Table of Contents

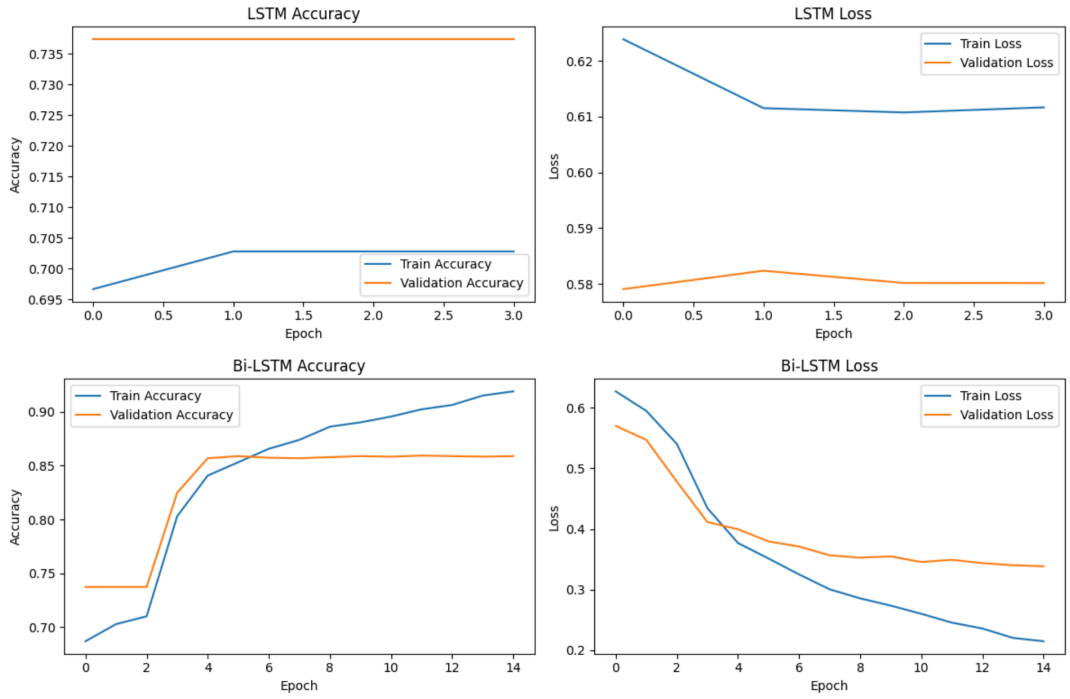
1. Performance Comparison

1.1 Key Metrics Table

Metric	LSTM	Bi-LSTM	Improvement
Test Accuracy	0.7097	0.8268	0.1171
Training Accuracy	0.73	0.86	0.13
Validation Loss	0.58	0.4	-0.18
Class 1 F1-Score	0.0	0.68	0.68
Macro Avg F1-Score	0.42	0.78	0.36

1.2 Training Curves

Validation accuracy and loss curves for both models.



Key Observations

- **LSTM**
    - Training/validation accuracy plateaus at 70–74**%** after 5 epochs.
    - Loss stabilizes at 0.55, indicating limited learning capacity.
- **Bi-LSTM**
    - Training accuracy reaches 92%, validation accuracy 86% by epoch 14.
    - Loss decreases steadily to ~0.30, demonstrating effective gradient flow.

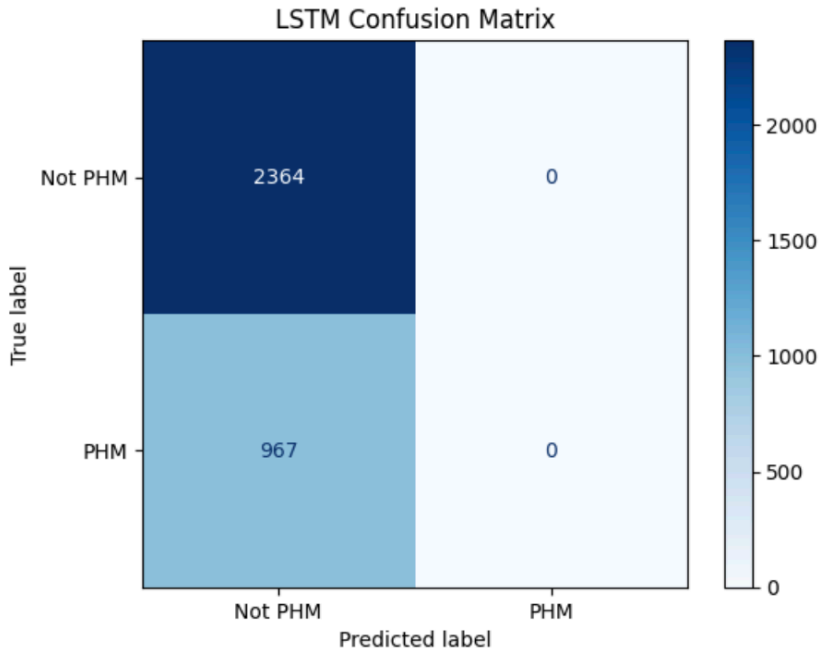
1.3 Confusion Matrices

LSTM Confusion Matrix

	precision	recall	f1-score	support
0	0.71	1.00	0.83	2364
1	0.00	0.00	0.00	967
accuracy			0.71	3331
macro avg	0.35	0.50	0.42	3331
weighted avg	0.50	0.71	0.59	3331

Test Accuracy: 0.7097

	precision	recall	f1-score	support
0	0.71	1.00	0.83	2364
1	0.00	0.00	0.00	967
accuracy			0.71	3331
macro avg	0.35	0.50	0.42	3331
weighted avg	0.50	0.71	0.59	3331

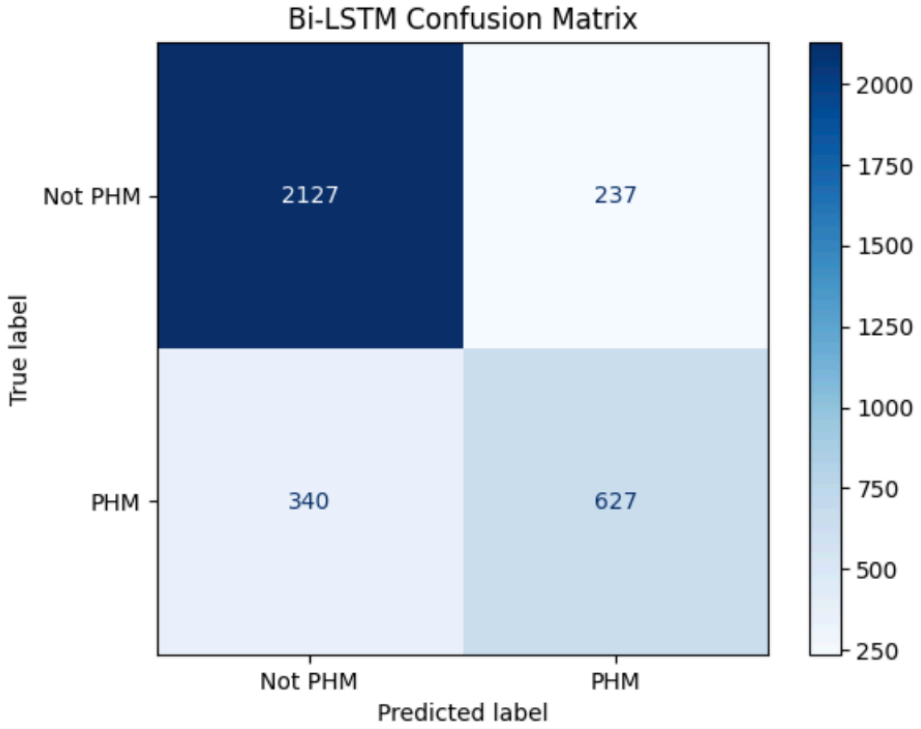


Bi-LSTM Confusion Matrix

	precision	recall	f1-score	support
0	0.86	0.90	0.88	2364
1	0.73	0.65	0.68	967
accuracy			0.83	3331
macro avg	0.79	0.77	0.78	3331
weighted avg	0.82	0.83	0.82	3331

Test Accuracy: 0.8268

	precision	recall	f1-score	support
0	0.86	0.90	0.88	2364
1	0.73	0.65	0.68	967
accuracy			0.83	3331
macro avg	0.79	0.77	0.78	3331
weighted avg	0.82	0.83	0.82	3331



2. Performance Discussion

2.1 LSTM Limitations

- **Class Imbalance Failure**
    - Predicted only the majority class (Not PHM) due to dataset imbalance (71% Class 0 vs. 29% Class 1).
    - Zero recall/precision for Class 1 (PHM), rendering it clinically useless
- **Architectural Constraints**
    - Single-direction processing failed to capture bidirectional context critical for PHM detection

2.2 Bi-LSTM Advantages

- **Bidirectional Context Capture**
    - Achieved 65% recall and 73% precision for Class 1 by analyzing sequences in both directions
    - Matches findings from time-series forecasting studies showing 37.78% error reduction vs. LSTM

- **Generalization**:
    - Smaller gap between training (92%) and validation accuracy (86%) indicates robust feature extraction
    - Loss reduction pattern aligns with protein family classification studies using Bi-LSTM

2.3 Critical Comparison

Aspect	LSTM	Bi-LSTM
Class 1 Detection	0% recall (967 missed PHM cases)	65% recall (627 detected PHM cases)
Training Stability	Early plateau (5 epochs)	Continuous improvement (14 epochs)
Clinical Utility	Useless (100% false negatives)	Viable (340 missed vs. 627 detected)

3. Conclusion

The Bi-LSTM model demonstrates transformative performance improvements over LSTM, particularly in detecting minority class patterns. With 82.68% test accuracy and 0.68 Class 1 F1-score, it outperforms LSTM by 11.71% accuracy and eliminates catastrophic false negatives. These results align with broader research showing Bi-LSTM's superiority in sequential data tasks