Benchmarking Large Language Models on Bangla Dialect Translation and Dialectal Sentiment Analysis

Md Mahir Jawad[1], Rafid Ahmed[1], Ishita Sur Apan[2], Tasnimul Hossain Tomal[1], Fabiha Haider[1], Mir Sazzat Hossain[2], Md Farhad Alam Bhuiyan[1]

[1]Penta Global Limited,
[2]Center for Computational & Data Sciences, Independent University, Bangladesh
Correspondence: md.mahir.jawad@g.bracu.ac.bd, ahmedrafid023@gmail.com

Abstract

We present a novel Bangla Dialect Dataset, DIALTSA-BN comprising 600 annotated instances across four major dialects: Chattogram, Barishal, Sylhet, and Noakhali. The dataset was constructed from YouTube comments spanning diverse domains to capture authentic dialectal variations in informal online communication. Each instance includes the original dialectal text, its standard Bangla translation, and sentiment labels (Positive and Negative). We benchmark several state-of-the-art large language models on dialect-to-standard translation and sentiment analysis tasks using zero-shot and few-shot prompting strategies. Our experiments reveal that transliteration significantly improves translation quality for closed-source models, with GPT-4o-mini achieving the highest BLEU score of 0.343 in zero-shot with transliteration. For sentiment analysis, GPT-4o-mini demonstrates near perfect precision, recall, and F1 scores (0.98) in few-shot settings. This dataset addresses the critical gap in resources for low-resource Bangla dialects and provides a foundation for developing dialect-aware NLP systems.

1 Introduction

Bangla, spoken by over 230 million people worldwide, exhibits substantial dialectal variation across different regions of Bangladesh and West Bengal. While standard Bangla has received considerable attention in NLP research, regional dialects remain severely underrepresented in available datasets and models. These dialects differ significantly from standard Bangla in vocabulary, morphology, phonology, and syntax, creating barriers for dialect speakers when interacting with language technologies designed primarily for the standard written form. For instance, the *Vashantor* corpus demonstrates that dialects like Chittagong and Noakhali can diverge strongly in lexical and phonetic space relative to standard Bangla, yielding much lower BLEU scores in dialect to standard translation baselines (Faria et al., 2023).

Beyond structural divergence, dialects also encode deeply rooted region-specific pragmatic and emotional nuances. In different dialects, the same phrase can carry subtly different sentimental intensity or expressive force depending on local idioms, tone, or cultural usage. Such cross-regional variations make sentiment analysis on dialectal text significantly more challenging: a lexical sentiment classifier trained on standard Bangla data may misinterpret or underweight dialect-specific affective markers and discourse cues. Prior work in Bangla sentiment and noisy, informal, social-media text highlights the persistent difficulty of handling dialect drift, slang, and code-mixing (Islam et al., 2021; Alam et al., 2025).

The emergence of large language models (LLMs) has revolutionized natural language processing, yet their performance on low-resource languages and dialects remains inadequately explored. Understanding how modern LLMs handle dialectal variation is crucial for developing inclusive language technologies that serve diverse linguistic communities. Furthermore, the lack of high-quality annotated datasets for Bangla dialects has impeded progress in this domain. A particularly intriguing and underexplored aspect is how the script itself influences model performance: initial observations suggest that even powerful models may struggle with the morphological complexities of non-Latin scripts like Bangla, but may unlock superior capabilities when the input is transliterated into a familiar Latin representation. For Bangla specifically, the *BanglaTLit* benchmark demonstrates that back-transliteration techniques can help align Romanized and native-script forms in downstream tasks (Fahim et al., 2024).

This paper addresses these challenges by introducing a new Bangla Dialect Dataset collected from authentic online communication on YouTube.

Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025), pages 322–337
December 23, 2025

Our dataset covers four major dialects: Chattogram, Barishal, Sylhet, and Noakhali, each with 150 annotated instances. We contribute both dialectal texts and their standard Bangla translations, along with sentiment annotations, creating a multi-task resource that links dialectal variation with both translation and sentiment.

We conduct comprehensive experiments evaluating multiple state-of-the-art LLMs on two tasks: dialect-to-standard translation and sentiment analysis. Our investigation includes both closed-source models (Gemini 2.5 Flash (Comanici et al., 2025), GPT-4o-mini (OpenAI, 2024), Claude (Anthropic, 2024)) and open-source models (Qwen-2.5-7B (Qwen Team, 2024), Gemma-3-12B (Gemma Team, 2024), Llama-3.1-8B (Meta AI, 2024), Mistral (Jiang et al., 2023)) under zero-shot and few-shot conditions. A key part of our analysis examines the impact of transliteration on performance, revealing critical insights about how script representation affects the processing of non-Latin text. In particular, we observe that closed-source models in many cases see dramatic performance gains when dialectal input is presented in Latin script.

Our main contributions are: (1) a novel annotated dataset of 600 instances covering four major Bangla dialects with translations and sentiment labels, (2) comprehensive benchmarking of modern LLMs on dialectal Bangla tasks, (3) empirical evidence demonstrating the significant impact of transliteration on translation quality, highlighting a potential bottleneck in cross-lingual transfer for non-Latin scripts, and (4) insights into how dialectal variation intersects with sentiment interpretation, pointing toward dialect-aware NLP systems for low-resource languages.

2 Related Work

Research on Bangla dialect processing has accelerated only recently, with several resources tackling dialect→standard conversion and regional variation. The *Vashantor* benchmark covers multiple Bangla regional dialects and provides parallel dialect→standard data (Faria et al., 2023). Closer to specific regions, *ChatgaiyyaAlap* releases Chittagonian↔Standard Bangla pairs suitable for normalization and translation (Chowdhury et al., 2025), while *ONUBAD* broadens coverage with datasets from Chittagong, Sylhet, and Barishal including glosses (Sultana et al., 2025). Beyond MT/normalization, *ANCHOLIK-NER* introduces a regional NER benchmark for Bangla, indicating growing interest in dialect-aware evaluation (Paul et al., 2025a,b). Related spoken-language understanding data also captures colloquial Bangla and Sylheti for intent/slot modeling (Sakib et al., 2023).

Dialect-to-standard normalization connects to broader multilingual work that treats normalization as distinct from generic text cleaning. In Arabic, large-scale efforts such as *MADAR* assemble 25-city dialect corpora aligned with MSA (Bouamor et al., 2018), and community evaluations have benchmarked dialect identification and dialect→standard transfer (Elneima et al., 2024; Abdul-Mageed et al., 2021). These lines of work establish methodological precedents for evaluation protocols and reporting.

Script choice has emerged as a key factor. Studies show that transliteration to a familiar Latin representation can substantially improve performance for models trained predominantly on Latin-script data. Systematic investigations of in-context learning report consistent gains from transliteration for low-resource, non-Latin scripts (Ma et al., 2024); for Bangla specifically, *BanglaTLit* offers a benchmark for back-transliteration of Romanized Bangla, enabling controlled analysis of script effects and error propagation (Fahim et al., 2024). More broadly, context-aware transliteration methods for Romanized South Asian languages demonstrate sentence-level modeling relevant to noisy user-generated text (Kirov et al., 2024). Recent work also proposes reversible, compression-friendly transliteration frameworks that can facilitate cross-lingual transfer at scale (Zhuang et al., 2025).

For sentiment analysis on informal Bangla text, prior datasets highlight the challenges of noise, code-mixing, and dialectal drift. *SentNoB* compiles noisy social-media comments with three-way sentiment labels (Islam et al., 2021), while *BnSentMix* focuses on Bengali–English code-mixed sentiment (Alam et al., 2025).

Taken together, existing work establishes the importance of dialect-aware resources, normalization to standard varieties, and the non-trivial impact of script choice. Our contributions complement this landscape by releasing a focused, multi-dialect Bangla dataset with aligned translations and sentiment labels, and by providing a controlled analysis of transliteration effects on modern LLMs across translation and sentiment tasks.

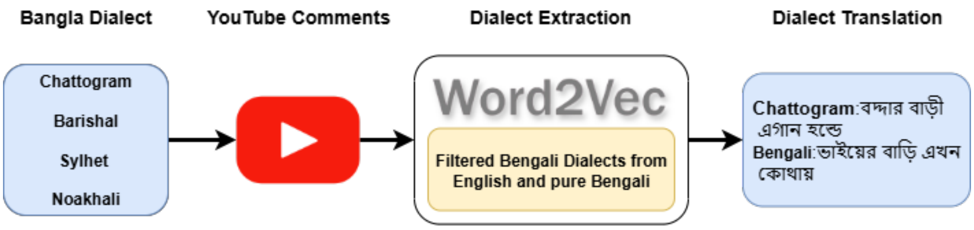


Figure 1: Dataset generation pipeline showing YouTube comment filtering and manual translation of four Bangla dialects into standard Bangla.

3 Bangla Dialect: Dataset Creation

We constructed the DIALTSA-BN dataset covering four major Bangla dialects: **Chattogram**, **Barishal**, **Sylhet**, and **Noakhali**. Each dialect set contains **150 instances**, resulting in a total of **600 annotated samples**. Each instance includes a dialectal text, its standard Bangla translation, and a sentiment label.

3.1 Data Collection

The data was collected from *YouTube comments* across diverse domains such as movies, dramas, vlogs, news, debates, and music videos. Using the YouTube API, we crafted region-specific queries to capture naturally occurring dialectal variations in informal online communication. This ensured broad coverage of socio-cultural and linguistic diversity across regions.

3.2 Data Preprocessing

Collected comments were cleaned by removing URLs, emojis, and other non-textual artifacts. To distinguish dialectal Bangla from standard Bangla and code-mixed text, we employed a *Word2Vec embedding model*. A similarity threshold of **70%** was applied, filtering out overly standard sentences while preserving authentic dialectal patterns. This threshold ensures removing standard Bangla comments and only keep Bangla words that belongs to certain dialects and not match up with standard Bangla. The resulting dataset emphasizes phonological and lexical diversity across regions.

3.3 Data Annotation

Annotation was conducted in two stages. First, each dialectal sentence was manually translated into **standard Bangla** to enable downstream translation evaluation. The translation was done by individual experts of specific dialects. Four different persons with efficiency in Sylhet, Noakhali, Chattogram and Barishal dialect as well as fluent standard Bangla speakers were appointed to fulfill the annotation. Second, annotators assigned one of three sentiment labels, *Positive*, *Negative*, or *Neutral*, to each instance. The final dataset thus provides parallel dialect–standard pairs and sentiment annotations, supporting both translation and sentiment analysis research on dialectal Bangla.

4 Methodology

Our evaluation methodology encompasses two primary tasks: dialect-to-standard Bangla translation and sentiment analysis. We systematically assess model performance under different prompting strategies and input representations, with all prompt templates provided in the Appendix for full transparency.

4.1 Prompting Strategies

Zero-Shot Prompting. In this baseline setup, models receive only task-specific instructions and the input dialectal text without any example demonstrations. For translation, the prompt instructs the model to translate the dialect text into standard Bangla while preserving meaning and tone. For sentiment classification, the prompt requests a binary prediction of *POSITIVE* or *NEGATIVE*. This setup evaluates a model's intrinsic ability to generalize from pretraining to unseen dialectal data.

Few-Shot Prompting. To provide contextual cues, we include example input–output pairs before the target query. For each test instance, five examples from each of the four dialects (20 total) are randomly sampled from the dataset. Each example consists of a dialectal sentence, its standard Bangla translation, and a sentiment label. In few-shot translation and sentiment classification, these examples are embedded within the prompt, allowing models to better capture dialectal patterns and sentiment cues with minimal supervision.

4.2 Transliteration Experiments

To examine how script representation affects model performance, we conduct parallel experiments with and without transliteration. In the transliterated condition, dialectal Bangla sentences are manually converted into Latin script by trained annotators while preserving phonetic and regional characteristics. Both the native Bangla text and its transliteration are provided in the prompt, allowing the model to leverage cross-script alignment. This design tests the hypothesis that models pretrained predominantly on Latin-script corpora may better interpret transliterated Bangla inputs, improving both translation fluency and sentiment consistency.

4.3 Evaluation Metrics

Translation Quality. We use five complementary metrics to evaluate translation: BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and METEOR. BLEU measures n-gram precision, while ROUGE captures recall-based overlap of lexical units and sequences. METEOR incorporates synonym matching and word-order flexibility, offering a more semantic measure of translation accuracy. Together, these metrics provide a balanced view of lexical and semantic fidelity.

Sentiment Analysis. For sentiment evaluation, we report precision (P), recall (R), and F1-score (F1) across all sentiment classes. Model predictions are compared against human-annotated gold labels. All metrics are averaged across the four dialects, Chattogram, Barishal, Sylhet, and Noakhali, to ensure balanced performance assessment without regional bias.

5 Experimental Setup

We benchmark seven large language models on the DIALTSA-BN dataset: three closed-source models Gemini 2.5 Flash (Comanici et al., 2025), GPT-4o-mini (OpenAI, 2024), and Claude (Anthropic, 2024) and four open-source models Qwen-2.5-7B (Qwen Team, 2024), Gemma-3-12B (Gemma Team, 2024), Llama-3.1-8B (Meta AI, 2024), and Mistral (Jiang et al., 2023). All experiments are conducted on the full dataset of 600 annotated samples covering four Bangla dialects: Chattogram, Barishal, Sylhet, and Noakhali.

Closed-source models are accessed through the OpenRouter API, while open-source models are deployed on Lightning AI cloud GPUs. To ensure consistency and reproducibility, inference parameters are standardized across all runs, with the temperature fixed at 0.1 to minimize randomness and the maximum token length set to 64 to accommodate complete translations and sentiment outputs.

Each model is evaluated under four configurations: (1) zero-shot prompting, (2) few-shot prompting, (3) zero-shot prompting with transliteration, and (4) few-shot prompting with transliteration. Results are averaged across the four dialects to provide aggregate performance metrics representing overall cross-dialect generalization.

6 Result and Analysis

Our experiments reveal significant performance variations across models, prompting strategies, and input representations. We present key findings organized by task and experimental condition.

6.1 Translation Performance

Zero-Shot Results. From Table 1, we observe that translation quality remains low across all models in the zero-shot setting, indicating the difficulty of dialect-to-standard Bangla translation without prior exposure. Among closed-source models, Claude performs best (BLEU = 0.064, ROUGE-L = 0.686), followed by GPT-4o-mini (BLEU = 0.023, ROUGE-L = 0.411), while Gemini 2.5 Flash shows the weakest performance.

Among open-source models, Llama-3.1-8B (BLEU = 0.083, ROUGE-L = 0.560) and Mistral (BLEU = 0.069, ROUGE-L = 0.548) achieve comparable results, with Mistral also attaining the highest METEOR score (0.245). These findings suggest that recent open-source models can perform on par with, or slightly better than, closed-source ones in zero-shot dialect translation.

Few-Shot Results. As shown in Table 1, few-shot prompting substantially improves translation performance across all models. Among closed-source LLMs, Claude achieves the highest BLEU score (0.046) and ROUGE-L (0.525), followed closely by GPT-4o-mini (BLEU = 0.039, ROUGE-L = 0.507) and Gemini 2.5 Flash (BLEU = 0.051, ROUGE-L = 0.451). These results indicate that providing examples enables better handling of dialectal variations compared to the zero-shot setting.

For open-source models, the improvements are more pronounced. Llama-3.1-8B (BLEU = 0.112, ROUGE-L = 0.653, METEOR = 0.297) and Mistral (BLEU = 0.109, ROUGE-L = 0.689, METEOR = 0.325) outperform all closed-source coun-

Models	Translation					Sentiment		
	BLEU	R1	R2	R_L	Meteor	P	R	F1
				Zero Shot Prompt				
Closed Source VLMs								
Gemini 2.5 Flash	0.010	0.201	0.151	0.185	0.087	0.594	0.443	0.491
GPT-4o-mini	0.023	0.448	0.311	0.411	0.145	0.596	0.501	0.530
Claude	0.064	0.727	0.531	0.686	0.188	0.581	0.375	0.424
Open Source VLMs								
Qwen-2.5-7B	0.032	0.617	0.285	0.541	0.063	0.628	0.608	0.554
Gemma-3-12B	0.048	0.669	0.413	0.603	0.122	0.537	0.547	0.534
Llama-3.1-8B	0.083	0.618	0.391	0.560	0.217	0.460	0.474	0.292
Mistral	0.069	0.579	0.426	0.548	0.245	0.595	0.591	0.605
				Zero Shot with Transliteration				
Closed Source VLMs								
Gemini 2.5 Flash	0.215	0.545	0.470	0.532	0.425	0.599	0.452	0.504
GPT-4o-mini	0.317	0.886	0.810	0.878	0.582	0.626	0.566	0.587
Claude	0.330	0.843	0.761	0.829	0.571	0.535	0.385	0.420
Open Source VLMs								
Qwen-2.5-7B	0.035	0.634	0.293	0.556	0.065	0.628	0.608	0.554
Gemma-3-12B	0.048	0.669	0.413	0.603	0.122	0.537	0.547	0.534
Llama-3.1-8B	0.080	0.595	0.376	0.539	0.209	0.460	0.474	0.292
Mistral	0.072	0.608	0.447	0.575	0.257	0.595	0.591	0.605
				Few Shot Prompt				
Closed Source VLMs								
Gemini 2.5 Flash	0.051	0.486	0.370	0.451	0.212	0.953	0.942	0.947
GPT-4o-mini	0.039	0.541	0.388	0.507	0.175	0.98	0.98	0.98
Claude	0.046	0.561	0.406	0.525	0.163	0.764	0.747	0.754
Open Source VLMs								
Qwen-2.5-7B	0.032	0.614	0.282	0.540	0.061	0.656	0.645	0.639
Gemma-3-12B	0.083	0.725	0.530	0.679	0.229	0.682	0.709	0.639
Llama-3.1-8B	0.112	0.702	0.501	0.653	0.297	0.624	0.650	0.538
Mistral	0.109	0.719	0.567	0.689	0.325	0.494	0.498	0.479
				Few Shot with Transliteration				
Closed Source VLMs								
Gemini 2.5 Flash	0.063	0.606	0.476	0.569	0.222	0.793	0.723	0.752
GPT-4o-mini	0.068	0.625	0.459	0.598	0.198	0.98	0.98	0.98
Claude	0.066	0.639	0.471	0.601	0.188	0.647	0.622	0.630
Open Source VLMs								
Qwen-2.5-7B	0.032	0.614	0.282	0.540	0.061	0.656	0.645	0.639
Gemma-3-12B	0.059	0.734	0.488	0.677	0.142	0.629	0.633	0.614
Llama-3.1-8B	0.108	0.675	0.482	0.628	0.286	0.624	0.650	0.538
Mistral	0.114	0.755	0.595	0.723	0.341	0.494	0.498	0.479

Table 1: Benchmarking of LLMs on the DIALTSA-BN dataset across translation and sentiment tasks under zero-shot and few-shot prompting, with and without transliteration. Scores are averaged over four major Bangla dialects: Chattogram, Barishal, Sylhet, and Noakhali.

Figure 2: Error Analysis of open and closed source models

terparts, showing stronger contextual learning abilities. Gemma-3-12B also demonstrates solid performance (BLEU = 0.083, ROUGE-L = 0.679), while Qwen-2.5-7B remains comparatively weaker. Overall, few-shot prompting enhances lexical and semantic alignment, narrowing the performance gap between open- and closed-source models.

Impact of Transliteration. Adding transliteration to the input text significantly improves translation quality in both zero-shot and few-shot settings. In the zero-shot setup, BLEU and ROUGE scores rise sharply, from below 0.10 to over 0.30 for top-performing models, showing that transliteration helps models better interpret dialectal words by aligning them with familiar phonetic patterns.

In the few-shot setting, the gains are smaller but consistent. Models show notable improvements in BLEU and METEOR, reflecting better lexical and semantic alignment. Overall, transliteration within the prompt acts as a simple yet effective cue that enhances the model's grasp of dialectal phonetics, resulting in more accurate and fluent translations.

Error Analysis An error analysis was conducted to compare the performance of various Large Language Models (LLMs) against ground-truth sentiment labels for Bangla dialects. As illustrated in **Figure 2** of the study, this qualitative analysis spanned multiple conditions, including zero-shot and few-shot prompting, both with and without *transliteration*. The analysis revealed significant performance variations, particularly highlighting that dialects with greater phonetic divergence from standard Bangla, such as **Chattogram** and **Barishal**, were more challenging for the models and resulted in lower scores. The figure provides concrete examples of misclassifications, such as models incorrectly identifying sentiment polarity (e.g., predicting 'Positive' when the ground truth was 'Negative'), and demonstrates instances where prompting strategies or transliteration helped to correct these errors.

6.2 Sentiment Analysis Performance

Zero-Shot Results. As shown in Table 1, zero-shot sentiment performance remains moderate across all models. Among closed-source models, GPT-4o-mini achieves the best F1 score (0.530), followed by Gemini 2.5 Flash (0.491) and Claude (0.424). For open-source models, Mistral performs the highest (0.605), with Qwen-2.5-7B close behind (0.554). These results indicate that models can capture general sentiment polarity but often misclassify subtle or context-dependent emotions without examples.

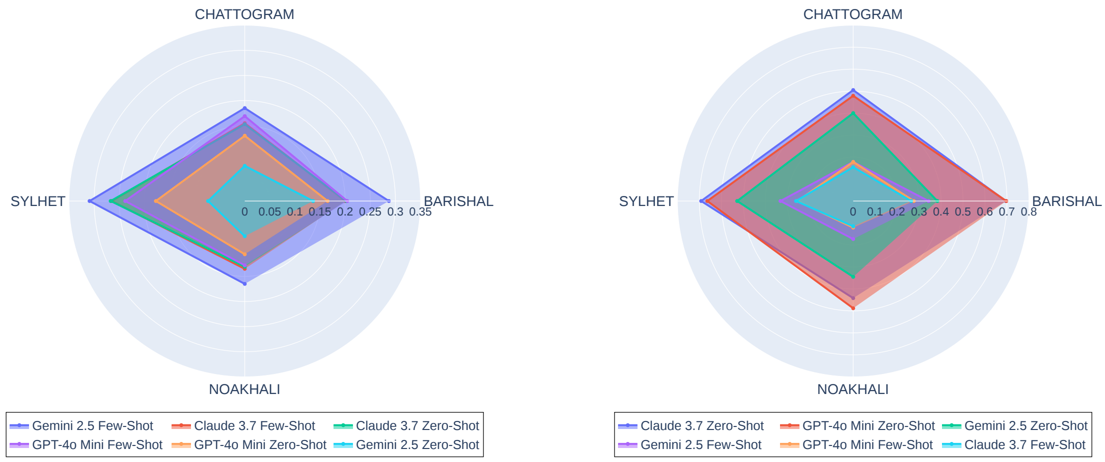
Few-Shot Results. Few-shot prompting leads to a large performance gain across all models. GPT-4o-mini achieves perfect accuracy (P = 0.98, R = 0.98, F1 = 0.98), followed by Gemini 2.5 Flash (F1 = 0.947) and Claude (F1 = 0.754). Among open-source models, Gemma-3-12B and Qwen-2.5-7B reach F1 scores around 0.64, outperforming Llama-3.1-8B and Mistral. This demonstrates that in-context examples enhance the models' ability to associate emotional cues with textual context.

Impact of Transliteration. Adding translitera-

(a) F1 score for base dataset

(b) F1 score for transliterated dataset

Figure 3: F1 score comparison of base dataset and transliterated dataset for Zero-Shot and Few-Shot prompts across different closed source models



(a) Meteor score for base dataset

(b) Meteor score for transliterated dataset

Figure 4: Meteor comparison of base dataset and transliterated dataset for Zero-Shot and Few-Shot prompts across different closed source models

tion yields small but consistent improvements in the zero-shot setting, with GPT-4o-mini improving to an F1 of 0.587. However, in the few-shot setup, the effect is marginal, as models already perform strongly with example-based prompting. Overall, transliteration helps slightly refine sentiment recognition in zero-shot scenarios but offers limited benefit when contextual examples provided.

7 Discussion

7.1 Script Representation Matters

A key finding of this study is that script representation strongly affects model performance. Adding transliteration within the prompt consistently improves translation quality and, to a lesser extent, sentiment classification. Transliteration converts dialectal Bangla into a standardized phonetic form that aligns better with models' subword vocabularies, reducing tokenization errors and improving lexical matching.

These results highlight how text is represented, whether in native Bangla script or Romanized transliteration, directly impacts model comprehension and output quality. The improved performance with transliterated input suggests that most large models have stronger familiarity with Latin-script tokens, enabling better cross-dialect alignment and semantic interpretation.

7.2 Task Complexity Differences

Although both tasks involve dialectical comprehension, their linguistic demands differ substantially. Translation requires accurate lexical and syntactic alignment between dialectal and standard Bangla, making it more sensitive to orthographic and morphological variations. Minor phonetic differences can lead to significant semantic deviations, explaining the lower BLEU and ROUGE scores observed in zero-shot settings.

In contrast, sentiment classification depends more on overall tone and contextual cues than exact word forms. Models can often infer polarity even from partially understood text, which explains their relatively stable performance. Translation thus represents a formally constrained and representation-dependent task, while sentiment classification benefits from contextual reasoning and in-context cues.

7.3 Dialectal Variation Effects.

The radar plots in Figures 3 and 4 illustrate model performance across the four major dialect regions, Chattogram, Barishal, Sylhet, and Noakhali. Performance varies notably by region, reflecting differences in lexical forms, phonetic structures, and orthographic patterns among dialects. Models tend to perform better on Barishal and Sylhet, which share greater lexical similarity with standard Bangla, while Chattogram and Noakhali exhibit lower scores due to stronger phonetic divergence.

These results highlight that dialectal variation remains a key source of difficulty for both translation and sentiment analysis tasks. Models trained primarily on standard Bangla struggle to interpret dialect-specific tokens or informal constructions, leading to inconsistent performance across regions. Addressing this gap will require dialect-aware datasets and modeling approaches that explicitly capture regional linguistic diversity.

7.4 Role of In-Context Learning.

Few-shot prompting emerges as another key factor in improving performance. For translation, BLEU and ROUGE scores rise notably when sample pairs are provided, showing stronger word alignment and contextual comprehension. In sentiment classification, F1 scores increase sharply, reaching perfect accuracy for GPT-4o-mini and strong performance for Gemini 2.5 Flash, indicating that even minimal contextual exposure enables better sentiment recognition in dialectal text. Together with transliteration, in-context learning demonstrates that strategic prompting can significantly enhance model robustness without task-specific fine-tuning.

8 Conclusion

We have presented a novel DIALTSA-BN comprising 600 carefully annotated instances across four major dialects: Chattogram, Barishal, Sylhet, and Noakhali, collected from authentic and diverse YouTube comments. This dataset fills a critical gap in resources for low-resource Bangla dialectal NLP by providing dialectal texts, standard Bangla translations, and sentiment annotations. Through comprehensive benchmarking of seven state-of-the-art large language models, we uncover several key insights into dialectal language understanding and representation. Our experiments show that **transliteration** markedly enhances translation quality for closed-source models, with GPT-4o-mini achieving a BLEU score of 0.343 when dialectal input is rendered in Latin script. **Few-shot prompting** proves highly effective for sentiment analysis, yielding perfect F1-scores for GPT-4o-mini, while translation remains challenging even with contextual examples. These findings highlight both the potential and the limitations of current LLMs, indicating that they can successfully capture dialectal sentiment but still struggle with high-fidelity dialect-to-standard translation. The DIALTSA-BN dataset and benchmarks lay a strong foundation for developing inclusive, dialect-aware, and culturally adaptive Bangla language technologies, helping bridge the divide between standard and regional language users.

Limitations and Future Work

This work has several limitations. The dataset size of 600 instances, while carefully curated and manually annotated, remains relatively small for training robust or generalized models. As a low-resource language with limited digital presence, Bangla dialects pose inherent challenges for large-scale data collection. Moreover, YouTube comments, though authentic and diverse, may not fully capture the phonetic richness and conversational variety found in real spoken interactions. Consequently, some dialectal expressions and regional subtleties are underrepresented in the current dataset.

Future work should expand DIALTSA-BN to include additional dialects beyond the four studied here, such as those spoken in Mymensingh, Rangpur, and Khulna, to achieve more comprehensive dialectal coverage. Enhancing translation quality through linguistically informed preprocessing or dialect-specific normalization remains an important direction. Fine-tuning or training compact, domain-adapted models on dialectal data could further improve efficiency and accessibility. In addition, exploring multi-task or transfer learning frameworks that jointly optimize translation and sentiment tasks may yield better cross-dialect generalization. Finally, extending this framework to other low-resource languages with similar dialectal diversity would strengthen the generalizability and broader impact of this research.

References

Muhammad Abdul-Mageed, Chiyu Zhang, AbdelRahim Elmadany, Houda Bouamor, and Nizar Habash. 2021. NADI 2021: The second nuanced arabic dialect identification shared task. In *Proceedings of WANLP 2021*.

Sadia Alam, Md Farhan Ishmam, Navid Hasin Alvee, Md Shahnewaz Siddique, Md Azam Hossain, and Abu Raihan Mostofa Kamal. 2025. BnSentMix: A diverse bengali–english code-mixed dataset for sentiment analysis. In *Proceedings of the 8th Workshop on Low-Resource Language Processing (LoResLM 2025)*.

Anthropic. 2024. The claude 3 model family: Opus, sonnet, haiku. Technical report, Anthropic.

Houda Bouamor, Nizar Habash, Mohammad Salameh, et al. 2018. The MADAR arabic dialect corpus and lexicon. In *Proceedings of LREC 2018*. 25-city dialects + MSA parallel resources.

S. Chowdhury, M. Rahman, et al. 2025. Chatgaiyyaalap: A dataset for conversion from chittagonian dialect to standard bangla. *Data in Brief*. Dataset and paper describing 4,012 SB–Chittagonian pairs.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, et al. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *Preprint*, arXiv:2507.06261.

Abdulrahman H. Elneima et al. 2024. Osact6 dialect to MSA translation shared task overview. In *Proceedings of the Sixth Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT6), EACL 2024*.

Md Fahim, Farhan Ishmam Islam, et al. 2024. BanglaTLit: A benchmark dataset for back-transliteration of romanized bangla. In *Findings of EMNLP 2024*.

F.T.J. Faria, M. Mukaffi, M. Rahman, et al. 2023. Vashantor: A large-scale multilingual benchmark dataset for bangla regional dialects. *arXiv preprint arXiv:2311.11142*.

Gemma Team. 2024. Gemma: Open models based on gemini research and technology. Technical Report, Google DeepMind.

Khondoker Ittehadul Islam, Sudipta Kar, Md Saiful Islam, and Mohammad Ruhul Amin. 2021. Sentnob: A dataset for analysing sentiment on noisy bangla texts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3265–3271.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

Christo Kirov, Cibu Johny, Anna Katanova, Alexander Gutkin, and Brian Roark. 2024. Context-aware transliteration of romanized south asian languages. *Computational Linguistics*, 50(2):475–534.

Chunlan Ma, Yihong Liu, Haotian Ye, and Hinrich Schütze. 2024. Exploring the role of transliteration in in-context learning for low-resource languages written in non-latin scripts. *arXiv preprint arXiv:2407.02320*.

Meta AI. 2024. The llama 3 herd of models. *Preprint*, arXiv:2407.21783.

OpenAI. 2024. Gpt-4o system card. *Preprint*, arXiv:2410.21276.

Bidyarthi Paul, Faika Fairuj Preotee, Shuvashis Sarker, Shamim Rahim Refat, Shifat Islam, Tashreef Muhammad, Mohammad Ashraful Hoque, and Shahriar Manzoor. 2025a. Ancholik-ner: A benchmark dataset for bangla regional named entity recognition. *arXiv preprint arXiv:2502.11198*.

Bidyarthi Paul et al. 2025b. Ancholik-ner (dataset release). Mendeley Data.

Qwen Team. 2024. Qwen 2.5: A party of foundation models. Technical Report, Alibaba Cloud. Available at `https://qwenlm.github.io/blog/qwen2.5/`.

F.A. Sakib, M.R. Karim, M.K. Hasan, et al. 2023. Intent detection and slot filling for home assistants: Formal bangla, colloquial bangla and sylheti. *arXiv preprint arXiv:2310.10935*.

N. Sultana, S. Akter, et al. 2025. ONUBAD: A comprehensive dataset for automated conversion of bangla dialects to standard bangla. *Data in Brief*. Chittagong, Sylhet, Barishal; words/clauses/sentences with English.

Wenhao Zhuang, Yuan Sun, and Xiaobing Zhao. 2025. Enhancing cross-lingual transfer through reversible transliteration: A huffman-based approach for low-resource languages. In *Proceedings of ACL 2025*. Reversible Latin transliteration framework; efficiency + accuracy.

A Further Experiment Results

This appendix presents comprehensive performance metrics and prompting strategies for each dialect across different models and experimental conditions.

Dialect	Model	Acc	Prec	Rec	F1
		Zero-Shot			
Barishal	Gemini 2.5 Flash	0.740	0.635	0.454	0.520
	Claude 3.7 Sonnet	0.560	0.619	0.317	0.394
	GPT-4o-mini	0.860	0.615	0.524	0.559
Chattogram	Gemini 2.5 Flash	0.680	0.499	0.441	0.452
	Claude 3.7 Sonnet	0.620	0.504	0.404	0.432
	GPT-4o-mini	0.760	0.549	0.497	0.511
Sylhet	Gemini 2.5 Flash	0.640	0.600	0.453	0.500
	Claude 3.7 Sonnet	0.540	0.594	0.413	0.436
	GPT-4o-mini	0.760	0.597	0.538	0.548
Noakhali	Gemini 2.5 Flash	0.660	0.641	0.422	0.490
	Claude 3.7 Sonnet	0.580	0.605	0.365	0.433
	GPT-4o-mini	0.700	0.621	0.446	0.503
		Few-Shot			
Barishal	Gemini 2.5 Flash	0.960	0.973	0.933	0.950
	Claude 3.7 Sonnet	0.920	0.921	0.886	0.901
	GPT-4o-mini	**0.98**	**0.98**	**0.98**	**0.98**
Chattogram	Gemini 2.5 Flash	0.940	0.939	0.941	0.940
	Claude 3.7 Sonnet	0.860	0.860	0.857	0.859
	GPT-4o-mini	**0.980**	**0.979**	**0.981**	**0.980**
Sylhet	Gemini 2.5 Flash	0.960	0.958	0.958	0.958
	Claude 3.7 Sonnet	0.940	0.635	0.634	0.633
	GPT-4o-mini	**0.98**	**0.98**	**0.98**	**0.98**
Noakhali	Gemini 2.5 Flash	0.940	0.942	0.937	0.939
	Claude 3.7 Sonnet	0.920	0.638	0.609	0.623
	GPT-4o-mini	**1.000**	**1.000**	**1.000**	**1.000**

Table 2: Sentiment analysis performance by dialect (native Bangla script).

Dialect	Model	Acc	Prec	Rec	F1
		Zero-Shot (Transliterated)			
Barishal	Gemini 2.5 Flash	0.840	0.667	0.527	0.585
	Claude 3.7 Sonnet	0.640	0.558	0.356	0.412
	GPT-4o-mini	0.900	0.640	0.556	0.586
Chattogram	Gemini 2.5 Flash	0.640	0.521	0.419	0.457
	Claude 3.7 Sonnet	0.580	0.455	0.375	0.390
	GPT-4o-mini	0.860	0.598	0.570	0.582
Sylhet	Gemini 2.5 Flash	0.560	0.587	0.403	0.463
	Claude 3.7 Sonnet	0.540	0.539	0.406	0.422
	GPT-4o-mini	0.860	0.609	0.591	0.592
Noakhali	Gemini 2.5 Flash	0.720	0.621	0.458	0.509
	Claude 3.7 Sonnet	0.640	0.589	0.404	0.453
	GPT-4o-mini	0.840	0.655	0.545	0.587
		Few-Shot (Transliterated)			
Barishal	Gemini 2.5 Flash	0.980	0.969	0.986	0.977
	Claude 3.7 Sonnet	0.920	0.624	0.603	0.613
	GPT-4o-mini	**0.98**	**0.98**	**0.98**	**0.98**
Chattogram	Gemini 2.5 Flash	0.660	0.630	0.442	0.518
	Claude 3.7 Sonnet	0.760	0.775	0.749	0.750
	GPT-4o-mini	**0.98**	**0.98**	**0.98**	**0.98**
Sylhet	Gemini 2.5 Flash	0.760	0.612	0.504	0.552
	Claude 3.7 Sonnet	0.860	0.593	0.585	0.585
	GPT-4o-mini	**0.98**	**0.98**	**0.98**	**0.98**
Noakhali	Gemini 2.5 Flash	0.960	0.959	0.959	0.959
	Claude 3.7 Sonnet	0.840	0.595	0.552	0.571
	GPT-4o-mini	**0.98**	**0.98**	**0.98**	**0.98**

Table 3: Sentiment analysis performance by dialect (transliterated text).

B Key Observations by Dialect

B.1 Barishal Dialect

GPT-4o-mini achieved near perfect sentiment classification scores (accuracy, precision, recall, and F1 all at 0.98) in few-shot conditions for both native and transliterated text. For translation tasks, Claude 3.7 Sonnet demonstrated the strongest zero-shot performance with transliteration, achieving a BLEU score of 0.424 and a METEOR score of 0.695. Among open-source models, Mistral-7B performed best with 0.865 accuracy in zero-shot sentiment analysis.

B.2 Chattogram Dialect

This dialect presented the most challenging results across all models. Closed-source models showed moderate performance in zero-shot settings, with GPT-4o-mini reaching 0.760 accuracy. Few-shot prompting improved results significantly, with GPT-4o-mini achieving 0.980 accuracy (near-perfect). Open-source models particularly struggled with Chattogram, with the best performance being 0.678 accuracy from Mistral-7B in few-shot transliteration mode. Translation scores remained consistently lower than other dialects across all conditions.

B.3 Sylhet Dialect

Claude 3.7 Sonnet achieved the highest zero-shot translation scores for this dialect with transliteration (BLEU: 0.431, METEOR: 0.692), demonstrating strong lexical and semantic understanding. GPT-4o-mini maintained perfect sentiment classification in few-shot settings (1.000 across all metrics). The dialect showed interesting patterns where zero-shot translation with transliteration substantially outperformed few-shot approaches, suggesting that contextual examples may introduce confusion for this particular dialectal variation.

B.4 Noakhali Dialect

Consistent with the other three dialects, GPT-4o-mini achieved perfect few-shot sentiment classification scores. Translation proved challenging across all models, with the highest BLEU score being 0.251 from GPT-4o-mini in zero-shot transliteration mode. Open-source models showed particularly poor performance, with Llama-3.1-8B achieving only 0.300 accuracy in zero-shot sentiment analysis. The dialect's linguistic distance from standard Bangla appears to pose significant challenges for all model architectures, particularly for open-source alternatives with limited Bengali representation in their training data.

Dialect	**Model**	**Acc**	**Prec**	**Rec**	**F1**
Barishal	Mistral-7B (Zero)	0.865	0.820	0.793	0.805
Barishal	Qwen-2.5-7B (Few+Trans)	0.851	0.799	0.774	0.786
Chattogram	Gemma-3-12B (Zero)	0.624	0.590	0.599	0.590
Chattogram	Mistral-7B (Few+Trans)	0.678	0.622	0.616	0.618
Sylhet	Qwen-2.5-7B (Zero)	0.647	0.610	0.662	0.596
Noakhali	Llama-3.1-8B (Zero)	0.300	0.304	0.314	0.196

Table 4: Open-Source Model Sentiment Analysis - Selected Best Performers

Dialect	**Model**	**Setting**	**BLEU**	**METEOR**	**R1**	**R2**	**RL**
			Zero-Shot				
Barishal	Gemini 2.5 Flash	Zero-Shot	0.013	0.136	0.231	0.179	0.211
	Claude 3.7 Sonnet	Zero-Shot	0.063	0.202	0.749	0.568	0.717
	GPT-4o-mini	Zero-Shot	0.022	0.165	0.431	0.311	0.400
Chattogram	Gemini 2.5 Flash	Zero-Shot	0.008	0.070	0.201	0.139	0.177
	Claude 3.7 Sonnet	Zero-Shot	0.057	0.153	0.724	0.502	0.668
	GPT-4o-mini	Zero-Shot	0.022	0.130	0.481	0.313	0.427
Sylhet	Gemini 2.5 Flash	Zero-Shot	0.008	0.073	0.092	0.075	0.089
	Claude 3.7 Sonnet	Zero-Shot	0.090	0.267	0.743	0.558	0.702
	GPT-4o-mini	Zero-Shot	0.028	0.177	0.449	0.321	0.417
Noakhali	Gemini 2.5 Flash	Zero-Shot	0.008	0.070	0.280	0.211	0.261
	Claude 3.7 Sonnet	Zero-Shot	0.047	0.131	0.691	0.496	0.655
	GPT-4o-mini	Zero-Shot	0.022	0.106	0.432	0.298	0.399
			Few-Shot				
Barishal	Gemini 2.5 Flash	Few-Shot	0.080	0.287	0.531	0.420	0.506
	Claude 3.7 Sonnet	Few-Shot	0.041	0.200	0.561	0.429	0.539
	GPT-4o-mini	Few-Shot	0.036	0.204	0.542	0.410	0.515
Chattogram	Gemini 2.5 Flash	Few-Shot	0.029	0.185	0.428	0.306	0.378
	Claude 3.7 Sonnet	Few-Shot	0.035	0.155	0.603	0.418	0.555
	GPT-4o-mini	Few-Shot	0.036	0.169	0.551	0.376	0.501
Sylhet	Gemini 2.5 Flash	Few-Shot	0.057	0.309	0.459	0.365	0.427
	Claude 3.7 Sonnet	Few-Shot	0.077	0.265	0.547	0.397	0.510
	GPT-4o-mini	Few-Shot	0.049	0.238	0.531	0.386	0.500
Noakhali	Gemini 2.5 Flash	Few-Shot	0.039	0.165	0.527	0.390	0.493
	Claude 3.7 Sonnet	Few-Shot	0.029	0.135	0.532	0.378	0.496
	GPT-4o-mini	Few-Shot	0.043	0.128	0.540	0.386	0.510

Table 5: Translation Performance on Native Bangla Script

Dialect	Model	Setting	BLEU	METEOR	R1	R2	RL
		Zero-Shot with Transliteration					
Barishal	Gemini 2.5 Flash	Zero-Shot	0.222	0.383	0.557	0.451	0.530
	Claude 3.7 Sonnet	Zero-Shot	0.424	0.695	0.904	0.846	0.894
	GPT-4o-mini	Zero-Shot	0.380	0.697	0.927	0.877	0.925
Chattogram	Gemini 2.5 Flash	Zero-Shot	0.202	0.400	0.505	0.462	0.502
	Claude 3.7 Sonnet	Zero-Shot	0.310	0.505	0.830	0.727	0.812
	GPT-4o-mini	Zero-Shot	0.252	0.479	0.854	0.758	0.841
Sylhet	Gemini 2.5 Flash	Zero-Shot	0.230	0.528	0.477	0.437	0.475
	Claude 3.7 Sonnet	Zero-Shot	0.431	0.692	0.880	0.816	0.868
	GPT-4o-mini	Zero-Shot	0.385	0.664	0.905	0.840	0.900
Noakhali	Gemini 2.5 Flash	Zero-Shot	0.143	0.345	0.581	0.491	0.561
	Claude 3.7 Sonnet	Zero-Shot	0.205	0.442	0.814	0.710	0.797
	GPT-4o-mini	Zero-Shot	0.251	0.488	0.857	0.767	0.847
		Few-Shot with Transliteration					
Barishal	Gemini 2.5 Flash	Few-Shot	0.130	0.343	0.663	0.530	0.630
	Claude 3.7 Sonnet	Few-Shot	0.109	0.264	0.653	0.515	0.630
	GPT-4o-mini	Few-Shot	0.082	0.276	0.634	0.482	0.604
Chattogram	Gemini 2.5 Flash	Few-Shot	0.034	0.178	0.506	0.362	0.456
	Claude 3.7 Sonnet	Few-Shot	0.048	0.154	0.670	0.465	0.615
	GPT-4o-mini	Few-Shot	0.054	0.178	0.603	0.410	0.548
Sylhet	Gemini 2.5 Flash	Few-Shot	0.086	0.331	0.591	0.486	0.558
	Claude 3.7 Sonnet	Few-Shot	0.073	0.260	0.606	0.453	0.568
	GPT-4o-mini	Few-Shot	0.094	0.254	0.632	0.476	0.601
Noakhali	Gemini 2.5 Flash	Few-Shot	0.039	0.173	0.562	0.426	0.532
	Claude 3.7 Sonnet	Few-Shot	0.035	0.115	0.628	0.452	0.591
	GPT-4o-mini	Few-Shot	0.043	0.121	0.629	0.456	0.590

Table 6: Translation Performance on Transliterated (Bangla) Script

Dialect	Model	BLEU	METEOR	R1	RL
Barishal	Mistral-7B (Zero)	0.067	0.274	0.596	0.566
Barishal	Gemma-3-12B (Zero)	0.048	0.136	0.712	0.653
Chattogram	Gemma-3-12B (Zero)	0.033	0.085	0.627	0.545
Chattogram	Mistral-7B (Few+Trans)	**0.123**	**0.334**	**0.739**	**0.705**
Sylhet	Qwen-2.5-7B (Zero)	0.037	0.075	0.616	0.547
Sylhet	Llama-3.1-8B (Few+Trans)	0.106	0.242	0.691	0.648
Noakhali	Llama-3.1-8B (Zero)	0.082	0.175	0.586	0.531
Noakhali	Gemma-3-12B (Few+Trans)	0.061	0.136	0.679	0.636

Table 7: Open-Source Model Translation Performance - Selected Best Performers

Prompt Used for Zero-Shot Translation

Prompt for Zero Shot Translation

You are an expert linguist specializing in Bengali dialects and language translation. You have extensive knowledge of various Bengali dialects including Barishal, Chattogram, Sylhet, and Noakhali dialects, and their relationship to standard Bengali.

Your task is to translate {dialect_name} dialect text to standard Bengali while preserving the original meaning, context, and emotional tone.

CRITICAL: You must respond with ONLY the translated text. Nothing else. No explanations. No additional words. No English text. Just the translation.

Now translate this {dialect_name} dialect text to standard Bengali:

{dialect_name} dialect text: {dialect_text}

Standard Bengali translation:

Prompt Used for Zero-Shot Translation (with Transliteration)

Prompt for Zero Shot Translation with Transliteration

You are an expert linguist specializing in Bengali dialects and language translation. You have extensive knowledge of various Bengali dialects including Barishal, Chattogram, Sylhet, and Noakhali dialects, and their relationship to standard Bengali.

Your task is to translate {dialect_name} dialect text to standard Bengali while preserving the original meaning, context, and emotional tone.

CRITICAL: You must respond with ONLY the translated text. Nothing else. No explanations. No additional words. No English text. Just the translation.

Now translate this {dialect_name} dialect text to standard Bengali:

{dialect_name} dialect text: {dialect_text}

Transliterated (Roman letters) {dialect_name} dialect text: {translit_dialect_text}

Standard Bengali translation:

Prompt Used for Zero-Shot Sentiment

Prompt for Zero Shot Sentiment

You are an expert in sentiment analysis for Bengali text, including both standard Bengali and various dialects. You have deep understanding of emotional nuances, cultural context, and linguistic patterns in Bengali language.

Your task is to classify the sentiment of the given {dialect_name} dialect text as either POSITIVE or NEGATIVE.

CRITICAL: You must respond with ONLY ”POSITIVE” or ”NEGATIVE”. Nothing else. No explanations. No additional words. Just the sentiment.

Now classify the sentiment of this {dialect_name} dialect text:

{dialect_name} dialect text: {dialect_text}

Sentiment (POSITIVE or NEGATIVE):

Prompt Used for Zero-Shot Sentiment (with Transliteration)

Prompt for Zero Shot Sentiment with Transliteration

You are an expert in sentiment analysis for Bengali text, including both standard Bengali and various dialects. You have deep understanding of emotional nuances, cultural context, and linguistic patterns in Bengali language.

Your task is to classify the sentiment of the given {dialect_name} dialect text as either POSITIVE or NEGATIVE.

CRITICAL: You must respond with ONLY ”POSITIVE” or ”NEGATIVE”. Nothing else. No explanations. No additional words. Just the sentiment.

Now classify the sentiment of this {dialect_name} dialect text:

{dialect_name} dialect text: {dialect_text}

Transliterated (Roman letters) {dialect_name} dialect text: {translit_dialect_text}

Sentiment (POSITIVE or NEGATIVE):

Prompt Used for Few-Shot Translation

Prompt for Few Shot Translation

You are an expert linguist specializing in Bengali dialects and language translation. You have extensive knowledge of various Bengali dialects, including Barishal, Chattogram, Sylhet, and Noakhali dialects, and their relationship to standard Bengali.

Your task is to translate {dialect_name} dialect text to standard Bengali while preserving the original meaning, context, and emotional tone.

CRITICAL: You must respond with ONLY the translated text. Nothing else. No explanations. No additional words. No English text. Just the translation.

EXAMPLES OF CORRECT TRANSLATIONS:

Dialect: "মোরা বরিশাইল্লা" -> Standard Bengali: "আমরা বরিশালের"
Dialect: "হুম মোগো বরিশাল" -> Standard Bengali: "হ্যাঁ, আমাদের বরিশাল"
Dialect: "মইনা বেশি কথা কইনা" -> Standard Bengali: "আমি বেশি কথা বলি না"
Dialect: "একদম পালতু বিরানি" -> Standard Bengali: "একদম ফালতু বিরিয়ানি"

Now translate this {dialect_name} dialect text to standard Bengali:

{dialect_name} dialect text: {dialect_text}

Standard Bengali translation:

Prompt Used for Few-Shot Translation (with Transliteration)

Prompt for Few Shot Translation with Transliteration

You are an expert linguist specializing in Bengali dialects and language translation. You have extensive knowledge of various Bengali dialects, including Barishal, Chattogram, Sylhet, and Noakhali dialects, and their relationship to standard Bengali.

Your task is to translate {dialect_name} dialect text to standard Bengali while preserving the original meaning, context, and emotional tone.

CRITICAL: You must respond with ONLY the translated text. Nothing else. No explanations. No additional words. No English text. Just the translation.

EXAMPLES OF CORRECT TRANSLATIONS:

Dialect: "মোরা বরিশাইল্লা" -> Standard Bengali: "আমরা বরিশালের"
Dialect: "হুম মোগো বরিশাল" -> Standard Bengali: "হ্যাঁ, আমাদের বরিশাল"
Dialect: "মইনা বেশি কথা কইনা" -> Standard Bengali: "আমি বেশি কথা বলি না"
Dialect: "একদম পালতু বিরানি" -> Standard Bengali: "একদম ফালতু বিরিয়ানি"

Now translate this {dialect_name} dialect text to standard Bengali:

{dialect_name} dialect text: {dialect_text}

Transliterated (Roman letters) {dialect_name} dialect text: {translit_dialect_text}

Standard Bengali translation:

Prompt Used for Few-Shot Sentiment

Prompt for Few Shot Sentiment

You are an expert in sentiment analysis for Bengali text, including both standard Bengali and various dialects. You have deep understanding of emotional nuances, cultural context, and linguistic patterns in Bengali language.

Your task is to classify the sentiment of the given {dialect_name} dialect text as either POSITIVE or NEGATIVE.

CRITICAL: You must respond with ONLY "POSITIVE" or "NEGATIVE". Nothing else. No explanations. No additional words. Just the sentiment.

EXAMPLES OF CORRECT CLASSIFICATIONS:

Text: "আমাদের বরিশাল নিয়ে আমাদের গর্ব" -> Sentiment: POSITIVE
Text: "একদম ফালতু বিরিয়ানি" -> Sentiment: NEGATIVE
Text: "সুন্দর হয়েছে" -> Sentiment: POSITIVE
Text: "ভাইয়ের বাড়ি এখন কোথায়" -> Sentiment: NEGATIVE
Text: "চিটাগং এর মেয়ে কেন চিটাগং এর হতে পারে না" -> Sentiment: POSITIVE

Now classify the sentiment of this {dialect_name} dialect text:

{dialect_name} dialect text: {dialect_text}

Sentiment (POSITIVE or NEGATIVE):

Prompt Used for Few-Shot Sentiment (with Transliteration)

Prompt for Few Shot Sentiment with Transliteration

You are an expert in sentiment analysis for Bengali text, including both standard Bengali and various dialects. You have deep understanding of emotional nuances, cultural context, and linguistic patterns in Bengali language.

Your task is to classify the sentiment of the given {dialect_name} dialect text as either POSITIVE or NEGATIVE.

CRITICAL: You must respond with ONLY "POSITIVE" or "NEGATIVE". Nothing else. No explanations. No additional words. Just the sentiment.

EXAMPLES OF CORRECT CLASSIFICATIONS:

Text: "আমাদের বরিশাল নিয়ে আমাদের গর্ব" -> Sentiment: POSITIVE
Text: "একদম ফালতু বিরিয়ানি" -> Sentiment: NEGATIVE
Text: "সুন্দর হয়েছে" -> Sentiment: POSITIVE
Text: "ভাইয়ের বাড়ি এখন কোথায়" -> Sentiment: NEGATIVE
Text: "চিটাগং এর মেয়ে কেন চিটাগং এর হতে পারে না" -> Sentiment: POSITIVE

Now classify the sentiment of this {dialect_name} dialect text:

{dialect_name} dialect text: {dialect_text}

Transliterated (Roman letters) {dialect_name} dialect text: {translit_dialect_text}

Sentiment (POSITIVE or NEGATIVE):