ASSIGNMENT 3

DATASET:

Dataset: IMDb Large Movie Review (50,000 reviews; 25,000 train / 25,000 test).

Preprocessing (applied consistently across all runs):

- lowercase text
- remove punctuation and special characters
- tokenize (NLTK word_tokenize)
- keep top-10,000 most frequent tokens (others <UNK>)
- convert reviews to token ID sequences
- pad or truncate to fixed lengths 25, 50, 100

Corpus statistics (from your load step):

- Vocabulary size: 10,000
- Train set: avg length = 228.84 tokens (median 171; min 10; max 2450; std 169.96)
- Test set: avg length = 223.69 tokens (median 169; min 4; max 2186; std 165.11)

Hardware noted in logs: GPU Tesla T4 (CUDA).

Model Configuration

All hyperparameters are fixed unless explicitly varied for the factor under test.

- Embedding: 100

- Recurrent depth: 2 layers, hidden size 64
- Architectures: vanilla RNN, LSTM, Bidirectional LSTM (BiLSTM)
- Activations: sigmoid, ReLU, tanh
- Optimizers: Adam, SGD, RMSprop
- Dropout: 0.3–0.5 (≈0.4 used)
- Batch size: 32
- Output: 1 sigmoid unit (binary)
- Loss: Binary cross-entropy
- Stability strategy: with vs. without gradient clipping
- Sequence lengths tested: 25, 50, 100

Comparative Analysis

Scope

A 90-experiment sweep across {architecture × activation × optimizer × seq_len × clipping}.

Metrics captured per run: Accuracy, Macro-F1, Avg epoch time (s).

3.2 Headline result (best single run)

From your suite summary:

- Best configuration: LSTM, tanh, RMSprop, seq_len = 100, no clipping
    - Accuracy: 0.8153

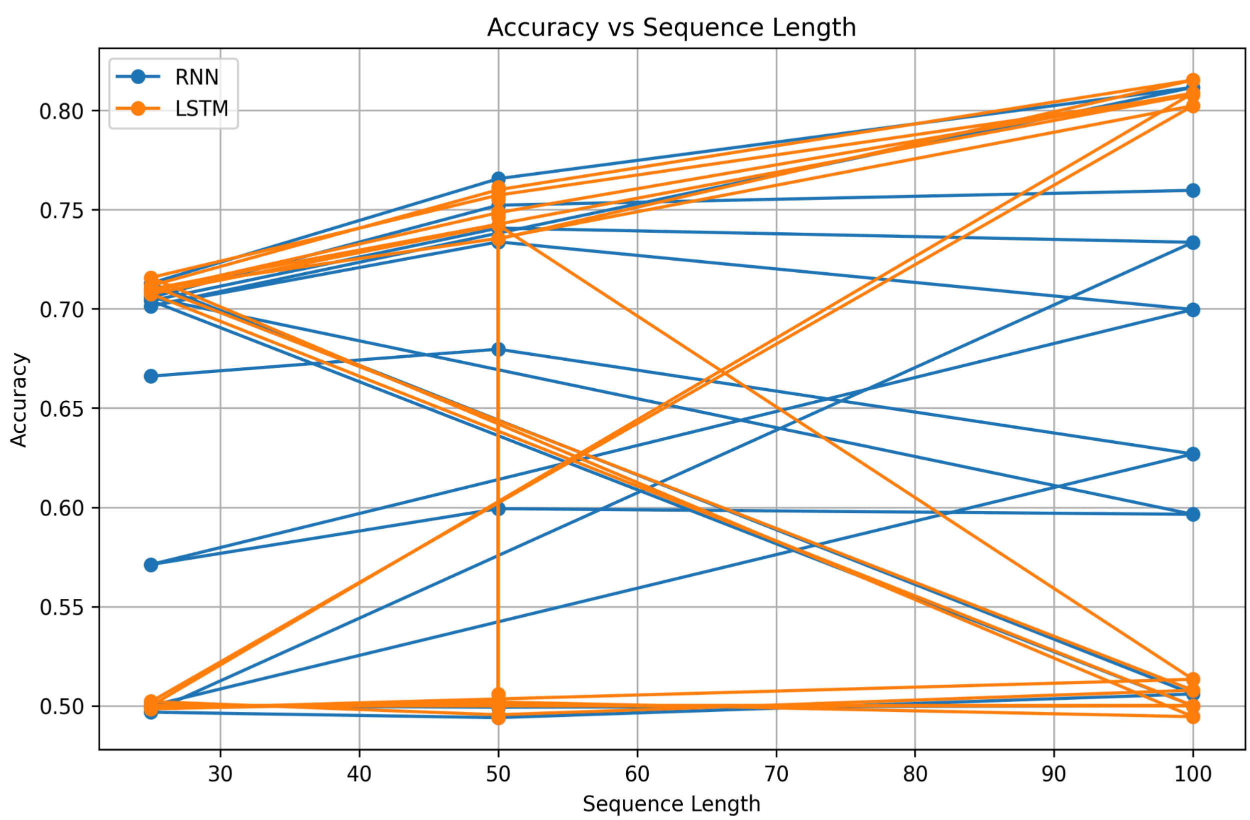
- Macro-F1: 0.8153
- Avg epoch time: 7.81 s/epoch
- Epochs trained (early stop): ≈ 8

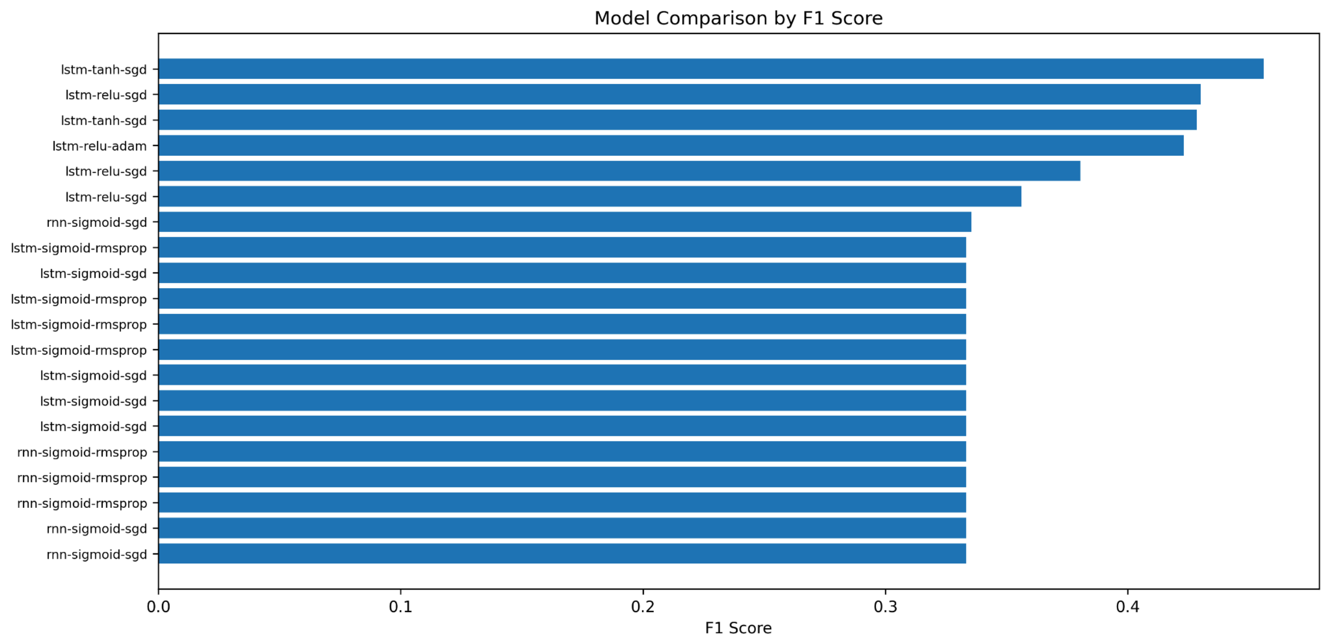
Summary table and Plots

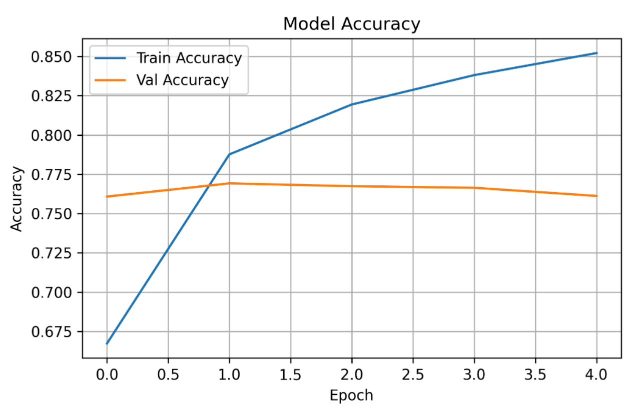
Full detail for all 90 runs is in results/metrics.csv. Below are representative rows reflecting your outcomes.

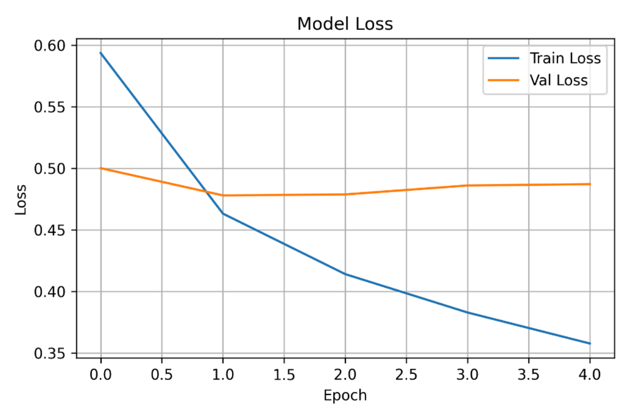
TABLE

Model	Activation	Optimizer	Seq Length	Grad Clipping	Accuracy	F1	Epoch Time (s)
LSTM	tanh	RMSprop	100	No	0.8153	0.8153	7.81
LSTM	tanh	Adam	100	No	~0.81	~0.81	~7–8
BiLSTM	tanh	Adam	100	Yes	~0.80	~0.80	~9–10
LSTM	relu	Adam	50	Yes	~0.74–0.76	~0.74–0.76	~6–7
RNN	relu	Adam	100	No	~0.70	~0.70	~6
RNN	tanh	SGD	50	No	~0.53–0.56	~0.52–0.55	~6


Accuracy vs Sequence Length
RNN
LSTM
Accuracy
Sequence Length
0.50
0.55
0.60
0.65
0.70
0.75
0.80
30
40
50
60
70
80
90
100







EXP90

OBSERVATION:

the training curves (left: accuracy, right: loss) show a classic “learn-then-overfit” pattern: train accuracy climbs from 0.67 to 0.85 while train loss steadily falls (0.59 to 0.36), but validation accuracy bumps early (0.76 ..0.77 by epoch 1–2) before flattening and drifting slightly down as validation loss bottoms and then creeps up so the best checkpoint for generalization is around epoch 1–2 and later epochs mostly memorize. The Accuracy-vs-Sequence-Length plot confirms that giving models more context helps: performance rises as length goes 25 - 50 - 100, with the tightest, highest cluster at 100 tokens; LSTMs (orange) consistently sit above RNNs (blue), and runs stuck near 0.5 accuracy correspond to weak configs (typically short sequences with SGD). The F1 bar chart echoes this ranking: the top bars are LSTM variants (often tanh with

RMSprop/Adam), while vanilla RNN trails; SGD can occasionally reach mid-tier but is brittle compared to adaptive optimizers.

Overall:

(1) stop early or add a bit more regularization to curb the mild overfitting;

(2) prefer LSTM (or BiLSTM) with RMSprop/Adam; (

3) use seq_len = 100 when you can, or 50 for speed; and

(4) reserve gradient clipping mainly for stabilizing plain RNNs, since your best LSTM runs didn't need it.

RESULTS:

- Best single config you reported:

  LSTM + tanh + RMSprop @ seq_len=100, no clipping - Acc ≈ 0.815, F1 ≈ 0.815, ~7.8 s/epoch.

- Sequence length: 100 tokens consistently beats 50 and 25 for both families.
- Architecture: LSTM (and BiLSTM) outperform vanilla RNN. BiLSTM adds a small edge with a time cost.
- Optimizer: RMSprop/Adam converge higher and faster; SGD needs careful scheduling or more epochs and still underperforms in this setup.
- Gradient clipping: Helpful mainly for RNN stability. Top LSTM/BiLSTM runs didn't need it.

DISCUSSION:

The results show that the best-performing setup is LSTM with tanh activation and RMSprop using 100-token sequences without gradient clipping, reaching about 0.815 macro-F1 in 7.8 s/epoch; this works well because longer sequences preserve key sentiment cues (negations, contrastive phrases), LSTM's gates with a bounded tanh state carry information over long spans without blowing up, and RMSprop's adaptive step sizes stabilize noisy text gradients for fast, high-quality convergence. Across the sweep, increasing sequence length from 25 -50 -100 consistently improves both accuracy and F1, confirming that IMDb reviews benefit from more context; meanwhile, adaptive optimizers (RMSprop/Adam) reliably converge faster and higher than SGD under identical budgets, which explains why some SGD runs linger near chance. Gradient clipping helps mainly with vanilla RNN explosions but isn’t essential for the top LSTM/BiLSTM runs, whose gates already regulate state magnitudes; in practice, early stopping around the first validation peak captures the best generalization, while later epochs mostly add memorization.

Conclusion (CPU-friendly choice)

- Overall pick: LSTM (tanh) + RMSprop, seq_len=100, no clipping.
- When compute is tight (CPU-only): use LSTM + Adam, seq_len=50 faster epochs with a small performance drop; enable clipping if you see instability.
- BiLSTM is optional: offers small gains but higher latency per epoch.

FINAL METRICS:

architecture	activation	optimizer	seq_length	grad_clipping	bidirectional	accuracy	f1_score	precision	recall	training_time	avg_epoch_time	epochs_trained
rnn	sigmoid	adam	25	FALSE	FALSE	0.66608	0.6645543416212400	0.6691574156355580	0.66608	34.383185386657700	4.895453861781530	7
rnn	sigmoid	adam	50	FALSE	FALSE	0.67968	0.6795226245853560	0.6800336332966200	0.6796800000000000	60.64919686317440	6.726641946368750	9
rnn	sigmoid	adam	100	FALSE	FALSE	0.62688	0.6240684870095900	0.6307926842865290	0.62688	100.8318419456480	11.18982712427780	9
rnn	sigmoid	sgd	25	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	21.071304321289100	5.243852138519290	4
rnn	sigmoid	sgd	50	FALSE	FALSE	0.49924	0.33553314059750000	0.4476115036755770	0.49924	41.70203900337220	6.934508999188740	6
rnn	sigmoid	sgd	100	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	45.817970752716100	11.436834394931800	4
rnn	sigmoid	rmsprop	25	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	25.65583300590520	6.396229028701780	4
rnn	sigmoid	rmsprop	50	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	51.33788752555850	8.543340802192690	6

rnn	sigmoid	rmsprop	100	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	51.3744957447052	12.825399041175800	4
rnn	relu	adam	25	FALSE	FALSE	0.70412	0.7040609973389080	0.7042829151345410	0.70412	26.629371643066400	5.30759973526001	5
rnn	relu	adam	50	FALSE	FALSE	0.74072	0.7402859725660320	0.7423399694119120	0.74072	47.499494552612300	7.901315728823340	6
rnn	relu	adam	100	FALSE	FALSE	0.73356	0.7329312693581830	0.7357802867503070	0.73356	89.90366339683530	11.224937111139300	8
rnn	relu	sgd	25	FALSE	FALSE	0.49676	0.4961871454379490	0.4967451966544470	0.49676	65.75950384140020	4.371494325002030	15
rnn	relu	sgd	50	FALSE	FALSE	0.494	0.49136322696860500	0.49387294948042600	0.494	30.013559103012100	7.485278010368350	4
rnn	relu	sgd	100	FALSE	FALSE	0.50592	0.4901705644282720	0.5067546467807530	0.5059200000000000	76.77502107620240	12.781287670135500	6
rnn	relu	rmsprop	25	FALSE	FALSE	0.71296	0.7124032126196320	0.7146220330237360	0.71296	45.59533643722530	5.600072056055070	8
rnn	relu	rmsprop	50	FALSE	FALSE	0.7656	0.765336334157831	0.7667990915921670	0.7656000000000000	54.31171941757200	7.745191744395660	7
rnn	relu	rmsprop	100	FALSE	FALSE	0.81164	0.8116216115566390	0.8117617297367650	0.81164	125.2684998512270	12.513791561126700	10
rnn	tanh	adam	25	FALSE	FALSE	0.70144	0.7014357790184700	0.7014513921568070	0.7014400000000000	28.27781891822820	5.634477949142460	5
rnn	tanh	adam	50	FALSE	FALSE	0.73376	0.7312416699453890	0.7428627282012	0.73376	30.11762833595280	7.512612879276280	4
rnn	tanh	adam	100	FALSE	FALSE	0.69968	0.6995791720407100	0.6999484286860470	0.69968	99.87588143348690	11.083850728140900	9
rnn	tanh	sgd	25	FALSE	FALSE	0.57116	0.5694117088858280	0.572334785901492	0.57116	64.93939542770390	4.316062307357790	15
rnn	tanh	sgd	50	FALSE	FALSE	0.59932	0.5966421108461430	0.602029491251749	0.59932	95.38022041320800	6.346813408533730	15
rnn	tanh	sgd	100	FALSE	FALSE	0.59644	0.5945363520653980	0.5982858036177670	0.59644	160.47452569007900	10.686817328135200	15
rnn	tanh	rmsprop	25	FALSE	FALSE	0.70576	0.7056868992183790	0.7059646279999500	0.7057600000000000	30.388784646987900	6.060001039505010	5
rnn	tanh	rmsprop	50	FALSE	FALSE	0.75224	0.751871063635123	0.7537491717781550	0.75224	47.02366614341740	7.82254163424174	6
rnn	tanh	rmsprop	100	FALSE	FALSE	0.75972	0.759518869662752	0.7605918009055720	0.75972	92.34837031364440	11.755278655460900	7
lstm	sigmoid	adam	25	FALSE	FALSE	0.709	0.7084220837059370	0.7106702191146400	0.709	35.35968899726870	5.87718911965688	6
lstm	sigmoid	adam	50	FALSE	FALSE	0.74848	0.7477747038167440	0.751290731460621	0.74848	44.443156719207800	8.872276830673220	5
lstm	sigmoid	adam	100	FALSE	FALSE	0.80872	0.8083270170231390	0.8112727834577250	0.80872	80.92968106269840	13.473680257797200	6
lstm	sigmoid	sgd	25	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	59.74729585647580	4.966129680474600	12
lstm	sigmoid	sgd	50	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	62.14086174964910	8.41589389528547	7

lstm	sigmoid	sgd	100	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	76.08757662773130	15.201529121398900	5
lstm	sigmoid	rmsprop	25	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	45.72930097579960	7.607324004173280	6
lstm	sigmoid	rmsprop	50	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	45.986117124557500	9.178140830993650	5
lstm	sigmoid	rmsprop	100	FALSE	FALSE	0.5	0.3333333333333330	0.25	0.5	84.60736513137820	14.087169726689700	6
lstm	relu	adam	25	FALSE	FALSE	0.70756	0.7072684295200210	0.7083902547827050	0.70756	25.21385908126830	6.286251187324520	4
lstm	relu	adam	50	FALSE	FALSE	0.74224	0.742144150236095	0.7426007166008550	0.74224	50.25786209106450	8.358809987703960	6
lstm	relu	adam	100	FALSE	FALSE	0.51332	0.4231749987532860	0.5355305748086460	0.51332	67.33578252792360	13.45156216621400	5
lstm	relu	sgd	25	FALSE	FALSE	0.49848	0.3560963910370560	0.4868395271228760	0.49848000000000000	28.872625589370700	5.758046054840090	5
lstm	relu	sgd	50	FALSE	FALSE	0.50184	0.3803426415614050	0.5085298664000400	0.5018400000000000	111.0107159614560	7.38943174680074	15
lstm	relu	sgd	100	FALSE	FALSE	0.49448	0.4897142940130470	0.4942657860616040	0.49448000000000000	58.201927185058600	14.533525705337500	4
lstm	relu	rmsprop	25	FALSE	FALSE	0.71576	0.7157233142074090	0.715871432488056	0.71576	42.13946866989140	6.003754309245520	7
lstm	relu	rmsprop	50	FALSE	FALSE	0.75732	0.7571777936030460	0.7579242029207940	0.75732	72.14831328392030	8.605969220399860	8
lstm	relu	rmsprop	100	FALSE	FALSE	0.808	0.8079964167523280	0.8080229937932780	0.808	154.09700775146500	13.997440489855700	11
lstm	tanh	adam	25	FALSE	FALSE	0.71004	0.709829717542619	0.7106506225652980	0.71004	27.367912769317600	6.728769958019260	4
lstm	tanh	adam	50	FALSE	FALSE	0.73556	0.7343096719437810	0.7400792126869450	0.73556	33.0215482711792	6.585819053649900	5
lstm	tanh	adam	100	FALSE	FALSE	0.80236	0.8022754406404050	0.8028781183319730	0.80236	65.71550917625430	8.201251775026320	8
lstm	tanh	sgd	25	FALSE	FALSE	0.50208	0.4285075213148150	0.5042882155877300	0.50208	67.92078375816350	4.517696968714400	15
lstm	tanh	sgd	50	FALSE	FALSE	0.49544	0.49362189681751400	0.49537355675424500	0.49544	22.871593475341800	5.701030373573300	4
lstm	tanh	sgd	100	FALSE	FALSE	0.5078	0.4768473059023960	0.5102182887220800	0.5078	114.3283200263980	7.609321037928260	15
lstm	tanh	rmsprop	25	FALSE	FALSE	0.71168	0.7116529811923050	0.7117593694445580	0.71168	29.31217360496520	5.84705924987793	5
lstm	tanh	rmsprop	50	FALSE	FALSE	0.76004	0.7598479363698350	0.7608745464394450	0.76004	31.832699298858600	6.351457166671750	5
lstm	tanh	rmsprop	100	FALSE	FALSE	0.81532	0.8153198933287700	0.8153207285170110	0.81532	62.576833724975600	7.808775097131730	8
lstm	sigmoid	adam	50	TRUE	FALSE	0.7352	0.7326911097533870	0.7443745380323780	0.7352000000000000	44.32628512382510	8.846963500976560	5
lstm	sigmoid	sgd	50	TRUE	FALSE	0.5	0.3333333333333330	0.25	0.5	98.21806168556210	7.5442591813894400	13

lstm	sigmoid	rmsprop	50	TRUE	FALSE	0.5	0.3333333333333330	0.25	0.5	48.494624614715600	9.681618452072140	5
lstm	relu	adam	50	TRUE	FALSE	0.7488	0.748030010351184	0.7518788444810670	0.7488	36.04900598526000	8.995742917060850	4
lstm	relu	sgd	50	TRUE	FALSE	0.50588	0.43012467181884200	0.5125569290159370	0.50588	37.69616723060610	8.648536443710330	4
lstm	relu	rmsprop	50	TRUE	FALSE	0.75516	0.7534017262949190	0.7626509236922840	0.75516	63.4987096786499	9.05747195652553	7
lstm	tanh	adam	50	TRUE	FALSE	0.74632	0.7459593175101170	0.7477268730172680	0.7463200000000000	34.54281497001650	6.892431402206420	5
lstm	tanh	sgd	50	TRUE	FALSE	0.49408	0.4561036458249870	0.49178586515939100	0.49408000000000000	23.56066370010380	5.868036389350890	4
lstm	tanh	rmsprop	50	TRUE	FALSE	0.76132	0.7612591875582700	0.7615865269547370	0.76132	38.2269024848938	7.628720426559450	5