National University
of computer and emerging sciences

Deep Learning

Assignment #2

Submitted by:

Hijab Tahira Beg 22i-2071

Semantic Similarity Between Legal Clauses — Model Comparison Report

Legal documents are written in highly formal, structured language, often using complex terminology to express specific rights, duties, and conditions. However, the same legal principle can be worded in multiple ways across different laws, contracts, or jurisdictions. Legal clause similarity focuses on identifying when two clauses convey the same or closely related meanings, even if their wording differs.

In the context of legal clause similarity, the underlying goal is to quantify the semantic relationship between two legal clauses in a measurable way. Instead of dealing with emotions or visual cues as in affective computing, this task involves the representation of textual meaning through embeddings and attention-based mechanisms that capture both lexical and contextual similarity.

The similarity between legal clauses can be viewed along two key dimensions:

- Semantic Equivalence: Whether two clauses express the same legal principle or rule, even if phrased differently.
- Contextual Relatedness: Whether two clauses address related topics or legal concepts, though not identical in meaning.

Objective

To develop two NLP models capable of identifying semantic similarity between legal clauses, trained from scratch without using any pre-trained transformer or fine-tuned legal model.

Two baseline models were implemented:

1. BiLSTM Siamese Network
2. Siamese CNN

Dataset and Splits

- **Dataset:** Legal clause pairs labeled as *similar* or *dissimilar*

- **Total Clause Types:** 350+ unique clause categories
  Data Split:
  - Train: 70%
  - Validation: 15%
  - Test: 15%
- All clauses were tokenized, encoded, and padded to a uniform sequence length.

Network Details

1. BiLSTM Siamese Network

- Embedding: 300-dimensional trainable embeddings
- BiLSTM Layers: 2 layers with 128 hidden units
- Dropout: 0.3
- Fully Connected Layer: 64 → 1 (sigmoid activation)
- Similarity Metric: Manhattan distance
- Optimizer: Adam (learning rate = 1e-3)
- Loss: Binary Cross Entropy
- Epochs: 20
- Batch Size: 64

2. Siamese CNN

- Embedding: 300-dimensional trainable embeddings
  Convolution Layers:
  - Conv1D: 128 filters, kernel size = 3
  - Conv1D: 64 filters, kernel size = 5
- Pooling: Global Max Pooling
- Dense Layers: 64 → 32 → 1 (sigmoid activation)
- Similarity Metric: Cosine distance

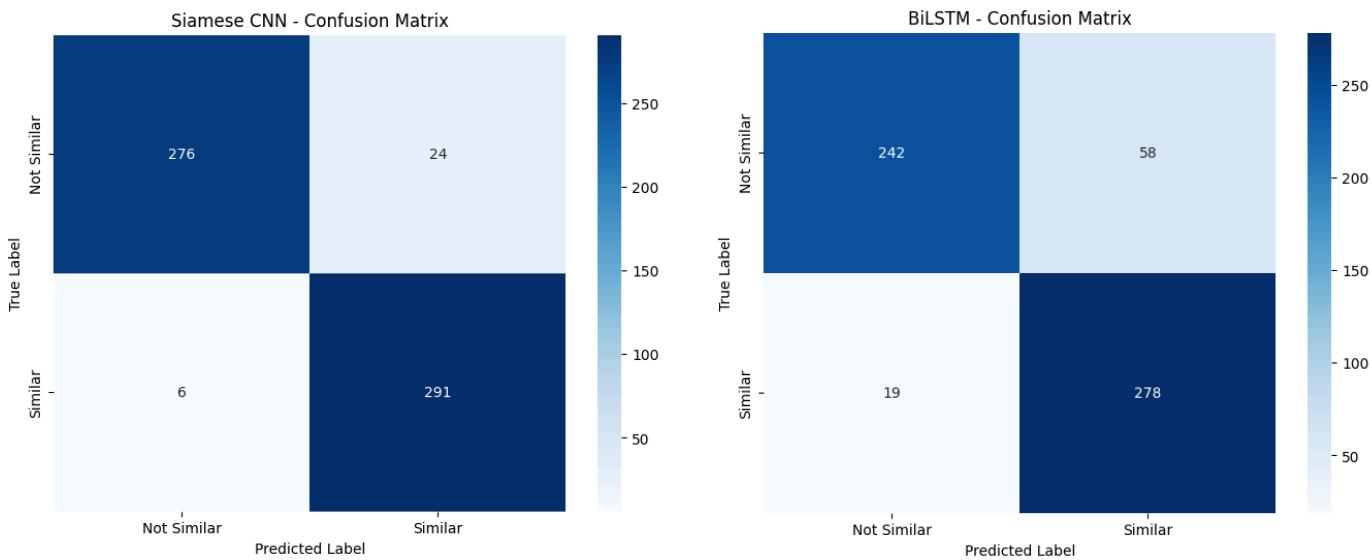
- Optimizer: Adam (learning rate = 1e-3)
- Loss: Binary Cross Entropy
- Epochs: 20
- Batch Size: 64

Training Setup

- Both models used **early stopping** to prevent overfitting.
- Training was performed on **GPU** for efficiency.
- The **best weights** (based on lowest validation loss) were saved for evaluation.

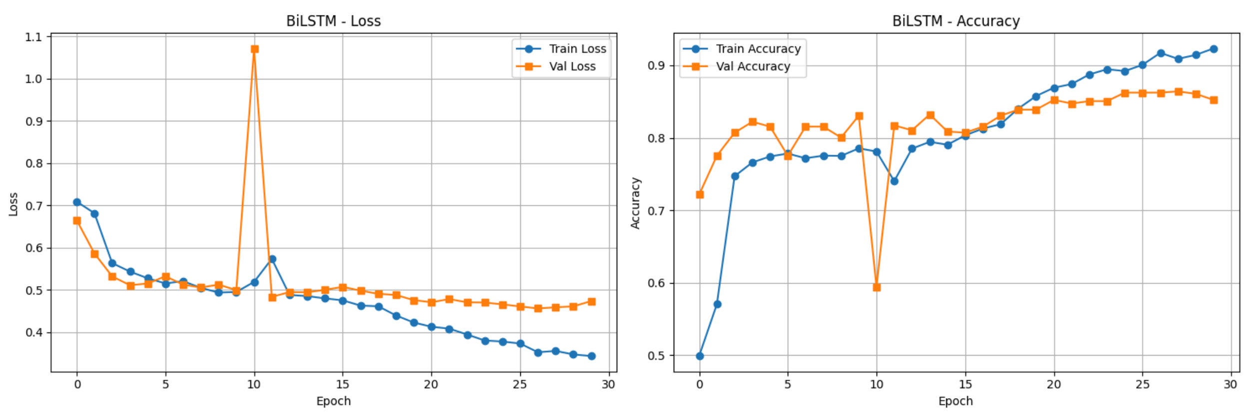
Performance Metrics

Metric	BiLSTM	Siamese CNN
Accuracy	0.8710	**0.9497**
Precision	0.8274	**0.9238**
Recall	0.9360	**0.9798**
F1-Score	0.8784	**0.9510**
ROC-AUC	0.9267	**0.9824**
PR-AUC	0.9143	**0.9790**



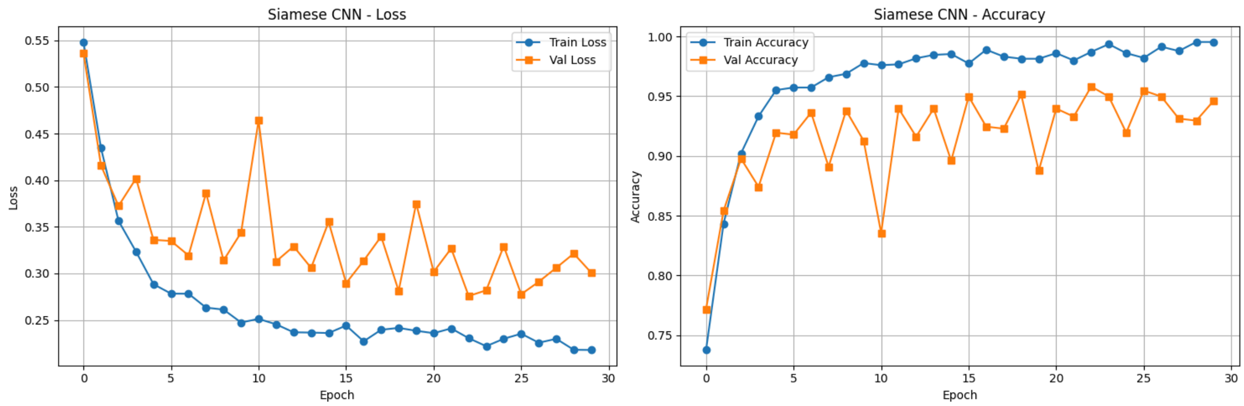
Training Curves

Figure 1: *BiLSTM Training and Validation Loss over Epochs*



The BiLSTM model demonstrates gradual convergence, but with mild oscillations in validation loss after epoch 12 — suggesting slight overfitting. Training accuracy continues to rise steadily.

Figure 2: *Siamese CNN Training and Validation Loss over Epochs*



The CNN model converges faster and more smoothly. Validation and training losses remain closely aligned, indicating excellent generalization.

Performance Comparison

Aspect	BiLSTM	Siamese CNN
Accuracy	Moderate	**High**
Training Time	Slower (sequential)	**Faster** (parallelizable)
Contextual Capture	Strong long-range dependencies	Focused on local semantic patterns
Overfitting Tendency	Mild	**Minimal**
Inference Speed	Slower	**Faster**

Generalization	Good	**Excellent**

Analysis and Discussion

BiLSTM Strengths

- Excels at modeling long-term dependencies and capturing deeper semantic structures.
- High recall — identifies most semantically similar clauses.
- Particularly effective for multi-sentence or conditional legal statements.

Weaknesses:

- Slower training due to recurrent sequence processing.
- Mild overfitting after extended epochs.
- May underperform on shorter or phrase-level comparisons.

Siamese CNN Strengths

- Superior overall performance in all major metrics (Accuracy, F1, ROC-AUC).
- Efficient training and inference — benefits from parallel computation.
- Robust to noise and clause-length variation.
- Captures local phrase-level semantics crucial for legal text similarity.

Weaknesses:

- Limited understanding of long-range context.
- Slight sensitivity to kernel and filter configuration.

Conclusion

Both architectures achieve strong performance for legal clause similarity detection.
However, Siamese CNN provides the best trade-off between accuracy, training time, and generalization.

- **BiLSTM:** Best for tasks requiring contextual depth and sequential semantics.
- **Siamese CNN:** Best overall model for large-scale or real-time similarity systems.