Bengali Text to International Phonetic Alphabet (IPA) Transcription: A Comprehensive UMT5 Based Approach

1st Niaz Mohaiman Abtahi
Dept.of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Email: tahiniaz@gmail.com

2ndAbrar Jahin Niloy
Dept.of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Email: abrarjahan.niloy@gmail.com

***Abstract*—In the rapidly evolving digital era, the use of International Phonetic Alphabet for Bangla language is crucial for maintaining linguistic precision, aiding language learning, fostering research, preserving linguistic diversity, and facilitating communication and speech therapy. It plays a vital role in various aspects of Bangla language study and related fields. In our work, we have trained UMT5-base transformer model architecture on the competition dataset to convert text into IPA notation from Bangla texts. Our work extends beyond training, delving into the exploration of potential enhancement avenues. We have investigated the impact of Bangla numerals, abbreviations, English texts and out-of-vocabulary words. Through these explorations, we observe a spectrum of outcomes, where some modifications result in tangible performance improvements, while others offer unique insights for future endeavors.**

I. Introduction

The International Phonetic Alphabet (IPA) is an alphabetic system of phonetic notation primarily based on the Latin script. It was devised by the International Phonetic Association in the late 19th century as a standardized representation of speech sounds in written form. The IPA is widely used by lexicographers, foreign language students and teachers, linguists, speech–language pathologists, singers, actors, constructed language creators, and translators. It is a vital tool for the study of language, language learning, and various applications related to speech and communication. It ensures precision and consistency in the representation of speech sounds, making it an invaluable resource for linguists, educators, and speech professionals.IPA serves as a phonetic notation system, utilizing symbols to represent each distinct sound found in human spoken language. Its scope encompasses all languages spoken on Earth.

However, despite the remarkable strides made in IPA analysis across various languages, the Bengali language has remained relatively unexplored in this realm. The DataVerse Challenge - ITVerse 2023 competition [4] on Bangla text to IPA transcription is a timely initiative that aims to address this gap. The competition provides a challenging dataset of Bangla texts. This will allow researchers to develop and evaluate new systems for Bangla language.

Traditionally, Transformers [6] have been used in a wide range of natural language processing applications. Examples include machine translation, text generation, question answering, automatic summarization, text classification, and sentiment analysis. In this paper, we propose a novel approach for Bangla text to IPA conversion. UMT5 [1] model is used to fulfill our purpose which is an encoder-decoder based transformer model. We focus on enhancing multiple aspects of the task, such as data preprocessing and model architecture. To evaluate our model, we utilize the metric WER [7]. Our approach achieves a public score of 0.13264 on the competition leaderboard.

II. Methodology

A. Model Selection

In order to generate the transcriptions, pretrained MT5-small [9] deep learning architecture from huggingface is employed at first. After that, MT5-base, M2M100 [3], BERT [2], ByT5-small [8], BanglaT5 [5] and UMT5-base models are finetuned on the competition dataset. UMT5-base outperforms other models by a significant margin. Hence, we have chosen the UMT5-base model architecture, an encoder-decoder based transformer model with SentencePiece Tokenizer as the cornerstone of our IPA transcription.

B. Preprocessing

Our preprocessing workflow encompassed several crucial steps to ensure the quality and consistency of input data for our transcription. Texts containing English alphabets and numerals are removed from provided datasets. We have chosen 10% of the labelled dataset as validation dataset and the rest as training dataset with reproducibility ensured. One of the difficulties we have faced in the competition is dealing with Bangla numerals, dates, signs and abbreviations. We have tried to convert these nonlinear texts into linear ones.

C. Training

Our training was conducted for a total of 15 epochs. Optimization was achieved through the AdamW optimizer, with an initial learning rate of 5e-4 and a weight decay of 1e-2. We started with a slow learning rate with the intention to avoid overshooting and promote model stability. The warmup steps were set to be 1000.

We structured data loading and processing using a batch size of 8 for training. During training, crucial metrics were logged every 2000 steps and model checkpoints were saved every 2000 iterations to facilitate potential model recovery.

III. Results

After a total training of 15 epochs in 1 phases, our model scored a Wer score of 0.13264 on the public test set and a training loss of 0.0024 and a validation loss of 0.034363 and wer score 0.018953. The loss metrics gradually decreased throughout the training. From figure 1 we can see that the wer score reached about 0.018953 after 36000 steps. However, the training loss slowly continued to improve over the training period.

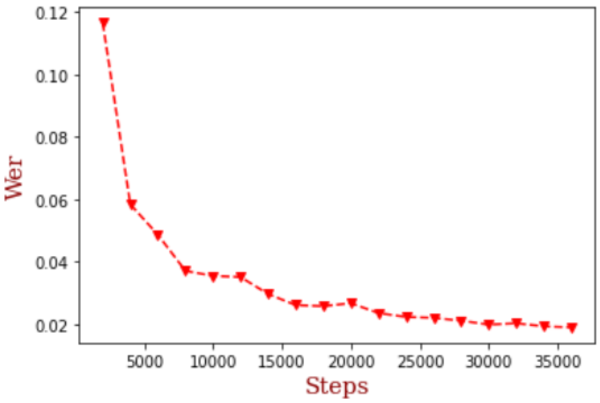
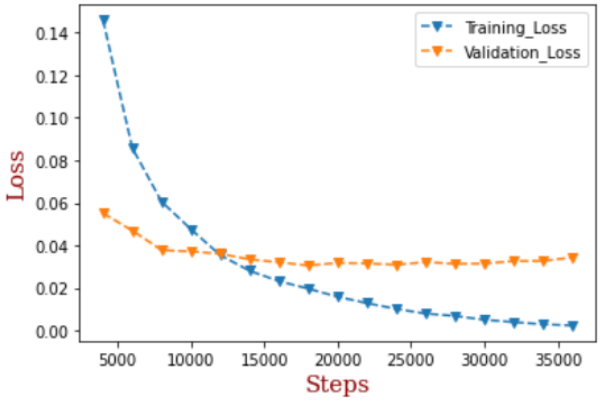


Fig. 1. Wer vs Epoch



We evaluate competition dataset using a MT5-small, MT5-base, UMT5-base, ByT5-small, Bangla-t5, M2M100 model architecture.

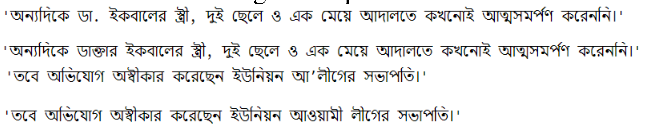
Table 3 summarizes the overall obtained results for each of the transformer model architecture

Model	Epoch	Batch Size	Validation loss	Wer
MT5-small	10	8	0.041791	0.023235
MT5-base	10	8	0.035327	0.174815
UMT5-base	15	8	0.034363	0.018953
ByT5-small	11	4	0.018967	0.156055
Bangla-t5	15	4	0.742986	0.909394
M2M100	15	4	0.018967	0.28967

IV. Discussion

A. Abbreviations

There are many abbreviations in the given dataset that we had to deal with. The full form is replaced for better results and better understanding. Examples:



'অন্যদিকে ডা. ইকবালের স্ত্রী, দুই ছেলে ও এক মেয়ে আদালতে কখনোই আত্মসমর্পণ করেননি।'

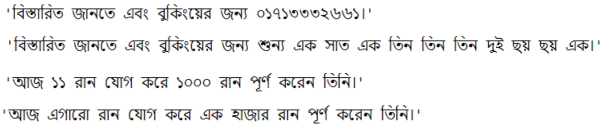
'অন্যদিকে ডাক্তার ইকবালের স্ত্রী, দুই ছেলে ও এক মেয়ে আদালতে কখনোই আত্মসমর্পণ করেননি।'

'তবে অভিযোগ অস্বীকার করেছেন ইউনিয়ন আ'লীগের সভাপতি।'

'তবে অভিযোগ অস্বীকার করেছেন ইউনিয়ন আওয়ামী লীগের সভাপতি।'

B. Bangla Numerals

We had to face much difficulty converting the numerals as these are pronounced differently according to the usage. The way a phone number is pronounced is not the same for others such as the amount of money because phone number is pronounced digit by digit and the amount of money is pronounced as a whole. Examples:



'বিস্তারিত জানতে এবং বুকিংয়ের জন্য ০১৭১৩৩৩২৬৬১।'

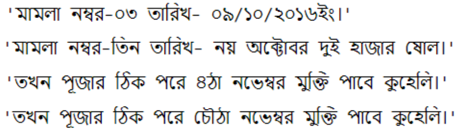
'বিস্তারিত জানতে এবং বুকিংয়ের জন্য শুন্য এক সাত এক তিন তিন তিন দুই ছয় ছয় এক।'

'আজ ১১ রান যোগ করে ১০০০ রান পূর্ণ করেন তিনি।'

'আজ এগারো রান যোগ করে এক হাজার রান পূর্ণ করেন তিনি।'

C. Dates

The dates are converted according the way they are pronounced. Examples:



'মামলা নম্বর-০৩ তারিখ- ০৯/১০/২০১৬ইং।'

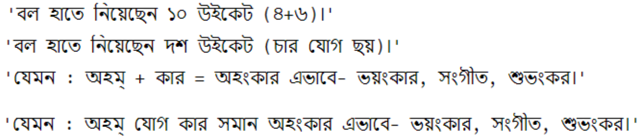
'মামলা নম্বর-তিন তারিখ- নয় অক্টোবর দুই হাজার ষোল।'

'তখন পূজার ঠিক পরে ৪ঠা নভেম্বর মুক্তি পাবে কুহেলি।'

'তখন পূজার ঠিক পরে চৌঠা নভেম্বর মুক্তি পাবে কুহেলি।'

D. Signs

Also, some texts contain special signs such as '+', '-'. Examples:



'বল হাতে নিয়েছেন ১০ উইকেট (৪+৬)।'

'বল হাতে নিয়েছেন দশ উইকেট (চার যোগ ছয়)।'

'যেমন : অহম্ + কার = অহংকার এভাবে- ভয়ংকার, সংগীত, শুভংকর।'

'যেমন : অহম্ যোগ কার সমান অহংকার এভাবে- ভয়ংকার, সংগীত, শুভংকর।'

E. Decimal Separator

Also, Numbers containing decimal separator are pronounced in a different way. Examples:

'স্ট্রাইক রেট ১৬৯.৭৯।'

'স্ট্রাইক রেট একশো উনসত্তর দশমিক সাত নয়।'

V. Conclusion

We present an effective scheme to finetune the transfermer based UMT5 model for Bengali texts to IPA transcription. Our approach has achieved state-of-the-art result on the competition leaderboard, with a public score of 0.13264. We believe that our approach is a significant step forward in the field of Bangla language.

References

[1] Hyung Won Chung, Xavier Garcia, Adam Roberts, Yi Tay, Orhan Firat, Sharan Narang, and Noah Constant. Unimax: Fairer and more effective language sampling for large-scale multilingual pretraining. In *The Eleventh International Conference on Learning Representations*, 2023.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[3] Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. Beyond english-centric multilingual machine translation, 2020.

[4] Shafiq-us Saleheen Sushmit Tahsin Nazmus Sakib Ahmed, saad noor. Dataverse challenge - itverse 2023, 2023.

[5] H. A. Z. Sameen Shahgir and Khondker Salman Sayeed. Bangla grammatical error detection using t5 transformer model, 2023.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[7] Thilo von Neumann, Christoph Boeddeker, Keisuke Kinoshita, Marc Delcroix, and Reinhold Haeb-Umbach. On word error rate definitions and their efficient computation for multi-speaker speech recognition systems. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, jun 2023.

[8] Linting Xue, Aditya Barua, Noah Constant, Rami Al-Rfou, Sharan Narang, Mihir Kale, Adam Roberts, and Colin Raffel. Byt5: Towards a token-free future with pre-trained byte-to-byte models, 2022.

[9] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mt5: A massively multilingual pre-trained text-to-text transformer, 2021.