

The National University of Computer and Emerging Sciences

Islamabad Campus



Deep Learning Assignment # 2

Hamza Asad 22i-1908 DS-C

10th November 2025

Architecture

Siamese BiLSTM

| Layer | Type |
|--------------------------|---|
| Embedding | 128-dim word embeddings |
| BiLSTM | 2 layers, 128 hidden units, dropout 0.3 |
| FC1 | Linear(512 → 128), ReLU |
| FC2 | Linear(128 → 64), ReLU |
| FC3 | Linear(64 → 1), Sigmoid |
| Total Parameters: | 4,336,385 |

Training Settings

- Optimizer: Adam ($lr = 1e-3$)
- Batch Size: 64
- Epochs: 15
- Loss: Binary Cross Entropy
- Dropout: 0.3
- Framework: PyTorch

Attention Encoder

| Layer | Type |
|---------------------------------------|------------------|
| Embedding | 128-dim |
| Multi-Head Self-Attention | 1 head |
| LayerNorm + FeedForward (128→128→128) | — |
| FC Layers | 256→128→64→1 |
| Activation | ReLU + Sigmoid |
| Total Parameters: | 3,756,545 |

Training Settings

- Optimizer: Adam ($lr = 1e-3$)
- Epochs: 15
- Batch Size: 64
- Dropout: 0.3
- Loss: Binary Cross Entropy

Dataset Splits

395 categories (e.g., liabilities, termination, confidentiality).

Each CSV file represented a clause category containing multiple textual clauses.

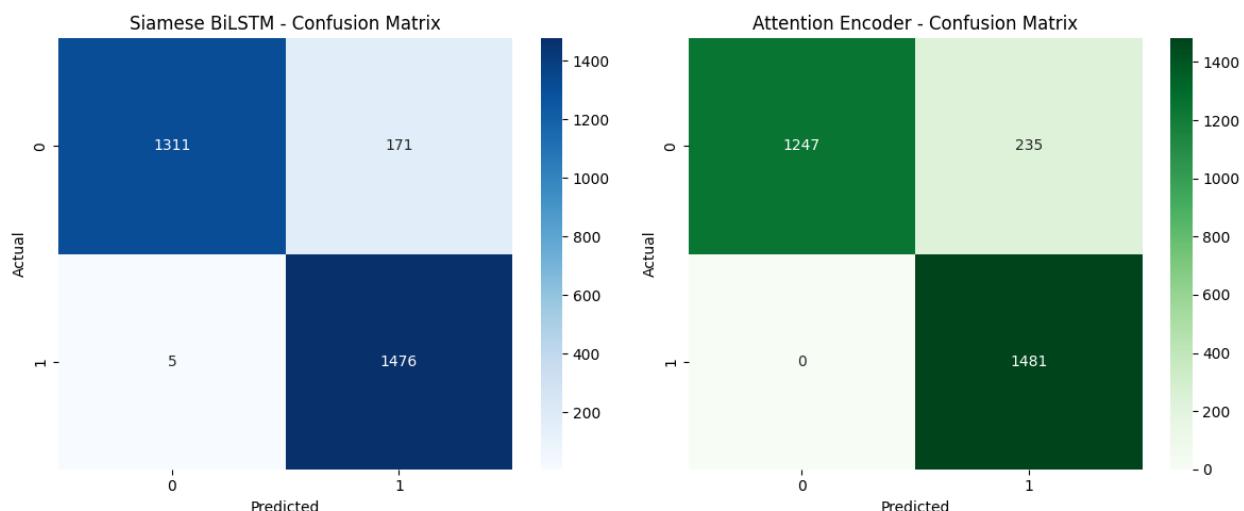
- **Total Clauses:** 150,878
- **Vocabulary Size:** 28,148 unique tokens
- **Categories:** 395

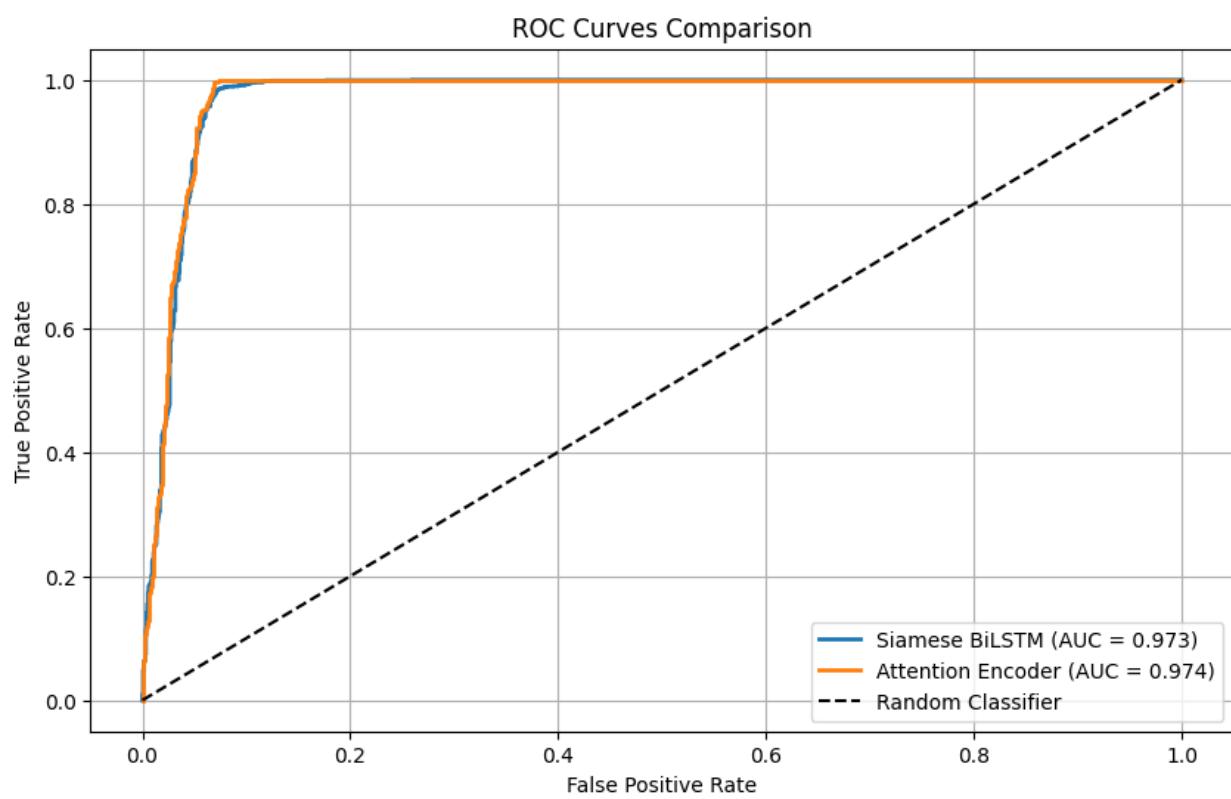
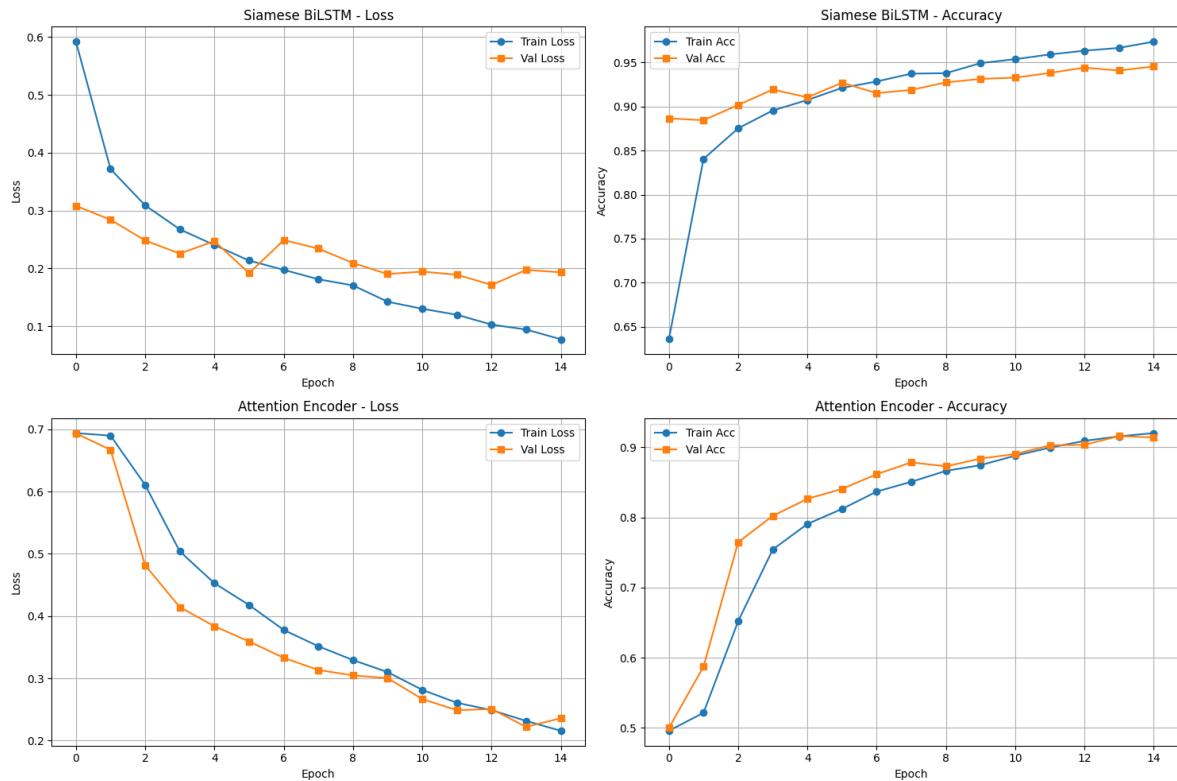
To train the models, a dataset of clause pairs was created:

- ~25 pairs per category
- **Positive pairs (similar):** 9,875
- **Negative pairs (non-similar):** 9,875
- **Total pairs:** 19,750

| Split | Pairs | Percentage |
|------------|--------|------------|
| Train | 13,825 | 70% |
| Validation | 2,962 | 15% |
| Test | 2,963 | 15% |

Training Graphs





Performance Measures

| Metric | Siamese BiLSTM | Attention Encoder |
|-----------|----------------|-------------------|
| Accuracy | 0.9406 | 0.9207 |
| Precision | 0.8962 | 0.8631 |
| Recall | 0.9966 | 1.0000 |
| F1-Score | 0.9437 | 0.9265 |
| ROC-AUC | 0.9726 | 0.9745 |

For this legal clause similarity task, accuracy, precision, recall, F1-score, and ROC-AUC were used to evaluate model performance. Accuracy provides a general correctness measure since the dataset is balanced, while precision ensures that only genuinely similar clauses are classified as such — crucial when false positives could lead to legal misinterpretation. Recall measures how effectively the model identifies all true similar clauses, important in retrieval systems where missing a relevant clause can be costly. F1-score balances precision and recall, making it the most reliable single metric for overall performance.

Finally, ROC-AUC reflects the model's ability to distinguish similar from dissimilar pairs across thresholds. In real-world (“in the wild”) deployments, **F1-score and ROC-AUC** are the most suitable metrics, as they capture robustness and threshold-independent performance crucial for legal document comparison systems.

Comparative analysis

| Aspect | Siamese BiLSTM | Attention Encoder |
|------------------------|--|---|
| Learning Speed | Faster convergence | Slower start |
| Generalization | Excellent | Moderate |
| Overfitting | Minimal | None |
| Semantic Understanding | Strong on lexical similarity | Slightly weaker without positional encoding |
| Precision | Higher (better false positive control) | Lower (more optimistic predictions) |
| Overall Winner | Siamese BiLSTM | |

E.g. Correct Predictions (Siamese BiLSTM)

1. *Further assurances* clauses — semantically identical, predicted correctly (confidence 0.992).
2. *Inspection rights* clauses — both about disclosure and inspection conditions (0.985).
3. *Non-discrimination* clauses — same intent, different wording (0.996).

E.g. Incorrect Predictions

1. *Financial information* vs *Non-competition* — predicted similar (1.0), semantically unrelated.
2. *Execution in counterparts* vs *Governing law* — structure similarity misled the model.
3. *Agreement consideration* clauses — very close wording but slightly different context; model missed (false negative).

Rest examples for other model are in the notebook.