 2021)	1.1M	12	News
M3LS (Ours)	1.1M	20	News
Multi-modal Summarization Datasets			
Dataset Name	**Dataset Size**	**Modalities**	**Domain**
MSMO (Zhu et al., 2018)	314K	Text + Image	News
E-DailyMail (Chen and Zhuge, 2018)	219K	Text + Image	News
How2 (Sanabria et al., 2018)	190K	Text + Video + Audio	Multiple Domains
MMSS (Li et al., 2018)	66K	Text + Image	News
VMSMO (Li et al., 2020)	185K	Text + Video + Audio	Social Network
M3LS (Ours)	1.1M	Text + Image	News

These languages, although very crucial, restrict people to communicate their thoughts to others who speak the same language. The gift of sight, however, is something that is universally shared by every human being on this plant, irrespective of their culture, ethnicity, or the language that they speak. Through this work we aim to instigate the research towards improving existing automatic summarization systems by leveraging information from multiple languages and visual modalities.

Various studies in the past have illustrated how unified summarization frameworks across multiple languages improve the summarization quality over mono-lingual frameworks (Wang et al., 2021). Similarly, there have been works in multi-modal summarization that illustrate how multi-modal input can help improve the quality of summarization over text summarization systems (Jangra et al., 2020a,b; Chen and Zhuge, 2018; Mukherjee et al.,

*These authors contributed equally to this work.

[1]The dataset and code used in this work are made available at https://github.com/anubhav-jangra/M3LS.

[2]https://www.ethnologue.com/guides/how-many-languages

[3]https://www.berlitz.com/en-uy/blog/most-spoken-languages-world

[4]https://en.wikipedia.org/wiki/Rotokas_language

Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pages 3620–3632
May 2-6, 2023

2022). Additionally, having multiple modalities in the output summary can help improve the overall satisfaction of the user (Zhu et al., 2018; Jangra et al., 2021b). Multiple modalities can also compensate for the inability of individual modalities to express various aspects of the summary. For instance, it is hard to express abstract concepts like "freedom", "gravity", etc. through images, while it can be expressed through text conveniently. Similarly, it is very difficult to describe a "Pangolin" to someone who hasn't seen one beforehand.

Hence, in this work we propose the task of Multi-modal Multi-lingual Summarization (M3LS), and also release the M3LS dataset[5] to facilitate the research in this direction. The dataset comprises 1.1M news articles, spanning 20 languages comprising *English*, *Chinese*, *Spanish*, *Russian*, *French*, *Ukrainian*, *Portuguese*, *Japanese*, *Tamil*, *Hindi*, *Marathi*, *Gujarati*, *Bengali*, *Sinhala*, *Urdu*, *Pashto*, *Indonesian*, *Telugu*, *Punjabi*, and *Nepali*; making it the largest language-spanning summarization dataset. To the best of our knowledge, the proposed dataset is the largest summarization dataset for 13 languages (*Russian*, *Ukrainian*, *Tamil*, *Hindi*, *Marathi*, *Gujarati*, *Bengali*, *Sinhala*, *Urdu*, *Pashto*, *Telugu*, *Punjabi*, and *Nepali*).

We hope that the proposed task and the dataset will instigate and inspire multi-modal and multi-lingual research in less-explored languages for solving various tasks including but not limited to automatic summarization (Nallapati et al., 2016; See et al., 2017), article headline generation (Jin et al., 2020; Gavrilov et al., 2019; Zhang et al., 2018), keyword extraction (Showrov and Sobhan, 2019; Lee and Kim, 2008; Yao et al., 2019), image caption generation (Xu et al., 2015; Bai and An, 2018), multi-modal embedding generation (Sun et al., 2019; Lu et al., 2019; Li et al., 2019; Zhou et al., 2020), large-scale language modeling (Raffel et al., 2020; Devlin et al., 2019) etc.

The major contributions of this work are as follows - *1) We have proposed the multi-modal multi-lingual summarization (M3LS) task. 2) We have released the largest multi-modal summarization dataset that spans 20 languages. 3) The proposed dataset is the largest text summarization dataset for 13 languages. 4) To the best of our knowledge, we present the first ever multi-modal cross-lingual dataset (consisting of Japanese-to-English and English-to-Japanese). 5) We have provided multi-modal summarization baseline results for our dataset and a detailed analysis of the dataset.*

[5]A sample of our dataset is available at https://github.com/zenquiorra/M3LS, the complete dataset will be released in the camera ready version of the work

2 Related Work

The field of text summarization is more than 5 decades old (Edmundson, 1969), and has evolved to a great extent in recent years. Prior to the advances in sequence-to-sequence frameworks (Sutskever et al., 2014), people mainly focused on extractive summarization techniques that aim to generate summary via extracting words, phrases, or sentences (Mihalcea and Tarau, 2004; Saini et al., 2019; Alguliev et al., 2010). See et al. (2017) proposed the Pointer-Generator Networks, an attentive recurrent neural network based framework (Bahdanau et al., 2015). Recent years have seen great progress in research in automatic summarization leveraging transformer based models (Zhang et al., 2020; Devlin et al., 2019) and attention mechanism (Vaswani et al., 2017).

In this section we discuss the related works showcasing multi-modal datasets and multi-lingual datasets. A detailed size comparison of these datasets with M3LS is shown in Table 1.

2.1 Multi-modal summarization datasets

Multi-modal summarization is the task of summarizing content comprising two or more input modalities. The output can be uni-modal or multi-modal depending on the task. In this section, we discuss existing large-scale multi-modal summarization datasets proposed in the community. We point the readers to Jangra et al. (2021a) for a comprehensive survey.

MSMO: Zhu et al. (2018) proposed a multi-modal summarization dataset that consists of text and images. The dataset is obtained from the DailyMail[6] website and contains 314,581 instances in English language. However, Hasan et al. (2021) illustrated that the DailyMail news highlights lack novel n-grams. Fabbri et al. (2021) also highlighted the inconsistency in quality of some reference summaries in the CNN/DailyMail dataset (Nallapati et al., 2016).

E-Dailymail Chen and Zhuge (2018) proposed the E-Dailymail dataset, which contains text and images extracted from the DailyMail website. The dataset consists of 219,100 instances in English,

[6]https://www.dailymail.co.uk/home/index.html

containing the input document, article title, images, and image captions.

How2: Sanabria et al. (2018) proposed a multi-modal summarization dataset consisting of text, video, and audio modalities; it contains over 2000 hours of videos accompanied by the corresponding audio and speech transcriptions.

MMSS: Li et al. (2018) proposed a multi-modal summarization dataset consisting of text and images with the aim of proposing an image-aided sentence summarization framework. The dataset has 66K instances in English language, that is generated by extracting sentence-headline pairs from the `Gigaword` corpus[7].

VMSMO: To the best of our knowledge, Li et al. (2020) proposed the first large-scale asynchronous text-audio-video summarization dataset. The dataset is generated from the famous microblogging platform `Sina Weibo`[8], and comprises of 184,920 instances in Chinese language.

Similar trends of incorporating multiple modalities in language tasks can also be noticed in several tasks like question answering (Singh et al., 2021), translation (Elliott and Kádár, 2017), sentiment analysis (Soleymani et al., 2017), lexico-semantic classification (Jha et al., 2022), keyword extraction (Verma et al., 2022) etc.

2.2 Multi-lingual Text Summarization Datasets

The popularity of studying the benefits of summarization in different languages to improve summarization qualities increased over the past few years. There have been a lot of research work in bi-lingual setting; however, in this work, we limit ourselves to discussing multi-lingual summarization datasets to be concise.

MLSUM : Scialom et al. (2020) proposed the `MLSUM` dataset that consists of 1.5 million news articles obtained from the `Dailymail/CNN` websites. The dataset spans five languages - French, German, Spanish, Russian and Turkish.

XL-Sum: Hasan et al. (2021) proposed the `XL-Sum` dataset that consists of 1.35 million articles in 44 languages obtained from `BBC news`, making it the most language-diverse summarization dataset to date. However, 25 of these 44 languages do not contain even 10,000 instances, making it incompetent to train any language model.

WikiLingua: Ladhak et al. (2020) proposed the `Wikilingua` dataset, which is the largest parallel multi-lingual summarization to date. The dataset consists of 770K instances in English language, and is extended to 17 other languages for varying number of English articles.

MLGSum: Wang et al. (2021) proposed the `MLGSum` dataset that consists of articles from various news providers such as `BBC`, `france243` and `select faz`. The dataset has five high-resource and seven low-resource languages, with a total of 1.1 million instances, and is a rich source for text summarization for German language with 500K instances.

We observe that multiple popular datasets (see Table 1) in multimodal summarization and multilingual summarization are useful for both technique evaluation and technique improvisation. However, the combined field of multilingual multimodal summarization has remained largely unexplored, and it can be attributed to the lack of dedicated high quality dataset and formalizing it as a problem statement. Hence, we formally define the `M3LS` task and discuss the dataset addressing the problem further.

3 M3LS Task

Given for each language $l_k \in L$ where L is the set of all languages, we have data $M^{l_k} =< T^{l_k}, I^{l_k} >$, where $T^{l_k} = \left\{t_1^{l_k}, t_2^{l_k}, \ldots, t_{|T|}^{l_k}\right\}$ is a set of documents, and $I^{l_k} = \{I^{t_1^{l_k}}, I^{t_2^{l_k}}, \ldots, I^{t_{|T|}^{l_k}}\}$ is a set of images, where $I^{t_j^{l_k}} = \left\{i_1, i_2, \ldots, i_{|I|}\right\}^{t_j^{l_k}}$ denotes the set of images belonging to the document $t_j^{l_k} \in T^{k_k}$ and $|.|$ denotes the cardinality of a set.

The task is to obtain a function F that maps documents $t_j^{l_{k_1}} \in T^{l_{k_1}}$ in language l_{k_1} along with their corresponding images, $I^{t_j^{l_{k_1}}} \in I^{l_{k_1}}$ to a set of multi-modal summaries in target language, l_{k_2}, comprising of text summaries (denoted by $O^{l_{k_2}}$) along with images from the input (denoted by $I^{l_{k_1}}$).

$$F :< T^{l_{k_1}}, I^{l_{k_1}} > \rightarrow < O^{l_{k_2}}, I^{l_{k_1}} > \quad (1)$$

When $k_1 \neq k_2$, we have multi-modal cross-lingual summarization, otherwise the task is multi-modal mono-lingual summarization, a graphic representation of the task is shown in Figure 1.

4 M3LS Dataset

Through the `M3LS` task, we motivate the need for a multi-modal multi-lingual dataset by studying the

[7] https://github.com/harvardnlp/sent-summary
[8] http://ir.weibo.com/

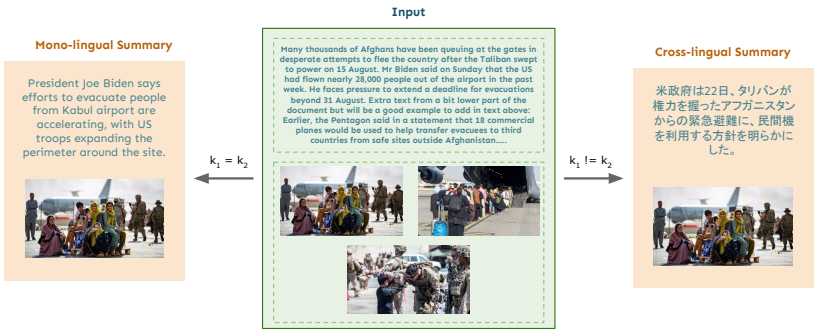

Figure 1: Proposed M3LS task.

developments in summarization techniques such as secondary enhancements using images with multi-modal output (Zhu et al., 2018), video-based multi-modal summarization (Li et al., 2020) and using multiobjective optimization (Jangra et al., 2020b). On the other hand, development of multi-lingual transformer based models like Xue et al. (2020) has publicly available checkpoints fine-tuned for multiple language modelling tasks, including multi-lingual summarization.

Development of such models requires high-quality heterogeneous data and improvements in various models utilizing multi-modal shared attention layers for annotated data with image-text pairs for a specific language task. To address these issues, we present M3LS and in this section we discuss various steps involved in its construction.

4.1 Dataset Construction

We explore the news domain, as it is one of the most abundant and readily available domains and covers articles in multiple topics, while describing the events and lacking extreme bias. We analyzed the structure of articles and surveyed multiple news providers before finalizing on BBC News, which provides full sentence summaries in a uniform structured format across multiple languages. The summaries are professionally created by the author which ensures the quality of the data. We explain the steps involved in creating the M3LS dataset and discuss various aspects of the data.

BBC News: BBC News[9] is a division of British Broadcasting Corporation responsible for gathering and broadcasting current news affairs. Each BBC news article has a text summary comprising complete sentences in the present tense, avoiding opinions and sensationalism. We cover 20 different languages with summaries written in corresponding languages. We extract data from various parts

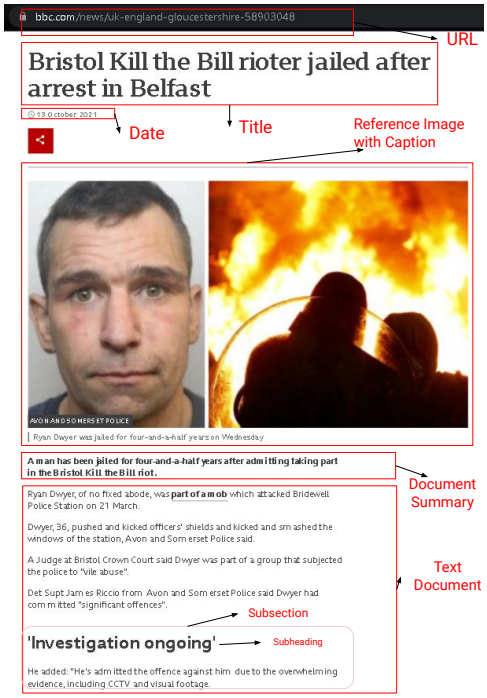

Figure 2: Snapshot format of a webpage used in development of M3LS, and various features extracted during the scraping procedure

of the webpage as shown in Figure 2.

Obtaining Articles: We obtain links to articles from the corresponding Twitter[10] pages for each BBC language news dataset. To extend the dataset, we scrape[11] valid links[12] obtained from the parsed articles of each language.

The final collection of links is scraped separately using scrapy[13] to obtain the final dataset. Since these links are showcased at the corresponding twitter page, these links ensure articles con-

[9] https://www.bbc.com/news

[10] https://twitter.com/bbc

[11] Data is collected in accordance with the terms and conditions mentioned on the website

[12] A link is valid if it contains a BBC article summary for the corresponding domain.

[13] https://scrapy.org

taining topics of interest and high popularity. We extend the dataset by recursively extracting links from suggestions and hyperlinks within a webpage.

Structuring the Data: We obtain various features from the webpage as shown in Figure (2), and compile them in a JSON format; we also provide a dedicated parser, instructions and a tutorial for the ease of access of features from any instance. The data is freely available for use in accordance with the terms and conditions of BBC News, we discuss this in detail on the same link where our dataset is uploaded.

Text Validation: In order to ensure high-quality text from the source, we manually read 10 instances from each language[14] from the collected links to verify if the articles are descriptive in nature and consist of text written in complete sentences.

Summary Validation: We manually checked the summary quality for 100 articles each in 4 languages[15] from our dataset and validated if the given summary captures the information represented in the text. For every article, after carefully reading the text, we assign the gold summary a score between 1-5 with 5 being the best possible summary which captures most of the important information from the given article and vice versa, and also including parameters like summary length and length of the article. We observe that for > 70 articles across the languages evaluated obtain a score of > 4 out of 5 in our analysis. Assuming the uniformity of articles published by BBC across multiple domains, we assume that this fact is true for every language in our dataset.

Final Dataset: In final dataset, each news article contains the text document, images with corresponding captions, keywords, links to related news articles, and a multi-modal summary comprising of a few sentences and an image.

Cross-lingual Dataset: Our cross-lingual dataset contains all features from our final dataset, along with multi-modal summaries consisting of text in another language. It is obtained from the links given by the author within the Japanese language article to the corresponding article in English language, we manually check the information provided in both articles using Google Translate[16] for 100 instances to verify the similarity of the content and summaries provided.

Train-Test-Validation split: The dataset has 1.2 million news articles which we split into 80% training, 10% test and 10% validation for languages having $\leq$ 50,000 articles, otherwise we select 90% data for training, 5% for testing and 5% as validation split.

5 Dataset Analysis

5.1 Overview

The M3LS dataset has 1.11M+ multi-lingual multi-modal instances across 20 languages and over 9K cross-lingual multi-modal instances for English-Japanese language pair. The dataset can be categorized into 8 high resource languages and 12 low resource languages[17] (refer to Appendix B for more details). The chosen languages originate from different parts of the globe, and belong to 5 different language roots: *Indo-European*, *Austronesian*, *Japanic*, *Dravidian*, and *Sino-Tibetan*.

M3LS dataset is quite diverse, with the least #articles for Sinhala (10,148) and greatest #articles for English (376,367). The dataset becomes even more complex and challenging with different sizes of input documents for different languages, with document size varying from $\tilde{3}30$ tokens to over 2800+ tokens. The dataset articles cover a wide time span, with articles from 2009 to 2021 (refer to Appendix A for more details).

We hope that the M3LS dataset will instigate and inspire research in less-explored languages, since 14 out of these 20 languages covered by the dataset are among the top-20 most spoken languages in the world[18]; this diversity helps in modelling tasks for both well-explored and less-explored languages.

5.2 Dataset Comparison

To study the size and span of our dataset, we compare M3LS with other summarization datasets extracted from the BBC News domain. We found that XSum contains 53% of the tokens from our dataset, while XL-Sum contains 58% of the tokens from our dataset across all languages present in M3LS. However, they are uni-modal in nature, while XSum is uni-lingual. We observe that M3LS is magnitudes larger when compared to XSum, while ex-

[14]For languages unknown to the authors, we use Google Translatehttps://translate.google.com to translate the content in English language

[15]We restrict ourselves to 4 languages (English, Hindi, Bengali and Marathi) due to the understanding of languages of the authors presenting this work

[16]https://google.com/translate

[17]The categorization is done based on a threshold value of 50k data instances.

[18]https://lingua.edu/the%2D20%2Dmost%2Dspoken%2Dlanguages%2Din%2Dthe%2Dworld%2Din%2D2022/

ceeding by times 2-3 in almost all individual language instances when compared to XL-Sum. Both of these datasets are used to train and fine-tune several state-of-the-art-summarization models like Pegasus, hence we believe that M3LS will offer a wider and better language modelling support in terms of size and diversity for the languages present in it, with the additional benefit of multi-modality.

6 Experiments

6.1 Setup

Depending upon the number of instances in each language within M3LS, we perform a train:test:validation split with a ratio of 80:10:10 if the number of instances is below 50K and 90:5:5 otherwise. To conduct our experiments in a multilingual setting, we survey publicly available tokenizers and sentence segmenters for multiple languages, and we combine them within one dedicated package for our experiments. We further define a set of rules for sentence segmentation for languages lacking such support from external packages within our package[19].

We compile our package using segtok[20] for the Indo-European language, IndicNLP[21] for Indian languages, fugashi (McCann, 2020) for Japanese (ja) and chinese[22] for Chinese.

For data pre-processing steps such as stopword removal, we collect stopwords from the nltk[23] package, and publicly available stopwords present in the spaCy[24] repository for all languages in a centralized pipeline for our experiments.

We evaluated the performance of various summarization techniques utilizing our dataset, including simpler techniques such as LEAD-3 and RANDOM which have proven to be quite useful in past (Ghalandari et al., 2020; Scialom et al., 2020; Sharma et al., 2019). We have also included statistics based CENTROID (Radev et al., 2004) and graph based TextRank (Mihalcea and Tarau, 2004) techniques.

To have a fair comparison across multiple languages using a shared dedicated model, we have evaluated the performance of an abstractive technique in a multi-lingual setting utilizing a pre-trained checkpoint[25] for summarization of the transformer-based MT5 (Xue et al., 2020) model. Finally to explore the multi-modal aspect of our dataset, we evaluate the performance of a multi-modal encoder-decoder based technique (Zhu et al., 2018) that utilizes images and text to generate a multi-modal text summary. However the publicly available implementation[26] for MSMO restricts us to evaluate it only for the English language. However, to compare this score, we evaluate the performance of three state-of-the-art transformer-based models - Pegasus (Zhang et al., 2020), BART (Lewis et al., 2020), and T5 (Xue et al., 2020) for summarization which are compatible with the English language.

Since, two of the pre-trained models we described above are fine-tuned on XSum and XL-Sum datasets which are extracted from the same source - BBC News - we avoid fine-tuning on models to have a fair comparison of the models and we explain the scores in discussions.

In all techniques, we set the generated summary length threshold as the average length of gold summary for the corresponding language in our corpus.

6.2 Baselines

Simpler Extractive Approaches

LEAD-3: In this baseline, the first three sentences of the source text are extracted as the final summary. This method is a robust baseline, as shown by (Sharma et al., 2019) for news summarization datasets.

RANDOM: We recursively extract words randomly from the source text until the threshold summary length is reached. The aim of this baseline is to understand and compare other baselines with an unbiased model as a point of reference.

Statistical Approach

CENTROID: We use the strategy proposed by Radev et al. (2004), which ranks sentences based on the centrality scores obtained by the words in the sentence. We use TF-IDF scores to measure each word's similarity, and extract top sentences from each ranking until the threshold summary length is obtained.

Graph Based Approach

TEXTRANK: TextRank (Mihalcea and Tarau,

[19] https://github.com/zenquiorra/TokSeg

[20] https://pypi.org/project/segtok/1.1.0/

[21] https://github.com/anoopkunchukuttan/indic_nlp_library

[22] https://pypi.org/project/chinese/

[23] https://nltk.org/

[24] https://github.com/explosion/spaCy/tree/master/spacy/lang

[25] https://huggingface.co/csebuetnlp/mT5_multilingual_XLSum

[26] We use the implementation provided by the authors, which is a multi-layered package, modification of which to be compatible for a multi-lingual setting isn't feasible based on the software complexity

2004) is an unsupervised graph-based ranking technique based on the relevance of sentences in the source text[27] We consider the sentences which are most central to the document based on the ranking as generated summaries.

RNN Based Approach

MSMO: MSMO (Zhu et al., 2018) is an encoder-decoder model trained for multi-modal summarization. It utilizes a multi-modal attention mechanism to generate multi-modal summaries utilizing text and images.

Transformer Based Approaches

MT5: MT5 (Xue et al., 2020) is a transformer-based seq2seq model pretrained for multiple natural language tasks. We use the publicly available checkpoint[28] pre-trained for text summarization on the XL-Sum dataset (Hasan et al., 2021) for a multi-lingual setting.

PEGASUS: Pegasus (Zhang et al., 2020) is a transformer-based model, pre-trained on a task to remove meaningful sentences from an input text, making it suitable for summarization. We used a checkpoint[29] of PEGASUS model pre-trained on the XSum dataset (Narayan et al., 2018) for summarization.

BART: BART (Lewis et al., 2020) uses a standard seq2seq architecture with a bi-directional encoder and a left-to-right decoder. We use a pre-trained model trained on the DailyMail/CNN (Nallapati et al., 2016) for our evaluation.

T5: T5 (Raffel et al., 2020) is an encoder-decoder model trained on a mixture of natural language tasks, including translation and summarization; it converts any task into a text-to-text format. We use the pre-trained T5-large model for the summarization task.

7 Results and Discussion

We evaluate the generated summaries against the gold summaries using the ROUGE (Lin, 2004) evaluation metric. We report the ROUGE-1, ROUGE-2, ROUGE-L f-scores across every baseline discussed above (Lin, 2004) (refer to Tables 1 and 3). We additionally report BERTSCORE for English baselines (Zhang et al., 2019) (refer to Table 1).

[27]We use the implementation provided by the gensimhttps://radimrehurek.com/gensim_3.8.3/summarization/summariser.html package and modify the segmentation and tokenizer part using our dedicated package.

[28]https://huggingface.co/csebuetnlp/mT5_multilingual_XLSum

[29]https://huggingface.co/google/pegasus-xsum

Table 2: Comparison of "ROUGE" fscores for summaries generated using Multi-modal baseline MSMO and Unimodal transformer based baselines against gold summaries from the English language dataset . "R-f1" denotes ROUGE-1 fscore, "R-f2" denotes ROUGE2 f-score, "R-fL" denotes the ROUGEL fscore, and "BrS" denotes BERTSCORE.

English	R-f1	R-f2	R-fL	BrS
BART	0.195	0.031	0.131	0.863
Pegasus	**0.389**	**0.181**	**0.321**	**0.910**
T5	0.197	0.0328	0.131	0.858
MSMO	0.217	0.046	0.158	0.851

7.1 Multi-lingual baseline scores

We observe that transformer based techniques used in our experiments perform significantly better compared to other techniques. However, for the "MT5" column, we observe very high scores and spikes of very low scores as observed in Table 3, this behavior maybe caused due to two factors:

- Relatively high scores can be attributed to the use of "MT5" checkpoint that is fine-tuned for the task of summarization on a dataset (XL-Sum) obtained from same source as ours.
- Very low scores for some languages can be attributed to the "ROUGE" evaluation metric which relies on token overlap[30]. Many of these languages, especially the ones with Dravidan and Indo-European origins have words which change their form significantly depending on their placement in the text and the context in which they appear, hence simple token overlap metrics show lower scores if the root form of the word isn't considered.

We observe that LEAD-3 performs better for the languages in which transformer-based baseline performs poorly, this can be attributed to two factors:

- As shown by Sharma et al. (2019) that LEAD-3 performs very well for summarization tasks when we consider the news domain, suggesting the idea that top sentences capture a lot of information within a news article.
- LEAD-3 considers top-3 sentences from the text, unlike abstractive summarization, new

[30]We are not implementing stemming of tokens during evaluation, due to the lack of support of multi-lingual stemming methods across various softwares which we have used for experimentation and to have an even comparison with the supported languages

Table 3: Performance of various techniques for summarization against the M3LS dataset gold summaries for every language. “Lang" refers to the language code for a language according to the ISO 639-1 standard, “R-f1" refers to the ROUGE-1 f-scores, “R-f2" refers to the ROUGE-2 f-scores, “R-fL" refers to the ROUGE-L f-scores

Base	**Random**			**LEAD-3**			**TextRank**			**CENTROID**			**MT5**		
Lang	R-f1	R-f2	R-fL	R-f1	R-f2	R-fL	R-f1	R-f2	R-fL	R-f1	R-f2	R-fL	R-f1	R-f2	R-fL
bn	0.003	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.004**	0.001	0.003
mr	0.013	0.000	0.012	0.041	0.005	0.040	0.025	0.002	0.025	0.006	0.001	0.006	**0.044**	0.005	**0.044**
gu	0.014	0.001	0.014	**0.039**	0.005	0.038	0.014	0.001	0.014	0.016	0.002	0.016	0.036	0.005	0.036
ps	0.002	0.000	0.001	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.000	0.000	**0.003**	0.000	0.001
uk	0.030	0.002	0.029	0.062	0.016	0.061	0.043	0.010	0.042	0.032	0.006	0.032	**0.094**	0.025	**0.094**
pt	0.179	0.009	0.114	0.204	0.033	0.124	0.199	0.030	0.128	0.089	0.008	0.075	**0.276**	0.085	0.193
id	0.118	0.001	0.083	0.172	0.037	0.117	0.144	0.030	0.104	0.104	0.014	0.080	**0.289**	0.115	0.233
ne	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pa	0.012	0.000	0.012	**0.038**	0.004	**0.038**	0.014	0.001	0.014	0.010	0.002	0.010	0.026	0.000	0.026
si	0.014	0.000	0.014	0.032	0.004	0.031	0.019	0.002	0.019	0.007	0.001	0.007	**0.039**	0.018	**0.039**
ur	0.006	0.000	0.006	0.023	0.001	0.023	0.006	0.000	0.005	0.024	0.001	0.023	**0.044**	0.000	**0.044**
fr	0.168	0.007	0.107	0.206	0.043	0.126	0.177	0.033	0.115	0.164	0.024	0.110	**0.209**	0.041	0.141
ru	0.032	0.001	0.032	0.071	0.017	0.069	0.041	0.012	0.040	0.036	0.008	0.036	**0.081**	0.011	**0.081**
ja	0.069	0.001	0.068	0.126	0.012	0.120	0.084	0.007	0.081	0.063	0.004	0.062	**0.306**	0.081	0.291
te	0.010	0.000	0.009	0.023	0.001	0.023	0.011	0.000	0.011	0.008	0.001	0.008	**0.026**	0.000	**0.026**
ta	0.014	0.001	0.014	**0.034**	0.005	**0.034**	0.023	0.003	0.022	0.012	0.001	0.012	0.026	0.000	0.026
zh	0.022	0.001	0.022	**0.053**	0.008	0.051	0.042	0.005	0.041	0.025	0.003	0.025	0.125	0.042	0.118
es	0.177	0.008	0.117	0.180	0.033	0.117	0.110	0.018	0.073	0.081	0.008	0.067	**0.280**	0.084	0.202
hi	0.010	0.000	0.010	**0.018**	0.002	0.018	0.013	0.001	0.013	0.005	0.000	0.005	0.002	0.000	0.001
en	0.146	0.002	0.102	**0.175**	0.026	0.114	0.100	0.014	0.071	0.140	0.016	0.102	**0.427**	0.182	0.345

tokens or new forms of existing tokens are not present in the given article. Since it is an extractive technique, the chances of token overlap are higher and hence better “f-scores".

7.2 Multi-modal baseline scores

Due to the limitation of lack of pre-trained frameworks in a multi-modal setting for most of the languages in the dataset, we were constrained to evaluate the multi-modal technique on the English dataset. On comparing the “f-scores" of various uni-modal techniques with the multi-modal technique, we notice that the transformer based model Pegasus outperforms other techniques. This is largely attributed to the fact that the pre-trained checkpoint we have used for evaluation of summaries through the Pegasus model is fine-tuned on the XSum dataset, which has data collected from the same source as ours. We observe that for other models which are not fine-tuned on a dataset extracted from same source as ours, the multi-modal technique MSMO is able to outperform other techniques.

7.3 Abstractiveness of the proposed dataset

We propose an abstractive summarization dataset where the target summaries are manually written by human beings. The M3LS dataset demands abstractive techniques since the percentage of novel uni-grams in the dataset is quite high (refer to “abs.gold" column in Appendix B). This fact is also observed in the results from the baseline techniques. For instance, MT5 performs consistently superior for multiple languages as observed in Table3, the abstractive baselines have thrice as good ROUGE scores as the extractive baselines.

8 Conclusion

In this work, we release a large-scale multi-modal multi-lingual summarization dataset comprising of over 1.1M+ news articles and spanning 20 languages, and motivate the problem statement of Multi-modal Multi-lingual summarization using M3LS. To the best of our knowledge, this is the first ever multi-modal summarization data set spanning several languages. The proposed dataset is the largest summarization dataset for 13 out of 20 languages. We have evaluated the performance of various baselines to establish the quality of the proposed dataset in both multi-modal and multi-lingual settings. Through this work, we hope to instigate research in various less-explored languages in the community for various research problems including but not limited to summarization, headline generation, keyword extraction, image caption generation, multi-modal embedding generation, etc. In future works, we plan to work on shared models which address the M3LS task utilizing our dataset.

Limitations

There are a few considerations to keep in mind in our work. **First**, the dataset currently has a multi-modal input, mapping to a textual summary. However, future work could involve annotating images to enhance the dataset with a multi-modal output. **Second**, the distribution for languages in the M3LS dataset is skewed due to the imbalanced number of articles published in BBC across languages and the late establishment of virtual print media in certain languages (as shown in Appendix A). **Third**, the current dataset uses an independent identically distributed split to create train and test sets, but more advanced techniques such as adversarial splits and likelihood splits could also be explored in future work. **Fourth**, while the current manuscript does not evaluate the dataset on both multi-modal and multi-lingual aspects simultaneously, we believe that this dataset has the potential to contribute to the development of such systems in the future.

Acknowledgements

This publication is an outcome of the R&D work undertaken in the project under the Visvesvaraya Ph.D. Scheme of Ministry of Electronics & Information Technology, Government of India, being implemented by Digital India Corporation (Formerly Media Lab Asia).

References

Rasim Alguliev, Ramiz Aliguliyev, and Makrufa Hajirahimova. 2010. Multi-document summarization model based on integer linear programming. *Intelligent Control and Automation*, 1(02):105.

Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*.

Shuang Bai and Shan An. 2018. A survey on automatic image caption generation. *Neurocomputing*, 311:291–304.

Jingqiang Chen and Hai Zhuge. 2018. Abstractive text-image summarization using multi-modal attentional hierarchical rnn. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4046–4056.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

H. P. Edmundson. 1969. New methods in automatic extracting. *J. ACM*, 16:264–285.

Desmond Elliott and Ákos Kádár. 2017. Imagination improves multimodal translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 130–141.

A. R. Fabbri, Wojciech Kryscinski, Bryan McCann, R. Socher, and Dragomir Radev. 2021. Summeval: Re-evaluating summarization evaluation. *Transactions of the Association for Computational Linguistics*, 9:391–409.

Daniil Gavrilov, Pavel Kalaidin, and Valentin Malykh. 2019. Self-attentive model for headline generation. In *European Conference on Information Retrieval*, pages 87–93. Springer.

Demian Gholipour Ghalandari, Chris Hokamp, John Glover, Georgiana Ifrim, et al. 2020. A large-scale multi-document summarization dataset from the wikipedia current events portal. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1302–1308.

Tahmid Hasan, Abhik Bhattacharjee, Md Saiful Islam, Kazi Samin, Yuan-Fang Li, Yong-Bin Kang, M Sohel Rahman, and Rifat Shahriyar. 2021. Xl-sum: Large-scale multilingual abstractive summarization for 44 languages. *arXiv preprint arXiv:2106.13822*.

Anubhav Jangra, Adam Jatowt, Mohammad Hasanuzzaman, and Sriparna Saha. 2020a. Text-image-video summary generation using joint integer linear programming. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II*, pages 190–198.

Anubhav Jangra, Adam Jatowt, Sriparna Saha, and Mohammad Hasanuzzaman. 2021a. A survey on multi-modal summarization. *arXiv preprint arXiv:2109.05199*.

Anubhav Jangra, Sriparna Saha, Adam Jatowt, and Mohammad Hasanuzzaman. 2020b. Multi-modal summary generation using multi-objective optimization. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1745–1748.

Anubhav Jangra, Sriparna Saha, Adam Jatowt, and Mohammed Hasanuzzaman. 2021b. Multi-modal supplementary-complementary summarization using multi-objective optimization. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 818–828.

Prince Jha, Gaël Dias, Alexis Lechervy, Jose G Moreno, Anubhav Jangra, Sebastião Pais, and Sriparna Saha. 2022. Combining vision and language representations for patch-based identification of lexico-semantic relations. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 4406–4415.

Di Jin, Zhijing Jin, Joey Tianyi Zhou, Lisa Orii, and Peter Szolovits. 2020. Hooks in the headline: Learning to generate headlines with controlled styles. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5082–5093.

Faisal Ladhak, Esin Durmus, Claire Cardie, and Kathleen McKeown. 2020. Wikilingua: A new benchmark dataset for multilingual abstractive summarization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 4034–4048.

Sungjick Lee and Han-joon Kim. 2008. News keyword extraction for topic tracking. In *2008 Fourth International Conference on Networked Computing and Advanced Information Management*, volume 2, pages 554–559. IEEE.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.

Haoran Li, Junnan Zhu, Tianshang Liu, Jiajun Zhang, and Chengqing Zong. 2018. Multi-modal sentence summarization with modality attention and image filtering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4152–4158. International Joint Conferences on Artificial Intelligence Organization.

Liunian Harold Li, Mark Yatskar, Da Yin, Cho-Jui Hsieh, and Kai-Wei Chang. 2019. Visualbert: A simple and performant baseline for vision and language. *arXiv preprint arXiv:1908.03557*.

Mingzhe Li, Xiuying Chen, Shen Gao, Zhangming Chan, Dongyan Zhao, and Rui Yan. 2020. Vmsmo: Learning to generate multimodal summary for video-based news articles. *arXiv preprint arXiv:2010.05406*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In *Advances in Neural Information Processing Systems*, pages 13–23.

Paul McCann. 2020. fugashi, a tool for tokenizing Japanese in python. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 44–51, Online. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411.

Sourajit Mukherjee, Anubhav Jangra, Sriparna Saha, and Adam Jatowt. 2022. Topic-aware multimodal summarization. In *Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022*, pages 387–398, Online only. Association for Computational Linguistics.

Ramesh Nallapati, Bowen Zhou, C. D. Santos, Çaglar Gülçehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. In *CoNLL*.

Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807.

Dragomir R Radev, Hongyan Jing, Małgorzata Styś, and Daniel Tam. 2004. Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6):919–938.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21:1–67.

Naveen Saini, Sriparna Saha, Anubhav Jangra, and Pushpak Bhattacharyya. 2019. Extractive single document summarization using multi-objective optimization: Exploring self-organized differential evolution, grey wolf optimizer and water cycle algorithm. *Knowledge-Based Systems*, 164:45–67.

Ramon Sanabria, Ozan Caglayan, Shruti Palaskar, Desmond Elliott, Loïc Barrault, Lucia Specia, and Florian Metze. 2018. How2: A large-scale dataset for multimodal language understanding. In *NeurIPS*.

Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, and Jacopo Staiano. 2020. Mlsum: The multilingual summarization corpus. *arXiv preprint arXiv:2004.14900*.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Eva Sharma, Chen Li, and Lu Wang. 2019. Bigpatent: A large-scale dataset for abstractive and coherent summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2204–2213.

Md Imran Hossain Showrov and Masrur Sobhan. 2019. Keyword extraction from bengali news. In *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, pages 658–662. IEEE.

Hrituraj Singh, Anshul Nasery, Denil Mehta, Aishwarya Agarwal, Jatin Lamba, and Balaji Vasan Srinivasan. 2021. Mimoqa: Multimodal input multimodal output question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5317–5332.

Mohammad Soleymani, David Garcia, Brendan Jou, Björn Schuller, Shih-Fu Chang, and Maja Pantic. 2017. A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65:3–14.

Chen Sun, Austin Myers, Carl Vondrick, Kevin Murphy, and Cordelia Schmid. 2019. Videobert: A joint model for video and language representation learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7464–7473.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Yash Verma, Anubhav Jangra, Sriparna Saha, Adam Jatowt, and Dwaipayan Roy. 2022. Maked: Multilingual automatic keyword extraction dataset. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6170–6179.

Danqing Wang, Jiaze Chen, Hao Zhou, Xipeng Qiu, and Lei Li. 2021. Contrastive aligned joint learning for multilingual summarization. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2739–2750.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2020. mt5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*.

Lu Yao, Zhang Pengzhou, and Zhang Chi. 2019. Research on news keyword extraction technology based on tf-idf and textrank. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, pages 452–455. IEEE.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*, pages 11328–11339. PMLR.

Ruqing Zhang, Jiafeng Guo, Yixing Fan, Yanyan Lan, Jun Xu, Huanhuan Cao, and Xueqi Cheng. 2018. Question headline generation for news articles. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 617–626.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Luowei Zhou, Hamid Palangi, Lei Zhang, Houdong Hu, Jason J Corso, and Jianfeng Gao. 2020. Unified vision-language pre-training for image captioning and vqa. In *AAAI*, pages 13041–13049.

Junnan Zhu, Haoran Li, Tianshang Liu, Yu Zhou, Jiajun Zhang, and Chengqing Zong. 2018. Msmo: Multimodal summarization with multimodal output. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4154–4164.

A Frequency of number of articles present in the M3LS dataset for a given year

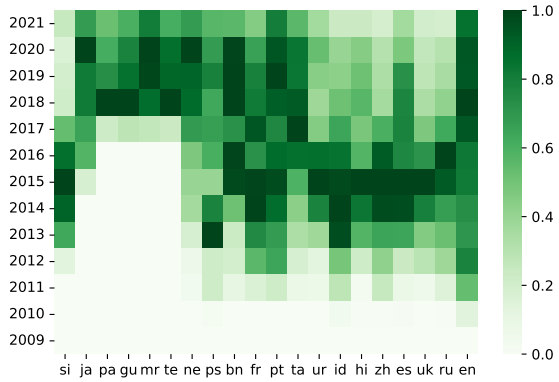

Figure 3: Temporal span of each language in the M3LS dataset. Darker colors correspond to a score of 1.0 and vice versa, which indicates a higher frequency of the number of articles published during the year for that language.

B Detailed Statistics of M3LS dataset

The detailed statistics of the curated M3LS dataset are presented in Table 4.

Table 4: “Lang." represents the language code from the IS0 639-1 standard, **articles** represents the number of articles for every language in the corpus, **toks.** represents the average number of tokens in an article within the corpus for the language, **tok.uni.** represents the average number of unique tokens in any article for the language, **sum.tok.** represents the average number of tokens in summary of an article for the language, **sum.uni** represents the average number of unique tokens within the summary of an article for the language, **sent.** represents the average number of sentences in a given article for the language, **abs.gold** represents the average percent of tokens in summary which are absent from the article for a given language, **images** represents the average number of images in an article for the given language, **i.c.r** represents the average ratio of the number of images consisting of a caption attached to them against the number of images which do not have a caption with them.

Lang.	**articles**	**tok.**	**tok.uni.**	**sum.tok.**	**sum.uni.**	**sent.**	**abs.gold**	**images**	**i.c.r**
bn	25283	464.78	254.73	23.12	21.97	28.42	43.18	2.44	0.58
mr	16161	871.36	404.82	27.88	25.02	63.32	48.33	0.00	0.00
gu	12175	868.75	404.84	25.87	23.63	50.11	40.87	4.94	0.30
ps	23205	523.35	207.95	32.85	27.52	18.29	33.32	2.10	0.49
uk	90846	471.90	216.57	24.59	22.78	23.49	54.07	1.56	0.22
pt	39454	2855.68	424.14	38.57	33.01	114.22	35.23	3.34	0.64
id	56108	587.88	225.12	24.59	22.89	28.54	37.07	2.47	0.57
ne	18953	402.17	229.71	21.50	20.95	23.79	45.00	2.07	0.24
pa	11600	843.74	319.75	30.96	27.35	38.87	30.42	4.57	0.37
si	10148	331.55	186.26	24.30	23.43	15.53	51.33	1.67	0.34
ur	55107	690.47	264.57	35.95	31.00	1.07	27.94	2.35	0.43
fr	25923	413.45	179.22	31.22	27.26	14.43	40.75	1.64	0.47
ru	95345	668.61	302.63	28.38	25.74	26.68	52.80	1.96	0.38
ja	11023	1052.94	282.58	48.59	36.97	33.92	29.86	2.90	0.56
te	15511	626.24	353.70	23.71	21.39	51.72	52.19	4.41	0.27
ta	38523	354.35	209.30	21.18	19.91	23.39	56.34	2.55	0.27
zh	60830	787.68	309.11	34.38	29.80	29.87	37.21	2.00	0.48
es	66816	2649.57	345.35	30.64	26.28	83.22	38.43	4.15	0.66
hi	61852	776.39	265.10	29.45	25.70	37.21	32.00	1.24	0.0
en	376367	657.95	268.60	25.27	23.40	22.28	32.86	1.39	0.35