LLMs as Zero-Shot Multi-Label Classifiers for Bangla Documents

Souvika Sarkar
Big Data Intelligence (BDI) Lab
Auburn University
Auburn, AL
szs0239@auburn.edu

Shubhra Kanti Karmaker Santu
Big Data Intelligence (BDI) Lab
Auburn University
Auburn, AL
sks0086@auburn.edu

Abstract

Bangla, the sixth most spoken language globally, poses NLP challenges due to its complexity and limited resources. Despite recent advances in LLMs, their performance on Bangla remains unexplored. We evaluated four LLMs on Bangla's Zero-Shot MLC task, revealing the need for further research and resources in Bangla NLP.

1 Introduction

Bangla, spoken by over 300 million native speakers, ranks as the sixth most spoken language globally[1]. Despite its extensive usage, Bangla remains underrepresented in Natural Language Processing (NLP) research (Joshi et al., 2020). Large Language Models (LLMs) like GPT, BLOOM, and LLaMA have transformed multilingual NLP tasks, their success primarily shines in widely spoken languages such as English and Chinese. The effectiveness of these models in low-resource languages like Bangla remains largely unexplored, motivating our study.

Our paper explores the performance of state-of-the-art sentence encoders, and four LLMs in Zero-Shot multi-label classification (Zero-Shot-MLC) tasks tailored for Bangla datasets. Our experimental results reveal the LLMs struggles to achieve satisfactory scores in Bangla, indicating areas for future research and improvement.

2 Background and Related Works

In Bangla NLP, researchers have explored various tasks such as Information Extraction (Rahman et al., 2008; Uddin et al., 2019; Sharif et al., 2016), Machine Translation (Hasan et al., 2019; Anwar et al., 2009; Ismail et al., 2014), Named Entity Recognition (Banik and Rahman, 2018; Chaudhuri and Bhattacharya, 2008), Question Answering (Islam et al., 2016; Sarker et al., 2019; Kowsher et al., 2019). Recent advancements include the introduction of Bangla-BERT (Kowsher et al., 2022) and BanglaBERT (Bhattacharjee et al., 2022), as well as the use of Transformer-based models in various tasks such as abusive comment detection (Aurpa et al., 2022b), text classification (Alam et al., 2020), and question answering (Aurpa et al., 2022a; Adnan and Anwar, 2022). Research on large language models like ChatGPT, Flan, and BLOOM has demonstrated their utility in diverse applications including healthcare education (Sallam, 2023), programming bug solving (Surameery and Shakor, 2023), and machine translation (Jiao et al., 2023). However, Our study addresses the gap in understanding the performance of recent LLMs on Bangla documents. We examine multiple state-of-the-art sentence encoders and LLMs in the context of zero-shot multi-label classification exclusively for Bangla documents, laying a foundation for future research in this area.

3 Problem Statement

We employ the Definition-Wild 0SHOT-TC methodology introduced by Yin et al. (2019) which further explored by Sarkar et al. (2023, 2022) for the English language.

3.1 *Encoder* & *LLM* Based 0-shotMLC

Our 0-shot-MLC approach comprises the following steps:

1. Input Document: The user provides the article text, custom-defined labels, and optional keywords.
2. Embedding Generation: We transform the article text, labels, and keywords into embeddings using language models and encoding methods.

[1] Source: *List of languages by the number of native speakers*, Wikipedia, https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers (accessed on 23 May 2023)

Topic+Keywords Based Label Embedding								
LASER			LaBSE			BanglaTransformer		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.162	0.750	0.267	0.282	0.477	**0.354**	0.224	0.648	0.334
Explicit-Mention Based Label Embedding								
LASER			LaBSE			BanglaTransformer		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.193	0.724	0.305	0.300	0.617	**0.404**	0.276	0.635	0.384

Table 1: Performance comparison of baseline sentence encoder-based approaches.

Topic+Keywords Based Label Embedding								
FLAN-UL2			BLOOM			GPT-NeoX		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.135	0.890	0.234	0.231	0.574	0.329	0.235	0.634	0.345
Explicit-Mention Based Label Embedding								
FLAN-UL2			BLOOM			GPT-NeoX		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.144	0.742	0.241	0.232	0.642	0.341	0.241	0.675	0.357

Table 2: Performance comparison of different large language models.

- **Article Embedding**: The entire article is embedded using sentence encoders and LLMs.
- **Label Embedding**: We employ two methods for label embedding: 1) *Label + Keyword*- Label embedding using label name and keywords, 2) *Explicit-Mentions*- Label embedding using article-text which contains explicit mentions of label names.

3. Threshold-based Label Assignment: We quantify the cosine similarity between the article embedding and the label embeddings and assign labels based on a specified threshold.
4. Zero-Shot multi-label classification: The classifier predicts relevant label(s) for the given article.

4 Experimental Setup

4.1 Dataset, Baselines and LLMs

We curated BanglaNewsNet, a benchmark corpus, by crawling news articles from `https://www.prothomalo.com/`. Each article is labeled with one or more categories. The dataset was cleaned by merging similar labels. See Table 3 for details.

Dataset Name	# of Articles	Avg. article length	Labels retained	Labels/ article
BanglaNewsNet	7245	≈2517 words	21	1.345

Table 3: An overview of the BanglaNewsNet dataset

Baselines include LASER, LaBSE, and Bangla sentence transformers. LLMs tested are BLOOM, FLAN-UL2, GPTNeoX, and ChatGPT. See Appendix for details.

5 Results & Discussion

Table 1 and 2 compares the performance of baseline sentence encoders and large language models (LLMs) respectively. LaBSE achieved the highest performance among sentence encoder models, with an approximate F_1 Score of 40%. BanglaTransformer performed slightly lower than LaBSE. Evaluation of LLMs on BanglaNewsNet (Table 2) revealed that they did not excel in zero-shot multi-label classification (0-Shot-MLC) on Bangla articles. Surprisingly, none of the LLMs surpassed LaBSE, underscoring the challenge of Zero-Shot-MLC for Bangla documents. Additionally, Table 4 showcases ChatGPT's performance, with an average F_1 score of 53%, indicating superiority over other LLMs but still falling short of optimality.

Precision	Recall	F_1 Score
0.515	0.573	0.537

Table 4: ChatGPT performance on BanglaNewsNet

In conclusion, this paper assesses the effectiveness of contemporary LLMs for the Zero-Shot-MLC task exclusively for a widely spoken yet low-resource language, i.e., Bangla, and identifies the limitations of current LLMs for this task. As such, our research contributes to the ongoing efforts to enhance the applicability and efficacy of LLMs for regional and low-resource languages, paving the way for future advancements in multilingual NLP research.

Limitations

As a limitation, it is worth mentioning that we had limited access to various large language models (LLMs) like LaMDA, Gato, LLaMA (downloadable by application, if approved) etc., which restricted their utilization in our experiments. Additionally, the availability of models like ChatGPT was constrained, requiring us to rely on the API for evaluating their performance. Consequently, the comparison between ChatGPT and the other three large language models is not an apple-to-apple comparison. Also, we experimented with only one dataset, and more data sets need to be evaluated in order to confirm whether our findings will hold in general.

References

Md Abdul Mazid Adnan and Md Musfique Anwar. 2022. Progressive guidance categorization using transformer-based deep neural network architecture. In *Hybrid Intelligent Systems: 21st International Conference on Hybrid Intelligent Systems (HIS 2021), December 14–16, 2021*, volume 420, page 344. Springer Nature.

Tanvirul Alam, Akib Khan, and Firoj Alam. 2020. Bangla text classification using transformers. *arXiv preprint arXiv:2011.04446*.

Md Musfique Anwar, Mohammad Zabed Anwar, and M Al-Amin Bhuiyan. 2009. Syntax analysis and machine translation of bangla sentences. *International Journal of Computer Science and Network Security*, 9(8):317–326.

Mikel Artetxe and Holger Schwenk. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Trans. Assoc. Comput. Linguistics*, 7:597–610.

Tanjim Taharat Aurpa, Richita Khandakar Rifat, Md Shoaib Ahmed, Md Musfique Anwar, and ABM Shawkat Ali. 2022a. Reading comprehension based question answering system in bangla language with transformer-based learning. *Heliyon*, 8(10):e11052.

Tanjim Taharat Aurpa, Rifat Sadik, and Md Shoaib Ahmed. 2022b. Abusive bangla comments detection on facebook using transformer-based deep learning models. *Social Network Analysis and Mining*, 12(1):24.

Nayan Banik and Md. Hasan Hafizur Rahman. 2018. Gru based named entity recognition system for bangla online newspapers. In *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, pages 1–6.

Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. 2022. BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1318–1327, Seattle, United States. Association for Computational Linguistics.

Sid Black, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, et al. 2022. Gpt-neox-20b: An open-source autoregressive language model. *arXiv preprint arXiv:2204.06745*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Bidyut Baran Chaudhuri and Suvankar Bhattacharya. 2008. An experiment on automatic detection of named entities in Bangla. In *Proceedings of the IJCNLP-08 Workshop on Named Entity Recognition for South and South East Asian Languages*.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2020. Language-agnostic bert sentence embedding. *arXiv preprint arXiv:2007.01852*.

Towhid Ahmed Foysal. 2020. Sentence transformers repository. `https://github.com/Foysal87/Bangla-sentence-embedding-transformer`.

Md. Arid Hasan, Firoj Alam, Shammur Absar Chowdhury, and Naira Khan. 2019. Neural machine translation for the bangla-english language pair. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pages 1–6.

Md. Aminul Islam, Md. Fasihul Kabir, Khandaker Abdullah-Al-Mamun, and Mohammad Nurul Huda. 2016. Word/phrase based answer type classification for bengali question answering system. In *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pages 445–448.

Sabir Ismail, M. Shahidur Rahman, and Md Abdullah Al Mumin. 2014. Developing an automated bangla parts of speech tagged dictionary. *16th Int'l Conf. Computer and Information Technology*, pages 355–359.

Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. 2023. Is chatgpt a good translator? a preliminary study. *arXiv preprint arXiv:2301.08745*.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

M. Kowsher, Abdullah As Sami, Nusrat Jahan Prottasha, Mohammad Shamsul Arefin, Pranab Kumar Dhar, and Takeshi Koshiba. 2022. Bangla-bert: Transformer-based efficient model for transfer learning and language understanding. *IEEE Access*, 10:91855–91870.

Md. Kowsher, M M Mahabubur Rahman, Sk Shohorab Ahmed, and Nusrat Jahan Prottasha. 2019. Bangla intelligence question answering system based on mathematics and statistics. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pages 1–6.

Mohammad Osiur Rahman, Fouzia Asharf Mousumi, Edgar Scavino, Aini Hussain, and Hassan Basri. 2008. Real time road sign recognition system using artificial neural networks for bengali textual information box. In *2008 International Symposium on Information Technology*, volume 2, pages 1–8.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3973–3983.

Malik Sallam. 2023. Chatgpt utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare*, volume 11, page 887. MDPI.

Souvika Sarkar, Dongji Feng, and Shubhra Kanti Karmaker Santu. 2022. Exploring universal sentence encoders for zero-shot text classification. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 135–147, Online only. Association for Computational Linguistics.

Souvika Sarkar, Dongji Feng, and Shubhra Kanti Karmaker Santu. 2023. Zero-shot multi-label topic inference with sentence encoders. *arXiv preprint arXiv:2304.07382*.

Sourav Sarker, Syeda Tamanna Alam Monisha, and Md Mahadi Hasan Nahid. 2019. Bengali question answering system for factoid questions: A statistical approach. In *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–5.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.

S M A Sharif, Nabeel Mohammed, Nafees Mansoor, and Sifat Momen. 2016. A hybrid deep model with hog features for bangla handwritten numeral classification. In *2016 9th International Conference on Electrical and Computer Engineering (ICECE)*, pages 463–466.

Nigar M Shafiq Surameery and Mohammed Y Shakor. 2023. Use chat gpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering (IJITC) ISSN: 2455-5290*, 3(01):17–22.

Abdul Hasib Uddin, Sumit Kumar Dam, and Abu Shamim Mohammad Arif. 2019. Extracting severe negative sentence pattern from bangla data via long short-term memory neural network. In *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, pages 1–6.

Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2021. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*.

Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 3912–3921. Association for Computational Linguistics.

A Appendix

A.1 "Entire Article" based embedding

Encode the entire article using sentence encoders or LLMs in a single shot, including articles that are long paragraphs and consist of more than one sentence.

A.2 Label Embedding Approaches

We have used 2 different approaches for computing label embedding. The consecutive sections discuss about different procedures for generating label embedding.

A.2.1 "Label name + Keywords" based embedding

Encode both label name and keywords, then average all embeddings to generate the final label embedding.

A.2.2 "Explicit-Mentions" based embedding

First, extract all the articles explicitly mentioning the label/phrase using algorithm 1 for all labels. Then, for each label, generate embeddings of all articles which are explicitly annotated/classified with that label, then average them to obtain the ultimate label embedding.

Algorithm 1 Article Annotation using Explicit Mention

```
1: Input: Article text, Label names and Keywords
2: Output: Articles annotated with explicit Label
3: for each article text do
4:     check whether the label name or set (at least 3) of
       the informative keywords are present or not in the cor-
       responding article text
5:     if present then annotate the article with the explicit
       label
6:     end if
7: end for
```

A.3 Baseline Sentence Encoders

This section presents a bird's-eye view of the sentence encoders and large language models we have used for our experiments.

- **Language-Agnostic SEntence Representations (LASER)**: LASER (Artetxe and Schwenk, 2019) is a sentence encoding model that generates language-agnostic representations. It is capable of encoding sentences from multiple languages into fixed-length vectors, enabling cross-lingual tasks and multilingual applications.
- **Language-agnostic BERT Sentence Embedding (LaBSE)**: LaBSE (Feng et al., 2020) is a language-agnostic model based on the BERT architecture. It provides sentence embeddings that capture the semantic meaning of sentences across different languages. LaBSE allows for cross-lingual understanding and transfer-learning tasks.
- **Bangla sentence embedding transformer**: This Bangla sentence transformer (Foysal, 2020) is specifically designed for the Bangla language. It utilizes a transformer-based architecture to encode Bangla sentences into meaningful representations, enabling various NLP tasks in Bangla text analysis. It was trained on 2,50,000 Bangla sentences(wiki) by sentence transformer. This work is inspired by Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (Reimers and Gurevych, 2019) technique.

A.4 Large Language Models

- **BLOOM**: Scao et al. (2022) introduce BLOOM, a massive language model with 176 billion parameters. BLOOM is trained on 46 natural languages and 13 programming languages and is the result of a collaborative effort involving hundreds of researchers. BLOOM is a causal language model trained to predict the next token in a sentence. This approach has been found effective in capturing reasoning abilities in large language models. BLOOM uses a Transformer architecture composed of an input embeddings layer, 70 Transformer blocks, and an output language-modeling layer. The sequential operation of predicting the next token involves passing the input tokens through each of the 70 BLOOM blocks. To prevent memory overflow, only one block is loaded into RAM at a time. The word embeddings and output language-modeling layer can be loaded on-demand from disk.
- **FLAN-UL2**: Wei et al. (2021) introduce a unified framework for pre-training models that demonstrate broad effectiveness across datasets and setups. They differentiate between architectural archetypes and pre-training objectives and propose a novel approach called Mixture-of-Denoisers (MoD) for combined pre-training. The new 20 billion parameter model achieves state-of-the-art performance on 50 diverse supervised NLP tasks, surpassing T5 and GPT-like models. Notably, it excels in language generation, understanding, classification, question answering, reasoning, and information retrieval. The model also performs well in zero-shot scenarios, outperforming GPT-3 and T5. The authors demonstrate its effectiveness in chain-of-thought prompting and reasoning, making it a valuable research option at a medium scale. Furthermore, applying FLAN instruction tuning enhances its performance, yielding competitive results in MMLU and Big-Bench scores compared to FLAN-PaLM 62B.
- **GPT-NeoX**: The GPT-NeoX-20B paper, authored by the Black et al. (2022), introduce an architecture similar to GPT-3 but with notable

differences. They utilize rotary positional embeddings for token position encoding instead of learned embeddings and parallelize the attention and feed-forward layers, resulting in a 15% increase in throughput. Unlike GPT-3, GPT-NeoX-20B exclusively employs dense layers. The authors trained GPT-NeoX-20B using EleutherAI's custom codebase (GPT-NeoX) based on Megatron and DeepSpeed, implemented in PyTorch. To address computational limitations, the authors reused the hyperparameters from the GPT-3 paper. In their evaluation, the researchers compared GPT-NeoX-20B's performance to their previous model, GPT-J-6B, as well as Meta's FairSeq 13B and different sizes of GPT-3 on various NLP benchmarks, including LAMBADA, WinoGrande, HendrycksTest, and MATH dataset. While improvements were desired for NLP tasks, GPT-NeoX-20B exhibited exceptional performance in science and math tasks.

- **ChatGPT**: ChatGPT (Brown et al., 2020) is an advanced language model developed by OpenAI. It is designed to generate human-like text responses in a conversational manner. ChatGPT is built upon the GPT (Generative Pre-trained Transformer) architecture, which is a state-of-the-art deep learning model for natural language processing. ChatGPT is trained on a massive amount of text data from the internet to learn patterns, grammar, and context in language. It utilizes a transformer-based neural network that consists of multiple layers of self-attention mechanisms and feed-forward neural networks. This architecture allows the model to understand and generate coherent and contextually relevant responses. The primary goal of ChatGPT is to provide natural and engaging interactions with users. It can be used in various applications, such as chatbots, virtual assistants, customer support systems, and more. By inputting a prompt or a message, users can receive a response generated by the model.

B A Case-Study with ChatGPT

Next, we thoroughly examined the performance of ChatGPT-3.5 on the Bangla Zero-Shot-MLC task. Due to the lack of access to the model's embeddings, we were unable to adopt the embedding-similarity-based classification approach as presented in Section 3.1. Instead, by utilizing the API, we adopted a prompting approach to perform the Zero-Shot-MLC task. Details of the ChatGPT prompts are presented in Table 5. For evaluation, we simply recorded the responses of ChatGPT as we provided these prompts. One interesting observation is that, initially, the ChatGPT responses included out-of-scope labels such as, "মিডিয়া" (Media), "জাতিসংঘ" (United Nations), "দুর্যোগ" (Disaster), "পুলিশ" (Police),"পরিবহন" (Transportation), etc. As Table 5 demonstrates, none of these labels were expected by the prompts, but ChatGPT took the liberty to include them in its response. More interestingly, after a few iterations, we noticed a decline in such occurrences. Unfortunately, due to the lack of access to the model's architecture and parameters, we were unable to investigate this context-sensitive behavior further.

<table>
<tr><th colspan="2">Prompt Design</th></tr>
<tr><td colspan="2">System setup</td></tr>
<tr><td colspan="2">The AI assistant has been designed to understand and categorize user input by the given labels. When processing user input, the assistant must predict the labels from one of the following pre-defined options: 'চাকরিবাকরি' (Job market), 'করোনাভাইরাস' (Coronavirus), 'চলচ্চিত্র ও তারকা' (Movies and celebrities), 'স্বাস্থ্য' (Health), 'ব্যাংক' (Banking), 'অর্থনীতি' (Economy), 'শিক্ষা' (Education), 'প্রাকৃতিক দুর্যোগ' (Natural disasters), 'আইন ও আদালত' (Law and justice), 'কূটনীতি' (Politics), 'শিল্প ও বাণিজ্য' (Industry and commerce), 'ভ্রমণ' (Travel), 'নকশা' (Design), 'ফুটবল' (Football), 'খাবারদাবার' (Food and dining), 'দেশ ও রাজনীতি' (Country and politics), 'আন্তর্জাতিক' (International), 'দেশের খবর' (Country news), 'রাশিয়া ইউক্রেন সংঘাত' (Russia-Ukraine conflict), 'ক্রিকেট' (Cricket), 'নারী' (Women). It is essential to note that an article may have multiple labels. If the user input is not relevant with any labels, the assistant should print nothing, indicating that the input does not align with the available categories. The agent MUST response with the following json format: {“Labels”: [“List of labels”]}</td></tr>
<tr><td>User</td><td>Taking into account the given Bangla article {ব্যাংক এশিয়া ২০১৪ সালে ব্যাংকিং সেবার বাইরে থাকা বিপুল জনগোষ্ঠীকে ব্যাংকিং সেবায় আনতে এজেন্ট ব্যাংকিং সেবা চালু করে। বর্তমানে রাষ্ট্রমালিকানাধীন ও বেসরকারি মিলিয়ে ৩১টি ব্যাংক এ সেবা দিচ্ছে। বর্তমানে এজেন্ট ব্যাংকিং সেবা গ্রহণকারীর সংখ্যা দেড় কোটির বেশি। এর মধ্যে ব্যাংক এশিয়ার গ্রাহক ৫৫ লাখের বেশি। এসব গ্রাহকের ৯২ শতাংশই গ্রামীণ জনগোষ্ঠী। আবার ৬২ শতাংশ গ্রাহকই নারী। সারা দেশে ব্যাংক এশিয়ার ৫ হাজার ৪০০-এর বেশি এজেন্ট আউটলেট রয়েছে, যাদের মধ্যে নারী এজেন্ট ৫৪০ জন। ..... (In 2014, Bank Asia introduced agent banking services to bring banking services to a large population that was outside the reach of traditional banking. At present, a total of 31 banks, including both public and private, are providing these services. The number of users availing agent banking services has surpassed 140 million. Among them, Bank Asia has more than 5.5 million customers, of which 92% are from rural areas. Furthermore, 62% of the customers are women. Throughout the country, there are more than 5,400 agent outlets of Bank Asia, including 540 outlets managed by female agents.....)}, predict the category or labels of this article from the list of mentioned labels.</td></tr>
<tr><td>ChatGPT</td><td>{“Labels”: “ব্যাংক” (Banking), “শিল্প ও বাণিজ্য” (Industry and commerce)}</td></tr>
<tr><td colspan="2">Directive: Taking into account the given Bangla article {article text}, predict the category or labels of this article from the list of mentioned {labels}. Please remember to only respond in the predefined JSON format without any additional information.</td></tr>
</table>

Table 5: Prompt design details for the Zero-Shot-MLC task on BanglaNewsNet.