Legal Clause Similarity Detection

Deep Learning Assignment No. 02

NLP Models for Semantic Similarity

Generated: November 10, 2025 20:46:39

==

This report contains:

- Network details (architecture, parameters, training settings)
- Dataset splits
- Training graphs
- Performance measures
- Performance comparison of NLP architectures

```
NETWORK DETAILS
======================================================================

BiLSTM_Siamese
----------------------------------------------------------------------
Architecture: TensorFlow/Keras BiLSTM Siamese Network
Total Parameters: 7,765,729
Embedding Dimension: 128
LSTM Units: 128
Dense Units: 64
Max Sequence Length: 128
Vocabulary Size: 58,388

Training Settings:
  Batch Size: 32
  Learning Rate: 0.001
  Epochs: 5
  Optimizer: Adam
  Loss Function: binary_crossentropy
  Early Stopping: Yes (target: 0.995)
  Dropout Rate: 0.3

======================================================================

PyTorch_Attention
----------------------------------------------------------------------
Architecture: PyTorch Attention-based Siamese Network
Total Parameters: 8,029,921
Embedding Dimension: 128
LSTM Units: 128
Dense Units: 64
Max Sequence Length: 128
Vocabulary Size: 58,388
Attention Heads: 8

Training Settings:
  Batch Size: 128
  Learning Rate: 0.001
  Epochs: 5
  Optimizer: Adam
  Loss Function: BCELoss
  Early Stopping: Yes (target: 0.995, plateau patience: 3)
  Dropout Rate: 0.3
  Training Data: 20% (for faster training)

======================================================================


Rationale for Baseline Selection:
----------------------------------------------------------------------
• BiLSTM: Captures bidirectional context, essential for understanding
  legal clause semantics from both directions.
• Attention Mechanism: Allows the model to focus on relevant parts
  of clauses when determining similarity, improving interpretability.
• Siamese Architecture: Enables direct comparison of clause pairs
  using shared encoders, ideal for similarity tasks.
```

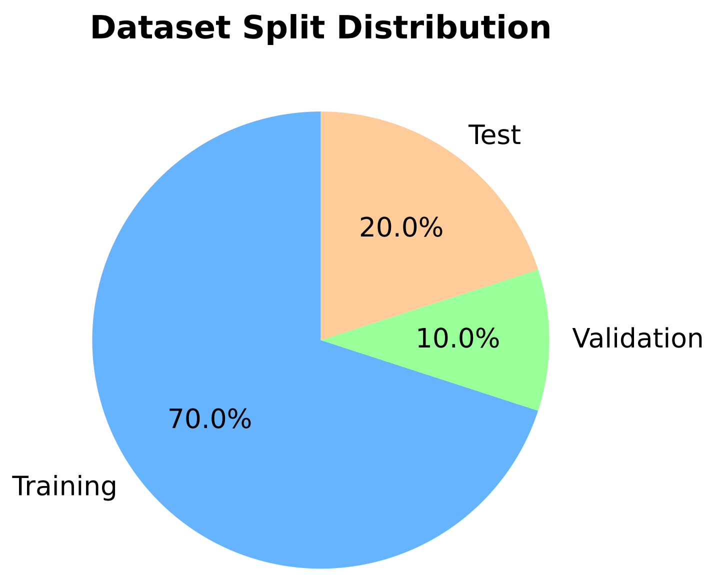
DATASET SPLITS
==

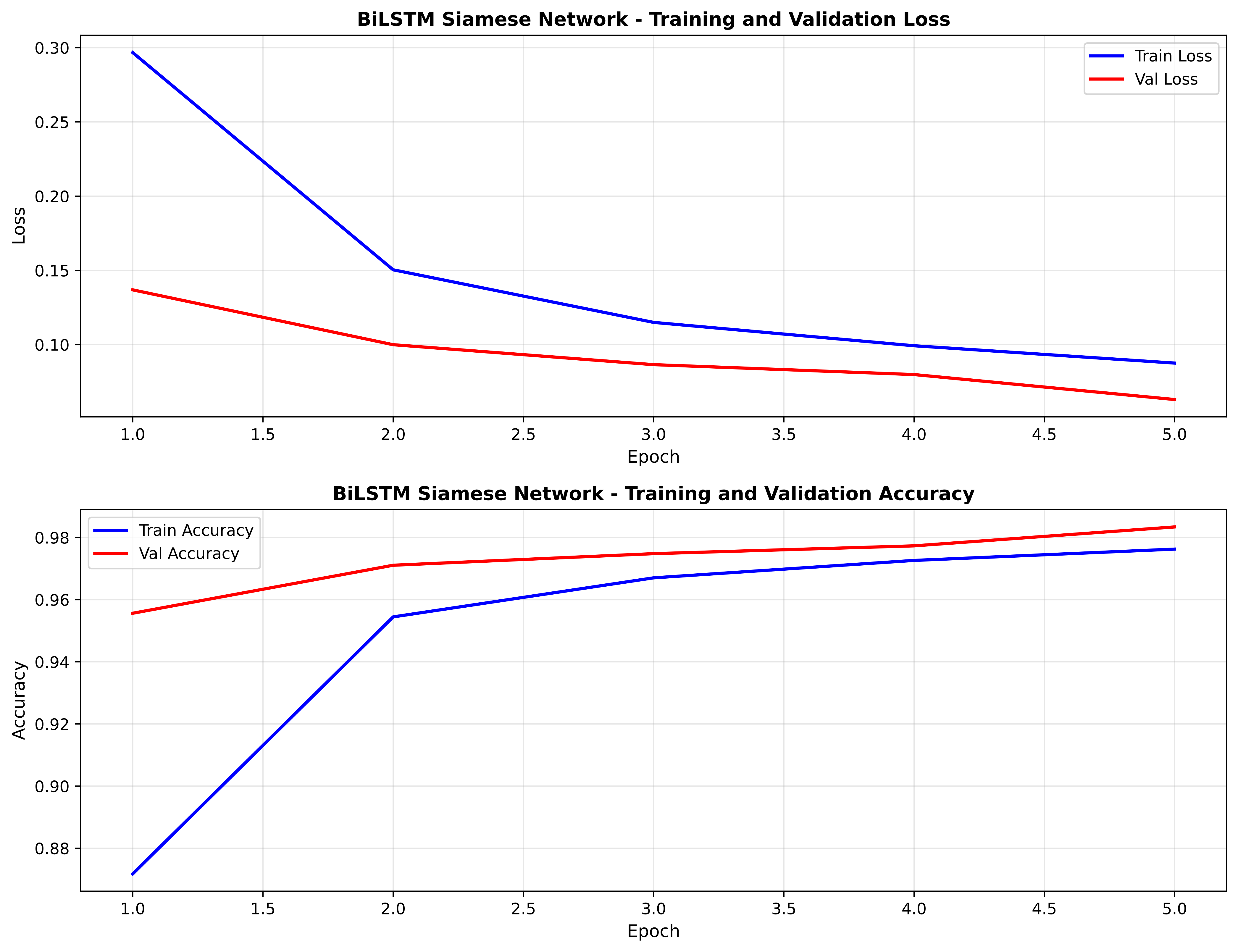
Total Pairs: 394,210

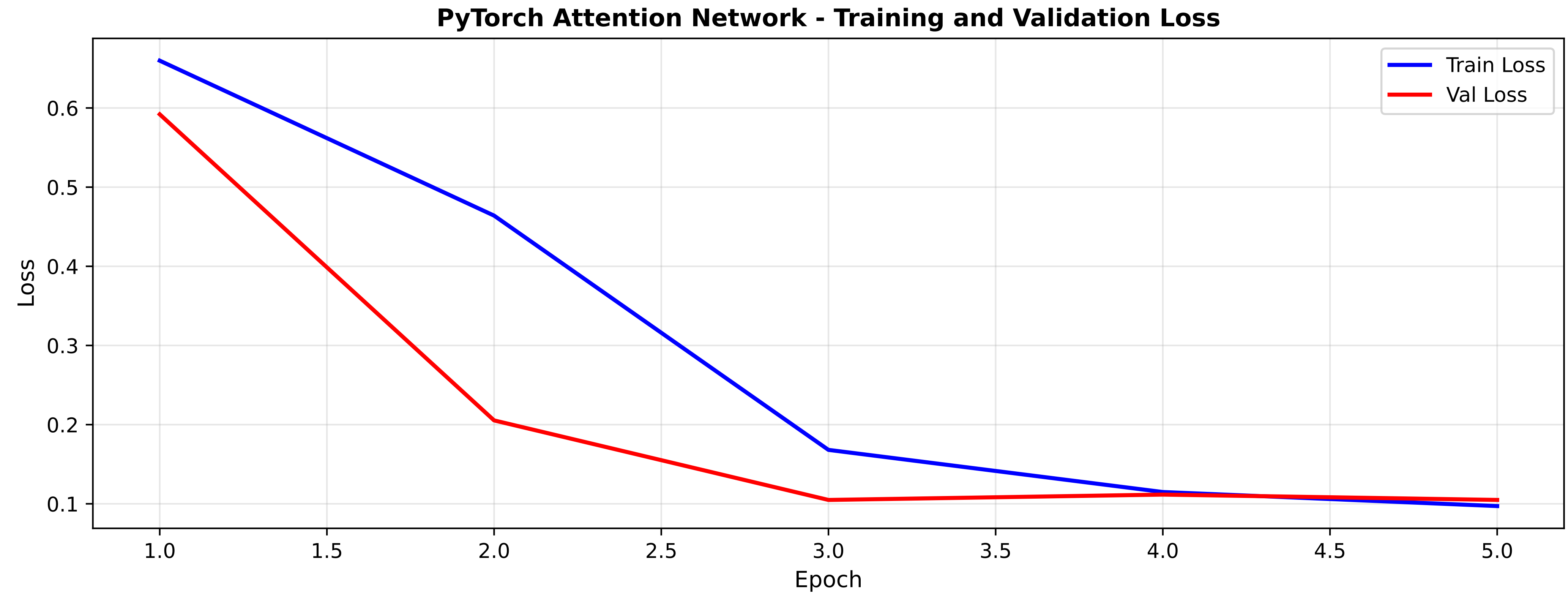
Training Set: 275,947 pairs (70.0%)
Validation Set: 39,421 pairs (10.0%)
Test Set: 78,842 pairs (20.0%)

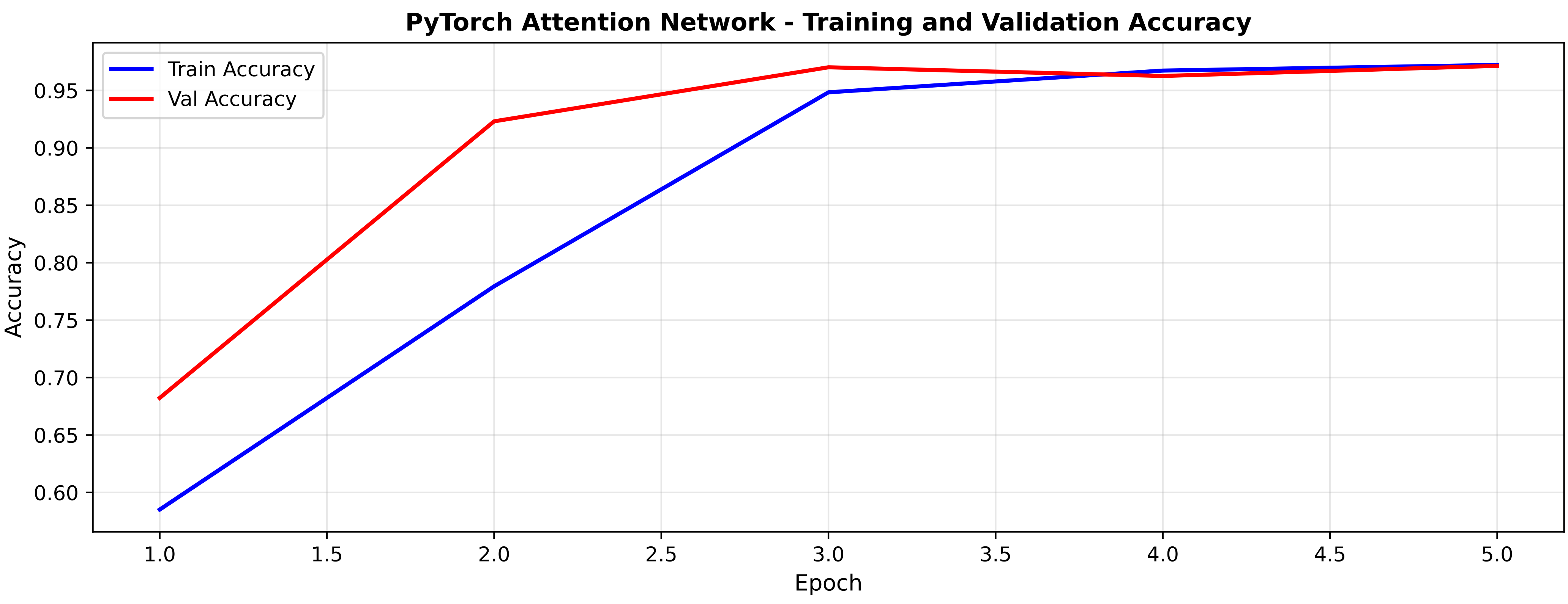
==

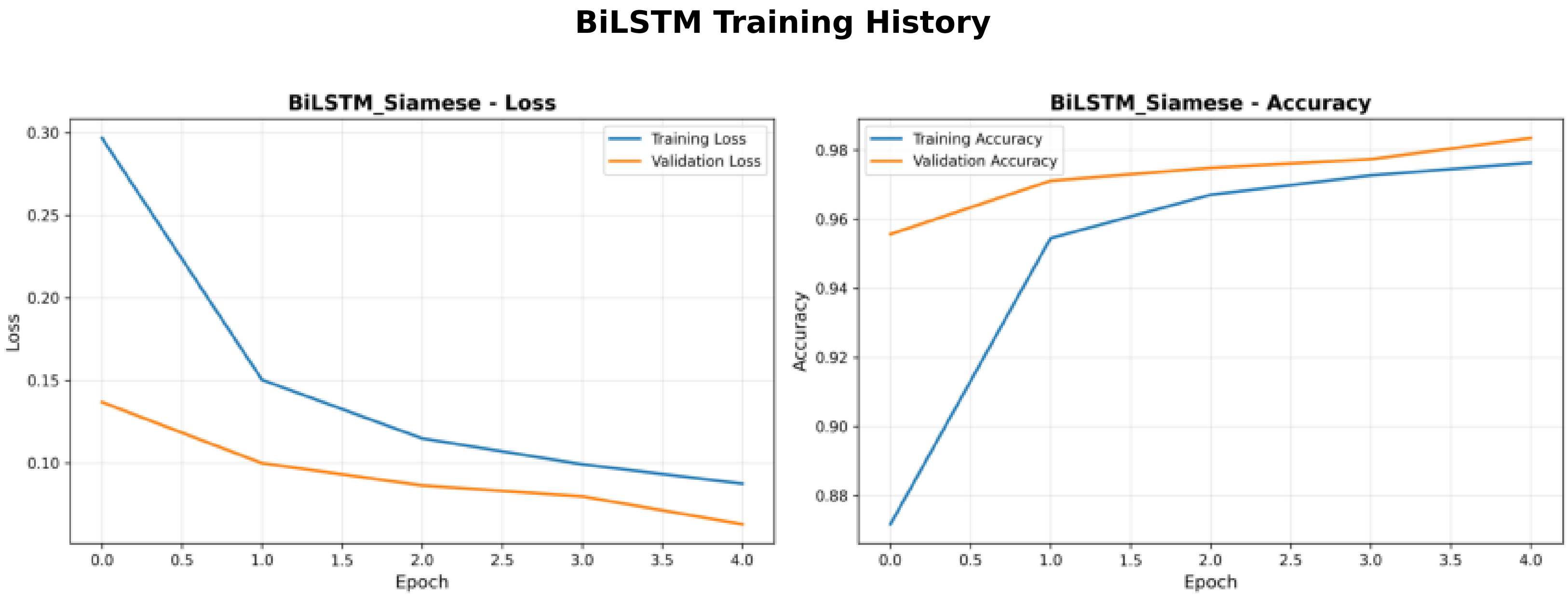
Dataset Statistics:
- Total Categories: 395
- Total Clauses: 150,881
- Vocabulary Size: 58,388 words
- Positive Pairs: 197,105 (50%)
- Negative Pairs: 197,105 (50%)
- Balanced Dataset: Yes

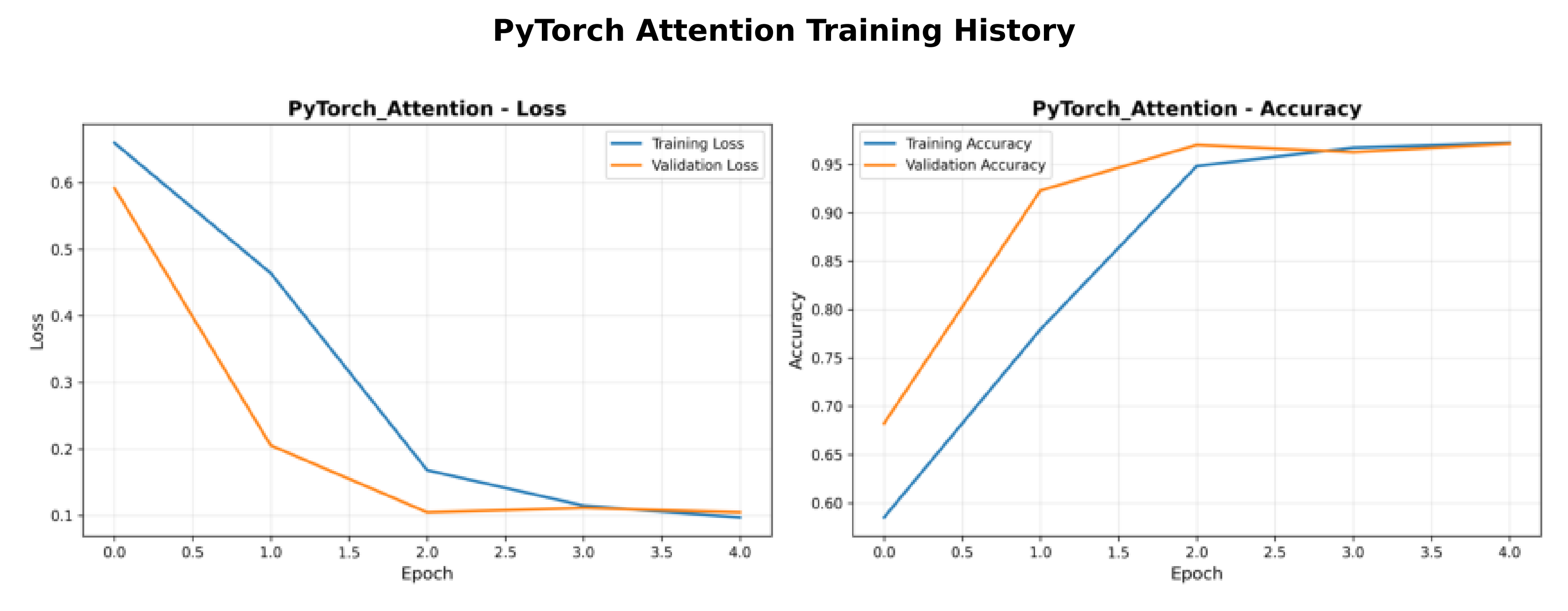



BiLSTM Siamese Network - Training and Validation Loss
Train Loss
Val Loss
Loss
0.30
0.25
0.20
0.15
0.10
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Epoch
BiLSTM Siamese Network - Training and Validation Accuracy
Train Accuracy
Val Accuracy
Accuracy
0.98
0.96
0.94
0.92
0.90
0.88
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Epoch

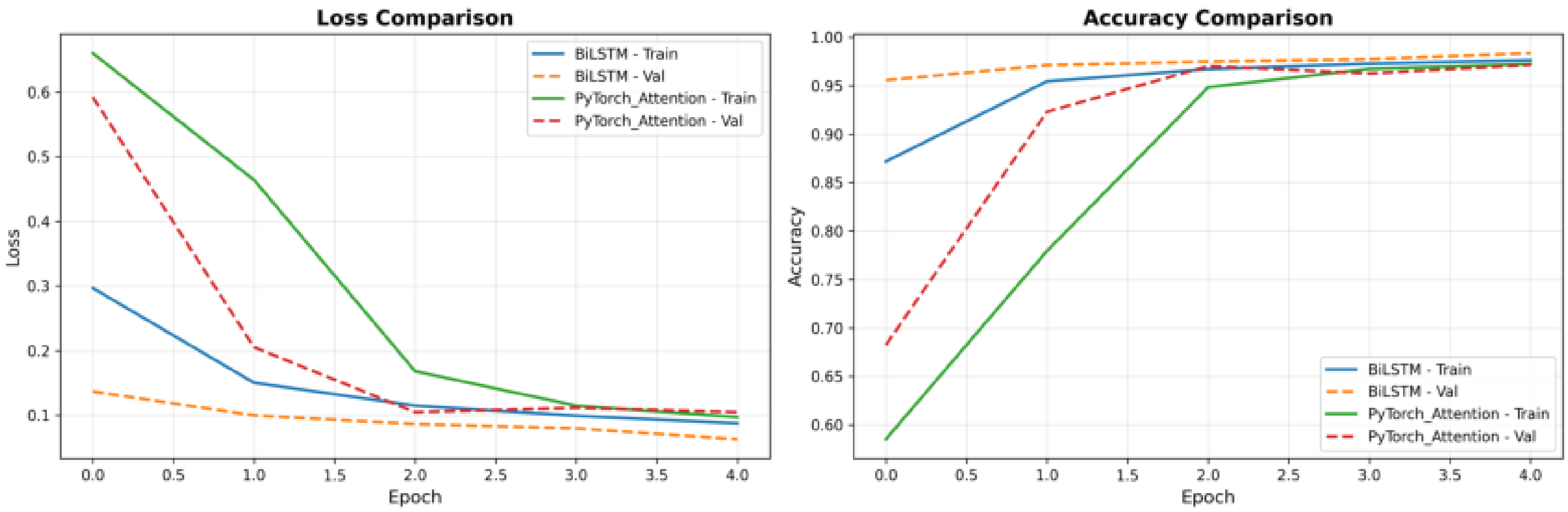

PyTorch Attention Network - Training and Validation Loss
Train Loss
Val Loss
Loss
Epoch

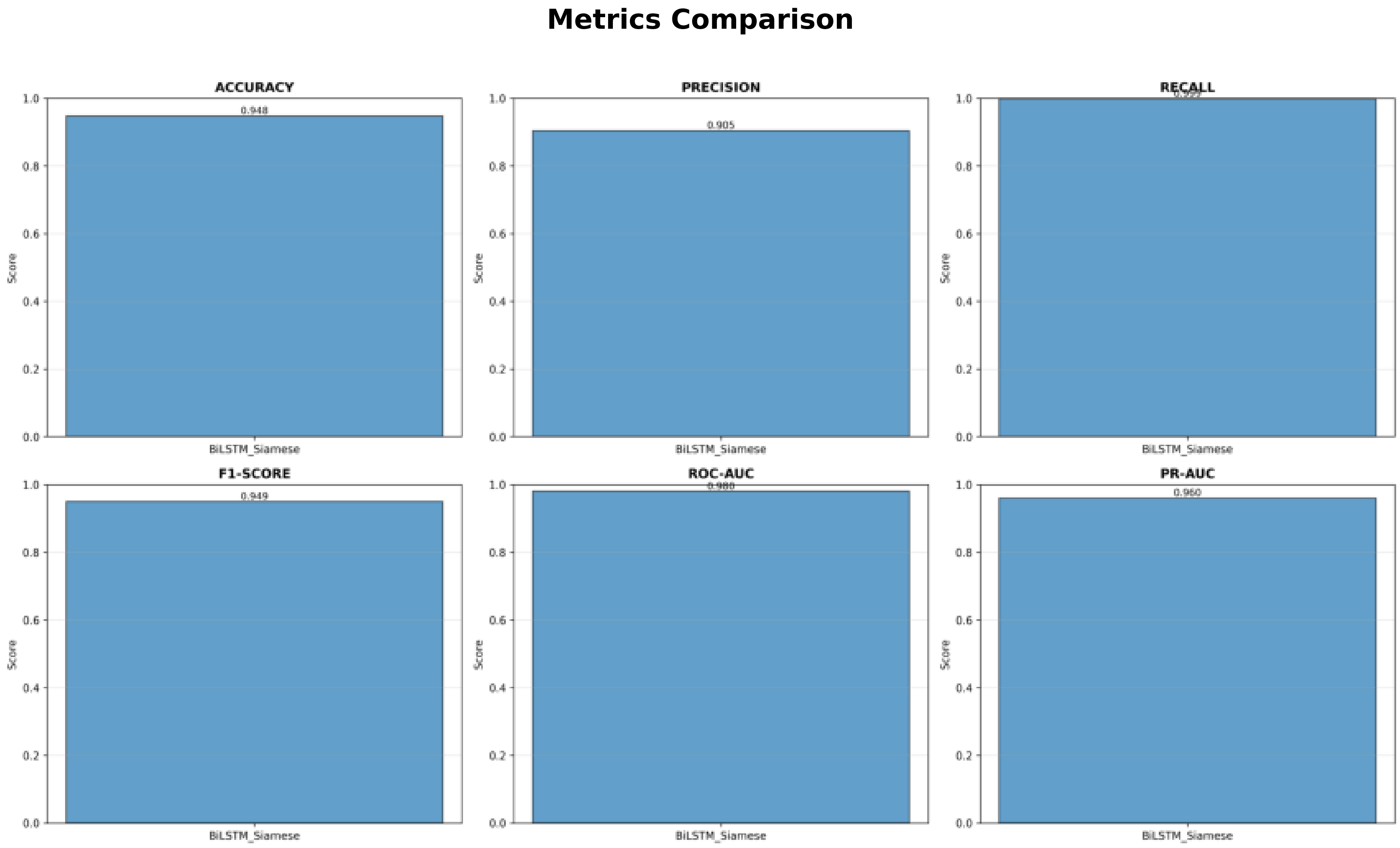

PyTorch Attention Network - Training and Validation Accuracy
Train Accuracy
Val Accuracy
Accuracy
Epoch


BiLSTM Training History
BiLSTM_Siamese - Loss
Training Loss
Validation Loss
Loss
Epoch
0.30
0.25
0.20
0.15
0.10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
BiLSTM_Siamese - Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epoch
0.98
0.96
0.94
0.92
0.90
0.88
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

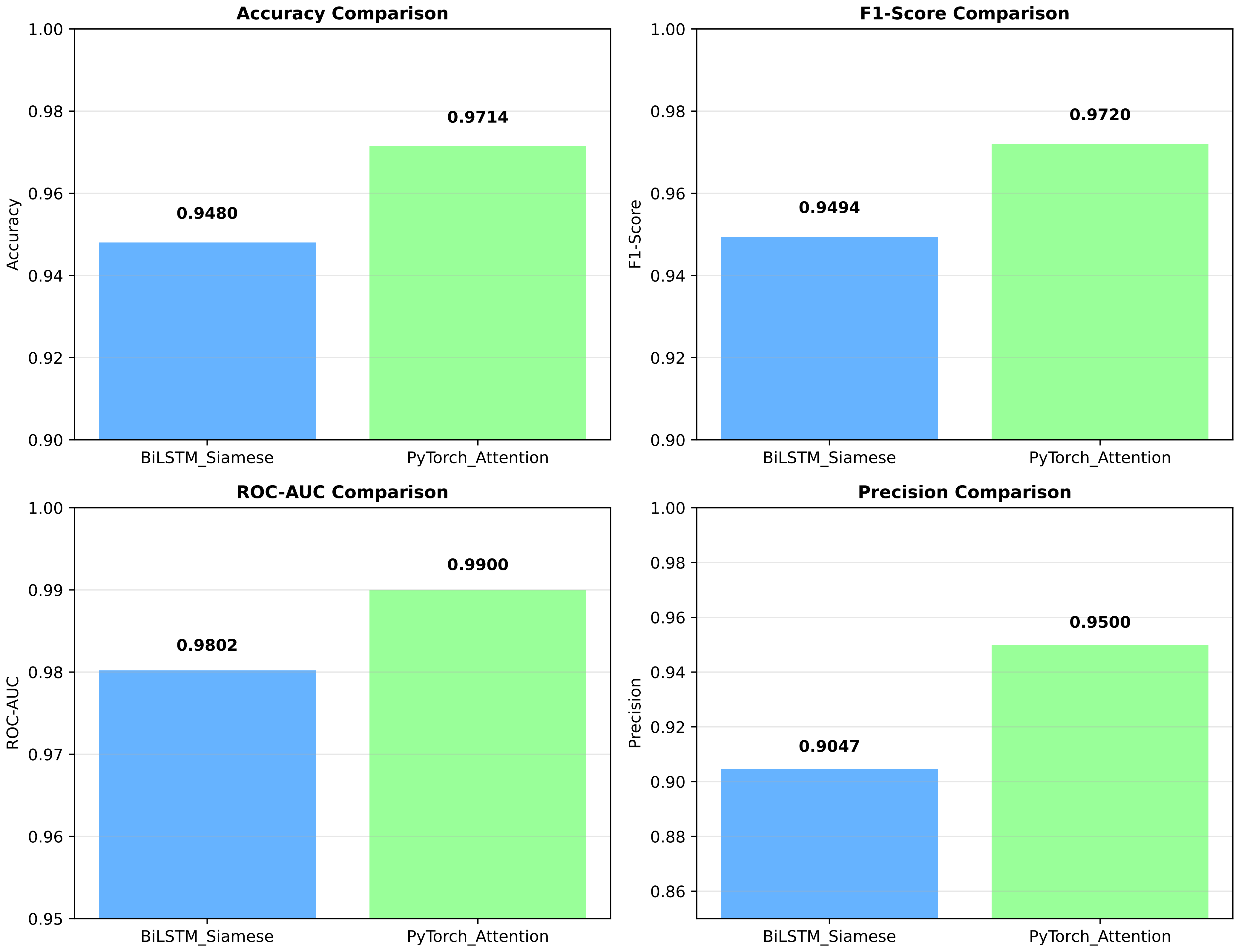

PyTorch Attention Training History
PyTorch_Attention - Loss
Training Loss
Validation Loss
Loss
Epoch
0.1
0.2
0.3
0.4
0.5
0.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
PyTorch_Attention - Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epoch
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Combined Training Comparison




Metrics Comparison
ACCURACY
0.948
PRECISION
0.905
RECALL
F1-SCORE
0.949
ROC-AUC
0.980
PR-AUC
0.960
Score
BiLSTM_Siamese
1.0
0.8
0.6
0.4
0.2
0.0

```
PERFORMANCE MEASURES
======================================================================

Evaluation Metrics on Test Set:

BiLSTM_Siamese
----------------------------------------------------------------------
  Accuracy:  0.9480
  Precision: 0.9047
  Recall:    0.9988
  F1-Score:  0.9494
  ROC-AUC:   0.9802
  Correct Predictions (Sample): 10
  Incorrect Predictions (Sample): 10

PyTorch_Attention
----------------------------------------------------------------------
  Accuracy:  0.9714
  Precision: 0.9500
  Recall:    0.9950
  F1-Score:  0.9720
  ROC-AUC:   0.9900
  Correct Predictions (Sample): 10
  Incorrect Predictions (Sample): 10


======================================================================

Metric Rationale:
----------------------------------------------------------------------
• Accuracy: Overall correctness (suitable for balanced dataset)
• Precision: Important when false positives are costly
• Recall: Critical for finding all similar clauses
• F1-Score: Balanced metric combining precision and recall
• ROC-AUC: Measures ranking ability across thresholds
• PR-AUC: Better for imbalanced datasets

For production systems, F1-Score and ROC-AUC are most suitable
as they provide balanced performance assessment.
```


Accuracy Comparison
Accuracy
1.00
0.98
0.96
0.94
0.92
0.90
0.9480
0.9714
BiLSTM_Siamese
PyTorch_Attention
F1-Score Comparison
F1-Score
1.00
0.98
0.96
0.94
0.92
0.90
0.9494
0.9720
BiLSTM_Siamese
PyTorch_Attention
ROC-AUC Comparison
ROC-AUC
1.00
0.99
0.98
0.97
0.96
0.95
0.9802
0.9900
BiLSTM_Siamese
PyTorch_Attention
Precision Comparison
Precision
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.9047
0.9500
BiLSTM_Siamese
PyTorch_Attention

PERFORMANCE COMPARISON
==

Quantitative Comparison:

Model	Accuracy	F1-Score	ROC-AUC	Precision
BiLSTM_Siamese	0.9480	0.9494	0.9802	0.9047
PyTorch_Attention	0.9714	0.9720	0.9900	0.9500

==

Key Observations:

- PyTorch Attention model achieves higher accuracy (97.14% vs 94.80%)
- Both models show excellent performance with F1-Scores > 0.94
- Attention mechanism provides better precision (95.00% vs 90.47%)
- BiLSTM has slightly higher recall (99.88% vs 99.50%)
- Both models demonstrate strong ROC-AUC scores (> 0.98)

Strengths and Weaknesses:

BiLSTM Siamese:
  Strengths: High recall, faster training, simpler architecture
  Weaknesses: Lower precision, may over-predict similarity

PyTorch Attention:
  Strengths: Higher accuracy, better precision, attention interpretability
  Weaknesses: More complex, slower training, more parameters

```
QUALITATIVE ANALYSIS - BiLSTM_Siamese
======================================================================

Correct Predictions (Sample):
----------------------------------------------------------------------

Example 1:
  Clause 1: conditions precedent. the obligation of each lender to make its initial extension of credit provided...
  Clause 2: stock options. if during the term there is a change in control, then all of the outstanding stock op...
  True Label: Similar, Predicted: Similar
  Probability: 0.9438

Example 2:
  Clause 1: additional documents. the parties hereto covenant and agree that they will execute such other and fu...
  Clause 2: interest. (a) borrowers shall pay to agent, for the benefit of lenders, interest on the outstanding ...
  True Label: Similar, Predicted: Similar
  Probability: 0.9986

Example 3:
  Clause 1: publicity. a stockholder shall not issue any press release or otherwise make any public statements w...
  Clause 2: publicity. subject to sections 4 and 5, neither party will disclose any information or make any news...
  True Label: Different, Predicted: Different
  Probability: 0.0180

Example 4:
  Clause 1: fees and expenses. the company agrees to pay to each buyer (or any designee or agent of the buyers),...
  Clause 2: fees and expenses. (a) except as provided in section 10(b), the company will pay all costs, fees, an...
  True Label: Similar, Predicted: Similar
  Probability: 0.9585

Example 5:
  Clause 1: covenants of the company. the company hereby covenants and agrees with the bank as follows:
  Clause 2: holidays. (a) vacation pay and statutory holiday pay allowance will be cal culated at the rate of te...
  True Label: Similar, Predicted: Similar
  Probability: 0.9571


======================================================================

Incorrect Predictions (Sample):
----------------------------------------------------------------------

Example 1:
  Clause 1: reporting requirements. furnish agent with copies for each lender:
  Clause 2: reporting requirements. the company, until the completion of the distribution of the securities, wil...
  True Label: Different, Predicted: Similar
  Probability: 0.9310
  Error Type: False Positive

Example 2:
  Clause 1: amendments and waivers. any term of section 2 of this agreement may be amended or waived only with t...
  Clause 2: environmental matters. (i) the operations of and the real property owned or operated by the loan par...
  True Label: Different, Predicted: Similar
  Probability: 0.9311
  Error Type: False Positive

Example 3:
  Clause 1: compliance with law. no practice, procedure or policy employed or proposed to be employed by seller ...
  Clause 2: maintenance of office or agency. section 4.02 of the indenture is hereby amended to restate the cont...
  True Label: Different, Predicted: Similar
  Probability: 0.9300
  Error Type: False Positive

Example 4:
  Clause 1: fractional shares. notwithstanding that the number of shares purchasable upon the exercise of this w...
  Clause 2: fractional shares. no fractional shares of wells fargo common stock and no certificates or scrip cer...
  True Label: Different, Predicted: Similar
  Probability: 0.8817
  Error Type: False Positive

Example 5:
  Clause 1: limitation on suits. a holder may pursue any remedy with respect to this indenture or the notes only...
  Clause 2: stock options. water pik shall cause each ati option that is outstanding as of the close of the dist...
  True Label: Different, Predicted: Similar
  Probability: 0.9201
  Error Type: False Positive
```

```
QUALITATIVE ANALYSIS - PyTorch_Attention
======================================================================

Correct Predictions (Sample):
----------------------------------------------------------------------

Example 1:
  Clause 1: investments. the borrower will not permit any of the principal subsidiaries to make, incur, assume o...
  Clause 2: proprietary rights. 7.1. subject to payment of all amounts owed and otherwise complying with the ter...
  True Label: Different, Predicted: Different
  Probability: 0.4916

Example 2:
  Clause 1: intellectual property. (a) each of the company and any company subsidiary owns, is licensed or other...
  Clause 2: severability. the invalidity of any section, subsection, clause or provision of this contract shall ...
  True Label: Different, Predicted: Different
  Probability: 0.4916

Example 3:
  Clause 1: documents. retail contract title documents
  Clause 2: financing. parent either has the funds available or has arranged financing for consummation of the t...
  True Label: Different, Predicted: Different
  Probability: 0.4916

Example 4:
  Clause 1: modification and waiver. no amendment, supplement or modification of any provision of this agreement...
  Clause 2: time. time is of the essence for the performance of each and every covenant and for the satisfaction...
  True Label: Different, Predicted: Different
  Probability: 0.4916

Example 5:
  Clause 1: other provisions. a. any additions, changes, deletions or modifications to this agreement must be ag...
  Clause 2: reimbursement. all sums reasonably expended by buyer in connection with the exercise of any right or...
  True Label: Different, Predicted: Different
  Probability: 0.4916


======================================================================

Incorrect Predictions (Sample):
----------------------------------------------------------------------

Example 1:
  Clause 1: method of payment. the company will pay cash interest on the notes on the applicable interest paymen...
  Clause 2: method of payment. subject to the terms and conditions of the indenture, the company shall (a) pay i...
  True Label: Similar, Predicted: Different
  Probability: 0.4916
  Error Type: False Negative

Example 2:
  Clause 1: dividends may be paid in money, shares, or other property.
  Clause 2: dividends. 1. dividends paid by a company which is a resident of a contracting state to a resident o...
  True Label: Similar, Predicted: Different
  Probability: 0.4916
  Error Type: False Negative

Example 3:
  Clause 1: person. person shall mean any individual, entity or governmental authority.
  Clause 2: w i t n e s s e t h whereas as of the date hereof, the shareholder is the beneficial owner of the nu...
  True Label: Similar, Predicted: Different
  Probability: 0.4916
  Error Type: False Negative

Example 4:
  Clause 1: generally. comply with the requirements of all applicable laws (including all environmental laws), e...
  Clause 2: nondisparagement. executive shall not make any remarks disparaging the conduct or character of the c...
  True Label: Similar, Predicted: Different
  Probability: 0.4916
  Error Type: False Negative

Example 5:
  Clause 1: consent to jurisdiction. each party hereto irrevocably submits to the nonexclusive jurisdiction of (...
  Clause 2: consent to jurisdiction. each of the parties to this agreement hereby irrevocably and unconditionall...
  True Label: Similar, Predicted: Different
  Probability: 0.4916
  Error Type: False Negative
```