BanglaIPA: Towards Robust Text-to-IPA Transcription with Contextual Rewriting in Bengali

Jakir Hasan[1], Shrestha Datta[1], Md Saiful Islam[1], Shubhashis Roy Dipta[2], Ameya Debnath[1],

[1]Shahjalal University of Science and Technology, BD
[2]University of Maryland, Baltimore County, USA

Correspondence: jakirhasan718@gmail.com

Abstract

Despite its widespread use, Bengali lacks a robust automated International Phonetic Alphabet (IPA) transcription system that effectively supports both standard language and regional dialectal texts. Existing approaches struggle to handle regional variations, numerical expressions, and generalize poorly to previously unseen words. To address these limitations, we propose **BanglaIPA**, a novel IPA generation system that integrates a character-based vocabulary with word-level alignment. The proposed system accurately handles Bengali numerals and demonstrates strong performance across regional dialects. BanglaIPA improves inference efficiency by leveraging a precomputed word-to-IPA mapping dictionary for previously observed words. The system is evaluated on the standard Bengali and six regional variations of the DUAL-IPA dataset. Results show that BanglaIPA outperforms baseline IPA transcription models by **58.4-78.7%** and achieves an overall mean word error rate of 11.4%, highlighting its robustness in phonetic transcription generation for the Bengali language.[1]

1 Introduction

The International Phonetic Alphabet (IPA) is a widely accepted notation system for representing the phonetic structure of languages (Association, 1999), providing precise pronunciation guidelines for learners, linguists, and speech-processing applications (Daniels and Bright, 1996). IPA plays a critical role in text-to-speech systems (Zhang et al., 2021), enabling accurate and consistent speech synthesis across diverse languages. However, many low-resource languages, including Bengali, still lack a standardized and efficient IPA transcription framework (Islam et al., 2024), which poses a significant bottleneck for real-time speech generation and computational linguistic research. Despite being spoken by approximately 260 million people worldwide (Islam et al., 2025), Bengali continues to exhibit many inconsistencies in phonetic representation due to unresolved phonetic analyses and persistent linguistic ambiguities (Kamal, 2025).

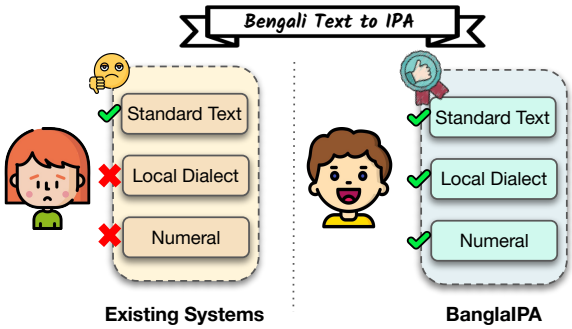

Figure 1: Existing IPA transcription systems (left) are limited to standard Bengali and struggle to process texts containing regional dialectal variations and numerical expressions. In contrast, BanglaIPA (right) successfully transcribes both **standard and regional dialects** and accurately **handles numerical expressions** through the incorporation of a **contextual rewriting** procedure.

The rich diversity of regional dialects in the Bengali language (Hasan and Dipta, 2025) makes the development of an automated IPA transcription system even more challenging. These dialects often diverge significantly from the standard language in terms of phonology, vocabulary, and syntactic structure (Hasan et al., 2024). In addition, Bengali numerals exhibit context-dependent pronunciations, which existing IPA transcription systems fail to transcribe accurately (Al Zubaer et al., 2020). The widespread use of the Bengali language highlights the need for a specialized IPA transcription framework, particularly to support high-quality text-to-speech synthesis (Nath and Sarma, 2024).

In this work, we introduce BanglaIPA, the first end-to-end IPA transcription system designed to support standard Bengali, six regional dialects, and numerical expressions. BanglaIPA integrates multiple processing modules that combine data-driven

[1]https://github.com/Jak57/BanglaIPA

machine learning with algorithmic techniques to generate accurate and robust phonetic transcriptions of Bengali text (Zhou, 2021). As illustrated in Fig. 1, existing systems are primarily developed for standard Bengali, which severely limits their applicability for users of regional dialects. Moreover, these systems are unable to properly transcribe numbers, instead converting individual Bengali digits into English forms[2]. BanglaIPA addresses these limitations through the incorporation of a contextual rewriting (Bao and Zhang, 2021) mechanism and by training the transcription generation model on both standard and regional dialectal data.

In the BanglaIPA system, input text containing numbers is first rewritten into word form by leveraging the full textual context. Like Bengali, in English, the digit '1' is pronounced differently in "1 dollar" and "1st place" despite sharing the same numeric form. The contextual rewriting ensures the preservation of the correct pronunciation of numbers. For each word in the rewritten text, IPA transcription is generated using a Transformer-based model (Vaswani, 2017) trained on the DUAL-IPA (Fatema et al., 2024) dataset. The resultant word-IPA pairs are cached in a dictionary for efficient reuse. We develop a State Alignment (STAT) algorithm to specify which subword segments require model-based transcription, enabling robust handling of out-of-vocabulary characters and symbols. Results demonstrate that our approach achieves a mean word error rate of 11.4% on the constructed test set, which is a substantial improvement of 58.4-78.7% over the baseline MT5 (Xue, 2020) and UMT5 (Chung et al., 2023) models. Moreover, BanglaIPA maintains a strong and consistent performance across all regional dialects, with word error rates remaining close to 11% for four of the six regions.

In summary, our main contributions are:

- We propose BanglaIPA, the first end-to-end system capable of generating phonetic representation (IPA) of the standard Bengali language, its six regional dialects, and numbers.
- BanglaIPA substantially outperforms existing baseline models on the DUAL-IPA dataset and demonstrates strong performance across regional variations with low word error rates.
- We introduce the State Alignment (STAT) algorithm, which enables efficient handling of out-of-vocabulary characters and symbols.

[2] https://ipa.bangla.gov.bd/

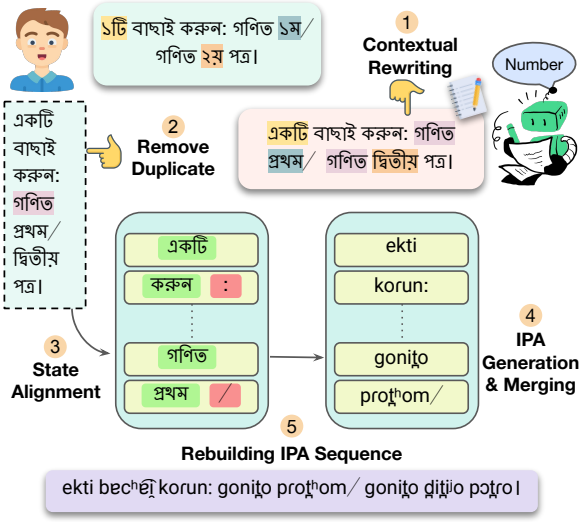

Figure 2: Overview of the **end-to-end processing** pipeline of the BanglaIPA system for generating the **International Phonetic Alphabet (IPA)** transcription.

2 Methodology

The **BanglaIPA** system comprises multiple processing modules designed to accurately transcribe Bengali text into the International Phonetic Alphabet (IPA). An overview of the complete processing pipeline is shown in Fig. 2 using an example script.

Contextual Rewriting The pronunciation of Bengali numerals is highly context-dependent, and the same numeric expression may correspond to multiple IPA transcriptions, posing significant challenges for rule-based approaches. Examples illustrating this pronunciation variability are provided in Appendix B. To address this issue, we incorporate a large language model (LLM), which offers strong contextual reasoning capabilities (Zhu et al., 2024). Specifically, we prompt the GPT-4.1-nano model (prompt in Fig. 6) to rewrite only the numerical expressions in the input text into their corresponding word forms based on the surrounding textual context. This rewriting step ensures that numerals are converted into linguistically appropriate lexical forms before generating IPA transcription.

De-duplication The contextually rewritten text is used to construct a dictionary in which each unique word serves as a key and its corresponding IPA transcription is stored as the value. For each newly encountered word, an empty IPA entry is initially assigned and subsequently populated by downstream processing modules. This de-duplication strategy ensures that an IPA transcription is generated only

Algorithm 1 The State Alignment algorithm segments a word into subwords using a predefined vocabulary. It assigns a state to each segment to indicate the need for model-based IPA generation.

Require: token t, set of characters C
Ensure: state and subtoken lists: S, T

Initialize $S \leftarrow []$, $T \leftarrow []$
$N \leftarrow$ length of t
$i \leftarrow 0$
while $i < N$ **do**
    $st \leftarrow$ **false**
    $t_s \leftarrow$ ""
    if $t[i] \in$ C **then**
        $st \leftarrow$ **true**
        while $i < N$ **and** $t[i] \in$ C **do**
            Append $t[i]$ to t_s
            $i \leftarrow i + 1$
        end while
    else
        while $i < N$ **and** $t[i] \notin$ C **do**
            Append $t[i]$ to t_s
            $i \leftarrow i + 1$
        end while
    end if
    Append st to S
    Append t_s to T
end while

once per unique word and reused thereafter, thereby reducing redundant computations of the system.

State Alignment The unique words extracted from the rewritten text may contain characters or symbols that do not belong to Bengali language. Such out-of-vocabulary characters and symbols present significant challenges for accurate IPA generation. To address this issue, we propose the State Alignment (`STAT`) algorithm (see Alg. 1), which leverages a predefined Bengali character set to split every word into subword segments. Each segment is assigned a state to represent whether it needs model-based IPA generation or has to be kept unchanged. This mechanism ensures that any segment with foreign characters and symbols is not passed to the model that was not seen during training, enabling robust handling of unseen characters.

Generation of IPA & Merging For each subword identified as requiring IPA generation, we construct a synthetic sentence by inserting spaces between adjacent characters within the subword and then apply vectorization. The resulting vectorized input is passed to a Transformer-based model, which generates sequences of IPA characters. After generation, the spaces between the characters are removed to produce the final IPA representation of the subword. The complete phonetic transcription for each word is obtained by merging the model-generated IPA for the relevant subwords with the remaining segments that do not require direct transcription. This merging process ensures a coherent and accurate word-level phonetic transcription.

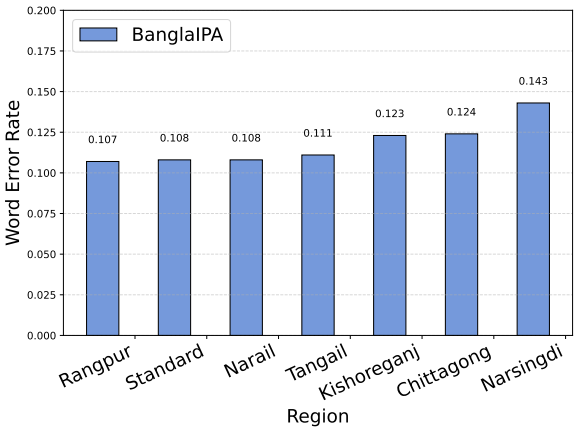

Figure 3: Region-wise word error rate distribution on the `DUAL-IPA` test dataset, with IPA transcriptions generated by the `BanglaIPA` system. The system consistently shows **very low word error rates** across most regions.

Rebuilding Sequence The merged IPA transcriptions are used to update the `Word-IPA` dictionary. The contextually rewritten text is then traversed from beginning to end, and the corresponding IPA entries are retrieved from the dictionary to reconstruct the final IPA transcription of the Bengali text.

3 Results & Discussion

3.1 Implementation Details

We train a small Transformer-based model comprising a single encoder-decoder architecture with approximately 8.6 million parameters. Model training is performed using the `RMSprop` optimizer (Elshamy et al., 2023) in conjunction with a sparse categorical cross-entropy loss function (Mao et al., 2023). The model is trained on the processed `DUAL-IPA` dataset (Fatema et al., 2024) as described in App. A for 40 epochs, using a learning rate of 0.001 and a batch size of 64.

3.2 Evaluating `BanglaIPA`

To evaluate the performance of our IPA transcription system, **`BanglaIPA`**, we conduct experiments

Model	Word Error Rate ↓							Mean
	Chittagong	Kishoreganj	Narail	Narsingdi	Standard	Rangpur	Tangail	
MT5	27.8	39.7	60.4	67.1	43.4	106.4	88.1	53.5
UMT5	31.6	22.8	19.5	28.5	28.6	29.0	27.8	27.4
BanglaIPA	**12.4**	**12.3**	**10.8**	**14.3**	**10.8**	**10.7**	**11.1**	**11.4**

Table 1: Performance comparison of different models on the DUAL-IPA dataset for Bengali text-to-IPA transcription.

on the prepared test set described in App. A from the DUAL-IPA dataset. This test set includes standard Bengali as well as six regional variations spoken in Bangladesh: Chittagong, Kishoreganj, Narail, Narsingdi, Rangpur, and Tangail. The distribution of text samples across these regions is detailed in Table 2. We compare the performance of BanglaIPA against two baseline models, MT5 (Xue, 2020) and UMT5 (Chung et al., 2023), using word error rate (WER) as the evaluation metric.

As shown in Table 1, BanglaIPA consistently outperforms both baseline models across all regions, achieving an overall word error rate improvement of 58.4-78.7%. Among the baselines, MT5 exhibits the weakest performance, with a WER of 53.5%. Although UMT5 performs better, achieving a WER of 27.4%, it still falls substantially short of the performance of BanglaIPA. The low overall WER of **11.4%** in our system represents a significant advancement in automated IPA transcription.

Performance across Regions In BanglaIPA, dialect-specific IPA is generated for dialectal vocabulary, while standard IPA is produced for standard vocabulary. We further analyze the system's regional performance variability. As illustrated in Fig. 3, the word error rate remains close to 11% for four of the seven evaluated regions. The lowest word error rate of 10.4% is observed for the Rangpur dialect. Compared to the baseline systems, BanglaIPA demonstrates more consistent performance across Bengali regional variations. This robustness can be attributed to training the Transformer-based model from scratch, whereas the baselines are fine-tuned on the text-IPA pairs.

Numerical Rewriting Analysis For assessing the impact of the LLM-based contextual rewriting stage on the IPA generation pipeline, we construct a dataset of eighteen sentences containing numerical expressions in diverse linguistic contexts, along with their corresponding human-validated rewritten forms. Experimental results show that, in the absence of the contextual rewriting stage, the word error rate between the original text containing numbers and the human-validated rewritten text is 27%. In contrast, incorporating the LLM-based rewriting stage reduces the WER to 1.3% when compared against the same human-validated references. This substantial reduction in error is critical for the downstream components of the BanglaIPA system, which operate on the rewritten text. Although the contextual rewriting stage is not perfectly accurate and may occasionally produce hallucinated outputs, its overall performance represents a significant improvement over the non-rewriting baseline and is sufficient for practical use within the system.

4 Related Work

4.1 Text-to-IPA in Bengali

In recent years, several International Phonetic Alphabet (IPA) transcription systems have been proposed for the Bengali language. A rule-based approach introduced by Al Zubaer et al. (2020) focuses on transcribing standard Bengali characters, words, and numerals. The District Guided Token (DGT) method proposed by Islam et al. (2024) is designed to handle only regional dialects by fine-tuning T5-based models (Raffel et al., 2020) with district-level information. The challenges inherent to Bengali IPA transcription are comprehensively analyzed by Fatema et al. (2024), who also introduced a novel transcription framework along with the DUAL-IPA dataset, which serves as a benchmark for evaluating transcription systems for Bengali.

4.2 Text-to-IPA in Foreign Languages

High-resource languages such as English (Engelhart et al., 2021), Mandarin (Odinye, 2015), German (Odom et al., 2023), and French have seen substantial advances in automatic IPA transcription in recent years. In particular, Yolchuyeva et al. (2020) investigates the effectiveness of attention-based Transformer architectures for the grapheme-to-phoneme (G2P) conversion task (Deri and Knight, 2016), demonstrating notable improvements over

earlier convolutional approaches (Yolchuyeva et al., 2019). More broadly, Cheng et al. (2024) presents a comprehensive survey of neural network-based methods for grapheme-to-phoneme conversion in both monolingual and multilingual settings, outlining current challenges and promising directions.

5 Conclusion & Future Work

In this work, we present BanglaIPA, the first end-to-end system designed to generate International Phonetic Alphabet (IPA) transcriptions of standard Bengali, six regional dialects, and numerals. BanglaIPA employs a contextual rewriting mechanism to handle context-dependent numerical pronunciations effectively. IPA transcriptions are generated on a word-by-word basis using a lightweight Transformer architecture trained on the DUAL-IPA dataset. To address out-of-vocabulary characters and symbols, we introduce the State Alignment (STAT) algorithm that applies model-based transcription generation only to valid subwords. Experimental results demonstrate that the proposed system outperforms baseline models, achieving a minimum word error rate of 11.4%. As future work, we plan to incorporate data from additional regional dialects and train a small language model for efficient contextual rewriting of numbers.

Limitations

The proposed system is trained on data covering six major regional dialects of Bengali. While this provides broad linguistic coverage, performance on dialects that are not represented in the training data may vary, and extending coverage to additional dialects remains an important direction for future work. In addition, the system employs an LLM-based module for context-aware rewriting of numerical expressions. Although this component improves robustness and consistency in numerical reasoning, it introduces modest additional computational cost, which we view as an acceptable trade-off given the gains in accuracy and generalization.

References

Nazmus Sakib Ahmed, saad noor, Shafiq us Saleheen, Sushmit, and Tahsin. 2023. Dataverse challenge - itverse 2023. https://kaggle.com/competitions/dataverse_2023. Kaggle.

Abdullah Al Zubaer, Iftikharul Islam, Md Manik Ahmed, Rohani Amrin, and Md Mehedi Hasan Naim. 2020. Development of phonetic transcription system for bangla to ipa. *Global Scientific Journal*, 8(9):146.

Shengnan An, Zexiong Ma, Zeqi Lin, Nanning Zheng, Jian-Guang Lou, and Weizhu Chen. 2024. Make your llm fully utilize the context. *Advances in Neural Information Processing Systems*, 37:62160–62188.

Nazmuddoha Ansary, Quazi Adibur Rahman Adib, Tahsin Reasat, Asif Shahriyar Sushmit, Ahmed Imtiaz Humayun, Sazia Mehnaz, Kanij Fatema, Mohammad Mamun Or Rashid, and Farig Sadeque. 2023. Unicode normalization and grapheme parsing of indic languages. *arXiv preprint arXiv:2306.01743*.

International Phonetic Association. 1999. Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet. Printed edition. Cambridge University Press.

Guangsheng Bao and Yue Zhang. 2021. Contextualized rewriting for text summarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12544–12553.

Shiyang Cheng, Pengcheng Zhu, Jueting Liu, and Zehua Wang. 2024. A survey of grapheme-to-phoneme conversion methods. *Applied Sciences*, 14(24):11790.

Hyung Won Chung, Noah Constant, Xavier Garcia, Adam Roberts, Yi Tay, Sharan Narang, and Orhan Firat. 2023. Unimax: Fairer and more effective language sampling for large-scale multilingual pretraining. *arXiv preprint arXiv:2304.09151*.

Peter T Daniels and William Bright. 1996. *The world's writing systems*. Oxford University Press.

Seniz Demir and Berkay Topcu. 2022. Graph-based turkish text normalization and its impact on noisy text processing. *Engineering Science and Technology, an International Journal*, 35:101192.

Aliya Deri and Kevin Knight. 2016. Grapheme-to-phoneme models for (almost) any language. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 399–408.

Reham Elshamy, Osama Abu-Elnasr, Mohamed Elhoseny, and Samir Elmougy. 2023. Improving the efficiency of rmsprop optimizer by utilizing nestrove in deep learning. *Scientific Reports*, 13(1):8814.

Eric Engelhart, Mahsa Elyasi, and Gaurav Bharaj. 2021. Grapheme-to-phoneme transformer model for transfer learning dialects. *arXiv preprint arXiv:2104.04091*.

Kanij Fatema, Fazle Dawood Haider, Nirzona Ferdousi Turpa, Tanveer Azmal, Sourav Ahmed, Navid Hasan, Mohammad Akhlaqur Rahman, Biplab Kumar Sarkar, Afrar Jahin, Md Rezuwan Hassan, and 1 others. 2024. Ipa transcription of bengali texts. *arXiv preprint arXiv:2403.20084*.

Jakir Hasan and Shubhashis Roy Dipta. 2025. Banglatalk: Towards real-time speech assistance for bengali regional dialects. *arXiv preprint arXiv:2510.06188*.

Md Nahid Hasan, Raiyan Azim, and Sadia Sharmin. 2024. Credibility analysis of robot speech based on bangla language dialect. In *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*, pages 1–6. IEEE.

Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M. Sohel Rahman, and Rifat Shahriyar. 2020. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for Bengali-English machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2612–2623, Online. Association for Computational Linguistics.

Ahmed Imtiaz Humayun, farigys, Md. Rezuwan Hassan, research bengaliai, Rubayet Sabbir Faruque, Sushmit, Tahsin, Tanvir Rahman Talha, and Yeasir Arafat. 2024. Bhashamul: Bengali regional ipa transcription. https://kaggle.com/competitions/regipa. Kaggle.

Md Fuadul Islam, Jakir Hasan, Md Ashikul Islam, Prato Dewan, and M Shahidur Rahman. 2025. Banglalem: a transformer-based bangla lemmatizer with an enhanced dataset. *Systems and Soft Computing*, page 200244.

SM Islam, Sadia Ahmmed, and Sahid Hossain Mustakim. 2024. Transcribing bengali text with regional dialects to ipa using district guided tokens. *arXiv preprint arXiv:2403.17407*.

Victoria Johansson. 2008. Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, 53:61–79.

SM Kamal. 2025. *Bengali text to international phonetic alphabet (IPA) transcription: a comprehensive BYT5 based approach*. Ph.D. thesis, Brac University.

Fan Liu and Yong Deng. 2020. Determine the number of unknown targets in open world based on elbow method. *IEEE Transactions on Fuzzy Systems*, 29(5):986–995.

David Malvern, Brian Richards, Ngoni Chipere, and Pilar Durán. 2004. *Lexical diversity and language development*. Springer.

Anqi Mao, Mehryar Mohri, and Yutao Zhong. 2023. Cross-entropy loss functions: Theoretical analysis and applications. In *International conference on Machine learning*, pages 23803–23828. pmlr.

C Nath and B Sarma. 2024. Ai enabled text-to-speech synthesis for unicode language. *Indian Journal of Science and Technology*, 17(42):4454–4461.

Mark Needleman. 2000. The unicode standard. *Serials review*, 26(2):51–54.

Sunny Ifeanyi Odinye. 2015. Phonology of mandarin chinese: a comparison of pinyin and ipa. *London: Bloomsburry*, pages 5–20.

William Odom, Benno Schollum, and Christina Balsam Curren. 2023. *German for singers: A textbook of diction and phonetics*. Plural Publishing.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Heike Wiese. 2003. *Numbers, language, and the human mind*. Cambridge University Press.

L Xue. 2020. mt5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*.

Sevinj Yolchuyeva, Géza Németh, and Bálint Gyires-Tóth. 2019. Grapheme-to-phoneme conversion with convolutional neural networks. *Applied Sciences*, 9(6):1143.

Sevinj Yolchuyeva, Géza Németh, and Bálint Gyires-Tóth. 2020. Transformer based grapheme-to-phoneme conversion. *arXiv preprint arXiv:2004.06338*.

Haitong Zhang, Haoyue Zhan, Yang Zhang, Xinyuan Yu, and Yue Lin. 2021. Revisiting ipa-based cross-lingual text-to-speech. *arXiv preprint arXiv:2110.07187*.

Zhi-Hua Zhou. 2021. *Machine learning*. Springer nature.

Yilun Zhu, Joel Ruben Antony Moniz, Shruti Bhargava, Jiarui Lu, Dhivya Piraviperumal, Yuan Zhang, Hong Yu, and Bo-Hsiang Tseng. 2024. Can large language models understand context? *arXiv preprint arXiv:2402.00858*.

Algorithm 2 SPLITDATASET: Construction of training and test sets from the DUAL-IPA dataset using the IPA novelty score.

Require: Dataset D, Elbow point score EP
Ensure: Training set D_{train}, Test set D_{test}

```
Initialize D_test ← ∅
Initialize D_train ← D
CS ← ∞
while CS ≥ 0 do
    for each sample S in D_train do
        Extract word IPA set S_S from S
        Extract word IPA set S_T from D_test
        CS(S) ← |S_S − S_T|
    end for
    Select S* ← arg max_S CS(S)
    CS ← CS(S*)
    if CS < EP then
        break
    end if
    Move S* from D_train to D_test
end while
return D_train, D_test
```

Appendix

A DUAL-IPA Dataset Detail

The DUAL-IPA dataset (Fatema et al., 2024) represents the first effort to standardize Bengali IPA transcription by providing parallel pairs of text and corresponding IPA annotations. Two subsets of this comprehensive dataset have been made publicly available through affiliated competitions[3]: the DataVerse Challenge - ITVerse 2023 (Ahmed et al., 2023), which contains a corpus of standard Bengali text, and Bhashamul: Bengali Regional IPA Transcription (Humayun et al., 2024), which includes corpora from six regional dialects. In this study, we utilize these publicly available subsets to train and evaluate our novel automatic IPA transcription system.

Preprocessing In Bengali, the same character can be represented using multiple Unicode encodings (Needleman, 2000), which introduces ambiguity into the text (Ansary et al., 2023). Since neural models typically perform better when trained on standardized and normalized inputs (Demir and Topcu, 2022), we apply a Bengali text normalization procedure using the normalizer proposed by Hasan et al. (2020). This preprocessing step ensures consistent character representations and improves the reliability of downstream modeling.

[3] https://www.bengali.ai/

Region	Number of Sample		
	Train	Test	Total
Chittagong	4157	605	4762
Kishoreganj	6089	642	6731
Narail	5449	573	6022
Narsingdi	3923	586	4509
Rangpur	3539	503	4042
Tangail	3732	513	4245
Standard	18882	3117	21999
Total	45771	6539	52310

Table 2: Distribution of the processed training and testing data in the DUAL-IPA dataset of parallel text corpus across the standard Bengali and its six regional dialects.

Elbow Method for Dataset Splitting We use the normalized dataset to construct the training and test sets, as the original test split of the DUAL-IPA dataset is not publicly available. To minimize lexical overlap between the two splits and to maximize the presence of unseen words in the test set, we employ a custom dataset splitting procedure, described in Alg. 2. At each iteration, the sample with the highest IPA novelty score is selected and assigned to the test set. This novelty score is computed based on the number of words in the sample that have not yet appeared in the test set. The test set construction continues until the novelty score falls below a predefined Elbow Point (EP) (Liu and Deng, 2020), after which the remaining samples are allocated to the training set.

Fig. 5 presents the novelty score progression across iterations for the Rangpur dialect. The Elbow Point (EP) is identified when the curve flattens and becomes approximately parallel to the x-axis, indicating that further iterations result in negligible changes to the score. In our experiments, the EP is observed at a score of approximately 3. Based on this criterion, we construct the training and test splits shown in Table 2. The resulting test set maximizes the number of unseen words and lexical diversity (Malvern et al., 2004; Johansson, 2008).

B Pronunciation Change of Number

Numbers are a fundamental component of any language, conveying critical information in everyday communication (Wiese, 2003). Accurate interpre-

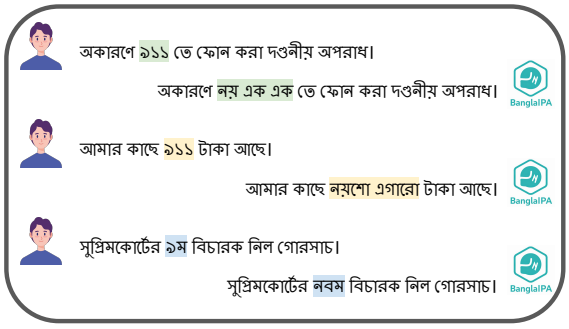

Figure 4: The pronunciation of Bengali numbers varies depending on the textual context (highlighted in colors). The `BanglaIPA` system converts numbers into word forms as an intermediate step in IPA transcription.

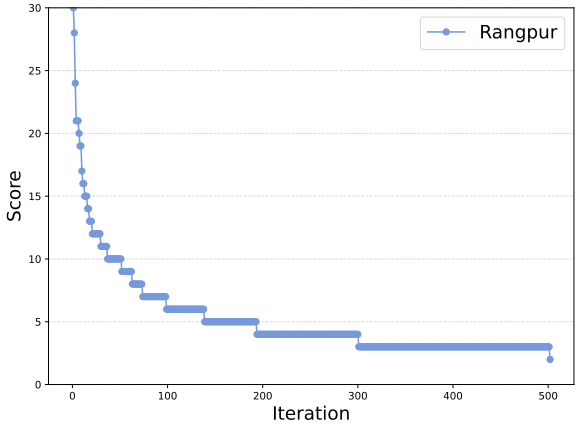

Figure 5: Identification of the optimal Elbow Point (EP) to terminate the dataset splitting procedure for the Rangpur dialect using the Elbow method.

tation of numerical expressions is essential for a robust IPA transcription system in Bengali. Numerical expressions representing quantities, years, dates, times, phone numbers, and similar constructs can have multiple verbal pronunciations and word forms, as illustrated in Fig. 4. In the `BanglaIPA` system, we address this challenge by converting numbers into intermediate word-form representations using `GPT-4.1-nano`. We provide the model with the prompt shown in Fig. 6, instructing it to leverage the contextual information of the full sentence (An et al., 2024) to generate word forms for numerical expressions while leaving other words unchanged. The resulting contextually enriched text is subsequently used for model-based IPA transcription, improving the accuracy and robustness.

System Prompt: You are a helpful chatbot who understands Bengali numerals in different contexts.

User Prompt: Please rewrite the provided text so that no Bengali digits are present. Convert the numbers to word form based on the context. Do not modify any words.

Here is the text: `{user_text}`.

Figure 6: Prompt used to instruct the large language model to generate contextually rewritten Bengali text by converting numerical expressions into word forms.