Comparative Study of Sequence-to-Label Models with Attention Mechanisms for Sentiment Classification

Jakkula Adishesh Balaji

July 2025

Abstract

In this report, we present a comprehensive evaluation of four sequence-to-label models (RNN, LSTM, BiRNN, BiLSTM) trained on a sentiment classification task. Each base model is evaluated both with and without four attention mechanisms: Bahdanau, LuongDot, LuongGeneral, and LuongConcat. Using GloVe embeddings, consistent preprocessing, and controlled hyperparameters, we analyze the effect of model architecture and attention on performance.

1 Data Preprocessing

- Datasets were sourced from HuggingFace, ignoring the unsupervised set.
- Text cleaning was done using Python's `re` module.
- A vocabulary was created from the cleaned data.
- Words were mapped to 100-dimensional GloVe embeddings.

2 Dataset Splitting and Loader

- Combined train and test datasets and split with a 70:10:20 ratio for train:val:test.
- Implemented a custom PyTorch `Dataset` class.
- Used `DataLoader` for efficient batching.

3 Training Loop

- Metrics tracked: accuracy, loss, precision, recall, and F1 score.
- Results logged to `../logs/`.
- Early stopping with patience of 3.
- Best model saved based on validation accuracy.
- Training/validation plots saved to `../plots/`.

4 Model Architectures

4.1 Vanilla RNN

- Embedding initialized with GloVe (trainable).
- RNN layer processes input sequence.
- Dropout applied to final hidden state.
- Optional attention using final hidden state.
- Output: classification logits.

4.2 Vanilla LSTM

- Embedding with GloVe (trainable).
- LSTM layer with hidden and cell states.
- Dropout applied.
- Attention (if present) uses final hidden state.
- Output:
  - With attention: $[\text{context}; \text{hidden}]$.
  - Without attention: $\text{hn}[-1]$.

4.3 Bidirectional RNN

- BiRNN processes sequences in both directions.
- Final hidden states are concatenated.
- Attention combines bidirectional outputs and hidden states.
- Output:
  - With attention: $[\text{context}; \text{forward hn}; \text{backward hn}]$.
  - Without attention: $[\text{forward hn}; \text{backward hn}]$.

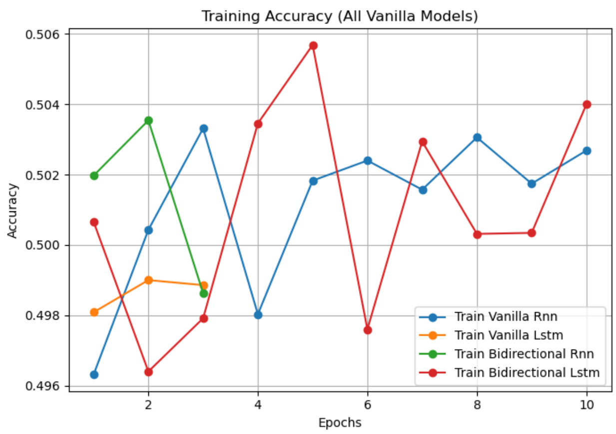
4.4 Bidirectional LSTM

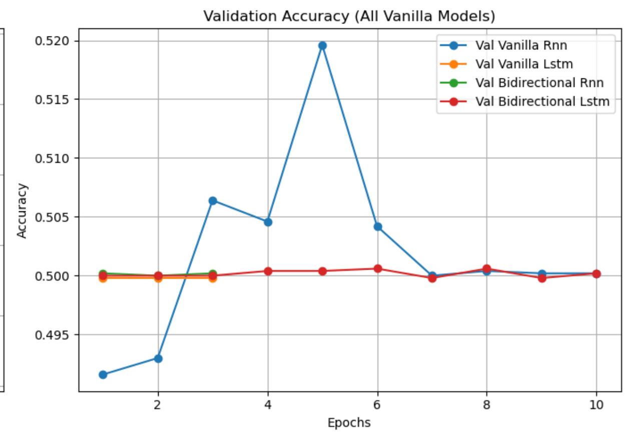
- BiLSTM captures bidirectional context.
- Dropout is applied.
- Attention combines bidirectional outputs and hidden states.
- Output:
  - With attention: $[\text{context}; \text{forward hn}; \text{backward hn}]$.
  - Without attention: $[\text{forward hn}; \text{backward hn}]$.

5 Baseline Models Evaluation (No Attention)

All four base models were trained with the same hyperparameters:

- Loss: CrossEntropyLoss
- Optimizer: Adam ($lr = 10^{-2}$)
- Epochs: 10 (with early stopping)

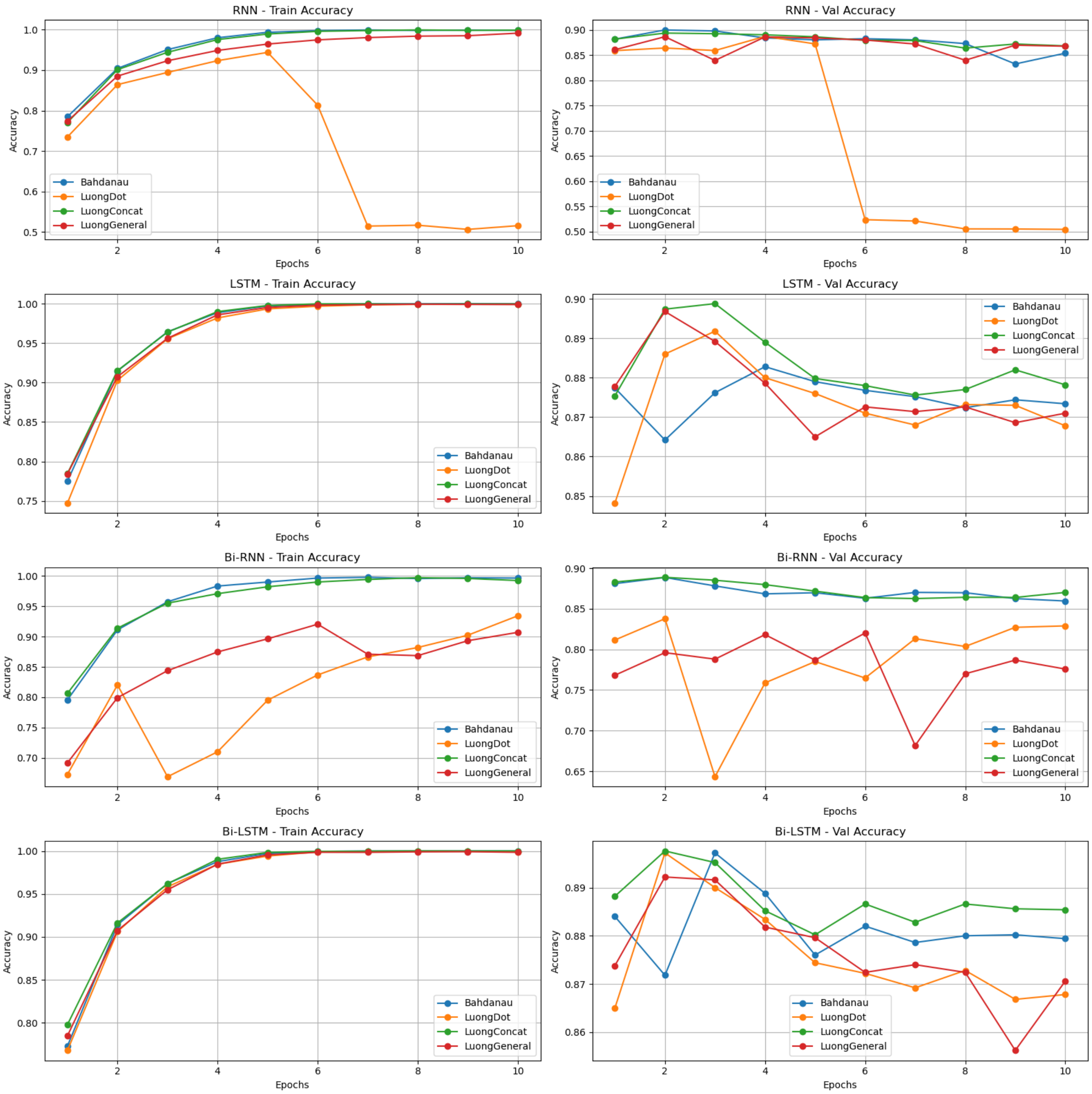




6 Attention Models Evaluation

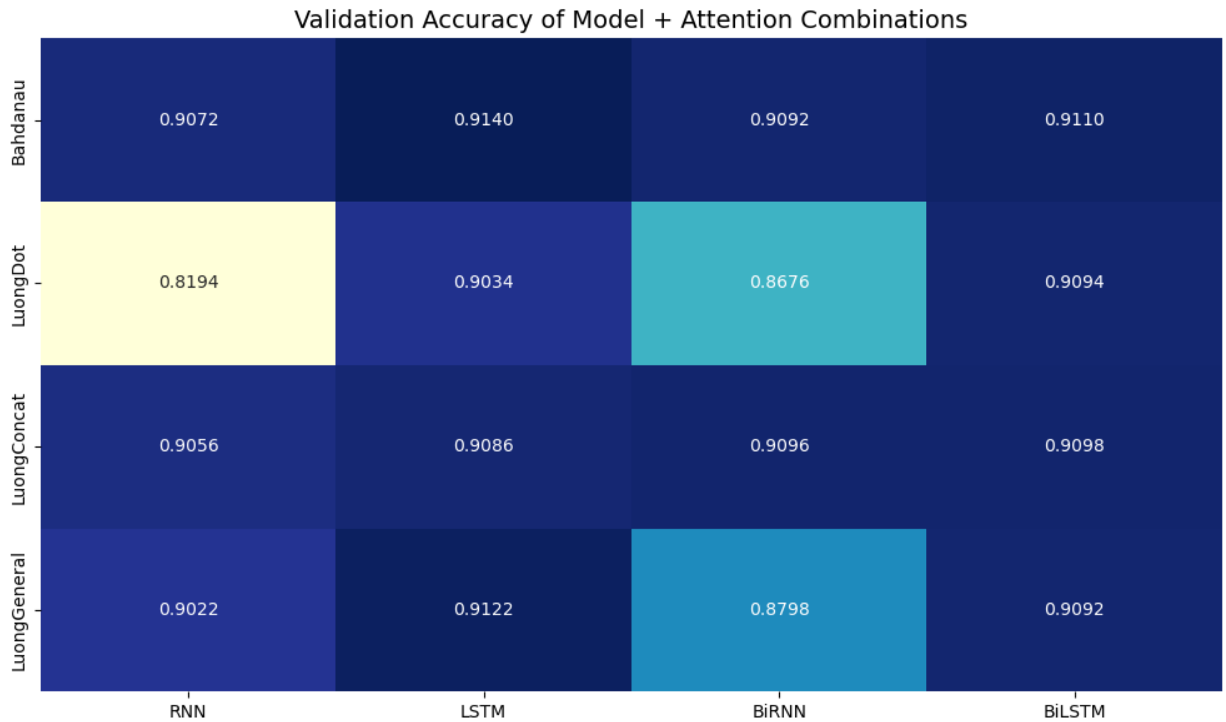
Each base model was trained with all four attention mechanisms.

- Loss: CrossEntropyLoss
- Optimizer: Adam ($lr = 10^{-3}$, weight decay $= 10^{-5}$)
- Epochs: 10 (early stopping)



7 Validation Accuracy Score Table

We plot a heatmap to compare validation accuracies across all 16 model-attention combinations.



8 Fine-Tuning Selected Models

Based on initial performance, the following combinations were fine-tuned:

- **Bidirectional LSTM:** Bahdanau, Luong Concat
- **Vanilla LSTM:** Bahdanau, Luong General

9 Best Model Performance

The best performing configuration was:

- **Model:** LSTM + Luong General Attention
- **Train Loss:** 0.0809
- **Train Accuracy:** 97.17%
- **Validation Loss:** 0.3041
- **Validation Accuracy:** 90.44%
- **Precision:** 0.9047
- **Recall:** 0.9044
- **F1 Score:** 0.9044

Final Test Set Evaluation

Metric	Value
Loss	0.3315
Accuracy	89.85%
Precision	0.8988
Recall	0.8985
F1 Score	0.8985

Conclusion

We observe that the inclusion of attention mechanisms significantly improves model performance. Notably, LSTM with Luong General Attention achieves the highest validation and test performance. Attention aids models in focusing on relevant input tokens, improving generalization.

Future Work: Additional investigation on Transformer-based baselines and multilingual datasets may yield further insights.