2024 IEEE International Conference on Computing, Applications and Systems (COMPAS),
September 25-26, 2024, Bangladesh

Improving Bangla Regional Dialect Detection Using BERT, LLMs, and XAI

Bidyarthi Paul
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka - 1208, Bangladesh
bidyarthipaul01@gmail.com

Faika Fairuj Preotee
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka - 1208, Bangladesh
faikafairuj2001@gmail.com

Shuvashis Sarker
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka - 1208, Bangladesh
shuvashisofficial@gmail.com

Tashreef Muhammad
Department of Computer Science and Engineering
Southeast University
Dhaka - 1208, Bangladesh
tashreef.muhammad@seu.edu.bd

2024 IEEE International Conference on Computing, Applications and Systems (COMPAS) | 979-8-3315-2976-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/COMPAS60761.2024.10796843

***Abstract*—This research addresses the challenge of identifying and categorizing Bangla regional dialects through the application of sophisticated natural language processing techniques. Automated translation, digital content personalization, and speech recognition systems are all improved by precise dialect detection, which is a result of the linguistic diversity in Bangladesh. Few-shot learning techniques were used to compare the performance of transformer-based models, specifically Bangla BERT, with state-of-the-art large language models such as GPT-3.5 Turbo and Gemini 1.5 Pro, and to fine-tune them. The methodology entailed the utilization of a diverse regional Bangla speech samples from the Vasantor dataset, which encompassed regions including Mymensingh, Chittagong, Barishal, Noakhali, and Sylhet. In order to enhance the interpretability of the model, implementation of Local Interpretable Model-agnostic Explanations (LIME) was done. The Bangla BERT model achieved the highest accuracy of 88.74%. GPT-3.5 Turbo's few-shot learning exhibited substantial potential, with an accuracy rate of 64%. These results underscore the significance of hyperparameter optimization and fine-tuning in the enhancement of regional dialect detection models.**

***Index Terms*—Bangla dialect detection, Bangla BERT, LLM, Explainable AI**

I. Introduction

Bengali, a language with deep roots in Sanskrit, is spoken by over 300 million people globally, making it the 6th most spoken language in the world. In Bangladesh, the linguistic landscape is notably diverse, with the majority of the population speaking two primary varieties of Bengali: mainstream Bangla, used by approximately 290 million people, and various regional dialects. Across Bangladesh's 64 districts, there are 55 regional languages or dialects. These dialects, learned organically without formal education in grammar, exhibit significant variations depending on the region. Examples include dialects spoken in Mymensingh, Barishal, Sylhet, Chittagong, and Noakhali, each characterized by unique phonological and lexical features.

The accurate recognition and categorization of these regional dialects are important for several reasons. First, it enhances the relevance and accuracy of automated translation services by ensuring that local linguistic details are considered. Second, it enables the development of digital content that is more aligned with the specific needs and preferences of users from different regions, thus improving user satisfaction. Third, regional language detection is crucial for the advancement of speech recognition systems, making them more effective and accessible to users from various geographical areas. However, conventional models often struggle with regional language detection due to their inability to grasp the subtleties and contextual complexities inherent in Bangla dialects.

This study addresses the communication challenges posed by Bangladesh's diverse regional dialects through the application of advanced natural language processing techniques. Transformer-based models, particularly Bangla BERT, which is trained on the Bangla language, are fine-tuned for the task of accurate detection and classification of regional dialects. Additionally, the study explores the effectiveness of state-of-the-art language models in dialect detection by employing prompting techniques with models like GPT-3.5 Turbo and Gemini 1.5 Pro. A comparative analysis is conducted to evaluate the performance of these models in few-shot learning scenarios versus fine-tuned approaches.

To enhance the interpretability of the models, explainable AI techniques are implemented. These techniques aim to identify specific textual features that indicate a particular regional dialect, thereby providing transparency in the models' predictions. This not only improves dialect detection but also offers valuable insights into the linguistic characteristics of each region, contributing to a better understanding of the diverse linguistic landscape in Bangladesh. The research ultimately seeks to promote linguistic inclusivity and cultural integration by addressing communication barriers within the country. The objectives of study are:

i Utilize transformer-based models, specifically Bangla BERT, fine-tuned for the detection and classification of regional dialects in Bangladesh.
ii Compare the performance of state-of-the-art language models, such as GPT-3.5 Turbo [16] and Gemini 1.5 Pro [15], using prompting techniques to assess whether few-shot learning can outperform fine-tuned approaches in this context.
iii Implement explainable AI techniques to identify and analyze specific textual features that indicate regional

dialects, thus enhancing the interpretability of the models' predictions and fostering a deeper understanding of linguistic diversity in Bangladesh.

II. Literature Review

A. Transformer Based Approaches

Recent advancements in Bengali natural language processing (NLP) have shown substantial progress in many areas. Jahan et al. (2022) [1] created BanglaHateBERT, a BERT model designed to find Bengali hate speech. It was very effective at what it did, scoring 94.3% accuracy with a new 15,000-sample offensive corpus and a large collection of Reddit comments that were banned. Anan et al. (2023) [2] focused on Bangla sarcasm detection, creating a BERT-based model with Stratified K-Fold Cross Validation. Using the BanglaSarc dataset of 5112 comments, their model reached an impressive validation accuracy of 99.60%, significantly outperforming traditional machine learning models. Banshal et al. (2023) [3] studied how to recognize emotions in Bengali using standard ML models and BERT-based models (BanglaBERT, mBERT, Bangla-Electra). They were able to get 83.23% accuracy on a dataset of 108,950 Facebook comments from Bangladeshi news media pages. The study showed that Bangla-BERT Base is better at predicting emotions than standard machine learning and deep learning models. Faria et al. (2023) [4] investigated translation and region detection for Bangla dialects using models like mT5 and BanglaT5. Their work, utilizing a dataset of 32,500 sentences covering Bangla, Banglish, and English, showed that BanglaT5 outperformed mT5 in translation tasks, achieving a BLEU score of 44.03 for the Chittagong dialect, while Bangla-bert-base achieved an accuracy of 85.86% in region detection. Jobair et al. (2024) [5] developed a new dataset for the detection of hate speech in Bangla. 8,600 comments from Facebook and YouTube were collected and divided into categories such as entertainment, politics, sports, and religion. They used five alternative models and discovered that, at 80%, the BERT model had the best accuracy. Furthermore, 97% accuracy was obtained through testing on an already-existing dataset with 30,000 records.

B. LLM Based Approaches

To utilize large language models (LLMs) to leverage extensive training on diverse and vast datasets for achieving impressive performance across various natural language processing tasks,Faria et al. (2024) [6] evaluated various pre-trained language models (PLMs) and large language models (LLMs) for Bengali natural language inference (NLI), with BanglaBERT achieving 82.04% accuracy and GPT-3.5 Turbo reaching 92.05% in a 15-shot setting. Alam et al.'s (2024) workflow [7] examined LLMs for NLP and voice tasks, such as GPT-4 and Bloom, highlighting their drawbacks in low-resource environments even if they showed appreciable gains with prompting and instruction adjustment. Hasan et al. (2024) [8] investigated sentiment analysis in a different study with models like Bloomz and Flan T5, with improved BanglaBERT obtaining an F1 score of 70.33. When Kuila and Sarkar (2024) [9] tested LLMs with improved BERT models, they discovered that Falcon-40B performed better in few-shot scenarios.These studies collectively demonstrate the potential of advanced models in enhancing the accuracy and effectiveness of Bengali NLP tasks.

C. Explainable AI in NLP

A context-based feature strategy, grounded in Explainable AI, has brought significant advancements in Bengali natural language processing (NLP) in recent works. In light of this, Anan et al. (2023) [2] improved the explainability of their BERT-based model for identifying Bangla sarcasm by using LIME (Local Interpretable Model-agnostic Explanations). This method highlighted the important factors impacting each decision, making the model's predictions easier for consumers to understand. For example, LIME's interpretation clarified why the model was able to correctly differentiate between sarcasm and non-sarcasm across different inputs. In their investigation, visual representations showed how LIME could give their model transparency and clarity. Moreover, Uddin et al. (2024) [10] expanded transparency and confidence in their transformer-based SMS spam detection models by utilizing Transformers Interpret and LIME. They overcame the black box issue by utilizing LimeTextExplainer, which improved the comprehensibility of model predictions. To test the model's sensitivity to changes in input, LIME modifies the content of SMS messages by adding or deleting words. Around each prediction, this procedure constructs a locally interpretable model focusing on the important features impacted by the perturbations and offers clear reasons for the model's logic. Additionally, Aporna, Amena Akter, et al. [11] proposed an Explainable AI approach for detecting offensive speech in Bengali using a CNN model on the Bengali Hate Speech Dataset, achieving 78% accuracy and providing quality explanations through completeness and sufficiency scores. These studies highlight the significant role of Explainable AI in advancing Bangla NLP by enhancing model transparency and effectiveness.

III. Methodology

A Schematic diagram of the methodology of the study is shown in Figure 1.

A. Dataset Collection

For this study, a variety of regional Bangla speech samples were collected from the VASANTOR dataset. 12,505 samples in all are included in the dataset, which is divided into regions named Mymensingh, Chittagong, Barishal, Noakhali, and Sylhet. Every sample in the dataset includes a sentence or phrase that is frequently used in everyday communication in these areas, capturing the distinct linguistic characteristics and dialectal variations of Bangla.

The dataset was obtained from regional speech collections and publicly accessible repositories[1]. It was carefully chosen to guarantee that various regions were fairly represented, giving the language detection models a strong basis on which to be trained and assessed. According to the original source, the samples were split into training, testing, and validation sets, consisting of 9,377 training samples, 1,877 testing samples, and 1,251 validation samples.

[1] https://data.mendeley.com/datasets/bj5jgk878b/2

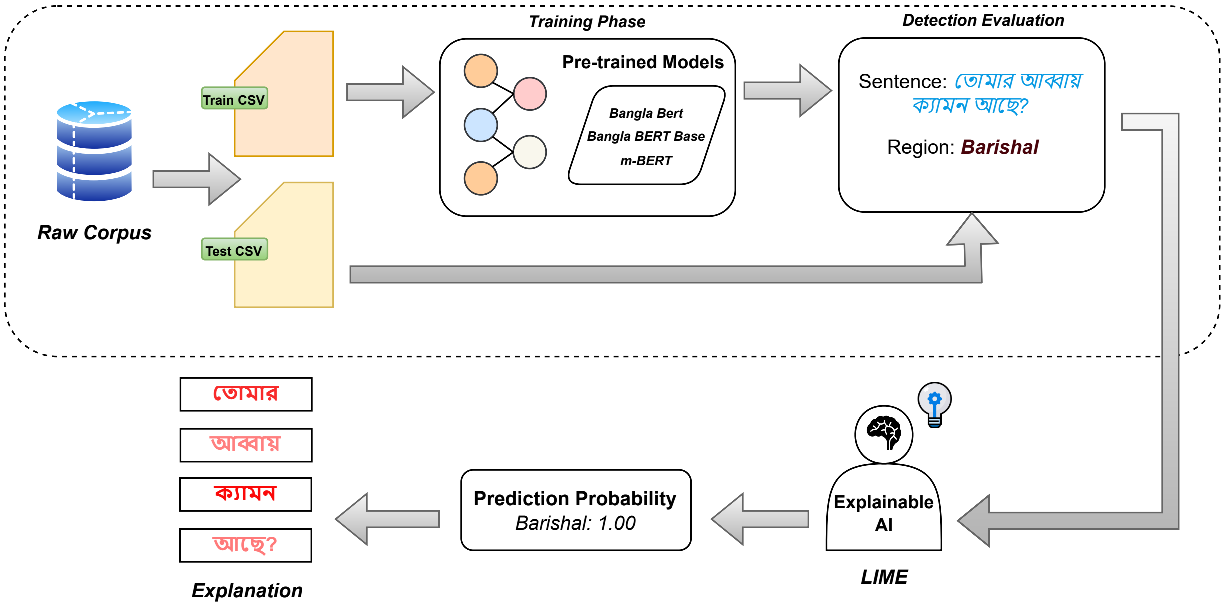

Fig. 1. Schematic Diagram of Bangladeshi Regional Language Detector

B. Pre-Trained language Models on Bengali

The challenging task of identifying regional dialects in Bangla speech was addressed by implementing multiple cutting-edge pre-trained models. These models were selected for their capacity to handle the specific linguistic traits and contextual complexities of the Bengali language, which are often overlooked by conventional models.

1) ELECTRA Based (Bangla BERT): The finer points and contextual complexities of Bengali are often beyond the capabilities of traditional models, particularly when dealing with a variety of regional dialects. Pre-trained on an extensive corpus of Bengali text, Bangla BERT [17] provides a profound contextual understanding of the language. This model improves the overall performance and dependability of detection system by strengthening the capacity to precisely identify and categorize regional variations in Bangla speech.

2) BERT Based (Bangla BERT): In order to take full advantage of the rich contextual information that is embedded in Bengali text, Bangla BERT Base [18] was incorporated into the proposed model. Pre-trained on large-scale Bengali corpora, Bangla BERT Base captures the intricate contextual relationships present in the language. Because of this, it works especially well for differentiating between the finer points of regional dialects. To improve regional language detection capabilities, Bangla BERT Base was utilized to ensure the model can reliably and efficiently determine the regional origins of speech samples.

3) BERT Based (m-Bert): Multilingual BERT (m-BERT) [19] was utilized to enhance the model's robustness and expand its contextual understanding. Designed for multilingual tasks, m-BERT is pre-trained in multiple languages, including Bengali. Its ability to comprehend and distinguish between diverse linguistic elements across languages makes it an excellent choice for the task. Incorporating m-BERT leverages its multilingual capabilities to improve regional dialect detection accuracy and efficacy, resulting in a more comprehensive and contextually aware solution.

C. Explainable AI

1) Local Interpretable Model-Agnostic Explanations (LIME): Regional language detection model was made more interpretable and transparent using Local Interpretable Model-agnostic Explanations (LIME) [20]. LIME approximates the complex model with an interpretable one around each prediction to generate local explanations. It tweaks the input data and observes changes in the model's predictions to train a simpler model that highlights the most important words and phrases. LIME helps us identify regional dialects, improve model trust by making its decision-making process transparent, and improve model performance by providing insights for debugging and performance. This makes the proposed approach accurate and understandable, boosting prediction confidence.

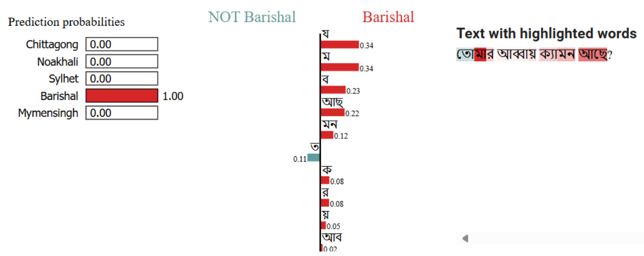

Fig. 2. Lime Predictions on Bangla Bert

D. Large Language Models (LLMs)

In addition to the pre-trained BERT models, two large language models (LLMs) were implemented for the regional language detection task: OpenAI GPT-3.5 and Gemini 1.5 Pro.

1) GPT-3.5 Turbo: OpenAI GPT-3.5, developed by OpenAI, is renowned for its advanced natural language understanding capabilities. Fine-tuning GPT-3.5 on the regional Bangla speech dataset involved using few-shot prompting to provide the model with contextual examples, which improved its ability to accurately detect and classify regional dialects. This approach leverages the model's ability to generalize from a few examples, enhancing its effectiveness for the task.

2) Gemini 1.5 Pro: Google Gemini 1.5 Pro, developed by Google, was fine-tuned and utilized with few-shot prompting to enhance its regional language detection capabilities. By providing the model with a few representative examples, its performance in identifying and differentiating between various Bangla dialects was improved. Gemini 1.5 Pro's robust language processing features contribute significantly to the effectiveness of the detection system.

IV. Experiments

A. Experimental Setup

Two different setups were used for the experiments. The first setup, utilized for lower epochs to stay within the free Colab GPU limits, was a Google Colab Notebook with Python 3.10.12, PyTorch 2.0.1, a Tesla T4 GPU (15 GB), 12.5 GB of RAM, and 64 GB of disk space. The second setup, designed for higher epochs, involved a more powerful configuration with a Ryzen 5 3600x processor, 32GB DDR4 RAM, and an Nvidia RTX 3060ti GPU.

B. Evaluation Metrics

When it comes to region detection, accuracy [14] is defined as the ratio of accurately classified regions to all of the regions in the dataset. It measures the degree to which the model can correctly place a text in its true region.

Apart from accuracy, Log Loss was implemented for the evaluation. While evaluating the effectiveness of classification models in the context of multiclass classification—where texts can be classified into multiple regions—logarithmic loss [13], also referred to as log loss or log loss. (Chittagong, Noakhali, Sylhet, Barishal, Mymensingh). The model's predicted class probabilities are quantified, with precise and confident predictions being rewarded and uncertain or inaccurate ones being penalized.

C. Hyperparameter Tuning

In the experiments, several key hyperparameters were tuned to optimize the performance of the regional language detection model, as summarized in Table I. Due to constraints in computational resources and time, batch sizes of 8 and 16 were selected, with the number of epochs varied from 5 to 35 in increments of 5. The learning rate was initially set at 0.0001 and then gradually decreased to 0.000001 to observe its impact on model accuracy. Since accuracy tended to decrease with higher learning rates, an additional experiment with a learning rate of 0.00002 was conducted to further fine-tune the model's performance. This systematic approach to hyperparameter tuning helped identify the optimal settings for achieving the best results within the given resource limitations.

TABLE I
Hyperparameter variations

Hyperparameters	Variations
Optimizer	AdamW
Batch Size	8,16
Learning Rate	0.0001, 0.00001, 0.000001, 0.00002
Epoch	5, 10, 15, 20, 25, 30, 35

V. Result Analysis

A. Pre-trained Models

The performance analysis of the pre-trained BERT models is illustrated in Table II, which include BanglaBert Base, BanglaBert, and m-BERT, reveals significant insights into the accuracy of these models across a variety of batch sizes, learning rates, and epochs. This configuration is effective, as evidenced by the fact that the highest accuracy achieved by the BanglaBert Base model with a batch size of 8 is 0.8638 at epoch 30 and a learning rate of 0.00002. In a similar vein, the best performance observed with a batch size of 16 is an accuracy of 0.8649 at epoch thirty, also with a learning rate of 0.00002; this demonstrates that the optimal learning rates for this model remain consistent across a variety of batch sizes. A batch size of eight and a learning rate of 0.00001 are the conditions under which the BanglaBert model reaches its highest level of accuracy, which is 0.8820 at epoch 35. The fact that this model is able to achieve the highest accuracy of 0.8874 at epoch 35 with a learning rate of 0.00002 demonstrates that it is able to benefit from lower learning rates in order to achieve better performance. With a batch size of sixteen and a learning rate of 0.00001, the m-BERT model achieves its highest level of accuracy, which is obtained at epoch 20 and is recorded as 0.8479. The significance of making adjustments to learning rates and batch sizes in order to achieve the highest possible level of performance from pre-trained BERT models in particular applications is highlighted by these findings.

Table III, indicates the performance overview of the best region detection model.

B. Large Language Models (LLMs)

1) Few Shot Prompting: For the analysis of large language model (LLM) implementations, experiments were conducted on a subset of the entire dataset to mitigate the costs associated with using LLM models. Specifically, 100 samples were selected for training, with 25 samples representing each region. Few-shot prompting was utilized, with a total of 25 shots—5 shots per region—to enhance the model's performance. This approach allowed the models to better comprehend and distinguish between the regional dialects.

To facilitate the identification of regions within the Bangla speech samples, a specific prompt was added. The few-shot prompting methodology employed in this work is depicted in Figures 3 and 4. The prompt was designed to assist the models in accurately identifying the region of each sample by providing examples and context.

2) Finetune: A total of 100 samples were used for the fine-tuning process, with 25 samples allocated to each region. Both large language models (LLMs) were adjusted using these examples to improve their performance in

TABLE II
PERFORMANCE (ACCURACY) OF THE PRE-TRAINED BERT MODELS

Models	Batch Size	Learning Rate	Epoch						
			5	10	15	20	25	30	35
Bangla Bert Base	8	0.0001	0.8361	0.8265	0.7614	0.7481	0.7620	0.7534	0.7598
		0.00001	0.8297	0.8345	0.8388	0.8468	0.8596	0.8521	0.8516
		0.000001	0.7091	0.7646	0.7876	0.8020	0.8110	0.8596	0.8543
		0.00002	0.8473	0.8553	0.8553	0.8585	0.8367	**0.8638**	0.8596
	16	0.0001	0.8527	0.8452	0.8452	0.8511	0.8207	0.7961	0.8180
		0.00001	0.8324	0.8495	0.8511	0.8495	0.8447	0.8521	0.8457
		0.000001	0.6616	0.7390	0.7710	0.7881	0.8004	0.8361	0.8180
		0.00002	0.8532	0.8521	0.8361	0.8505	0.8367	**0.8649**	0.8591
Bangla Bert	8	0.0001	0.8778	0.8692	0.8745	0.8740	0.8756	0.8569	0.8500
		0.00001	0.8724	0.8794	0.8687	0.8767	0.8751	0.8783	**0.8820**
		0.000001	0.6200	0.7582	0.8105	0.8255	0.8399	0.8479	0.8532
		0.00002	0.8826	0.8810	0.8740	0.8783	0.8778	0.8783	0.8724
	16	0.0001	0.8938	0.8639	0.8729	0.8788	0.8826	0.8762	0.8778
		0.00001	0.8585	0.8729	0.8745	0.8778	0.8815	0.8799	0.8815
		0.000001	0.5453	0.6787	0.7662	0.8020	0.8207	0.8303	0.8404
		0.00002	0.8676	0.8815	0.8804	0.8788	0.8868	0.8729	**0.8874**
m-Bert	8	0.0001	0.8293	0.8312	0.8471	0.8435	0.8240	0.8358	0.8456
		0.00001	0.8287	0.8383	0.8425	0.8425	0.8324	0.8303	**0.8436**
		0.000001	0.6696	0.6696	0.8116	0.8116	0.8052	0.8116	0.8169
		0.00002	0.8249	0.8431	0.8388	0.8308	0.8308	0.8308	0.8383
	16	0.0001	0.7956	0.8187	0.8356	0.8030	0.8198	0.8045	0.7961
		0.00001	0.8276	0.8351	0.8324	**0.8479**	0.8452	0.8404	0.8393
		0.000001	0.6648	0.7158	0.7620	0.7817	0.7945	0.8009	0.8009
		0.00002	0.8409	0.8431	0.8313	0.8383	0.8303	0.8383	0.8409

TABLE III
PERFORMANCE (ACC. OR ACCURACY) OVERVIEW OF THE BEST REGION DETECTION MODEL

Models	Acc.	Log Loss	Region	Precision	Recall	F1-Score
Bangla Bert	88.74%	0.904	Chittagong	0.8626	0.9040	0.8828
			Sylhet	0.8399	0.8533	0.8466
			Noakhali	0.9821	0.7307	0.8379
			Barishal	0.9821	0.9786	0.9594
			Mymensingh	0.8426	0.9707	0.9021

regional language detection. This fine-tuning enabled the models to better capture the subtleties of each regional dialect.

After fine-tuning, the models' accuracy was assessed. Gemini 1.5 Pro achieved an accuracy of 36%, while GPT-3.5 Turbo demonstrated a significantly higher accuracy of 64%. These results highlight the differences in capabilities between the two LLMs and the effectiveness of fine-tuning in enhancing model performance.

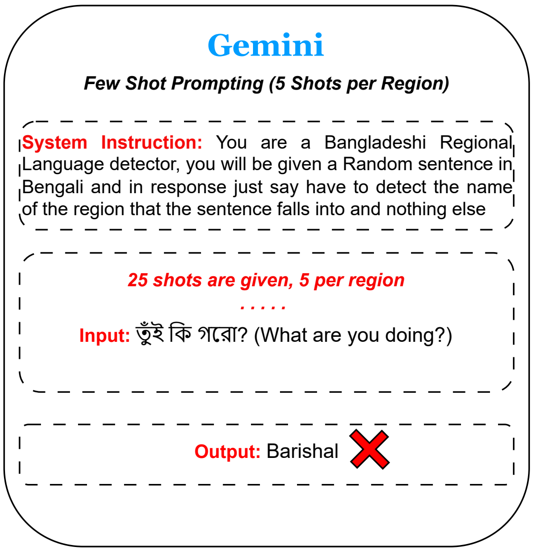

Fig. 3. Prompt Design for Few Shot Learning in Gemini 1.5 pro

C. Explainable AI

The Figure 2 demonstrates the use of Local Interpretable Model-agnostic Explanations (LIME) for enhancing the interpretability of the proposed regional language detection model. The model predicts the region as Barishal with a probability of 1.00, while other regions have a probability of 0.00. LIME highlights specific words that significantly contribute to this prediction. This visualization confirms that the model bases its prediction on distinct linguistic features, making its decision-making process transparent and understandable.

D. Discussion

The analysis showed that at epoch 30, BanglaBert Base, with a batch size of 16, and a learning rate of 0.00002, achieved the highest accuracy of 0.8649. At epoch 35,

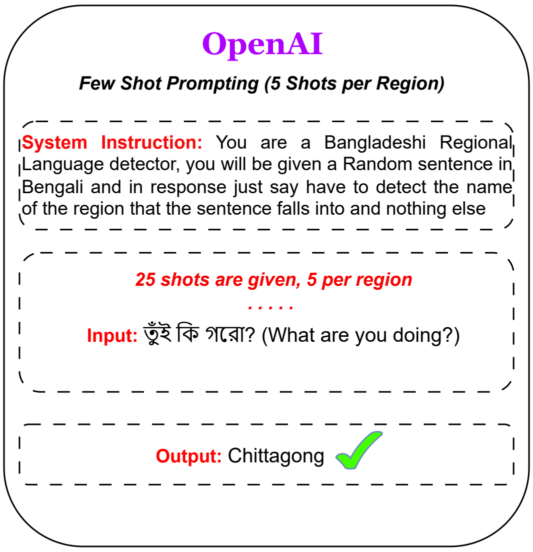

Fig. 4. Prompt Design for Few Shot Learning in Gpt 3.5 Turbo

BanglaBert achieved its maximum accuracy of 0.8874 while maintaining the same learning rate and batch size. At epoch 20, the m-BERT model achieved an accuracy of 0.8479 with a batch size of 16 and a learning rate of 0.00001. The significance of ideal learning rates and batch sizes is highlighted by these findings. Furthermore, LLM performance was greatly enhanced by few-shot prompting and fine-tuning; GPT-3.5 Turbo achieved 64% accuracy, proving its superiority over Gemini 1.5 Pro.

VI. Conclusion

This study demonstrates the effectiveness of advanced pre-trained language models, such as Bangla BERT, GPT-3.5 Turbo, and Gemini 1.5 Pro, in detecting and classifying Bangla regional dialects. Bangla BERT achieved the highest accuracy at 88.74%, with GPT-3.5 Turbo showing notable potential at 64%. The use of Explainable AI, particularly LIME, enhanced the transparency of the models by highlighting the key linguistic features influencing their decisions. Future work will include expanding the dataset to cover more dialects and exploring Named Entity Recognition (NER) to improve contextual understanding. Implementing these models in practical applications like automated translation and content personalization will help validate the findings and improve communication across different regions. This research provides valuable insights into Bangla dialect detection and offers practical benefits for improving digital services in linguistically diverse areas.

References

[1] Jahan, Md Saroar, et al. "BanglaHateBERT: BERT for abusive language detection in Bengali." Proceedings of the second international workshop on resources and techniques for user information in abusive language analysis. 2022

[2] Anan, Ramisa, et al. "Interpretable bangla sarcasm detection using bert and explainable ai." 2023 IEEE 13th annual computing and communication workshop and conference (CCWC). IEEE, 2023.

[3] Banshal, Sumit Kumar, et al. "MONOVAB: An Annotated Corpus for Bangla Multi-label Emotion Detection." arXiv preprint arXiv:2309.15670 (2023).

[4] Faria, Fatema Tuj Johora, et al. "Vashantor: a large-scale multilingual benchmark dataset for automated translation of bangla regional dialects to bangla language." arXiv preprint arXiv:2311.11142 (2023).

[5] Jobair, Md, et al. "Bengali Hate Speech Detection with BERT and Deep Learning Models." International Conference on Big Data, IoT and Machine Learning. Singapore: Springer Nature Singapore, 2023.

[6] Faria, Fatema Tuj Johora, et al. "Unraveling the Dominance of Large Language Models Over Transformer Models for Bangla Natural Language Inference: A Comprehensive Study." arXiv preprint arXiv:2405.02937 (2024)

[7] Alam, Firoj, et al. "LLMs for Low Resource Languages in Multilingual, Multimodal and Dialectal Settings." Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts. 2024.

[8] Hasan, Md Arid, et al. "Zero-and few-shot prompting with llms: A comparative study with fine-tuned models for bangla sentiment analysis." arXiv preprint arXiv:2308.10783 (2023).

[9] Kuila, Alapan, and Sudeshna Sarkar. "Deciphering Political Entity Sentiment in News with Large Language Models: Zero-Shot and Few-Shot Strategies." arXiv preprint arXiv:2404.04361 (2024).

[10] Uddin, Mohammad Amaz, et al. "ExplainableDetector: Exploring Transformer-based Language Modeling Approach for SMS Spam Detection with Explainability Analysis." arXiv preprint arXiv:2405.08026 (2024).

[11] Aporna, Amena Akter, et al. "Classifying offensive speech of bangla text and analysis using explainable ai." International Conference on Advances in Computing and Data Sciences. Cham: Springer International Publishing, 2022.

[12] Hossain, Prommy Sultana, et al. "Multi-label extreme learning machine (MLELMs) for bangla regional speech recognition." Applied Sciences 12.11 (2022): 5463.

[13] Rezaei, Ashkan, et al. "Fairness for robust log loss classification." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 04. 2020.

[14] Labatut, Vincent, and Hocine Cherifi. "Accuracy measures for the comparison of classifiers." arXiv preprint arXiv:1207.3790 (2012).

[15] Ye, Junjie, et al. "A comprehensive capability analysis of gpt-3 and gpt-3.5 series models." arXiv preprint arXiv:2303.10420 (2023).

[16] Pornprasit, Chanathip, and Chakkrit Tantithamthavorn. "GPT-3.5 for Code Review Automation: How Do Few-Shot Learning, Prompt Design, and Model Fine-Tuning Impact Their Performance?." arXiv preprint arXiv:2402.00905 (2024).

[17] Bhattacharjee, Abhik, et al. "BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla." arXiv preprint arXiv:2101.00204 (2021).

[18] S. Sarker, "Banglabert: Bengali mask language model for bengali language understanding," textsIGitHub, 2020.

[19] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[20] Tiwari, Ravi Shekhar. "Hate speech detection using LSTM and explanation by LIME (local interpretable model-agnostic explanations)." Computational Intelligence Methods for Sentiment Analysis in Natural Language Processing Applications. Morgan Kaufmann, 2024. 93-110.