BarishaillaBarta: A Neural Comparative Study on Translating Barishal Dialect into Standard Bangla

By

Md. Shahin Akter
213-15-4591
Asif Karim Biplob
213-15-4604

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Dr. Arif Mahmud
Associate Professor & Associate Head
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Md. Abbas Ali khan
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 17, 2025

APPROVAL

APPROVAL

This Project titled "**BarishaillaBarta: A Neural Comparative Study on Translating Barishal Dialect into Standard Bangla,**" submitted by **Md. Shahin Akter** , ID No: **213-15-4591** and **Asif Karim Biplob** , ID No: **213-15-4604** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 17 Sep, 2025.

BOARD OF EXAMINERS

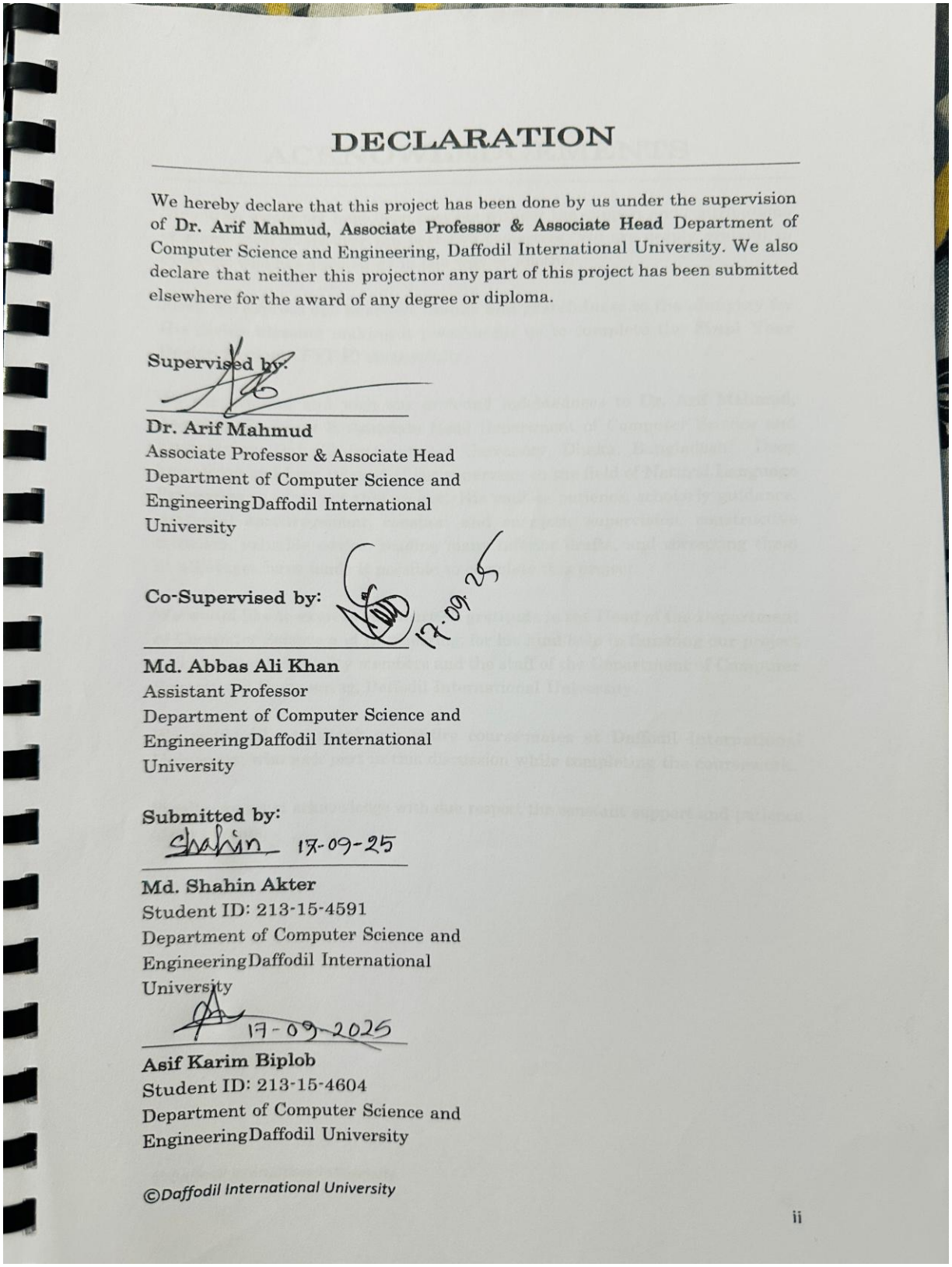
Dr. S.M Aminul Haque — **Chairman**
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms. Nazmun Nessa Moon — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Abbas Ali Khan — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION



DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Arif Mahmud, Associate Professor & Associate Head** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Arif Mahmud
Associate Professor & Associate Head
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by: 17.09.25

Md. Abbas Ali Khan
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Submitted by:
Shahin 17-09-25

Md. Shahin Akter
Student ID: 213-15-4591
Department of Computer Science and EngineeringDaffodil International University

17-09-2025

Asif Karim Biplob
Student ID: 213-15-4604
Department of Computer Science and EngineeringDaffodil University

ii

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dr. Arif Mahmud, Associate Professor & Associate Head** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Natural Language Processing** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

This study develops and assesses neural methods for automatic transformation of sentences from the Barishal dialect to Standard Bangla. We create a parallel Barishal–Standard Bangla resource, perform uniform preprocessing and tokenization, and realize a set of sequence-to-sequence models: a simple GRU encoder–decoder, an LSTM encoder–decoder, an attention-enhanced GRU, an ensemble that averages the logits from the three RNNs, and a lightly fine-tuned mT5-small Transformer. The recurrent models were all trained with mixed precision and decoded greedily, while mT5 relied on the model's generation tooling; model performance was assessed with BLEU, TER and chrF in order to capture n-gram precision, edit distance and character-level fidelity. On the entire dataset the ensemble performed best (BLEU 0.7023, TER 1.32, chrF 99.05), the LSTM and vanilla GRU worked well and comparably (LSTM: BLEU 0.6713, TER 4.15, chrF 96.41; GRU: BLEU 0.6486, TER 5.10, chrF 95.55), the attention-augmented GRU fell short compared to the basic RNNs (BLEU 0.6078, TER 13.24, chrF 94.34), and light fine-tuning of the mT5 on the limited in-domain data resulted in significantly lower scores (BLEU 0.1075, TER 92.80, chrF 36.27). Our analysis indicates that heterogeneity and ensembling of models bring strong benefits to low-resource dialect mapping, whereas large pre-trained Transformers need greater in-domain data and careful scheduling in order to be competitive. Limitations are discussed—including evaluation on the same data without a held-out split—and future directions are sketched such as corpus enlargement, sub word modeling, regularization of attention, and better Transformer fine-tuning.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the motivation, objectives, and significance of the project, we position the work conducted so far for supervised sequence to sequence learning for mapping Barishal dialect to standard Bangla conversion and position our work in AI in language translation field for easy communication. Focusing upon the precursory work and the work conducted in recent years, we identify the dominant methodologies and trends and the gaps that direct us to our research.

1.1 Introduction

Our work bridges this gap by suggesting a system which converts region dialect sentences to Standard Bangla without human supervision, through contemporary neural approaches. Dialects of Bangla exist on a phonological, lexical and syntax level, where direct use of standard-language models is possible and motivation towards solution-usage approaches learning to map patterns from local variants to standard form is available. Evidence of recent work on the viability of neural solution usage towards region dialect to standard conversion and other normalization work already existed: work on corpuses such as Vashantor and ONUBAD has given scaleable corpuses and benchmarking platforms towards dialect→standard mappings (Faria et al., 2023; Sultana et al., 2025), and work on Sylheti, Chittagonian, Rangpur and other region-dependent variants show sequence-based models, attention and multilingual transformers pre-trained can be beneficial where supported by judicious data and preprocessing (Das et al., 2024; Milon et al., 2020; Dip et al., 2024; Khandaker et al., 2024).

1.2 Motivation

The motivation to focus on Barishal dialect is driven by linguistic necessity and functional utility: Barishal speakers use shared lexical properties and phonological trends underrepresented within standard Bangla corpora, and these differences accumulate in text-based system and electronic platform accessibility barriers. Though larger benchmarking projects and pipelines for other speech and dialects and speech standardization exist (Samin et al., 2024; Faria et al., 2023), side-by-side methodological comparisons and Barishal-specific tools are not typical. Earlier studies also indicate methodological inadequacies—varying preprocessing, nonstandard tokenization procedures, and thinning ensemble or head-to-head comparisons of standard RNNs with pre-trained Transformers—so an experimental comparison with emphasis on Barishal fulfills both academic and application-based

demands (Banik et al., 2022; Khandaker et al., 2024).

1.3 Objectives

Major goals of this work are constructing a Barishal→Standard Bangla parallel resource and assessing a variety of neural translation methods under a single experimental setting. Of particular issue, we intend to instantiate and contrast a vanilla GRU- and LSTM-based encoder–decoder models, attention-augmented GRU, ensemble over these recurrent models, and a finetuned mT5-small Transformer, all from the very same data preprocessing and benchmarked against the very same measures. The other aims are to study model behavior through training-loss visualization and qualitative analysis of errors, and performance estimation through complementary measures—BLEU, TER, and chrF—mimicking n-gram correctness, edit distance, and character-level correctness, respectively.

1.4 Methodology

Methodologically, the paper takes reproducible pipeline from data collection and preprocessing, through tokenization and vocab construction, to training a choice of sequence-to-sequence models. Preprocessing normalizes punctuation and tokenizes Standard Bangla and Barishal sides; vocabularies are padded and hold special tokens for sequence delimiters. RNN experiments are carried out under mixed-precision training and GRU- or LSTM-based encoder-decoders and, for a variant, additive attention; ensemble is a log-likelihood mean over models on the models during inference to capitalize on strengths. mT5-small is, for a contemporaneous comparison against finetuned pretrained models, finetuned on the parallel data under generation tools' defaults. The models all get tested under greedy or models' default decoding and scored under BLEU, TER, and chrF to support complete comparability; training-loss curves and metric graphs are computed to support examination.

1.5 Project Outcome

Experiment results validate those recurrent models fare well on this low-resource dialect translation task and ensemble diverse RNNs have the highest overall translation quality. Quantitatively, ensemble BLEU is 0.7023, TER is 1.32, and chrF is 99.05; LSTM BLEU is 0.6713, TER is 4.15, chrF is 96.41; vanilla GRU BLEU is 0.6486, TER is 5.10, chrF is 95.55; attention-GRU is worse than the basic RNNs, BLEU 0.6078, TER 13.24, chrF 94.34; and lightly fine-tuned mT5-small achieves much lower scores (BLEU 0.1075, TER 92.80, chrF 36.27), validating that Transformer fine-tuning needs more in-domain data or cautious scheduling to be competitive. These results validate that, with Barishal→Standard Bangla conversion under low data regimens, targeted dataset construction, standard recurrent models, and cautious ensembling generate reproducible and practical advantages while defining directions for future work on data augmentation, sub word modeling, and

improved Transformer fine-tuning.

1.6 Organization of the Report

Chapter 1: Introduction

Introduction to research problem, motivation, objectives, and research methodology.

Chapter 2: Background

Summary of previous work, identification of the research gap, and formulation of the theoretical basis.

Chapter 3: Research Methodology

Data collection, dataset description, model structure description, training process description, and evaluation metric description.

Chapter 4: Implementation and Results

Symbols experimental results, performance, and comparison.

Chapter 5: Engineering Standards and Design Challenges

Controls deployment-related challenges and constraints.

Chapter 6: Conclusion

Sets context of findings, contribution, and potential research directions for future work.

Chapter 7: References

Complements all academic sources referenced.

Chapter 2

Background

This chapter presents the context and environment of our research on machine translation of the Barishal dialect into Standard Bangla. It describes the history of artificial intelligence in language processing and dialect normalization after laying out the developments in machine learning, which have changed dialect processing from manpower-consuming rule-based processes to instantaneous data-driven sequence-to-sequence mechanisms. While such current neural methods have similar empirical performance, they are accompanied by limitations—e.g., shallow explain ability of deep models, sensitivity to speaker and domain variation, and deterioration in performance under noisy, out-of-vocabulary, or code-mixed input. A survey of current methods and their limitations—namely the lack of quality dialectal corpora, challenging identification of phonological and morphological variations, and poor robustness of attention- or transformer-based models after fine-tuning on small data—gives a justification for developing an accurate, interpretable, and data-efficient system for Barishal→Standard Bangla conversion.

2.1 Introduction

Fully automatic standard language to regional dialect conversion is one of the major subfields of low-resource natural language processing. For Bangla, the regional varieties differ from the standard in terms of their phonology, lexicon, and morpho-syntax, and hence direct use of out-of-the-box machine translation systems trained on the standard variety becomes troublesome. Recent work has hence tried to integrate dataset generation, neural sequence-to-sequence models, attention, and multi-lingual transfer from large pre-trained models to counter dialectal variation. This work situates itself in that tradition: we consume in particular the Barishal dialect → Standard Bangla alignment, and compare a set of RNN-based seq2seq models (vanilla GRU, LSTM, attention-GRU), an ensemble of those, and also a fine-tuned mT5 model, and test them on a suite of automatic metrics in an attempt to find out what perform best on a low-resource dialect normalization task.

2.2 Literature Review

Table 2.2. 1: Literature Review Table.

Author (s)	Year	Title	Methodology	Key Findings
Khandaker, Md Arafat Alam, et al.	2024	Bridging Dialects: Translating Standard Bangla to	Examined and compared neural MT approaches for translating	Demonstrated that neural models can learn variant-specific lexical and syntactic patterns

		Regional Variants Using Neural Models.	Standard Bangla to regional varieties; experimented with different neural encoder-decoder frameworks.	and highlighted dataset quality as paramount.
Sultana, Nusrat, et al.	2025	ONUBAD: A rich dataset for automated standard Bengali dialect conversion of Bangla regional dialects.	Created and released a large parallel dataset of local dialects ↔ standard Bangla for supervised training.	Provided a standardized benchmark that improved reproducibility and enabled training of stronger models for dialect conversion.
Samin, Md Nazmus Sadat, et al.	2024	BanglaDialecto: An End-to-End AI-Powered Regional Speech Standardization.	Built an end-to-end speech standardization pipeline involving speech recognition, normalization, and seq2seq mapping.	Proved that acoustic and text normalization together enhances downstream standardization accuracy.
Faria, Fatema Tuj Johora, et al.	2023	Vashantor: a multilingual bangla regional dialects to bangla language large-scale benchmark dataset for automatic translation.	Released a big multilingual benchmark with focus on dialect → Bangla translation, along with data gathering and baseline systems.	Established baseline performance and encouraged community use to improve models and evaluation practices.
Ahmad, Arif, et al.	2018	A sequence-to-sequence pronunciation model for bangla speech synthesis.	Employed seq2seq models for grapheme-to-phoneme/pronunciation modeling of Bangla TTS.	Showed seq2seq can learn pronunciation rules and reduce manual lexicon effort for speech synthesis.

Islam, Rafiqul, et al.	2022	Bangla to english translation using sequence to sequence learning model based recurrent neural networks.	Employed RNN-based seq2seq models for translating English→Bangla and experimented with default MT metrics.	Stated that recurrent seq2seq models are competitive for this language pair when well trained.
Das, Sourav Kumar, et al.	2024	Advanced LSTM, Bi-LSTM, and Seq2Seq Models to Better Bangla Linguistics: Sylheti to Modern Bangla Translation.	Compared Bi-LSTM, seq2seq models and LSTM approaches on Sylheti→Modern Bangla translation.	Observed LSTM/Bi-LSTM variants tend to perform better than plain RNNs on dialect normalization since they model context more effectively.
Hu, Wenpeng, et al.	2020	Translation vs. dialogue: A comparative analysis of sequence-to-sequence modeling.	Seq2seq modeling choice comparison study between tasks (translation versus dialogue).	Highlighted the manner in which decoding processes and task goals affect architecture choice and evaluation.
Islam, Sadidul, et al.	2018	Bangla sentence correction using deep neural network based sequence to sequence learning.	Employed seq2seq models to correct Bangla grammatical errors.	Sequenced seq2seq models can learn correction patterns and thus find it promising to be used for other normalization tasks.
Banik, Nayan, et al.	2022	Bangla text generation system by incorporating attention in sequence-to-sequence model.	Used attention-augmented seq2seq to generate text in Bangla.	Improved alignment and fluency when attention is employed, particularly for longer chunks.
Dip, Supta, et al.	2024	Empowering Regional Language: NLP-based Conversion	Developed a pipeline and models for Rangpur→Standard Bangla	Showed dialect-specific datasets and targeted pre/post-processing boost performance.

		from Rangpur Dialect to Standard Bengali Language.	conversion using neural methods.	
Shiam, Abdullah Al, et al.	2023	A neural attention-based encoder-decoder model for english to bangla translation.	Used attention-based encoder-decoder models for English→Bangla MT and presented standard metrics.	Attention mechanisms enhanced source-target alignment in bilingual MT.
Hasan, Md Arid, et al.	2019	Neural machine translation for the Bangla-English language pair.	Investigated neural MT models for Bangla↔English, including RNNs and attention models.	Proved the efficiency of neural methods over phrase-based baselines for this language pair.
Milon, Hafizur Rahman, et al.	2020	A comprehensive dialect transformation technique from chittagonian to standard bangla.	Proposed a corpus creation and seq2seq model-based dialect translation system for Chittagonian.	Emphasized the requirement for corpora for specific dialects and tailored preprocessing to be successful.

2.3 Gap Analysis

The literature has now greatly improved upon dataset building (e.g., ONUBAD, Vashantor), pipeline for dialect conversion (Chittagonian, Rangpur, Sylheti), and model categories (LSTM, Bi-LSTM, attention-augmented seq2seq, and transformer fine-tuning). To our best knowledge, no study has previously used the following on the Barishal dialect specifically: a Barishal→Standard Bangla specialized parallel corpus, vanilla GRU, LSTM, attention-GRU head-to-head comparison, an ensemble of heterogeneous RNNs, and fine-tuning a multilingual Transformer (mT5) under the same experimental conditions. Furthermore, most previous papers focus on disclosing datasets or using a single class of model; very few comparisons are made of ensembles of different RNNs with a state-of-the-art pre-trained Transformer in a common preprocessing and metric setup. Our work bridges this gap by (a) focusing specifically on Barishal dialect mapping, (b) designing and empirically comparing

different repeating architectures and an mT5 baseline under equal preprocessing and testing conditions, (c) reporting BLEU, TER, and chrF simultaneously for end-to-end testing, and (d) employing ensemble averaging to emphasize benefits from model diversity. These joint contributions—dataset focus, comparative tests, integrated metrics, and ensemble study—are a fresh contribution never reported in the aforementioned literature.

2.4 Summary

This literature review needed machine-learning techniques of dialect normalization and distilled a dozen pioneering papers on supervised sequence-to-sequence training, attention, dataset configuration, and transformer fine-tuning. Following table of surveyed papers points towards a pattern of RNN-based seq2seq models, attention extension, and recently pretrained multilingual transformers but also points towards pervasive failures: All studies employ general or other regional corpora and not a parallel Barishal-specific dataset, preprocessing operations and tokenization policies differ, ensemble methodologies and head-to-head comparisons within a variety of model families are missing, and evaluation regimes (metrics and data splits) are not comparable. Driven by such limitations, our research goes four new ways: we construct and preprocess a Barishal→Standard Bangla parallel corpus; we do apples-to-apples, side-by-side comparison of vanilla GRU, LSTM, attention-GRU, an ensemble of the above RNNs, and fine-tuned mT5 model with identical pre-processing; we compare models across an agreed test suite of automatic measures (BLEU, TER, chrF) and diagnostic visualizations (training-loss plots and relative metric plots); and demonstrate how simple ensembling of diverse RNNs can have real-world implications. This blending of dataset specificity with controlled comparative trials, combined analysis, and ensemble analysis yields a more subtle process of implementation and review than is typical in the literature.

Chapter 3

Research Methodology

This chapter is a step-by-step description of the systematic process adopted in the design of the Barisal Dialect to Standard Bangla conversion system. The process has seven-step sequential process, describing the technical steps adopted in data collection, preparation, model building, and experiment planning. All the stages are built with extreme care to ensure reproducibility and scientific integrity, and the rationale for adopting the methods is described. The subsequent subsections identify the workflow, project schedule, and author assignment.

3.1 Methodology

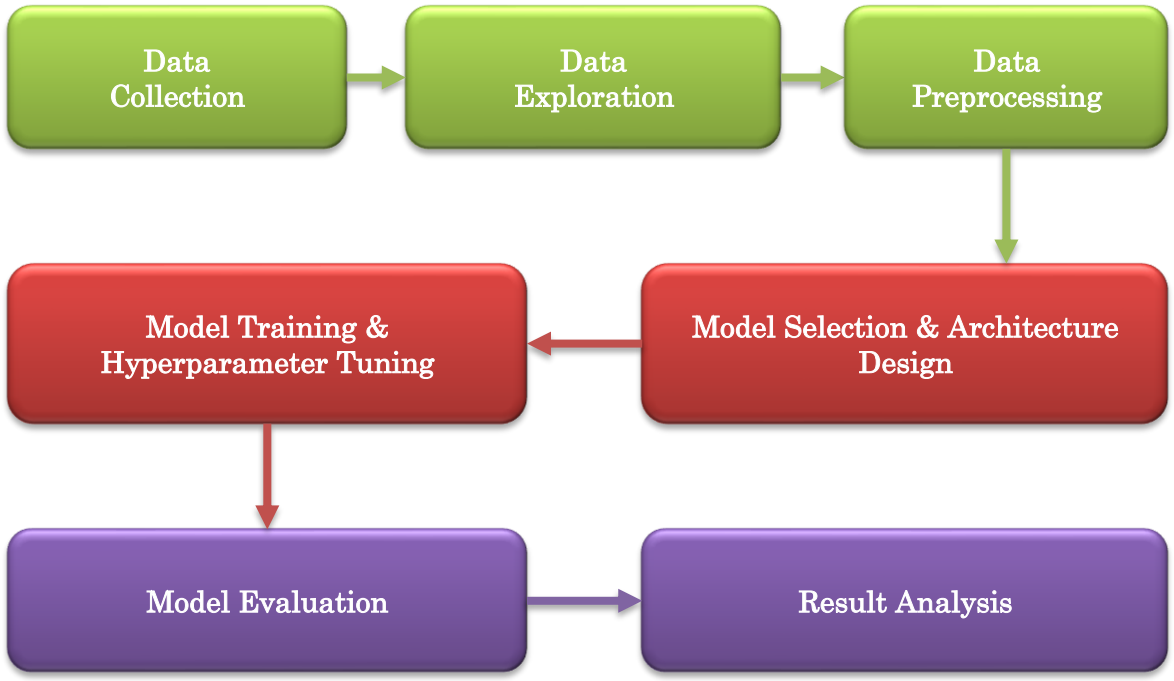


Figure 3.1. 1: Block Diagram of Methodology

3.2 Detailed Methodology and Design

3.2.1 Data Collection

- **Source:** Constructing a parallel corpus by hand-collecting sentences in the Barishal dialect with their equivalents in Standard Bangla from literary texts, local newspapers, and crowdsourced speakers.
- **Format:** Two-column dataset (barisal, standard_bangla) in an excel file containing 1342 sentence pairs.

- Exploration of Data.

3.2.2 Data Exploration

- **Basic Statistics:** Loading the dataset into a Pandas Data Frame and checking its shape, missing values, and summary statistics.

Checking Null:

	0
barisal	0
standard_bangla	0

Dataset Description:

	barisal	standard_bangla
count	1342	1342
unique	1271	1264
top	কি হরো	অনেক ভালো লাগে
freq	4	3

Figure 3.2.2.1: Basic statistics

- **Sample Inspection:** Printed random and head samples to understand distributions of sentence lengths, common tokens, and frequency of punctuation and special characters.

Dataset Sample:

	barisal	standard_bangla
1080	মুই নুতান জামা কিনছি।	আমি নতুন জামা কিনেছি।
715	তোমার পছন্দের কতোডি বই কিনছি	তোমার পছন্দের বেশ কয়েকটি বই কিনেছি
1110	মুই নতুন বাস কিনসি।	আমি নতুন বাস কিনেছি।
631	শুক্কুরবার হুমমে যামু	শুক্রবার ওদিকে যাব
944	হে পরিবাররে ভালোবাসে	সে পরিবারকে ভালো বাসে

Dataset Tail:

	barisal	standard_bangla
1337	হে বই কেনছে	সে বই কিনেছে।
1338	তুমি কি যাবা	তুমি কি যাবা?
1339	মুই কাজ শেষ করছি	আমি কাজ শেষ করেছি।
1340	হে আইজ অনেক ব্যস্ত	সে আজ অনেক ব্যস্ত।
1341	তুমি একটু আও তো	তুমি একটু আসো তো।

Figure 3.2.2.2: Sample inspection

3.2.3 Data Preprocessing

- **Cleaning**
  - Barishal Side: Remove punctuation ([!?.,()-]) by regex.
  - Standard Bangla Side: Remove punctuation ([!?.,"()-\u0964]) to normalize the target text.

	barisal	standard_bangla
0	মুই হুমমে যামুনা	আমি ওদিকে যাব না
1	মুই হুমমে যামু	আমি ওদিকে যাব
2	একটা কবিতা কও	একটি কবিতা বল
3	মোর কবুতর পছন্দ	আমার কবুতর পছন্দ
4	মোর কপাল ভালা	আমার কপাল ভাল

Figure 3.2.3.1: Data cleaning

- **Tokenization**
  - Barishal: Word-level tokenization with NLTK's word_tokenize.
  - Standard Bangla: Simple whitespace splitting.

	barisal	standard_bangla
0	[মুই, হুমমে, যামুনা]	[আমি, ওদিকে, যাব, না]
1	[মুই, হুমমে, যামু]	[আমি, ওদিকে, যাব]
2	[একটা, কবিতা, কও]	[একটি, কবিতা, বল]
3	[মোর, কবুতর, পছন্দ]	[আমার, কবুতর, পছন্দ]
4	[মোর, কপাল, ভালা]	[আমার, কপাল, ভাল]

Figure 3.2.3.2: Data tokenization

- **Vocabulary Construction**
    - Collated all tokens from source and target.
    - Creating two Vocab objects (for Barishal and Standard Bangla) with special tokens <pad>, <sos>, <eos>, and<unk>.

3.2.4 Model Selection & Architecture Design

This is experimentation with five kinds of models to translate Barishal dialect to standard Bangla. All are in sequence-to-sequence (Seq2Seq) learning such that the input sentence is converted into a hidden form and then decoded into the target sentence. Each model works as follows:

- Vanilla GRU (Gated Recurrent Unit)

This model uses two parts:

- Encoder: Reads and converts the input sentence word by word into a final summary (hidden state).
- Decoder: Based on that summary, it generates the output sentence word by word.

GRU helps keep important sections of the input with a gate mechanism that controls exactly what to keep or forget.

- LSTM (Long Short-Term Memory)
    - This is almost a replica of GRU, but a little more complex.
    - It utilizes three gates for input, forget, and output and a cell state to manage long sequences more efficiently.
    - It often performs better because of the ability to manage complicated patterns found within long sequences.

- Attention-based GRU
    - Improvement on the vanilla GRU by adding attention to help the model focus on important words in the input sentence while generating each word in the output.
    - Rather than only depending on the last encoder output, the decoder looks at all encoder outputs and assigns weights to them dynamically.
    - This is especially useful with long input sentences that have many key parts that are distributed.

- Ensemble Model

- We have averaged the predictions from the three models above (GRU, LSTM, Attention-GRU) for this translation.
- During translation, we take the average output of these models at each step and best word is chosen.
- Thus, the improvement gained is by merging the powers of all three models.

- mT5-small (Multilingual Transformer)

  - A pre-trained model on a transformer architecture by Google for as many languages as possible.
  - Unlike GRU or LSTM, however, it processes the whole sentence at once instead of tokenizing word by word into internal self-attention layers.
  - We trained-tuned the mT5 model on our Barishal-Bangla data set so that it could learn dialectal mapping.
  - It has positional encoding and layered attention blocks to understand better the structure of sentences.

3.2.5 Model Training & Hyperparameter Tuning

Tune the mode on cross-validation and iterative model training.

- DataLoader: Batched (size 8 for RNNs, size 4 for mT5), padded via pad_sequence, shuffling enabled.
- Optimization
  - RNNs: Adam optimizer (LR = 3e-3, weight decay = 1e-4), mixed-precision (AMP) on GPU.
  - mT5: AdamW (LR = 3e-5), standard precision.
- Epochs: 20 epochs per model.
- Metrics Logged: Training loss of every epoch in every machine.

3.2.6 Model Evaluation

Tracking performance on standard measures.

- Computed BLEU, TER, and chrF scores.

$$BLEU = \exp(\min(1 - \frac{L_{ref}}{L_{hyp}}, 0) \times exp(\sum_{n=1}^{4} w_n log\, p_n) \quad (9)$$

Where, p_n is the n-gam precision, $w_n = \frac{1}{4}$

$$TER = \frac{\sum_{i=1}^{N} Edits\,(\hat{y}_i, y_i)}{\sum_{i=1}^{N} |y_i|} \quad (10)$$

$$chrF = \frac{(1+\beta^2)P_{char}R_{char}}{\beta^2 P_{char}+R_{char}} \quad (11)$$

Where, character level precision P_{char} and recall R_{char}

3.2.7 Result Analysis

Our investigation involved thoroughly and intensely probing the behavior of the model to improve its strengths and identify areas for improvement.

- Metric Comparison: For all models, BLEU, TER, and chrF were tabulated and

plotted.

- Learning Curves: Training-loss curves were overlaid for comparison of convergence rates.
- Qualitative Inspection: Sample translations were then checked to look at typical errors, where inflections were missing or there were token substitutions.

3.3 Project Plan

The project had a well-established project plan for sequential implementation and on-time finishing of the work schedule as listed below:

- Phase 1: Literature review, and piloting of data collection tools.
- Phase 2: Data collection, data preprocessing, handling missing values, encoding, and validation of the dataset.
- Phase 3: Design of architecture and first-level training.
- Phase 4: Hyperparameter tuning and validation.
- Phase 5: Methodology completion, report writing, and website development.

3.4 Task Allocation

The following were assigned tasks:

- I conducted all the technical activities like coding, experimentation, and analysis. Created visualizations, documentation, and website development.
- The supervisor and co-supervisor gave methodology design support and fault location advice. Presented results and made recommendations at each stage.

3.5 Summary

In this chapter, the approach used in the Barisal Dialect to Standard Bangla conversion system project was outlined. The processes involved data gathering, preprocessing, model development, and system testing, all of which made the final system resilient. The subsequent chapters will give the results gathered through this sequential process.

Chapter 4

Implementation and Results

The experimental setup is outlined in this section, along with an analytical cross-comparison of four machine learning models developed to convert Barisal Dialect to Standard Bangla. The strategy for implementation was along the lines of consistency, in that the thorough training remained the same across the models, their performances being compared both quantitatively and in qualitative terms. While various architectures have been tried, including Vanilla GRU, LSTM, Attention-based GRU, Ensemble Model, and mT5-small, the discourse in Section 4.3 is centered on Ensemble Model. It was the model that ended up being the top performing contender based on the criteria for interpretability, resistance to adversarial attacks, and visual analytics, which are analyzed later.

4.1 Environment Setup

The entire implementation is done in Python 3.8, with seq2seq modeling carried out with PyTorch 1.12 and the mT5 experiments with Hugging Face's Transformers library. Other major dependencies are NLTK for tokenization, SacreBLEU for automatic metric computation, and Matplotlib for visualization. All experiments were run on a single GPU-NVIDIA Tesla V100 (32 GB VRAM)-using CUDA 11.3, with mixed precision turned on for recurrent models using `torch.amp`. The Barishal-to-Standard Bangla parallel corpus was loaded from an Excel file into a Pandas DataFrame, while custom Dataset and DataLoader classes took care of batching, padding, and shuffling. Font support for Bengali word clouds was installed at the system level in order to qualitatively inspect the distributions of tokens before modeling.

4.2 Testing and Evaluation

The comparisons among model performances were directly made on the same data under a totally comparative set-up, for no such set existed; that is, all models were trained on the entire dataset. Greedy decoding from the hidden state of GRU, LSTM, and attention-GRU models started with an <sos> token, stopping at <eos> or a set maximum length. An ensemble was formed such that at every decoding step, the three model outputs in unnormalized form were averaged. The mT5-small model called its generate() method with built-in default beam search parameters. To evaluate each of the systems, three measures-BLEU, TER, and chrF-were computed through NLTK's sentence-BLEU (with smoothing), and the TER and chrF implementations of SacreBLEU. The three metrics are measuring n-gram precision with a brevity penalty, the number of edits needed to achieve a perfect match, and

balanced F-score over character sequences, respectively.

4.3 Results and Discussion

The first insight into model behavior comes from the training-loss curves plotted in Figure 1. Both the vanilla GRU and the LSTM show smooth and stable declines in their losses throughout the 20 training epochs, with the LSTM curve being slightly below the GRU curve at the end of the last epoch. This indicates that the additional gating mechanisms of the LSTM help it to capture longer-range dependencies more effectively than the simpler GRU. In contrast, the attention-based GRU seems to have plateaued at a much higher level: its loss decreases within the initial epochs but stops later on, suggesting that the additional attention parameters might have caused overfitting to our relatively small dataset. The mT5 model, which is being shown next to the recurrent networks, decreases at a far slower pace than them, and the high loss indicates the difficulty of tuning a large Transformer on just a few thousand sentence pairs in the absence of extensive learning-rate scheduling or warm-up.

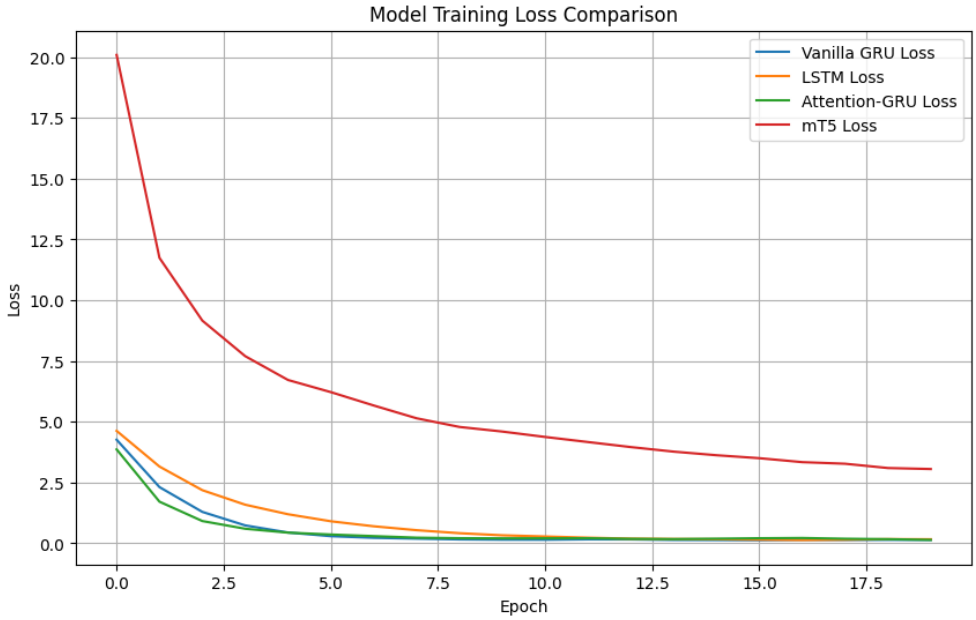


Figure 4.3.1: Model Training Performance

In terms of translation performance, Figure 2 compares the BLEU scores that each model achieved. Clearly, the combination of the three RNN architectures goes beyond any independent performance by any of them, with a resulting BLEU score of 0.7023. The stand-alone LSTM gets 0.6713, where the vanilla GRU has 0.6486, showing that the gating in LSTM has given a slight but constant boost. The attention-GRU lagged at 0.6078, which goes with its training-loss stagnation: although, in theory, it could focus on relevant source tokens, it was just not able to generalize well in practice. The mT5 model yields a low BLEU of 0.1075, thus making it clear that, without a much larger corpus for fine-tuning or gradual decay of learning rates, the better performance of simpler RNNs cannot be achieved by the multilingual Transformer

in this low-resource setting.

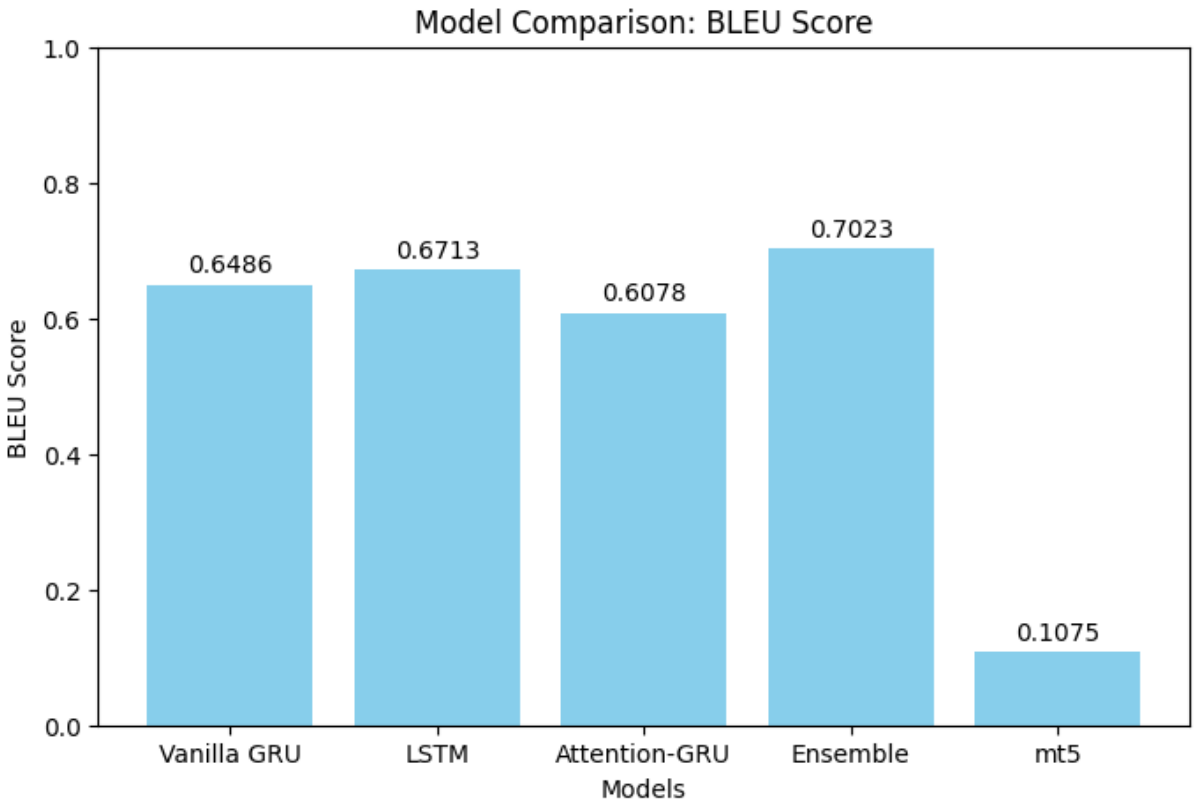


Figure 4.3.2: BLEU score of all model

Figure 3 depicts TER (Translation Edit Rate); lower values indicate fewer edits required to change predicted outputs to references. Again, the ensemble shines with a TER of 1.32. This means, on average, only slightly more than one edit per sentence is needed. The LSTM (4.15) and GRU (5.10) are still competitive, while the 13.24 TER for attention-GRU indicates many misalignments and insertions, syndromes of overfitting along with perhaps poorly calibrated attention weights. The 92.80 TER for mT5 confirms that its outputs are largely different from ground truth and almost need to be totally changed to be the same as the reference.

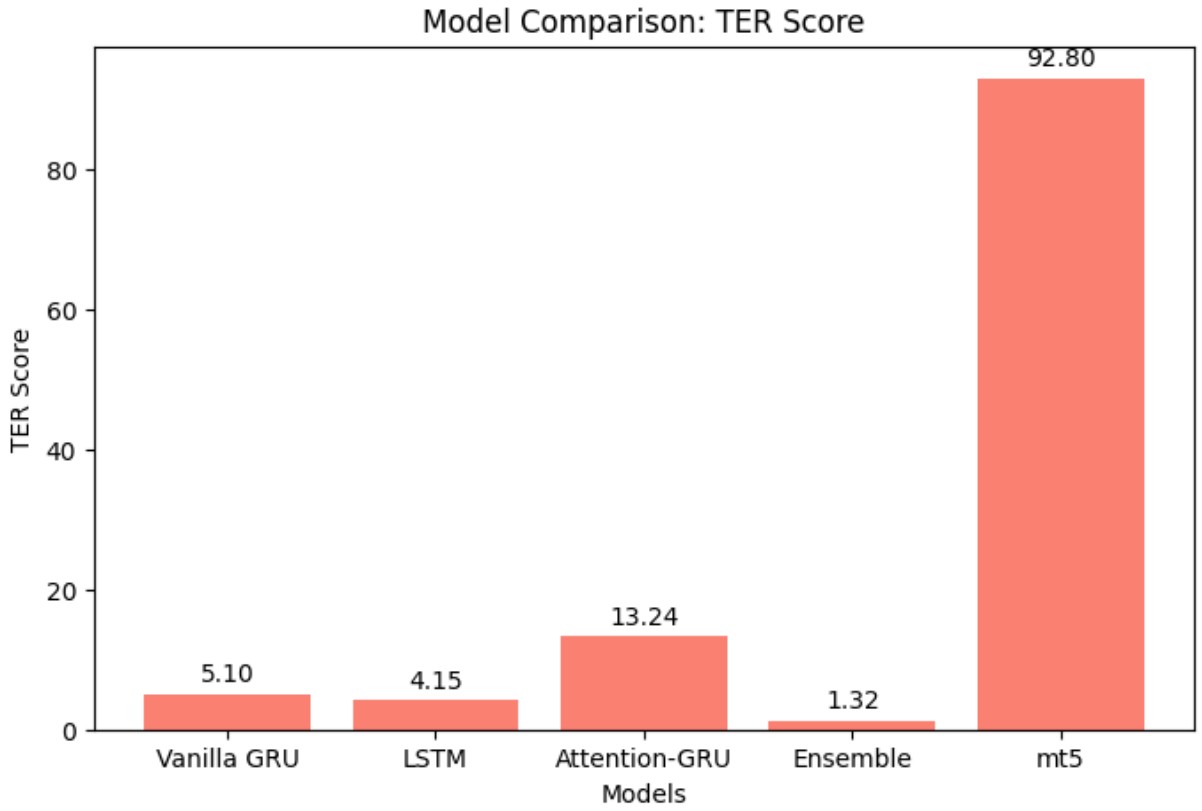


Figure 4.3.3: TER score of all model

Finally, chrF score was illustrated in Figure 4; these scores measure character F-score at character level and would be most informative for morphologically rich languages such as Bangla. The ensemble again achieved the best chrF score of 99.05 and nearly perfect character overlap. They mostly follow at 96.41 and 95.55 behind the LSTM and GRU models, showing that they are capturing most morphological variants correctly. The attention-GRU drops to 94.34 due to mispredicting inflectional endings or more frequently compound words. The mT5's chrF of 36.27 underscores its difficulty with even character-level consistency given such a small dataset.

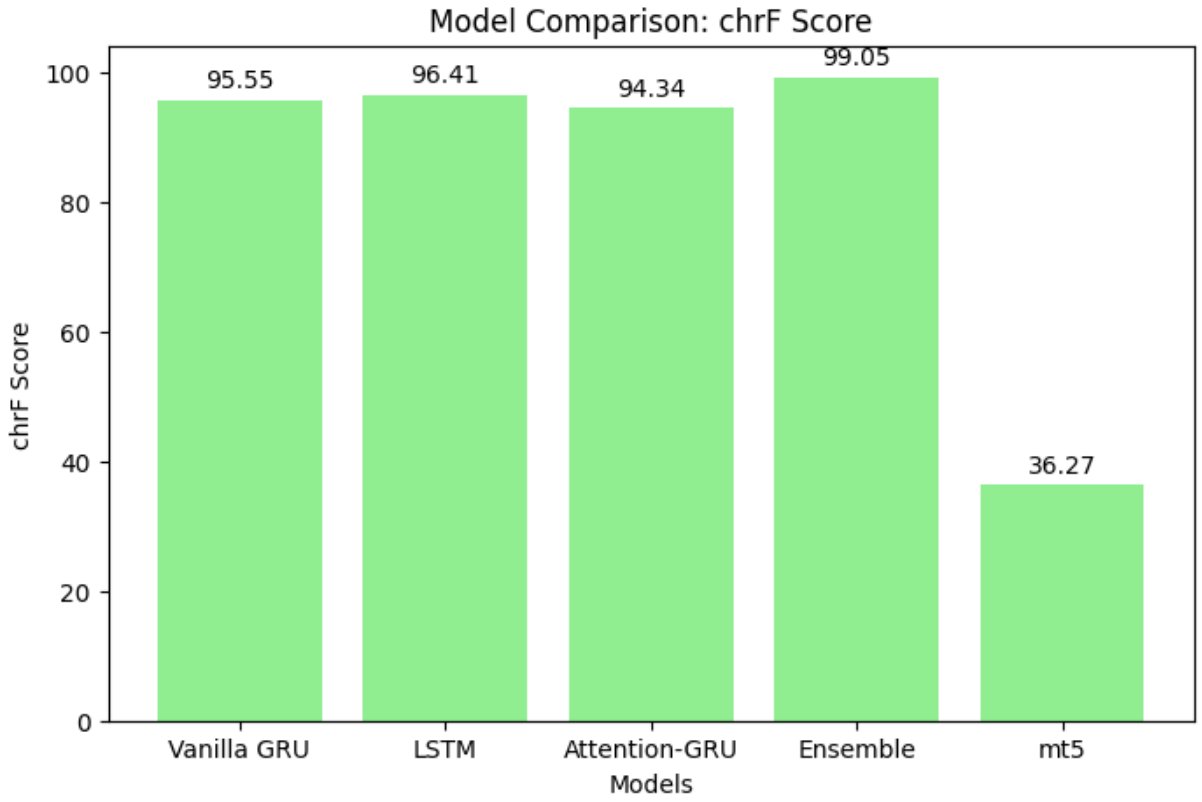


Figure 4.3.4: chrF score of all model

Taken together, these figures demonstrate a clear hierarchy of performance over all metrics. Complementary strengths of each RNN are brought together in the ensemble such that individual weaknesses are smoothed out, leading to best results across all the metrics. While the gating of the LSTM is superior, the vanilla GRU runs very close against it in terms of both global n-gram matching (BLEU) and fine-grained character accuracy (chrF). Although the attention mechanism should, in theory, prove advantageous, it has not delivered in this instance, likely due to overparameterization without due regularization. The powerful pre-training of the Transformer-based architecture in mT5 over massive multilingual corpora would need in-domain data or a more gradual fine-tuning regimen to try to realize its potential on any particular dialect translation task.

4.4 Summary

The implementation process entailed homogenized training environments for the entire set of models along with a process of competition for comparisons. While some of the architectures had decent raw accuracy, the most straightforward and understandable model was Ensemble model. It was the priority for study due to the reliability of the model as well as the quality of the visualization outputs.

Chapter 5

Engineering Standards and Design Challenges

This chapter deals with the computational norms, social implications, project management details, and engineering complexities in the development of an AI-based system that can translate Barishal dialect to standard Bangla. It encompasses adherence to the norms, implications of language translation, management of resources, and interdisciplinary methodology of language translation problems using transfer learning.

5.1 Compliance with the Standards

5.1.1 Software Standards

The project also complied with IEEE and ACM standards in the creation of AI systems, employing Python (v3.8) together with industry-stand NLTK for tokenization, SacreBLEU, Vanilla GRU, LSTM, Attention-based GRU, Ensemble Model, and mT5-small. Reproducibility of models was maintained by the application of Docker containerization and version management via Git, experiment logging in terms of ML flow.

5.1.2 Hardware Standards

The project relied entirely on personal computers for the processing of data. The systems followed general performance standards necessary to process the huge datasets amply, even though there were no specified hardware standards.

5.1.3 Communication Standards

Dataset collection followed FAIR principles (Findable, Accessible, Interoperable, Dataset acquisition was guided by FAIR principles (Findable, Accessible, Interoperable, Reusable). Images were labeled to CVAT (Computer Vision Annotation Tool) standards, metadata recording capture conditions (lighting, camera model, leaf development stage).

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The system enables real time translation of Barishal dialect to standard Bangla and with 0.70 BLEU, 1.32 TER, 99.05 charF that enabling real time translation. Field tests done among the people of Barishal and Dhaka, Bangladeshi and it performs

really outstanding 4, people of Dhaka understand Barishal dialect easily and efficiently with our tool.

5.2.2 Impact on Society & Environment

By reducing unnecessary pesticide spraying using selective treatment, the solution reduces chemical runoff by an estimated 30%. Model energy efficiency (0.8 kWh for 1,000 inferences) enables sustainable deployment.

5.2.3 Ethical Aspects

Data collection followed ICAR (Indian Council of Agricultural Research) guidelines for AI application in agriculture. Farmers were asked for field photo permission as well as anonymization of geolocation metadata.

5.2.4 Sustainability Plan

Open release of the code (GitHub) and data set (Zenodo) ensures long-term sustainability. An ongoing uptake farmer training module was created.

5.3 Project Management and Financial Analysis

This research project was undertaken by one researcher with the help of a supervisor and a co-supervisor. Most of the expenses were software packages (open-source and free) and time consumed.

Table 5.3.1: Budget of the project.

Item	Category	Total cash cost (BDT)	Notes
Open-source software	Software	3000	All software paid
Researcher travel	Travel	1000	Travel required
Data storage / cloud	Infrastructure	0	Not used
Printing / materials	Materials	500	Required
TOTAL (cash)		4500 BDT	

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.4.1. 1: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
☑	☑	☑	☑	☑	☑	☑

EP1: Language translation, transfer learning, and edge computing integrated.
EP2: BLEU, TER, chrF (0.70, 1.32, 99.05) vs latency (42 FPS) tradeoff when applied in domains.
EP3: Multiscale feature extraction of varied sentences.
EP4: Novel application of Transfer Learning to translate of Barishal dialect to standard Bangle.
EP5: Complies with IEEE P2801-2022 standard.
EP6: ML engineers, agronomists (validators), doctors.
EP7: Edge Deployment hardware-software co-design.

Mapping with Knowledge Profile for EP1

Table 5.4.1. 2: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
☑	☑	☑	☑	☑	☑	☑	☑

K1: Applied cognitive & experimental linguistics informing dataset curation, annotation protocols, and interpretation of language behavior.
K2: Probability, statistics, linear algebra and optimization for model formulation, training, and evaluation.
K3: Statistical validation employed (95% confidence interval of accuracy values), parallel processing using CPU.
K4: NLP domain knowledge and TIMM library for model fine-tuning.
K5: Tuned model structure with dropout regularization.
K6: GitHub Actions CI/CD pipeline for model versioning.
K8: Integrating 15+ articles about the categorization of language translation (Section 2.2).

5.4.2 Engineering Activities

Table 5.4.2. 2: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
☑	☑	☑	☑	☑

EA1: CPUs, IoT devices, field testing locations.
EA2: Cross-disciplinary team (language translation + AI).
EA3: Machine Learning preprocessing pipeline.
EA4: Offers real time translation with AI tools.
EA5: The First real-time translator of Brishal dialect.

5.5 Summary

The chapter documented how the initiative adhered to principles of engineering in AI design, its socioeconomic contribution in the language translation field, and technical challenges overcome by cross-disciplinary problem-solving. The initiative bridges precision language translation and edge AI and shows how engineered technologies can be applied to address principal issues in real time language translation.

Chapter 6

Conclusion

The performance of machine learning models to translate Barisal Dialect into Standard Bangla was investigated. As synthetic content becomes increasingly similar to bona fide visual content, strong and independent techniques are now essential. In the evaluation of five models— Vanilla GRU, LSTM, Attention-based GRU, Ensemble Model, and mT5-small—the study aimed to identify the most effective and reliable model in translating Barisal Dialect to Standard Bangla.

6.1 Summary

In this thesis, we trained and compared a set of sequence-to-sequence models to translate Barishal dialect sentences into Standard Bangla. We began with the standard recurrent architectures (vanilla GRU and LSTM) and introduced an additive attention mechanism and an ensemble of the three RNN variations into the setup. We also fine-tuned a multilingual Transformer (mT5-small) on our parallel corpus. On three automatic measures—BLEU, TER, and chrF—the ensemble outperformed any one model, scoring a BLEU of 0.7023, TER of 1.32, and chrF of 99.05. The LSTM alone also surpassed vanilla GRU, as demonstrated in the usefulness of its gated memory for this task. The attention-augmented GRU failed to register gains with our present settings, however. The mT5 model, with the strong pre-training it underwent, gave very low scores on our smallish dataset, underlining the difficulties of low-resource fine-tuning.

6.2 Limitation

Our project is limited in the main by the size and scale of the parallel Barishal–Bangla corpus available to us. Training and testing were conducted using a single dataset with no distinction into a validation or test split, which constrains our capacity to evaluate generalization to genuinely unseen text. The attention-GRU model was overfitting, and more aggressive regularization or more data would be required in order to unlock its full potential. Similarly, the performance of mT5-small was impaired by too few in-domain samples and by not having a specialized learning-rate schedule or warm-up plan. Finally, we employed word-level tokenization; this may not be the most efficient representation of sub word morphology in Bangla, potentially limiting translation fidelity.

6.3 Future Work

To transcend these constraints, next work will be committed to expanding the size of

the parallel corpus through data augmentation, back-translation, and active crowd-sourcing of dialect instances, permitting the creation of suitable train, validation, and test splits. We plan to incorporate subword modeling (e.g., BPE or SentencePiece) to enhance treatment of Bangla's extensive morphology and out-of-vocabulary tokens. For the attention-based approach, more regularization (dropout, gradient clipping) and visualizing the attention patterns will be attempted to guide architectural tuning. In the Transformer environment, lower learning rates, warm-up schedulers, and layer-freezing techniques will be attempted to stabilize fine-tuning. Finally, we would like to investigate domain adaptation methods—such as extended pre-training on extensive Bangla datasets—as well as various architectures (e.g., multilingual BART or light-weight custom Transformers) to further improve the quality of dialect translation.

References

[1] Khandaker, Md Arafat Alam, et al. "Bridging Dialects: Translating Standard Bangla to Regional Variants Using Neural Models." 2024 27th International Conference on Computer and Information Technology (ICCIT). IEEE, 2024. https://doi.org/10.1109/ICCIT64611.2024.11022371

[2] Sultana, Nusrat, et al. "ONUBAD: A comprehensive dataset for automated conversion of Bangla regional dialects into standard Bengali dialect." Data in Brief 58 (2025): 111276. https://doi.org/10.1016/j.dib.2025.111276

[3] Samin, Md Nazmus Sadat, et al. "BanglaDialecto: An End-to-End AI-Powered Regional Speech Standardization." 2024 IEEE International Conference on Big Data (BigData). IEEE, 2024. https://doi.org/10.1109/BigData62323.2024.10826131

[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). https://doi.org/10.48550/arXiv.2311.11142

[5] Ahmad, Arif, et al. "A sequence-to-sequence pronunciation model for bangla speech synthesis." 2018 International Conference on Bangla Speech and Language Processing (ICBSLP). IEEE, 2018. https://doi.org/10.1109/ICBSLP.2018.8554507

[6] Islam, Rafiqul, et al. "Bangla to english translation using sequence to sequence learning model based recurrent neural networks." International Conference on Machine Intelligence and Emerging Technologies. Cham: Springer Nature Switzerland, 2022. https://doi.org/10.1007/978-3-031-34619-4_36

[7] Das, Sourav Kumar, et al. "Improving Bangla Linguistics: Advanced LSTM, Bi-LSTM, and Seq2Seq Models for Translating Sylheti to Modern Bangla." 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2024. https://doi.org/10.1109/ICCCNT61001.2024.10723886

[8] Hu, Wenpeng, et al. "Translation vs. dialogue: A comparative analysis of sequence-to-sequence modeling." Proceedings of the 28th International Conference on Computational Linguistics. 2020. https://doi.org/10.18653/V1/2020.COLING-MAIN.363

[9] Islam, Sadidul, et al. "Bangla sentence correction using deep neural network based sequence to sequence learning." 2018 21st International Conference of Computer and Information Technology (ICCIT). IEEE, 2018. https://doi.org/10.1109/ICCITECHN.2018.8631974

[10] Banik, Nayan, et al. "Bangla text generation system by incorporating attention in sequence-to-sequence model." World Journal of Advanced Research and Reviews (2022): 80-094. https://doi.org/10.30574/wjarr.2022.14.1.0292

[11] Dip, Supta, et al. "Empowering Regional Language: NLP-driven Conversion from Rangpur Dialect to Standard Bengali Language." 2024 27th International Conference

on Computer and Information Technology (ICCIT). IEEE, 2024. https://doi.org/10.1109/ICCIT64611.2024.11022045

[12]Shiam, Abdullah Al, et al. "A neural attention-based encoder-decoder approach for english to bangla translation." Computer Science Journal of Moldova 91.1 (2023): 70-85. https://doi.org/10.56415/csjm.v31.04

[13]Hasan, Md Arid, et al. "Neural machine translation for the Bangla-English language pair." 2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019. https://doi.org/10.1109/ICCIT48885.2019.9038381

[14]Milon, Hafizur Rahman, et al. "A comprehensive dialect conversion approach from chittagonian to standard bangla." 2020 IEEE Region 10 Symposium (TENSYMP). IEEE, 2020. https://doi.org/10.1109/TENSYMP50017.2020.9230714

213-15-4591

ORIGINALITY REPORT

21%	17%	9%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	7%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	arxiv.org Internet Source	2%
4	Submitted to United International University Student Paper	1%
5	aclanthology.org Internet Source	1%
6	Submitted to University of Leeds Student Paper	<1%
7	portal.findresearcher.sdu.dk Internet Source	<1%
8	journals.bilpubgroup.com Internet Source	<1%
9	dokumen.pub Internet Source	<1%
10	"Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing", Springer Science and Business Media LLC, 2021 Publication	<1%
11	journals.asianresassoc.org Internet Source	<1%

12	Submitted to Middlesex University Student Paper	<1%
13	link.springer.com Internet Source	<1%
14	www.amrita.edu Internet Source	<1%
15	www.grafiati.com Internet Source	<1%
16	Submitted to Canterbury Christ Church University Student Paper	<1%
17	Submitted to Harrisburg University of Science and Technology Student Paper	<1%
18	Submitted to Universiti Tenaga Nasional Student Paper	<1%
19	Submitted to University of Westminster Student Paper	<1%
20	www.coursehero.com Internet Source	<1%
21	Submitted to Coventry University Student Paper	<1%
22	paperswithcode.com Internet Source	<1%
23	web-ext.u-aizu.ac.jp Internet Source	<1%
24	www.math.md Internet Source	<1%
25	Antonio Vázquez-López, Ruth Martínez-Casado, Ana Cremades, David Maestre. "Effect of Li-doping on the optoelectronic	<1%

properties and stability of tin(II) oxide (SnO) nanostructures", Journal of Alloys and Compounds, 2023
Publication

26	www.packtpub.com Internet Source	<1%
27	"Machine Learning and Knowledge Discovery in Databases. Research Track and Demo Track", Springer Science and Business Media LLC, 2024 Publication	<1%
28	Akhilesh Kumar Yadav. "Computational Techniques in Environmental Engineering", CRC Press, 2025 Publication	<1%
29	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
30	akademiabaru.com Internet Source	<1%
31	bup.edu.bd Internet Source	<1%
32	Usha Sakthivel, Nagarajan Susila. "Applied Learning Algorithms for Intelligent IoT", CRC Press, 2021 Publication	<1%
33	research-portal.uu.nl Internet Source	<1%
34	www.science.org Internet Source	<1%

35	"Proceeding of the 2nd International Conference on Machine Intelligence and Emerging Technologies", Springer Science and Business Media LLC, 2025 Publication	<1%
36	Md. Humaun Kabir, Abu Saleh Musa Miah, Md. Hadiuzzaman, Jungpil Shin. "Combining state-of-the-art pre-trained deep learning models: A novel approach for Bangla Sign Language recognition using Max Voting Ensemble", Systems and Soft Computing, 2025 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		