MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Comparative Study of Encoder-Decoder Architectures with Attention Mechanisms

Guided By :
Dr.Sunita S. Barve

Presented By :
Sakshi Aher(202201040089)
Rohit Dahale(202201070052)
Akhilesh Ukey(202201040136)

CONTENT

- Paper summary:Aim,Objectives,Problem statements,Methodology
- Model diagrams and architecture
- Dataset description
- Metric-wise performance comparison
- Graphs (training curves, attention maps)
- Final analysis table and discussion

Research Paper summary

Title of Research paper :

Automatic Grammar Error Correction Model Based on Encoder-Decoder Structure for English Texts

Aim:

To develop a grammar error correction (GEC) model for English texts that improves accuracy and efficiency by leveraging an enhanced encoder-decoder architecture, addressing common grammar issues with minimal manual rule intervention.

Objectives:

- Propose a deep learning-based encoder-decoder model for automatic grammar correction.
- Integrate attention mechanisms and BiGRU to improve accuracy.
- Train and evaluate the model on standard grammar correction datasets.
- Enhance generalization and semantic understanding of corrected text.

Methodology

Model Architecture
A Transformer-based encoder-decoder framework was employed. The encoder utilized the pre-trained UL2 model for rich contextual embeddings, while the decoder followed the standard Transformer design with self-attention and cross-attention layers.
Tokenization & Preprocessing
WordPiece tokenization was applied to input and output sequences. Special tokens like <CLS>, <SEP>, and <PAD> were introduced to format the input structure effectively.
Input Representation
Contextualized embeddings from UL2 were used as input. The decoder received shifted target sequences to facilitate autoregressive learning during training.

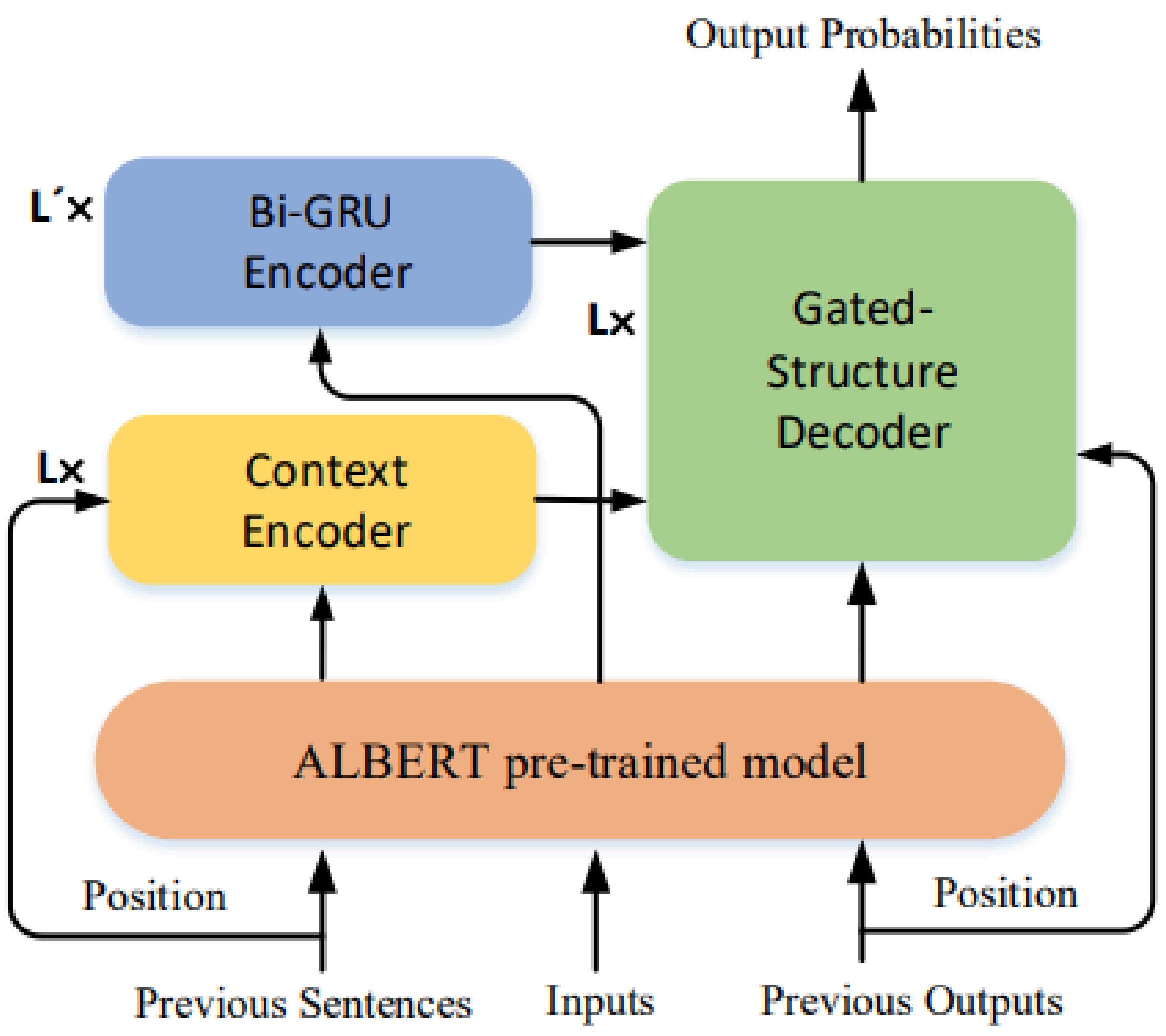


Figure 1. The structure of DCIM

Methodology

Training Strategy

Teacher forcing was used to guide decoder learning, and cross-entropy loss was employed as the objective function for optimization.

Dataset

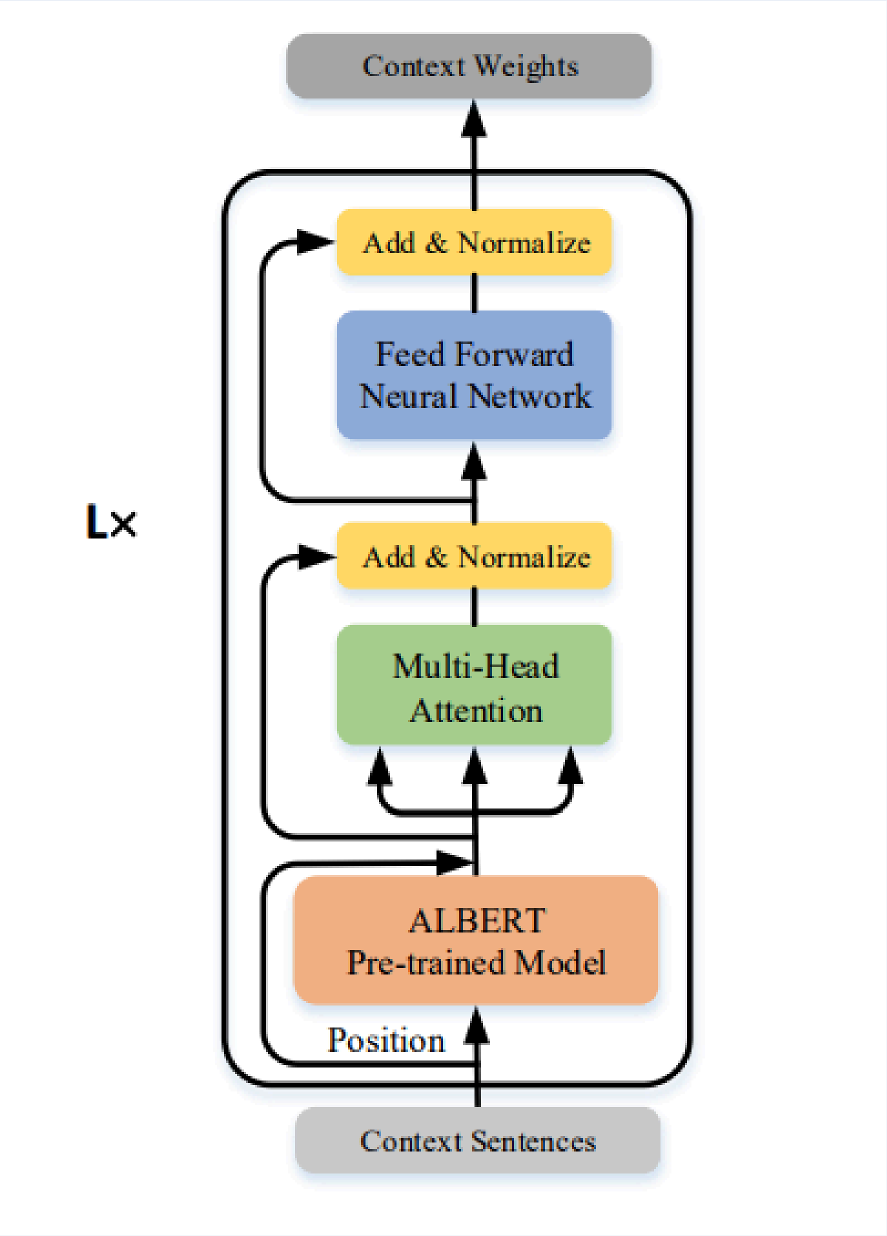
The model was trained on domain-specific parallel corpora, including CPL-30H and FLS, which provided aligned input-output sentence pairs.

Evaluation Metrics

Model performance was evaluated using Precision, Recall, F1-Score, and GLEU to assess both token-level accuracy and overall sequence quality.

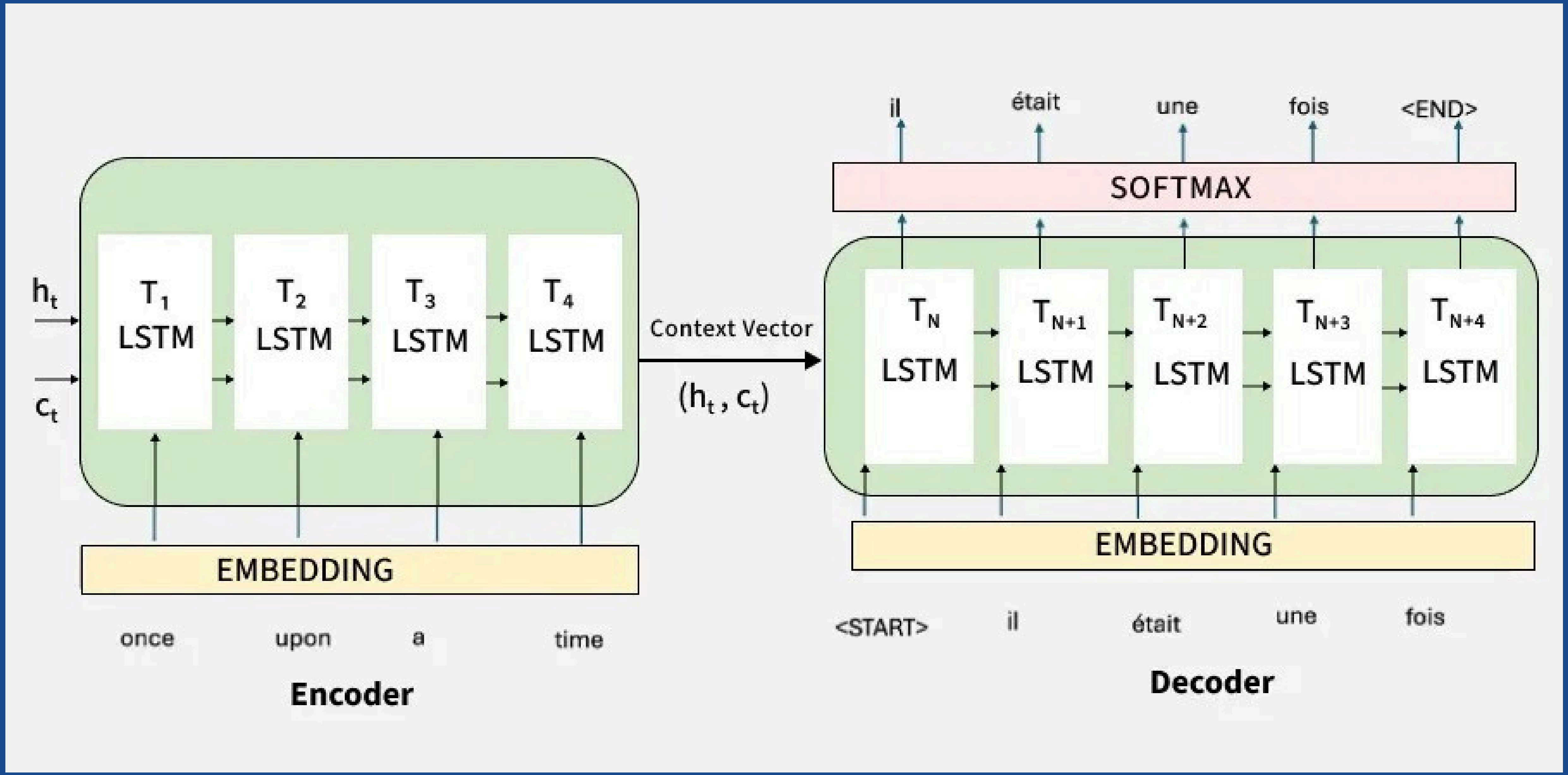
Inference

Beam search was applied during inference to generate fluent and accurate output sequences by considering multiple decoding paths.



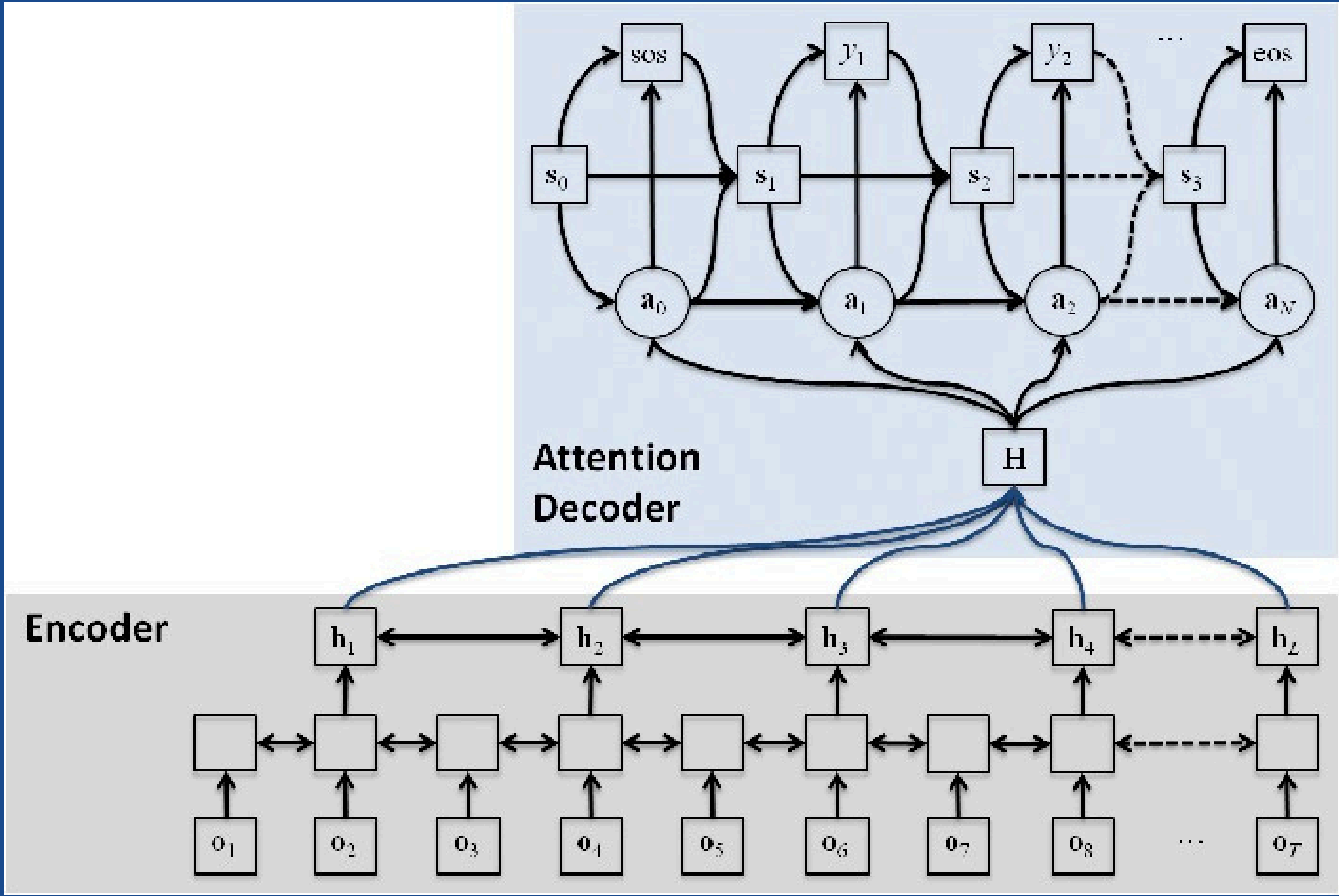
Model diagrams & Architecture

LSTM Without Attention



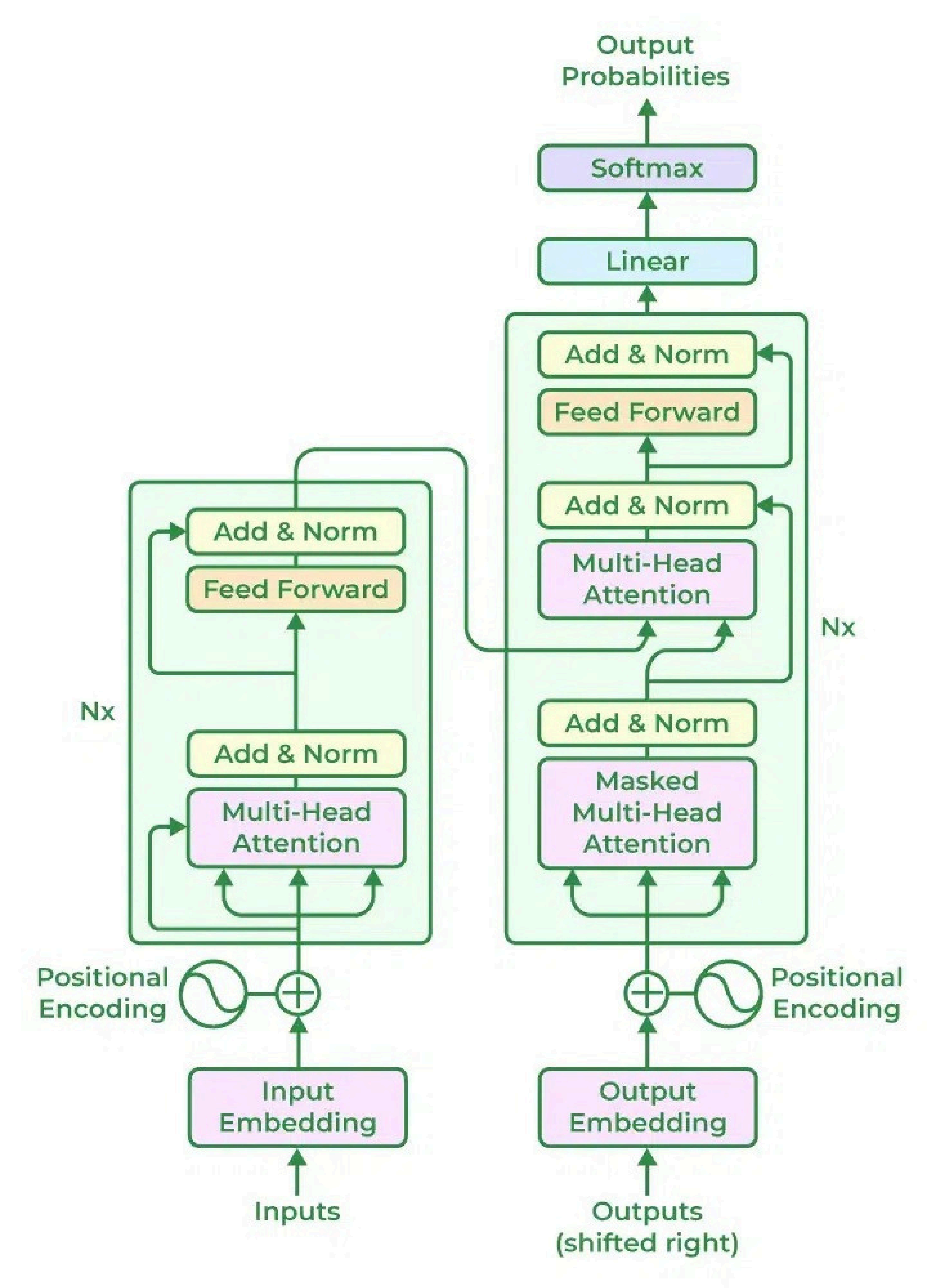
Model diagrams & Architecture

LSTM + Attention



Model diagrams & Architecture

Attention Layer in Transformer



Dataset description

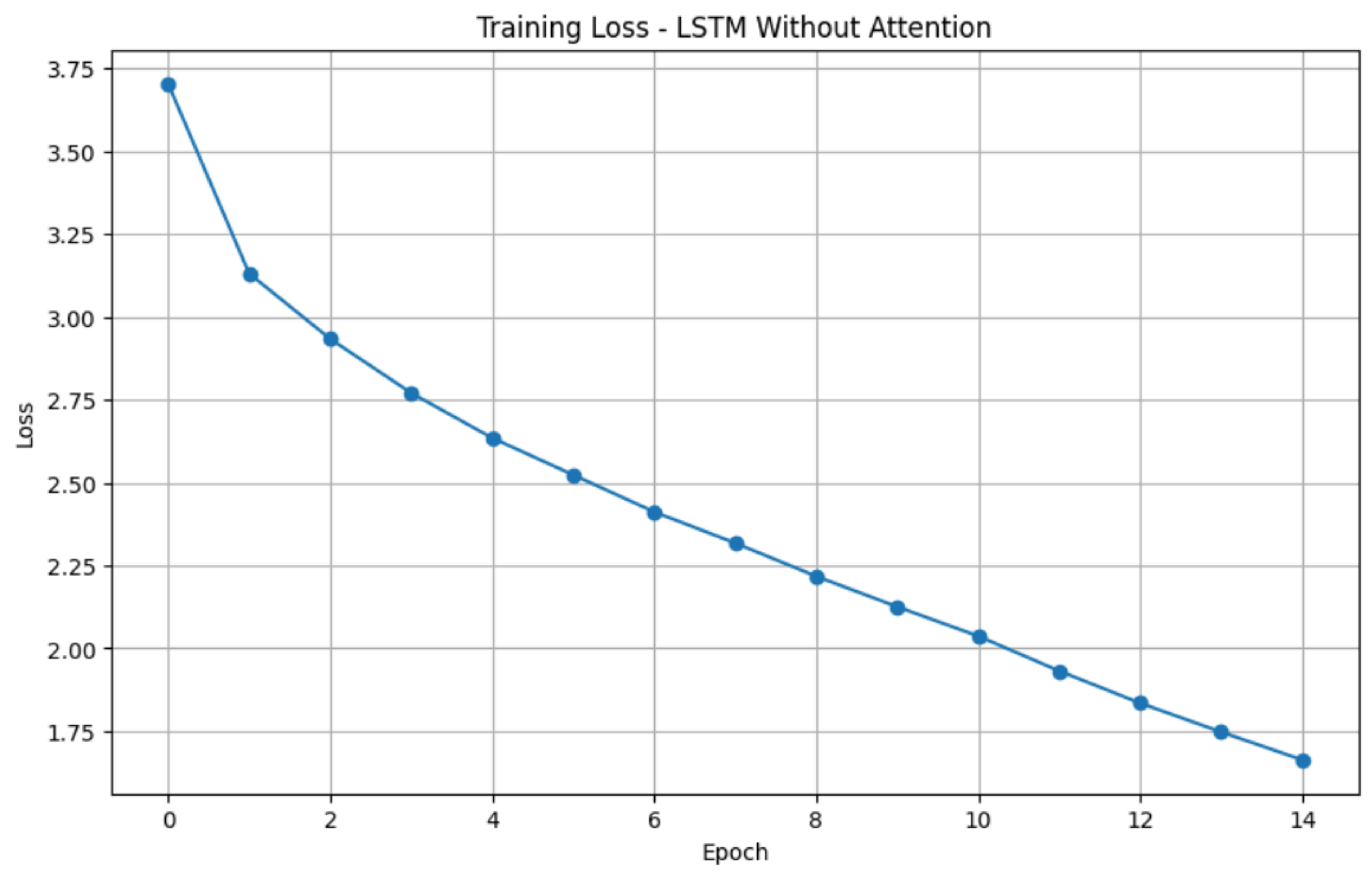
- Source: https://www.kaggle.com/datasets/yupiter/grammar-correction
- Data Type: Parallel sentences with grammatical mistakes ("wrong") and their corrected versions ("correct")
- Size: ~17,000 sentence pairs
- Split: 80% train / 10% validation / 10% test

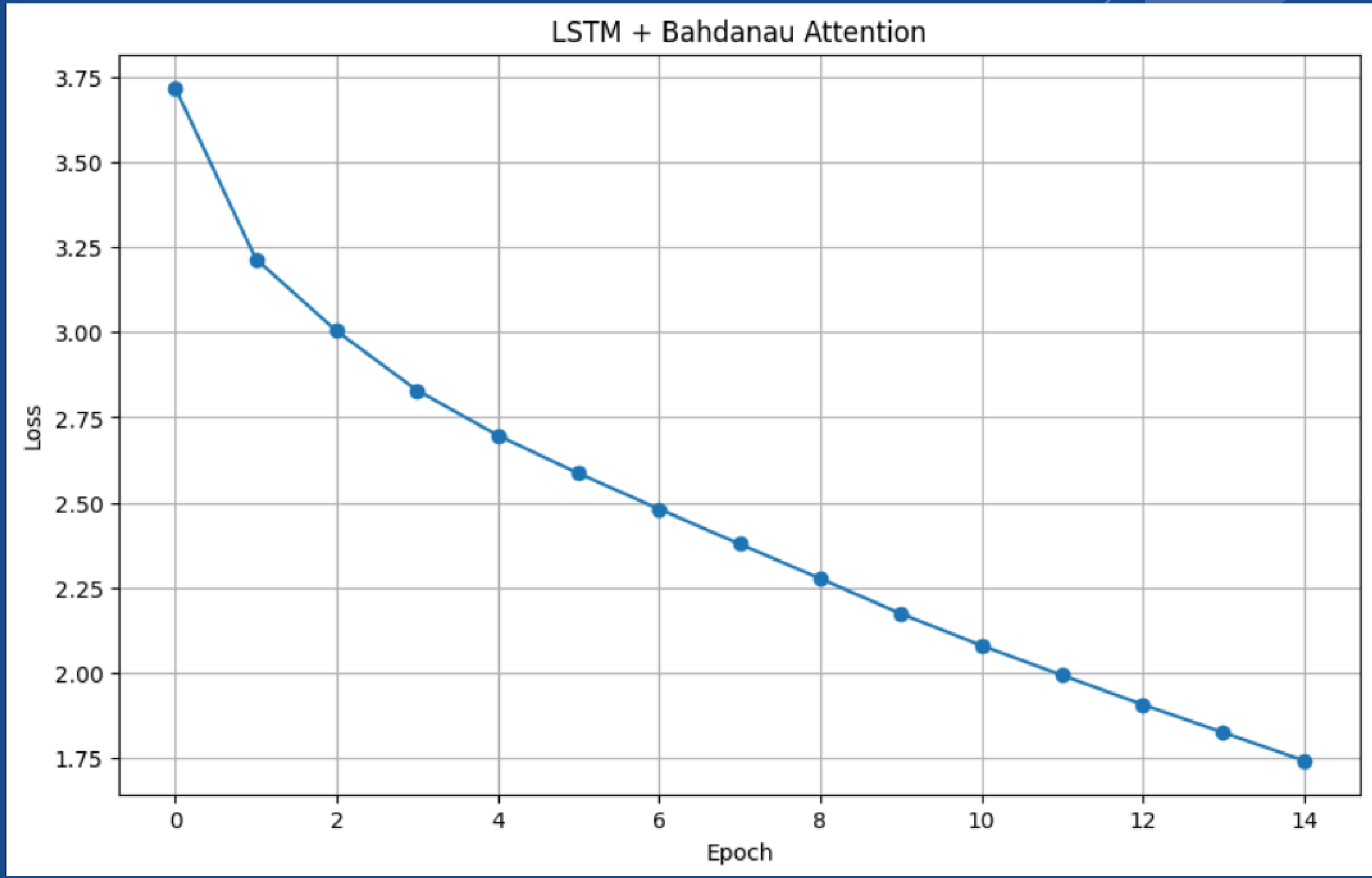
Metric-wise performance comparison

MODEL COMPARISON RESULTS:

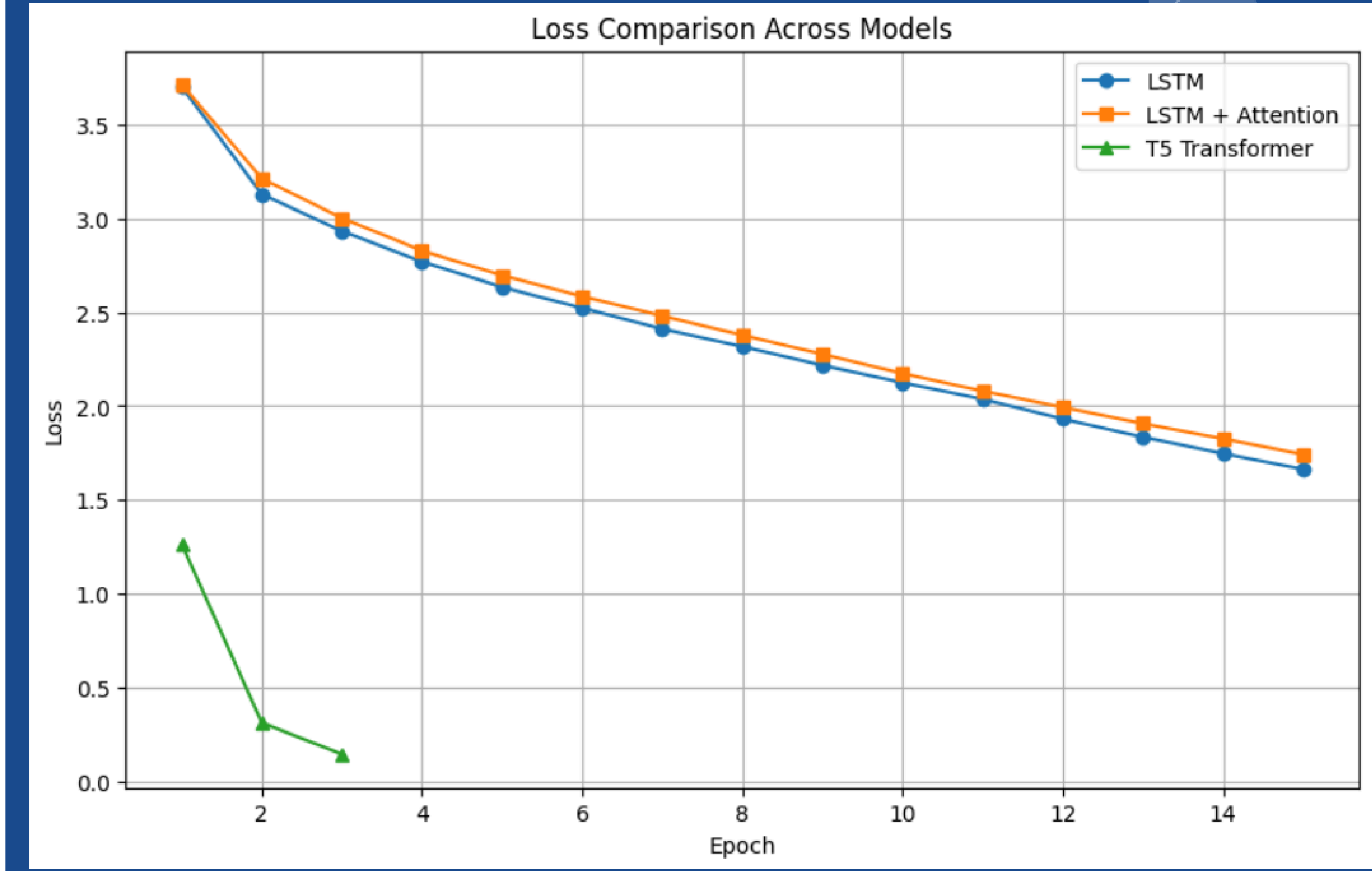
Model	BLEU	METEOR	ROUGE-L	Train Time (s)	Inference Time (s/sample)
LSTM	0.9225	0.9457	0.9410	122.45	0.1337
LSTM + Attn	0.8899	0.9207	0.9088	695.5	0.2097
T5 Transformer	0.8627	0.9680	0.9555	314.6	0.1385

Graphs (training curves, attention maps)





Graphs (training curves, attention maps)



Conclusion

- Encoder-decoder models effectively learn grammar corrections from paired data
- Attention mechanisms significantly improve correction quality
- Transformer achieves highest accuracy but requires more training time
- LSTM + Attention offers the best trade-off between performance and speed

Thank You!!!