DATA 641: Assignment 3

Name: Emily Hightower

Date: 11/12/2025

Code Implementation:

GitHub Repository:

Code Components & Structure: The code is structured as follows:

```
├── data/
│   ├── processed/        <- PyTorch tensors & Pickled tokenizer
│   └── dataset.csv       <- raw IMDb dataset
├── src/
│   ├── preprocess.py          <- process raw data
│   ├── models.py              <- create model framework
│   ├── train.py               <- iterative training loop
│   ├── evaluate.py       <- plots F1, accuracy, & training loss
│   └── utils.py               <- utilities for other source code
├── results/
│   ├── metrics.csv       <- results from each architecture
│   ├── preprocessing_stats.csv     <- pre-processed data statistics
│   ├── preprocessing_times.csv     <- pre-processing time cost
│   ├── batch_logs/       <- loss logs for each trained model
│   └── plots/            <- plots for evaluation
├── driver.py             <- execution (train, models, evaluate)
├── hardware.py                <- returns system hardware
├── report.pdf            <- this document
├── requirements.txt           <- system requirements
└── README.md             <- project overview
```

The following sections are included in the README.MD file in the GitHub repository.

Implementing the Code: To execute the full code, run the following commands:

```
pip install -r requirements.txt
python hardware.py
python src/preprocess.py \
```

```
    --input data/dataset.csv \
    --output_dir data/processed \
    --results_file results/preprocessing_times.csv \
    --stats_file results/preprocessing_stats.csv
python driver.py
```

This implementation will take 6.214 hours with the following hardware:

```
- CPU: arm
- CPU cores: 14
- Logical processors: 14
- RAM: 48.00 GB
- GPU: None
```

To train a single model with this code, run (for example):

```
python src/train.py
    --architecture LSTM
    --activation tanh
    --optimizer adam
    --sequence_length 50
    --clip
```

Report:

This project explores Sentiment Classification as a core Natural Language Processing (NLP) task used to categorize the emotional tone of a piece of text into classes like positive or negative. The project implements and evaluates Recurrent Neural Networks (RNN) architectures for sentiment classification, as treated as a sequence classification problem.

Dataset Summary:

This project considers Sentiment Classification using the IMDb Movie Review Dataset, which contains 50,000 labeled reviews. The raw data labels are ‘positive’ or ‘negative’ which are converted to a binary classifier (‘1’, ‘0’) for this project. The reviews are pre-processed to allow for machine learning analysis. The pre-processing begins with lowercasing the reviews and removing punctuation and special characters using a regex. This provides a clean string for tokenization. Using the Keras Tokenizer, the reviews are tokenized. Before additional processing, the average review length is 230 tokens and the full vocabulary size is 180,586 tokens. This vocabulary is reduced to the 10,000 most frequent words, or the most common vocabulary. The reviews are converted into a sequence of token IDs and truncated to the fixed lengths of 25, 50, and 100. This process is applied to a random separation of 50,000 training reviews and 50,000 testing reviews, with the tokenizer fit only on the training text. An out-of-vocabulary token (OOV) is included.

The final result of text pre-processing are four training tensors (labels, seq25, seq50, seq100), four testing tensors (labels, seq25, seq50, and seq100), and the tokenizer (if needed for later inference). At each step of the pre-processing procedure, the time cost was recorded. The full time to process the original 50,000 IMDb review dataset was 4.1209 seconds.

Model Configurations:

This project calls for experimentation with the following model architectures:

- **Architecture:** RNN, LSTM, Bidirectional LSTM
- **Activation Function:** sigmoid, ReLu, tanh
- **Optimizer:** adam, stochastic gradient descent (sgd), RMSProp
- **Sequence Length:** 35, 50, 100
- **Stability Strategy:** none, gradient clipping

In total, this is 162 unique model combinations. However, LSTM and Bidirectional LSTM (BiLSTM) do not allow for activation functions other than tanh. In both LSTM and BiLSTM, tanh stabilizes the cell state and final hidden output by naturally scaling between -1 and +1. Arbitrary activation functions (like ReLu) cause the gating logic to break. The sigmoid function is applied in three other gates (input, forget, and output). As tanh is required for the gating logic to function, PyTorch does not allow other activation functions. For this reason, the number of unique possible model configurations is 90.

The code is set-up to iterate through each possible unique architecture, as described in the “Homework 3” document (162). To prevent the code from breaking, the models.py code is set-up to allow the LSTM and BiLSTM functions to ignore activation function calls other than tanh. However, as the code logs these calls as a unique architectures, the LSTM and BiLSTM models are trained thrice, using the tanh activation function each time. The duplication of the LSTM and BiLSTM models is redundant but allows for comparison of training rates and evaluation metrics for each trained model.

In addition to these model configurations, the following model design notes were provided by "Homework 3":

- **Embedding Layer:** size 100
- **Hidden Layers:** 2, size 64
- **Dropout:** 0.3-0.5 (reduces overfitting)
- **Batch Size:** 32
- **Output Layer:** Fully Connected with Sigmoid Activation for Binary Classification
- **Loss:** Binary Cross-Entropy Loss

With these specifications, the parameters for each configuration of models is mostly fixed. The Embedding Layer is 1,000,000 parameters; the embedding layer parameters is the vocabulary size (10,000) times the embedding dimension size (100). With these embedding layer parameters, the following table highlights the increasing number of parameters for the RNN, LSTM, and BiLSTM architectures.

Model	Embedding Layer	Layer 1 (Input to Hidden)	Layer 2 (Hidden to Hidden)	Fully Connected Layer	Total Parameters
RNN	1,000,000	10,624	8,320	65	1,019,009
LSTM	1,000,000	42,496	33,280	65	1,075,841
BiLSTM	1,000,000	84,992	99,328	129	1,184,449

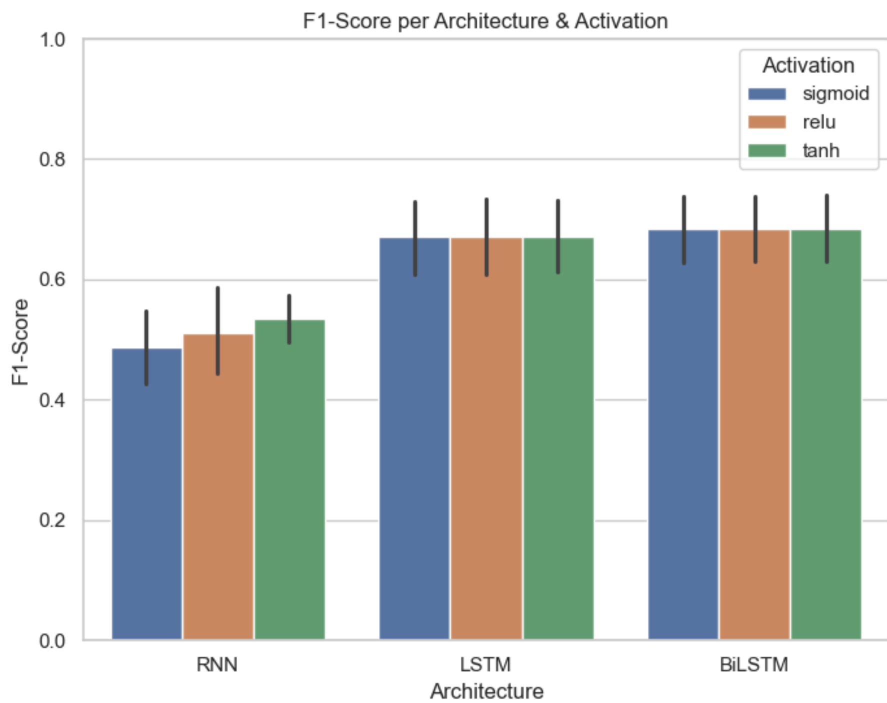
The models are trained in an iterative loop, going through a list of all possible combinations. Each model is saved and the batch loss at each training epoch is recorded. After training all models, the models are evaluated using the specified evaluation metrics:

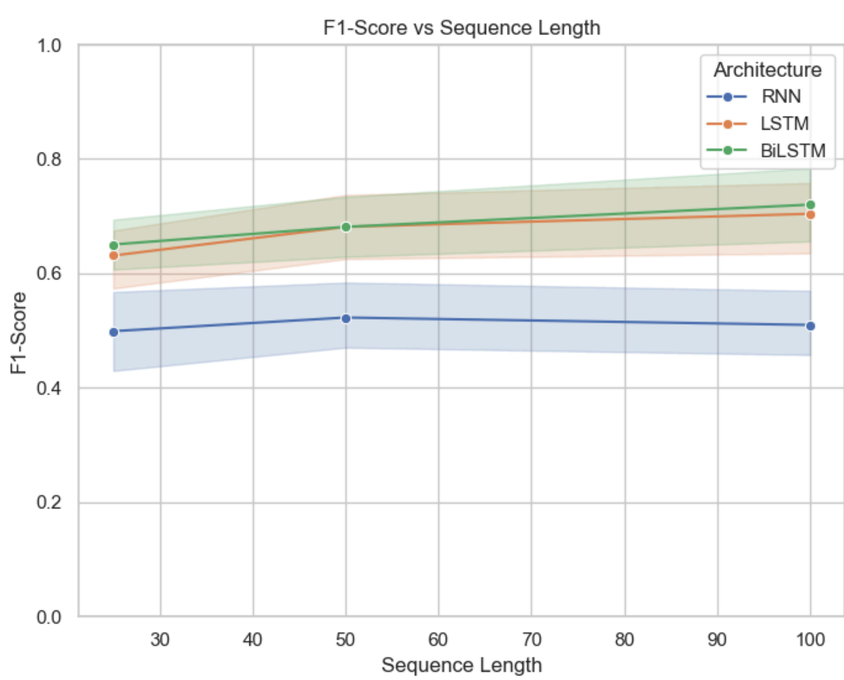
- Accuracy
- F1-Score (macro)
- Training time per Epoch (seconds)

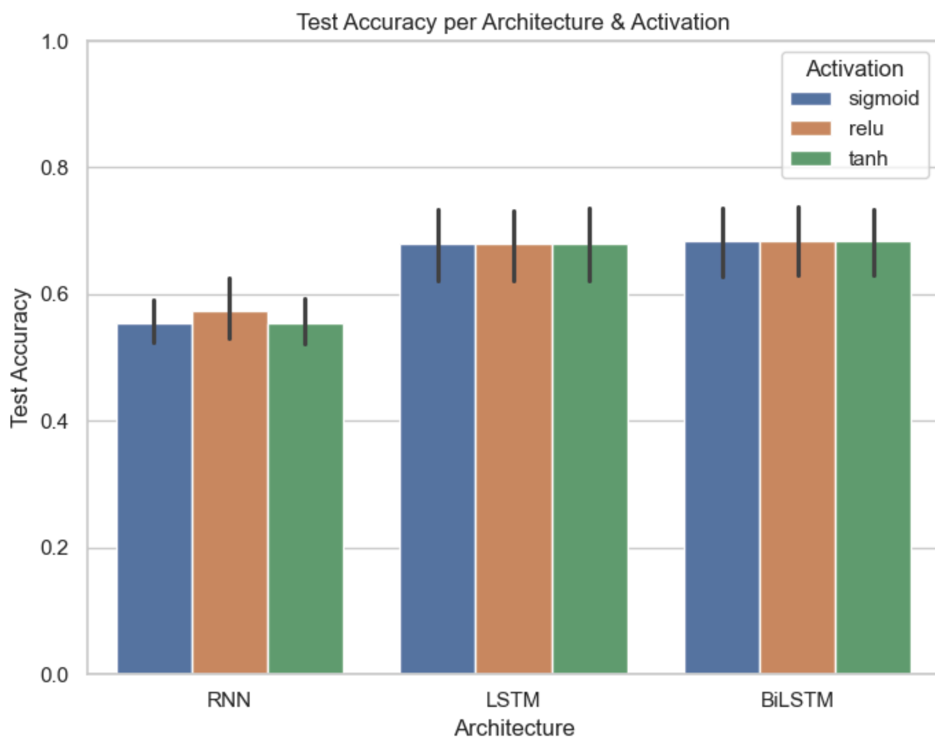
These metrics are recorded, and the best and worst models are identified from the total metric dataframe. From the recorded model metrics, plots for Accuracy vs. Sequence Length, F1-Score vs. Sequence Length, and Training Loss vs. Epochs for the best and worst models are created.

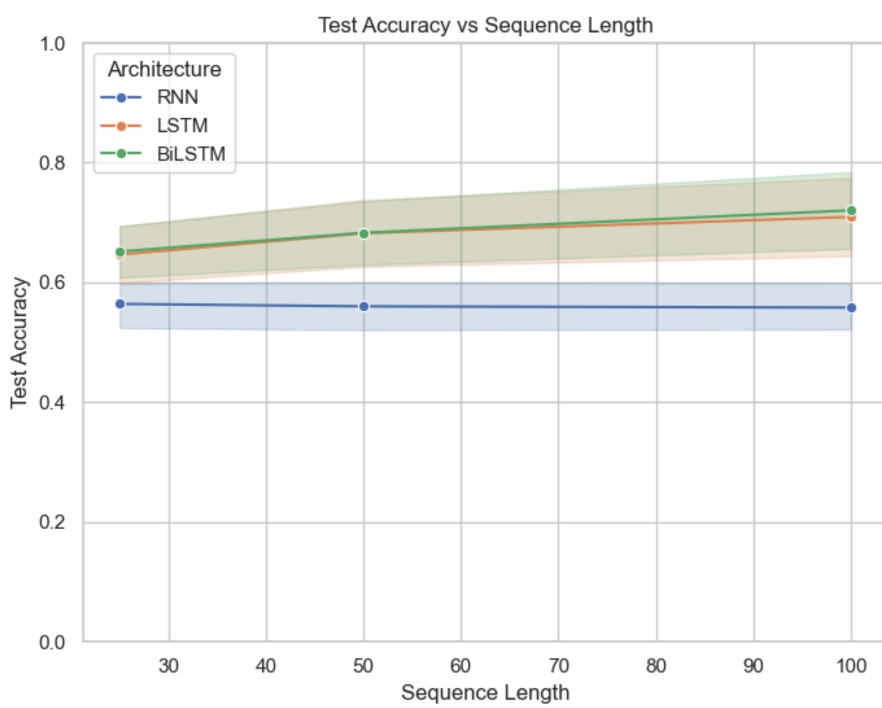
Comparative Analysis & Discussion:

The full list of results for all 162 trained models (90 unique) is included in **Appendix 1**. This table contains the configuration details (model, activation, optimizer, sequence length, gradient clipping) and the evaluation results (accuracy, F1, time per epoch, and total training time). The table also includes a “Real Activation” column to highlight the LSTM and BiLSTM application of tanh despite ‘sigmoid’ or ‘relu’ activation function parameter calls. The table is in order of execution in the code. The following plots highlight the results from these models as functions of evaluation metric by activation function and sequence length.









These graphs illustrate that LSTM and BiLSTM architectures generally perform better than RNN architectures, even with varying activation functions and sequence lengths. Changing activation functions and sequence length only marginally changes the model outcomes within the architecture category, highlighting that if computational costs are high, smaller models may be sufficient to effectively model the data within the IMDb dataset. The largest performance improvements occur between RNN and LSTM and between sequence size 25 and 50 for LSTM (and BiLSTM), suggesting that the medium sized parameter model may be the best model.

The following tabular display highlights the average, maximum, and minimum values for the accuracy, F1-score (macro) and time per epoch (in seconds) metrics for each model configuration. This display highlights the range of results for each parameter, regardless of other configuration parameters. Notably, measures of accuracy, F1 scores, and time per epoch increase moving from RNN to LSTM to BiLSTM, with only marginal gains between LSTM and BiLSTM model types. Additionally, the tanh activation function performs better in terms of accuracy and F1-scores in these models compared to the sigmoid and ReLu functions, but this may be a result of the number of tanh activation function models. The time per epoch for the tanh activation function is substantially higher than the other activation functions. This may be due to the skew in models trained with tanh towards longer LSTM and BiLSTM architectures. The adam optimizer outperformed SGD and RMSProp in accuracy, F1-score, and time per epoch. Finally and importantly, clipping and sequence length do not seem to indicate significant differences between the parameters, with accuracy and F1-scores in the same relative ranges for average, maximum, and minimum values. Differences between these parameters appear in the time cost, with the maximum time per epoch for longer sequences more than doubling shorter sequences. Interestingly, F1-scores and accuracy converge for this dataset. This may suggest that the dataset is balanced, or the binary classification task is almost always correct due to the forced binning into 0 or 1 classification.

Accuracy				
Option	**Configuration**	**Average Test Accuracy**	**Maximum Accuracy**	**Minimum Accuracy**
Architecture	RNN	0.5604644444	0.77752	0.49528
Architecture	LSTM	0.6791333333	0.81792	0.50448
Architecture	BiLSTM	0.6847133333	0.81796	0.52128
Activation	sigmoid	0.55396	0.70352	0.49932
Activation	relu	0.5734511111	0.77752	0.49932
Activation	tanh	0.6636460317	0.81796	0.49528
Optimizer	adam	0.7329807407	0.81796	0.53448
Optimizer	sgd	0.5137533333	0.52804	0.49932
Optimizer	rmsprop	0.677577037	0.8168	0.49528
Sequence_Length	25	0.6204718519	0.72056	0.49932

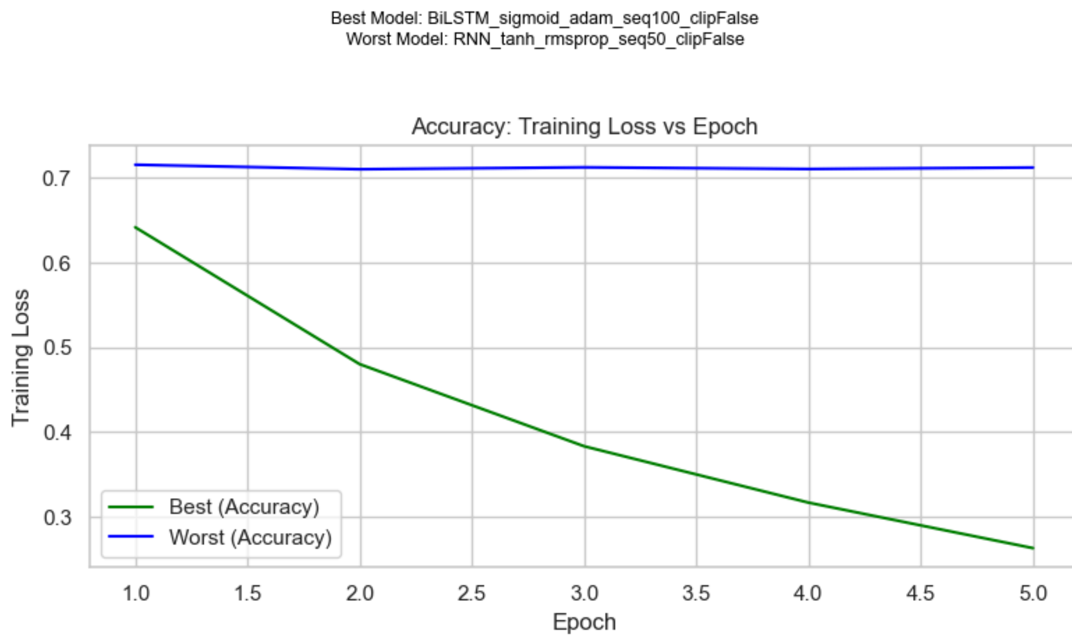
Sequence_Length	50	0.6414585185	0.7664	0.49528
Sequence_Length	100	0.6623807407	0.81796	0.4964
Clipping	TRUE	0.6458187654	0.8168	0.4964
Clipping	FALSE	0.6370553086	0.81796	0.49528
F1-Score (macro)				
Option	**Configuration**	**Average F1**	**Maximum F1**	**Minimum F1**
Architecture	RNN	0.5103483939	0.7771648885	0.333030974
Architecture	LSTM	0.6720509092	0.8177721931	0.4596622235
Architecture	BiLSTM	0.6838638868	0.8176594371	0.5192362541
Activation	sigmoid	0.4861542612	0.7018338585	0.333030974
Activation	relu	0.5105567644	0.7771648885	0.333030974
Activation	tanh	0.657439792	0.8177721931	0.4174620959
Optimizer	adam	0.7302870101	0.8177721931	0.5245665663
Optimizer	sgd	0.4928593197	0.5278188825	0.333030974
Optimizer	rmsprop	0.6431168601	0.8167898587	0.333030974
Sequence_Length	25	0.5933654018	0.7201859912	0.333030974
Sequence_Length	50	0.6283238581	0.7663341401	0.3336354186
Sequence_Length	100	0.6445739301	0.8177721931	0.333030974
Clipping	TRUE	0.6250899628	0.8167898587	0.333030974
Clipping	FALSE	0.6190854971	0.8177721931	0.333030974
Time per Epoch (seconds)				
Option	**Configuration**	**Average Time**	**Maximum Time**	**Minimum Time**
Architecture	RNN	4.209531958	7.126180792	1.85425415
Architecture	LSTM	27.64820865	200.9129839	4.241625929
Architecture	BiLSTM	30.30049609	282.2603531	8.438465261
Activation	sigmoid	4.432817798	7.126180792	2.090245628
Activation	relu	3.840587126	6.174719381	1.85425415
Activation	tanh	25.45732931	282.2603531	2.074538851
Optimizer	adam	15.00869385	200.9129839	2.719475365
Optimizer	sgd	26.68030918	282.2603531	1.85425415
Optimizer	rmsprop	20.46923368	188.8842692	2.406712008
Sequence_Length	25	15.43438824	188.8842692	1.85425415
Sequence_Length	50	11.86731185	126.0163296	2.879447412
Sequence_Length	100	34.85653661	282.2603531	5.003467226
Clipping	TRUE	21.33507624	282.2603531	2.023707676

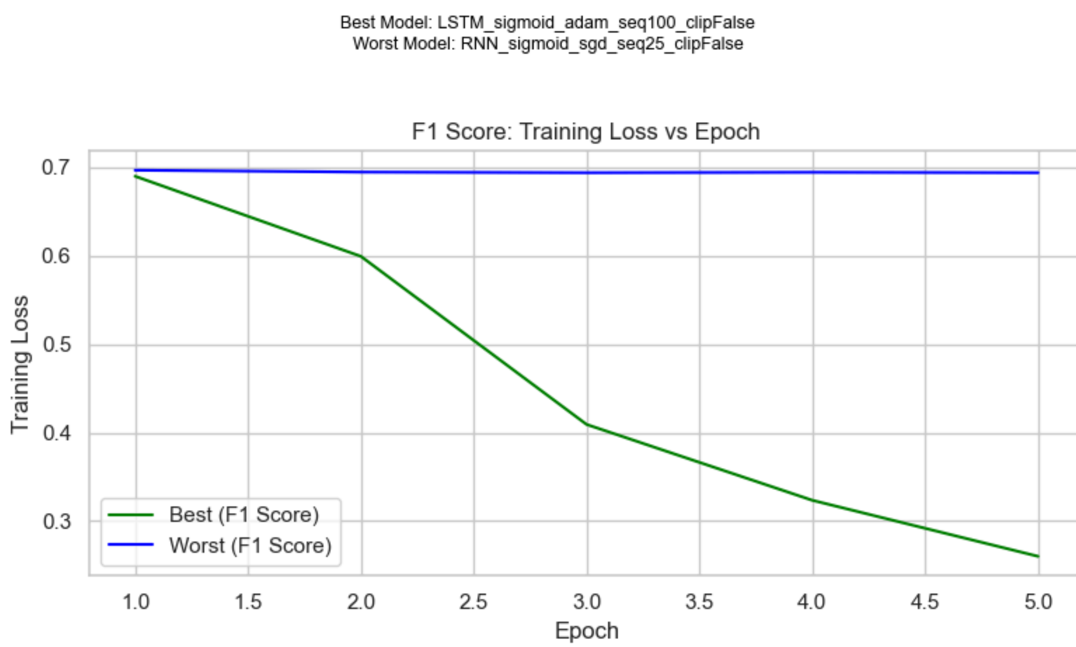
Clipping	FALSE	20.10374823	194.5273928	1.85425415

After calculating these metrics per model, the code identified the best and worst model in terms of F1-score and Accuracy. [The metrics per model is included in Appendix 1.] The best and worst models are as follows:

Accuracy		
	Best	**Worst**
Model	BiLSTM_sigmoid [tanh]_adam_seq100_clipFalse	RNN_tanh_rmsprop_seq50_clipFalse
Accuracy	0.81796	0.49528
F1-Score	0.8176594371	0.4614248893
Time per Epoch	32.30282459	3.874368429
Total Training Time	185.267163	3.874368429
F1-Score		
	Best	**Worst**
Model	LSTM_sigmoid [tanh]_adam_seq100_clipFalse	RNN_sigmoid_sgd_seq25_clipFalse
Accuracy	0.81792	0.49932
F1-Score	0.8177721931	0.333030974
Time per Epoch	15.85894113	2.090245628
Total Training Time	91.70776081	13.29218817

The following graphs highlight the loss per training epoch for the best and worst models. They illustrate that for the best models, F1-scores and accuracy improve over training epochs while the worst models stay relatively consistent over time. However, while the graphs highlight improvement across time periods, they do not highlight the amount of total training time required by the best models. The BiLSTM model takes approximately two times as long as the best LSTM model with equivalent results for F1-score and accuracy. This reinforces a conclusion from the earlier graphs and tables: the marginal improvement between LSTM and BiLSTM for this experiment may not be justified due to the increased computational costs.





Overall, the LSTM configuration performed the best (LSTM_sigmoid [tanh]_adam_seq100_clipFalse) due to the high F1-score, high accuracy, and lower training time. The table above highlights that sequence length did not noticeably improve performance when other parameters are held constant, but the graph highlights that moving from sequence length 25 to 50 provided the highest performance jump while sequence length 50 to 100 was minimal. The adam optimizer outperformed the SGD and RMSProp, but was also part of the worst performing model as well. Finally, gradient clipping did not show meaningful impact to stability with clipping versus non-clipping models performing nearly equal (holding other parameters constant). It is important to note that the best and worst performing models only include clippingFalse.

Conclusion:

This project trained 162 models with 90 unique model configurations. The code took over 6 hours to run using a laptop with arm CPU, 14 CPU cores, 48 GB of RAM, 14 logical processors, and no GPUs. With this configuration, the optimal model is as follows:

- **Architecture:** LSTM
- **Activation Function:** tanh
- **Optimizer:** adam
- **Sequence Length:** 100
- **Stability Strategy:** no gradient clipping

With less computational power, shorter sequence lengths with this architecture (clipping or no clipping) is justifiable as they will return similar but marginally lower results at lower cost.

Citations:

Hassan N. “Homework 3.” *ELMS*, 2025, https://umd.instructure.com/courses/1395714/assignments/7388552.

Lakshmipathi N. “IMDB Dataset of 50K Movie Reviews.” *Kaggle.com*, 2019, www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews/code.

Maas, Andrew L., Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. “Learning Word Vectors for Sentiment Analysis.” *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics*, June 2011, Portland, Oregon, pp. 142–150. http://www.aclweb.org/anthology/P11-1015

Appendices:

Appendix 1: Tabular Model Results

Architecture	Activation	Real Activation	Optimizer	Sequence Length	Gradient Clipping	Train Accuracy	F1 Score	Time per Epoch	Total Training Time
RNN	sigmoid	sigmoid	adam	25	FALSE	0.49868	0.6622353308	3.002891588	17.75739503
RNN	sigmoid	sigmoid	adam	25	TRUE	0.50016	0.6914388625	3.243323755	19.04986691
RNN	sigmoid	sigmoid	adam	50	FALSE	0.50388	0.5854415928	4.360077095	25.75948787
RNN	sigmoid	sigmoid	adam	50	TRUE	0.50024	0.6573972367	4.498740244	26.26622415
RNN	sigmoid	sigmoid	adam	100	FALSE	0.501	0.5535074052	6.939178658	40.62146688
RNN	sigmoid	sigmoid	adam	100	TRUE	0.49868	0.7018338585	7.126180792	41.52116299
RNN	sigmoid	sigmoid	sgd	25	FALSE	0.50068	0.333030974	2.090245628	13.29218817
RNN	sigmoid	sigmoid	sgd	25	TRUE	0.50068	0.3359102607	2.339451838	14.54074097
RNN	sigmoid	sigmoid	sgd	50	FALSE	0.49792	0.4979297701	3.347254276	20.46869111
RNN	sigmoid	sigmoid	sgd	50	TRUE	0.49952	0.3539373584	3.613185835	21.80709004
RNN	sigmoid	sigmoid	sgd	100	FALSE	0.50008	0.4305966367	5.978465891	35.50540924
RNN	sigmoid	sigmoid	sgd	100	TRUE	0.50088	0.4357470438	6.390799379	37.6341331
RNN	sigmoid	sigmoid	rmsprop	25	FALSE	0.50068	0.333030974	2.632794666	16.03041101
RNN	sigmoid	sigmoid	rmsprop	25	TRUE	0.50088	0.3335099051	2.864991903	17.22864509
RNN	sigmoid	sigmoid	rmsprop	50	FALSE	0.50124	0.4715287032	3.916500854	23.33573914
RNN	sigmoid	sigmoid	rmsprop	50	TRUE	0.49704	0.5541312161	4.139304352	24.44584608
RNN	sigmoid	sigmoid	rmsprop	100	FALSE	0.50072	0.4716742583	6.533453941	38.29907298
RNN	sigmoid	sigmoid	rmsprop	100	TRUE	0.49772	0.3478953142	6.773879671	39.57990408
RNN	relu	relu	adam	25	FALSE	0.49852	0.6738764588	2.719475365	15.99874687
RNN	relu	relu	adam	25	TRUE	0.50132	0.6814504174	2.993211079	17.38585091
RNN	relu	relu	adam	50	FALSE	0.50052	0.7400798337	3.718479824	21.36956
RNN	relu	relu	adam	50	TRUE	0.50128	0.7429768439	3.946530676	22.52814317
RNN	relu	relu	adam	100	FALSE	0.5034	0.6358197048	5.901535416	33.23644018
RNN	relu	relu	adam	100	TRUE	0.50428	0.7771648885	6.174719381	34.62900496
RNN	relu	relu	sgd	25	FALSE	0.49656	0.4910494876	1.85425415	11.59893227

RNN	relu	relu	sgd	25	TRUE	0.49616	0.4917472314	2.023707676	12.41493392
RNN	relu	relu	sgd	50	FALSE	0.50072	0.4743967292	2.879447412	17.14106393
RNN	relu	relu	sgd	50	TRUE	0.5014	0.4742615222	3.083028078	18.18554902
RNN	relu	relu	sgd	100	FALSE	0.49756	0.5043468398	5.003467226	28.72368097
RNN	relu	relu	sgd	100	TRUE	0.49656	0.5034570684	5.199058151	29.73008108
RNN	relu	relu	rmsprop	25	FALSE	0.50068	0.333030974	2.406712008	14.42445087
RNN	relu	relu	rmsprop	25	TRUE	0.50068	0.333030974	2.652814627	15.67790604
RNN	relu	relu	rmsprop	50	FALSE	0.49932	0.3336354186	3.466333866	20.11302614
RNN	relu	relu	rmsprop	50	TRUE	0.49932	0.3336354186	3.645373678	21.03711295
RNN	relu	relu	rmsprop	100	FALSE	0.50068	0.333030974	5.562927055	31.53115535
RNN	relu	relu	rmsprop	100	TRUE	0.50068	0.333030974	5.899492598	33.28600717
RNN	tanh	tanh	adam	25	FALSE	0.50304	0.6710592304	2.956381416	17.66001606
RNN	tanh	tanh	adam	25	TRUE	0.50032	0.6989555609	3.132901335	18.57454515
RNN	tanh	tanh	adam	50	FALSE	0.49508	0.5245665663	4.212598372	24.79703307
RNN	tanh	tanh	adam	50	TRUE	0.50012	0.7107065326	4.405457449	25.76753283
RNN	tanh	tanh	adam	100	FALSE	0.50292	0.5254017669	6.743455219	39.26810312
RNN	tanh	tanh	adam	100	TRUE	0.5026	0.6776551939	6.992983294	40.550704
RNN	tanh	tanh	sgd	25	FALSE	0.49756	0.509917749	2.074538851	13.19647574
RNN	tanh	tanh	sgd	25	TRUE	0.4972	0.5104897446	2.299548721	14.30744076
RNN	tanh	tanh	sgd	50	FALSE	0.50052	0.51107133	3.384439182	20.64453602
RNN	tanh	tanh	sgd	50	TRUE	0.50068	0.5120149755	3.542253256	21.4096272
RNN	tanh	tanh	sgd	100	FALSE	0.49976	0.5081790253	5.92268815	35.18368578
RNN	tanh	tanh	sgd	100	TRUE	0.4998	0.5078183719	6.098359871	36.03013682
RNN	tanh	tanh	rmsprop	25	FALSE	0.49676	0.4763055224	2.591219759	15.76564717
RNN	tanh	tanh	rmsprop	25	TRUE	0.50172	0.4174620959	2.829912806	16.93888831
RNN	tanh	tanh	rmsprop	50	FALSE	0.50292	0.4614248893	3.874368429	23.0687921
RNN	tanh	tanh	rmsprop	50	TRUE	0.50132	0.4680135242	4.106151056	24.20842505
RNN	tanh	tanh	rmsprop	100	FALSE	0.50204	0.4937382663	6.468196917	37.87905836
RNN	tanh	tanh	rmsprop	100	TRUE	0.49616	0.4332344654	6.757983017	39.40540314
LSTM	sigmoid	tanh	adam	25	FALSE	0.49684	0.7162555838	5.328887272	31.2467792

LSTM	sigmoid	tanh	adam	25	TRUE	0.49716	0.7127892494	5.579943848	32.52608299
LSTM	sigmoid	tanh	adam	50	FALSE	0.49572	0.766166755	8.898771858	51.94928908
LSTM	sigmoid	tanh	adam	50	TRUE	0.49876	0.7639771092	9.091915703	52.90944672
LSTM	sigmoid	tanh	adam	100	FALSE	0.49696	0.8177721931	15.85894113	91.70776081
LSTM	sigmoid	tanh	adam	100	TRUE	0.49796	0.8030660525	16.15519462	93.33997512
LSTM	sigmoid	tanh	sgd	25	FALSE	0.5032	0.4596622235	4.262915516	25.91278291
LSTM	sigmoid	tanh	sgd	25	TRUE	0.5032	0.4596622235	4.606488323	27.62815738
LSTM	sigmoid	tanh	sgd	50	FALSE	0.49868	0.513140567	7.743811846	45.78611302
LSTM	sigmoid	tanh	sgd	50	TRUE	0.49868	0.513140567	7.841633463	46.28558898
LSTM	sigmoid	tanh	sgd	100	FALSE	0.4996	0.4978356383	14.92021036	86.9522059
LSTM	sigmoid	tanh	sgd	100	TRUE	0.4996	0.4978356383	15.1167347	88.24968195
LSTM	sigmoid	tanh	rmsprop	25	FALSE	0.49436	0.7174018932	4.919689512	29.17807484
LSTM	sigmoid	tanh	rmsprop	25	TRUE	0.495	0.7201859912	5.127979231	30.22980189
LSTM	sigmoid	tanh	rmsprop	50	FALSE	0.49404	0.7642449102	8.218925667	48.03357506
LSTM	sigmoid	tanh	rmsprop	50	TRUE	0.49584	0.7661827767	8.6219203	50.45806885
LSTM	sigmoid	tanh	rmsprop	100	FALSE	0.5	0.7951851153	15.32045302	89.14200115
LSTM	sigmoid	tanh	rmsprop	100	TRUE	0.49888	0.8124118786	15.34901004	88.90907431
LSTM	relu	tanh	adam	25	FALSE	0.49684	0.7162555838	5.222302103	30.65103483
LSTM	relu	tanh	adam	25	TRUE	0.49716	0.7127892494	5.419797516	31.63113594
LSTM	relu	tanh	adam	50	FALSE	0.49572	0.766166755	8.273142004	48.17567301
LSTM	relu	tanh	adam	50	TRUE	0.49876	0.7639771092	8.764283371	51.04922295
LSTM	relu	tanh	adam	100	FALSE	0.49696	0.8177721931	15.41742902	89.58840585
LSTM	relu	tanh	adam	100	TRUE	0.49796	0.8030660525	15.79822454	91.49984813
LSTM	relu	tanh	sgd	25	FALSE	0.5032	0.4596622235	4.241625929	25.67121887
LSTM	relu	tanh	sgd	25	TRUE	0.5032	0.4596622235	4.464231634	26.78298903
LSTM	relu	tanh	sgd	50	FALSE	0.49868	0.513140567	7.315238523	43.44292307
LSTM	relu	tanh	sgd	50	TRUE	0.49868	0.513140567	8.037494087	47.44994998
LSTM	relu	tanh	sgd	100	FALSE	0.4996	0.4978356383	194.5273928	984.86536
LSTM	relu	tanh	sgd	100	TRUE	0.4996	0.4978356383	200.0792174	1012.469076
LSTM	relu	tanh	rmsprop	25	FALSE	0.49436	0.7174018932	188.8842692	948.7960241

LSTM	relu	tanh	rmsprop	25	TRUE	0.495	0.7201859912	5.109588671	283.6502569
LSTM	relu	tanh	rmsprop	50	FALSE	0.49404	0.7642449102	8.378630018	1037.885952
LSTM	relu	tanh	rmsprop	50	TRUE	0.49584	0.7661827767	8.63845892	50.29177594
LSTM	relu	tanh	rmsprop	100	FALSE	0.5	0.7951851153	63.00787063	327.3303239
LSTM	relu	tanh	rmsprop	100	TRUE	0.49888	0.8124118786	15.98477087	92.31772518
LSTM	tanh	tanh	adam	25	FALSE	0.49684	0.7162555838	5.211719084	30.46963
LSTM	tanh	tanh	adam	25	TRUE	0.49716	0.7127892494	5.425405455	31.53943014
LSTM	tanh	tanh	adam	50	FALSE	0.49572	0.766166755	8.700883007	50.45926976
LSTM	tanh	tanh	adam	50	TRUE	0.49876	0.7639771092	9.076536655	52.53227115
LSTM	tanh	tanh	adam	100	FALSE	0.49696	0.8177721931	15.60462351	90.23835111
LSTM	tanh	tanh	adam	100	TRUE	0.49796	0.8030660525	200.9129839	1944.223771
LSTM	tanh	tanh	sgd	25	FALSE	0.5032	0.4596622235	4.285694361	25.77157712
LSTM	tanh	tanh	sgd	25	TRUE	0.5032	0.4596622235	80.97807703	409.2318718
LSTM	tanh	tanh	sgd	50	FALSE	0.49868	0.513140567	126.0163296	637.0888913
LSTM	tanh	tanh	sgd	50	TRUE	0.49868	0.513140567	7.769250441	45.60438609
LSTM	tanh	tanh	sgd	100	FALSE	0.4996	0.4978356383	14.83741059	86.22267079
LSTM	tanh	tanh	sgd	100	TRUE	0.4996	0.4978356383	15.20678988	88.22017598
LSTM	tanh	tanh	rmsprop	25	FALSE	0.49436	0.7174018932	4.91218605	28.99340916
LSTM	tanh	tanh	rmsprop	25	TRUE	0.495	0.7201859912	5.162683249	30.2413702
LSTM	tanh	tanh	rmsprop	50	FALSE	0.49404	0.7642449102	8.304097319	48.40782714
LSTM	tanh	tanh	rmsprop	50	TRUE	0.49584	0.7661827767	8.660620403	50.36631918
LSTM	tanh	tanh	rmsprop	100	FALSE	0.5	0.7951851153	15.57706337	90.23204279
LSTM	tanh	tanh	rmsprop	100	TRUE	0.49888	0.8124118786	15.83351965	91.62536907
BiLSTM	sigmoid	tanh	adam	25	FALSE	0.50328	0.7137214784	9.779121065	56.53948116
BiLSTM	sigmoid	tanh	adam	25	TRUE	0.50404	0.7178626795	10.0527432	57.95730615
BiLSTM	sigmoid	tanh	adam	50	FALSE	0.5026	0.7663341401	17.40449915	100.3983178
BiLSTM	sigmoid	tanh	adam	50	TRUE	0.50248	0.762739391	17.67348619	101.799592
BiLSTM	sigmoid	tanh	adam	100	FALSE	0.49456	0.8176594371	32.30282459	185.267163
BiLSTM	sigmoid	tanh	adam	100	TRUE	0.494	0.8162996849	32.59640756	186.6776428
BiLSTM	sigmoid	tanh	sgd	25	FALSE	0.50424	0.5192362541	8.438465261	49.81666112

BiLSTM	sigmoid	tanh	sgd	25	TRUE	0.50424	0.5192362541	8.699323416	51.09109688
BiLSTM	sigmoid	tanh	sgd	50	FALSE	0.50008	0.5204291616	15.9295629	92.86494613
BiLSTM	sigmoid	tanh	sgd	50	TRUE	0.50008	0.5204291616	16.21054392	94.33483219
BiLSTM	sigmoid	tanh	sgd	100	FALSE	0.50604	0.5278188825	30.86074553	177.92259
BiLSTM	sigmoid	tanh	sgd	100	TRUE	0.506	0.5277404779	31.0124547	178.4458268
BiLSTM	sigmoid	tanh	rmsprop	25	FALSE	0.50228	0.7139376127	9.242186403	53.88484073
BiLSTM	sigmoid	tanh	rmsprop	25	TRUE	0.50632	0.7181152042	9.521737099	55.27708507
BiLSTM	sigmoid	tanh	rmsprop	50	FALSE	0.49964	0.7613325748	16.64259558	96.46165204
BiLSTM	sigmoid	tanh	rmsprop	50	TRUE	0.50092	0.755995845	16.9875484	97.53991985
BiLSTM	sigmoid	tanh	rmsprop	100	FALSE	0.49528	0.8138718653	30.98030791	178.1863902
BiLSTM	sigmoid	tanh	rmsprop	100	TRUE	0.49364	0.8167898587	31.4876163	180.6950078
BiLSTM	relu	tanh	adam	25	FALSE	0.50328	0.7137214784	9.56979394	55.52644825
BiLSTM	relu	tanh	adam	25	TRUE	0.50404	0.7178626795	9.75200243	56.41654587
BiLSTM	relu	tanh	adam	50	FALSE	0.5026	0.7663341401	16.707758	96.78463197
BiLSTM	relu	tanh	adam	50	TRUE	0.50248	0.762739391	17.01148577	98.37819219
BiLSTM	relu	tanh	adam	100	FALSE	0.49456	0.8176594371	33.9801867	193.1115849
BiLSTM	relu	tanh	adam	100	TRUE	0.494	0.8162996849	33.71329904	191.701431
BiLSTM	relu	tanh	sgd	25	FALSE	0.50424	0.5192362541	8.474189949	49.8914392
BiLSTM	relu	tanh	sgd	25	TRUE	0.50424	0.5192362541	8.699625254	50.98176503
BiLSTM	relu	tanh	sgd	50	FALSE	0.50008	0.5204291616	15.91109128	92.74129105
BiLSTM	relu	tanh	sgd	50	TRUE	0.50008	0.5204291616	16.10335889	93.66392493
BiLSTM	relu	tanh	sgd	100	FALSE	0.50604	0.5278188825	95.26965728	1443.404551
BiLSTM	relu	tanh	sgd	100	TRUE	0.506	0.5277404779	282.2603531	2469.849704
BiLSTM	relu	tanh	rmsprop	25	FALSE	0.50228	0.7139376127	155.8802512	787.0250278
BiLSTM	relu	tanh	rmsprop	25	TRUE	0.50632	0.7181152042	132.3699485	669.521312
BiLSTM	relu	tanh	rmsprop	50	FALSE	0.49964	0.7613325748	17.38854609	100.4392798
BiLSTM	relu	tanh	rmsprop	50	TRUE	0.50092	0.755995845	17.63831224	101.5748441
BiLSTM	relu	tanh	rmsprop	100	FALSE	0.49528	0.8138718653	32.77028089	187.847204
BiLSTM	relu	tanh	rmsprop	100	TRUE	0.49364	0.8167898587	32.7600018	187.7473369
BiLSTM	tanh	tanh	adam	25	FALSE	0.50328	0.7137214784	9.993169165	57.8385632

BiLSTM	tanh	tanh	adam	25	TRUE	0.50404	0.7178626795	10.30673504	59.45994186
BiLSTM	tanh	tanh	adam	50	FALSE	0.5026	0.7663341401	17.79039459	102.5688431
BiLSTM	tanh	tanh	adam	50	TRUE	0.50248	0.762739391	17.67474728	101.5186992
BiLSTM	tanh	tanh	adam	100	FALSE	0.49456	0.8176594371	32.88582969	188.445864
BiLSTM	tanh	tanh	adam	100	TRUE	0.494	0.8162996849	33.46587873	191.2834549
BiLSTM	tanh	tanh	sgd	25	FALSE	0.50424	0.5192362541	8.576922035	50.56398988
BiLSTM	tanh	tanh	sgd	25	TRUE	0.50424	0.5192362541	8.824629068	51.83060908
BiLSTM	tanh	tanh	sgd	50	FALSE	0.50008	0.5204291616	16.16858115	94.228544
BiLSTM	tanh	tanh	sgd	50	TRUE	0.50008	0.5204291616	16.55953317	95.99937296
BiLSTM	tanh	tanh	sgd	100	FALSE	0.50604	0.5278188825	31.60902953	181.9427431
BiLSTM	tanh	tanh	sgd	100	TRUE	0.506	0.5277404779	31.75388923	182.5245411
BiLSTM	tanh	tanh	rmsprop	25	FALSE	0.50228	0.7139376127	9.639084673	56.02822709
BiLSTM	tanh	tanh	rmsprop	25	TRUE	0.50632	0.7181152042	9.785176325	56.74246764
BiLSTM	tanh	tanh	rmsprop	50	FALSE	0.49964	0.7613325748	17.21993876	99.54152918
BiLSTM	tanh	tanh	rmsprop	50	TRUE	0.50092	0.755995845	17.32138944	99.85785699
BiLSTM	tanh	tanh	rmsprop	100	FALSE	0.49528	0.8138718653	32.25669899	185.0656421
BiLSTM	tanh	tanh	rmsprop	100	TRUE	0.49364	0.8167898587	32.31285062	185.3217769

Appendix 2: Printed Model Results

The following sections offer three examples of printed terminal results for each of the architectures.

RNN Models

Running: RNN-sigmoid-adam-seq25-clipFalse

Running: RNN-sigmoid-adam-seq25-clipFalse
Epoch 1/5 | Avg Loss: 0.6890 | Time: 2.95s
Epoch 2/5 | Avg Loss: 0.6628 | Time: 3.01s
Epoch 3/5 | Avg Loss: 0.6470 | Time: 3.04s
Epoch 4/5 | Avg Loss: 0.6195 | Time: 2.99s
Epoch 5/5 | Avg Loss: 0.5978 | Time: 3.02s

Final Results | Train Acc: 0.4987 | Test Acc: 0.6624 | F1: 0.6622
Batch loss log saved to:
results/batch_logs/batchloss_RNN_sigmoid_adam_seq25_clipFalse.csv

Total Training Runtime: 16.70s

Running: RNN-relu-rmsprop-seq100-clipFalse

Running: RNN-relu-rmsprop-seq100-clipFalse
Epoch 1/5 | Avg Loss: 49.6732 | Time: 5.57s
Epoch 2/5 | Avg Loss: 49.1808 | Time: 5.56s
Epoch 3/5 | Avg Loss: 50.2478 | Time: 5.55s
Epoch 4/5 | Avg Loss: 49.9121 | Time: 5.58s
Epoch 5/5 | Avg Loss: 49.9760 | Time: 5.54s

Final Results | Train Acc: 0.5007 | Test Acc: 0.4993 | F1: 0.3330

Batch loss log saved to:
results/batch_logs/batchloss_RNN_relu_rmsprop_seq100_clipFalse.csv

Total Training Runtime: 30.38s

Running: RNN-tanh-sgd-seq50-clipFalse

Running: RNN-tanh-sgd-seq50-clipFalse
Epoch 1/5 | Avg Loss: 0.6976 | Time: 3.37s
Epoch 2/5 | Avg Loss: 0.6945 | Time: 3.41s
Epoch 3/5 | Avg Loss: 0.6935 | Time: 3.38s
Epoch 4/5 | Avg Loss: 0.6927 | Time: 3.42s
Epoch 5/5 | Avg Loss: 0.6923 | Time: 3.34s

Final Results | Train Acc: 0.5005 | Test Acc: 0.5123 | F1: 0.5111
Batch loss log saved to:
results/batch_logs/batchloss_RNN_tanh_sgd_seq50_clipFalse.csv

Total Training Runtime: 19.47s

LSTM Models

Running: LSTM-sigmoid-adam-seq25-clipFalse

Warning: Activation argument 'sigmoid' is ignored for LSTM.
Warning: Activation argument 'sigmoid' is ignored for LSTM.

Running: LSTM-sigmoid-adam-seq25-clipFalse
Epoch 1/5 | Avg Loss: 0.6541 | Time: 5.32s
Epoch 2/5 | Avg Loss: 0.5415 | Time: 5.34s
Epoch 3/5 | Avg Loss: 0.4614 | Time: 5.32s
Epoch 4/5 | Avg Loss: 0.3917 | Time: 5.31s
Epoch 5/5 | Avg Loss: 0.3233 | Time: 5.36s

Final Results | Train Acc: 0.4968 | Test Acc: 0.7166 | F1: 0.7163

Batch loss log saved to:
results/batch_logs/batchloss_LSTM_sigmoid_adam_seq25_clipFalse.csv

Total Training Runtime: 30.09s

Running: LSTM-relu-rmsprop-seq100-clipTrue

Warning: Activation argument 'relu' is ignored for LSTM.

Warning: Activation argument 'relu' is ignored for LSTM.

Running: LSTM-relu-rmsprop-seq100-clipTrue

Epoch 1/5 | Avg Loss: 0.6893 | Time: 16.03s

Epoch 2/5 | Avg Loss: 0.5308 | Time: 15.96s

Epoch 3/5 | Avg Loss: 0.4159 | Time: 15.98s

Epoch 4/5 | Avg Loss: 0.3549 | Time: 15.99s

Epoch 5/5 | Avg Loss: 0.3206 | Time: 15.97s

Final Results | Train Acc: 0.4989 | Test Acc: 0.8131 | F1: 0.8124

Batch loss log saved to:
results/batch_logs/batchloss_LSTM_relu_rmsprop_seq100_clipTrue.csv

Total Training Runtime: 91.26s

Running: LSTM-tanh-sgd-seq50-clipFalse

Running: LSTM-tanh-sgd-seq50-clipFalse

Epoch 1/5 | Avg Loss: 0.6934 | Time: 7.77s

Epoch 2/5 | Avg Loss: 0.6932 | Time: 7.55s

Epoch 3/5 | Avg Loss: 0.6932 | Time: 7.48s

Epoch 4/5 | Avg Loss: 0.6931 | Time: 7.49s

Epoch 5/5 | Avg Loss: 0.6930 | Time: 599.79s

Final Results | Train Acc: 0.4987 | Test Acc: 0.5150 | F1: 0.5131

Batch loss log saved to:
results/batch_logs/batchloss_LSTM_tanh_sgd_seq50_clipFalse.csv

Total Training Runtime: 636.04s

BiLSTM Models

Running: BiLSTM-sigmoid-adam-seq25-clipFalse

Warning: Activation argument 'sigmoid' is ignored for BiLSTM.

Warning: Activation argument 'sigmoid' is ignored for BiLSTM.

Running: BiLSTM-sigmoid-adam-seq25-clipFalse

Epoch 1/5 | Avg Loss: 0.6350 | Time: 9.76s

Epoch 2/5 | Avg Loss: 0.5206 | Time: 9.74s

Epoch 3/5 | Avg Loss: 0.4375 | Time: 9.77s

Epoch 4/5 | Avg Loss: 0.3575 | Time: 9.79s

Epoch 5/5 | Avg Loss: 0.2734 | Time: 9.83s

Final Results | Train Acc: 0.5033 | Test Acc: 0.7138 | F1: 0.7137

Batch loss log saved to:
results/batch_logs/batchloss_BiLSTM_sigmoid_adam_seq25_clipFalse.csv

Total Training Runtime: 55.44s

Running: BiLSTM-relu-rmsprop-seq100-clipTrue

Warning: Activation argument 'relu' is ignored for BiLSTM.

Warning: Activation argument 'relu' is ignored for BiLSTM.

Running: BiLSTM-relu-rmsprop-seq100-clipTrue

Epoch 1/5 | Avg Loss: 0.6064 | Time: 32.95s

Epoch 2/5 | Avg Loss: 0.4230 | Time: 32.73s

Epoch 3/5 | Avg Loss: 0.3520 | Time: 32.70s

Epoch 4/5 | Avg Loss: 0.3144 | Time: 32.79s

Epoch 5/5 | Avg Loss: 0.2884 | Time: 32.63s

Final Results | Train Acc: 0.4936 | Test Acc: 0.8168 | F1: 0.8168

Batch loss log saved to:
results/batch_logs/batchloss_BiLSTM_relu_rmsprop_seq100_clipTrue.csv

Total Training Runtime: 186.58s

Running: BiLSTM-tanh-sgd-seq50-clipFalse

Running: BiLSTM-tanh-sgd-seq50-clipFalse

Epoch 1/5 | Avg Loss: 0.6930 | Time: 16.19s

Epoch 2/5 | Avg Loss: 0.6929 | Time: 16.17s

Epoch 3/5 | Avg Loss: 0.6927 | Time: 16.15s

Epoch 4/5 | Avg Loss: 0.6925 | Time: 16.15s

Epoch 5/5 | Avg Loss: 0.6919 | Time: 16.18s

Final Results | Train Acc: 0.5001 | Test Acc: 0.5233 | F1: 0.5204

Batch loss log saved to:
results/batch_logs/batchloss_BiLSTM_tanh_sgd_seq50_clipFalse.csv

Total Training Runtime: 93.07s