Predicting bug fix time

Rakesh Pavan (17IT154)

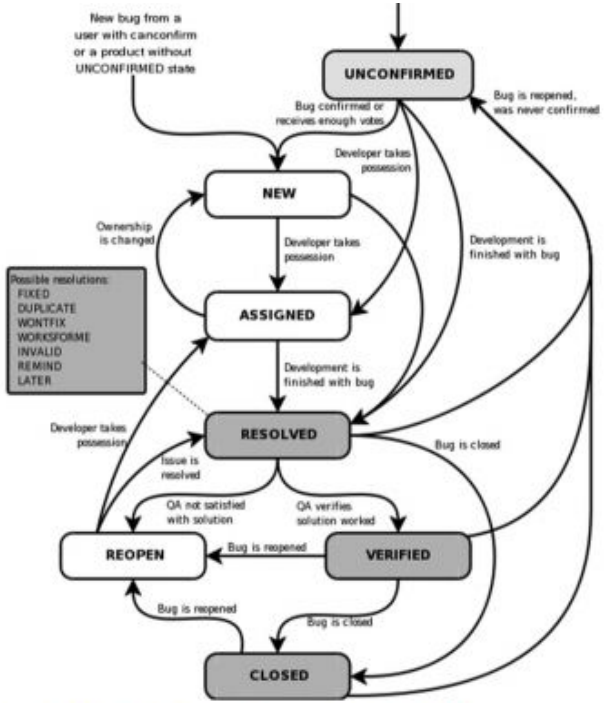
Problem Statement & Objectives

A significant amount of time is spent by software developers in investigating bug reports.

It is useful to indicate when a bug report will be closed, since it would help software teams to prioritise their work.

Thus, in this work we aim to develop a technique to predict the bug fix time using the temporal activity sequence of the bug given as input.

We will make use of the Mozilla Firefox BugZilla database for this purpose (as done in the base paper).



Bug fix life cycle

Literature

Base Paper:

M. Habayeb, S. S. Murtaza, A. Miranskyy and A. B. Bener, "**On the Use of Hidden Markov Model to Predict the Time to Fix Bugs**," in *IEEE Transactions on Software Engineering, vol. 44, no. 12, pp. 1224-1244, 1 Dec. 2018*.

In this paper, a novel approach using Hidden Markov Models and temporal sequences of developer activities is proposed. The approach is empirically demonstrated in a case study using eight years of bug reports collected from the Firefox project.

Other References:

Ardimento, Bilancia, Massimo. “Predicting Bug-Fix Time: Using Standard Versus Topic-Based Text Categorization Techniques". Springer International Publishing, Discovery Science 2016

Zhang, Hongyu & Gong, Liang & Versteeg, Steve. (2013). Predicting bug-fixing time: An empirical study of commercial software projects. Proceedings - International Conference on Software Engineering. 1042-1051. 10.1109/ICSE.2013.6606654.

Methodology

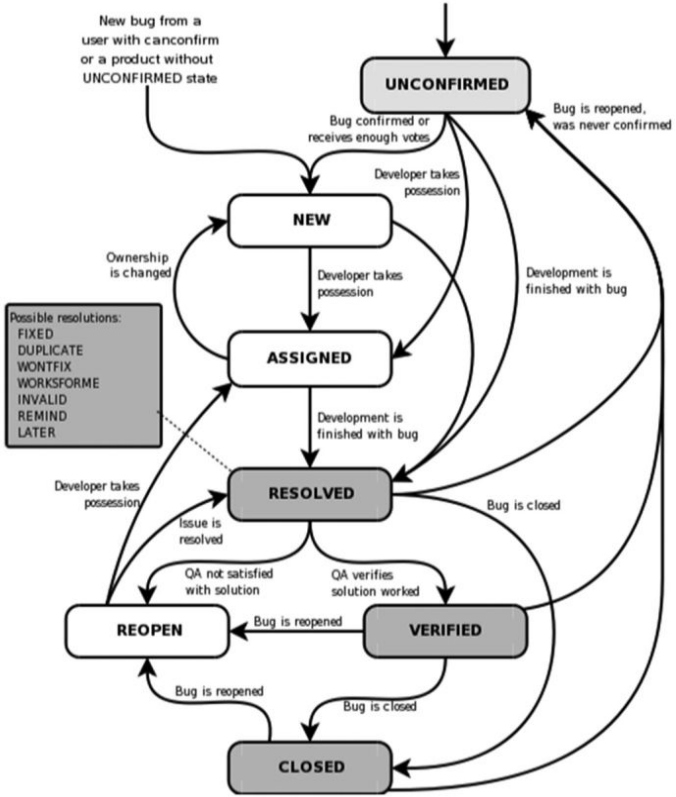
Most of the previous studies have used the frequency of occurrence of certain developer activities as input attributes in building their prediction models. However, these approaches tend to ignore the temporal nature of the occurrence of these activities.

The authors of the base paper take this into consideration, but the use of Markov model gives room for more improvement. Markov model is a class of statistical model which makes the **strong Markovian memorylessness assumption about the state transitions**. Thus not being able to fully utilize the temporal nature of the bug activity sequences.

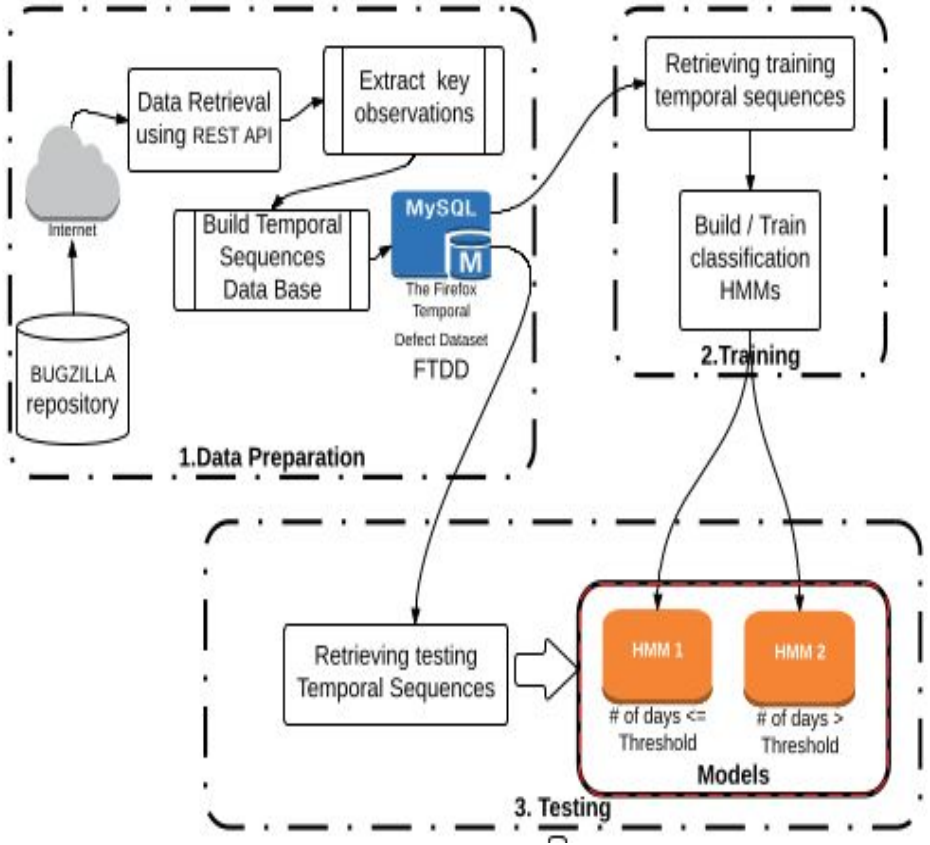
Therefore we propose using recurrent network architectures, starting with RNNs (Recurrent Neural Networks) and trying out LSTM (Long Short Term Memory) and GRU (Gated Recurrent Unit) as alternatives.

The problem of different length of bug life cycles as input will be fixed using temporal sequence padding approach.

Bug Activity Diagram



Extracting the bug activities



Sample Extracted Data - Unprocessed

id	who	when	field_name	added	removed
324056	annie.sullivan@gmail.com	2006-01-19T21:56:17Z	flagtypes.name	review? (bugs@bengoodger.com)	
324056	bugs@bengoodger.com	2006-01-20T01:43:52Z	flagtypes.name	review+	review? (bugs@bengoodger.com)
324056	annie.sullivan@gmail.com	2006-01-20T17:04:07Z	status	RESOLVED	NEW
324056	annie.sullivan@gmail.com	2006-01-20T17:04:07Z	resolution	FIXED	
324063	gavin.sharp@gmail.com	2006-01-19T23:05:08Z	cc	gavin.sharp@gmail.com	

Sample Extracted Data - After Processing

For bug reports resolved below 60 days			For bug reports resolved above 60 days		
Sequence of activities	No. of occurrences	Percentage	Sequence of activities	No. of occurrences	Percentage
NCZ	5,679	37.34%	NWCZ	2,712	15.95%
NCCZ	3,131	20.59%	NCCZ	2,520	14.82%
MCZ	1,465	9.63%	NCWCZ	2,145	12.62%
NCCCZ	918	6.04%	NCZ	2,105	12.38%
MCCZ	799	5.25%	NCCCZ	1,190	7.00%
ECZ	594	3.91%	NCCWCZ	739	4.35%
NDZ	449	2.95%	MCZ	629	3.70%
ECCZ	380	2.50%	MCCZ	585	3.44%
NDCZ	271	1.78%	MWCZ	569	3.35%
MCCCZ	241	1.58%	NCWZ	531	3.12%
NCCCCZ	233	1.53%	NCCCCZ	431	2.54%
NQCZ	159	1.05%	NWCCZ	398	2.34%
ECCCZ	137	0.90%	NWZ	388	2.28%
NCDZ	134	0.88%	MCWCZ	365	2.15%
MDZ	126	0.83%	NCCWZ	351	2.06%
NCWZ	110	0.72%	ECZ	318	1.87%
NQCCZ	107	0.70%	MCCCZ	301	1.77%
NCWCZ	98	0.64%	ECCZ	286	1.68%
NDCCZ	93	0.61%	NCWCCZ	219	1.29%
NCQCZ	86	0.57%	NCCCWCZ	217	1.28%

Observation to Symbol mapping used

Observation	Symbol	*Description*
Reporting	*N*	Reporter has only reported one bug as of bug creation date.
	M	Reporter has reported more than one and less than ten bugs prior to bug creation date.
	E	Reporter has reported more than ten bug reports prior to bug creation date.
Assignment	*A*	Bug confirmed and assigned to a named developer.
	R	Bug confirmed and put to general mailbox for volunteers to work on.
Copy	*C*	A certain person has been copied on the bug report.
	D	More than one person has been copied within one hour on the bug report.
Review	*V*	Developer requested code review.
	Y	Response to code review.
	S	Developer requested super review.
	H	Response to super code review.
File Exchange	*F*	File exchanged between developers and reporters.
Comments exchange	*W*	Comment exchanged on whiteboard.
Milestone	*L*	A Milestone has been set for solution deployment.
Priority	*P*	Priority changed for bug report.
Severity	*Q*	Severity changed for bug report.
Resolution	*Z*	Bug reached status resolved.

Median Number of Days Required for a Fix versus Number of Activities in Temporal Sequence

Length of sequence	Total count	% of bug reports resolved below 60 days	% of bug reports resolved above 60 days	Median No. of days
2	2,857	68.39%	31.61%	7
3	12,369	69.68%	30.32%	6
4	14,196	42.70%	57.30%	127
5	9,678	33.17%	66.83%	244
6	5,579	38.66%	61.34%	164
7	3,591	47.15%	52.85%	78
8	2,543	57.65%	42.35%	33
9	2,241	65.02%	34.98%	20
10	1,796	67.71%	32.29%	17
11	1,491	69.62%	30.38%	20
12	1,194	67.84%	32.16%	22
13	943	64.16%	35.84%	29
14	820	61.10%	38.90%	36
15	606	60.56%	39.44%	40
16	498	56.43%	43.57%	47
17	441	56.69%	43.31%	45
18	339	53.10%	46.90%	54
19	290	53.45%	46.55%	57
20	218	43.58%	56.42%	79
21	207	41.06%	58.94%	81
22	160	43.13%	56.88%	76
23	132	40.91%	59.09%	82
24	138	39.86%	60.14%	78
25	116	33.62%	66.38%	100
26-51	802	24.02%	75.98%	183
52-77	80	17.32%	82.68%	281
78-103	16	3.33%	96.67%	557
104-129	4	0.00%	100.00%	267
130-155	3	0.00%	100.00%	770
156-181	1	0.00%	100.00%	2,460
182-207	1	0.00%	100.00%	1,042

Activities and their frequencies

Activity/Observation	Symbol	Frequency of activities for bug reports resolved below 60 days		Frequency of activities for bug reports resolved above 60 days	
		Count	Percentage	Count	Percentage
Assignment to named developer	A	9,189	6.58%	4,410	3.06%
Assignment to volunteers	R	268	0.19%	721	0.50%
More than one person has been copied within one hour on the bug report	D	7,856	5.63%	5,698	3.95%
A certain person has been copied on the bug report	C	62,837	44.99%	85,266	59.17%
Developer requested code review	V	18,033	12.91%	9,747	6.76%
Response to code review	Y	12,687	9.08%	5,047	3.50%
Developer requested super review	S	157	0.11%	87	0.06%
Response to super code review	H	210	0.15%	82	0.06%
File exchanged between developers and reporters	F	1,000	0.72%	794	0.55%
Comment exchanged on whiteboard	W	13,699	9.81%	22,269	15.45%
Milestone has been set for solution deployment	L	8,827	6.32%	4,804	3.33%
Priority changed for bug report	P	2,817	2.02%	2,558	1.78%
Severity changed for bug report	Q	2,079	1.49%	2,609	1.81%

Number of Bugs Resolved by Year

Year	Total bug reports submitted	No. of new feature requests	No.of bug reports	No.of resolved bugs	No. of bugs resolved excluding bugs resolved on creation date (day 0)	Median No. of days till resolution	% of resolved bug reports
2006	11,571	809	10,762	10,531	7,125	179	97.85%
2007	9,962	662	9,300	8,918	6,526	194	95.89%
2008	16,070	1,012	15,058	14,114	10,557	230	93.73%
2009	11,967	659	11,308	10,290	7,538	203	91.00%
2010	14,402	861	13,541	10,633	7,854	97	78.52%
2011	12,381	720	11,661	8,362	5,834	28	71.71%
2012	11,609	441	11,168	7,024	4,659	10	62.89%
2013	12,485	449	12,036	8,508	6,974	17	70.69%
2014	15,652	373	15,279	8,064	6,283	7	52.78%
Totals	116,099	5,986	110,113	86,444	63,350		78.50%

More statistics and analysis for data

Resolution of Bugs Closed on Submission Date versus Reporter Expertise

Resolution	Reporter			Total bug reports	Total bug reports %
	Novice	Moderate	Experienced		
DUPLICATE	7,961	3,238	1,354	12,553	54.36%
FIXED	717	262	633	1,612	6.98%
INCOMPLETE	467	134	24	625	2.71%
INVALID	4,670	1,349	414	6,433	27.86%
WONTFIX	322	110	101	533	2.31%
WORKSFORME	979	261	98	1,338	5.79%
Bug reports count	15,116	5,354	2,624	23,094	
Bug reports %	65.45%	23.18%	11.36%	100%	

Median of Activities versus Median of Resolution Per Bug Report

Activity within the observation sequence	Median of No. of days for activity to occur	Median for No. of days till bug resolution
Second activity	2	60
Third activity	6	63
Fourth activity	8	44
Fifth activity	9	35
Sixth activity	12	33

Changes in Severity and Priority of Firefox Bug Reports

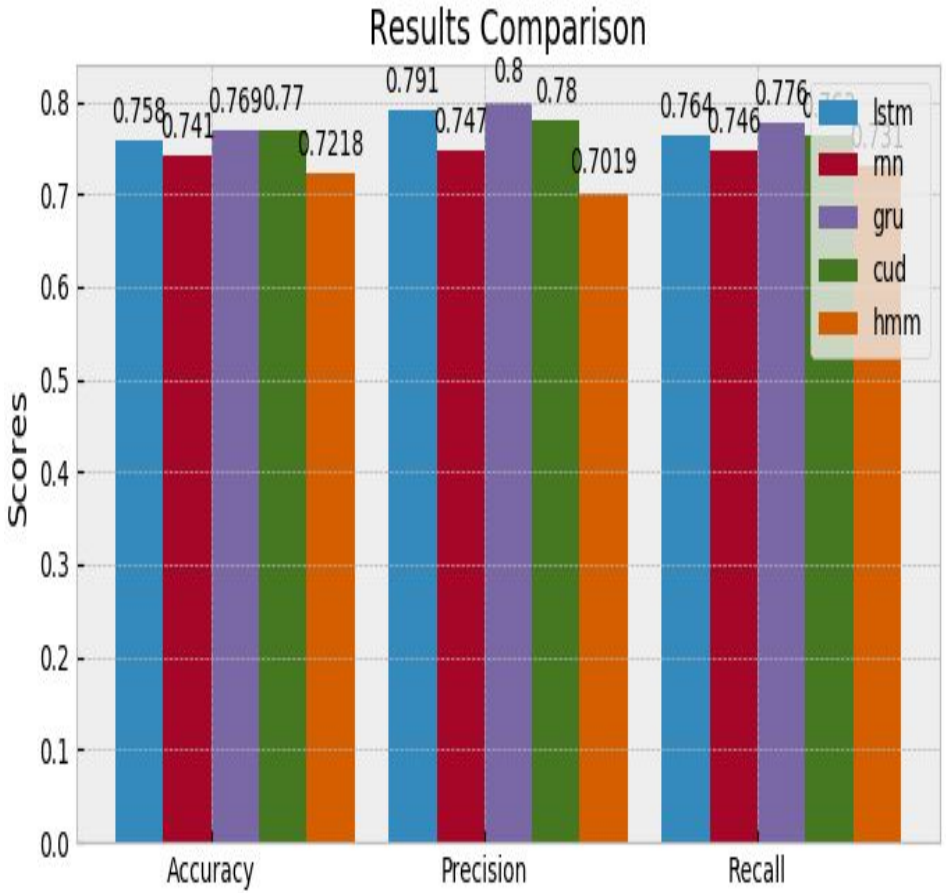
	Number of reports	% Of Total
Bug reports with only change in severity	3,451	5.45%
Bug reports with only change in priority	3,790	5.98%
Bug reports with both change in priority and change in severity	588	0.93%
Bug reports with unchanged severity and priority	55,521	87.64%
Total	63,350	100%

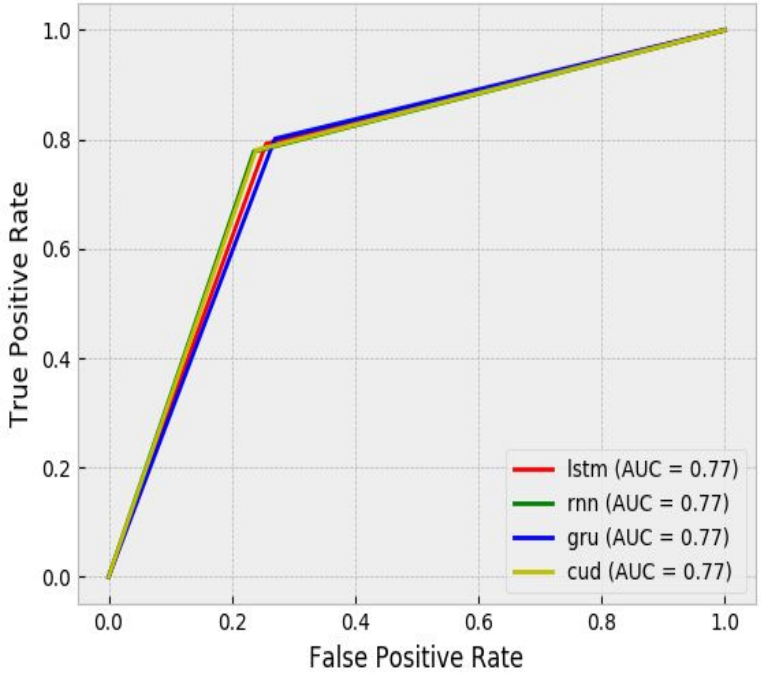
RESULTS:

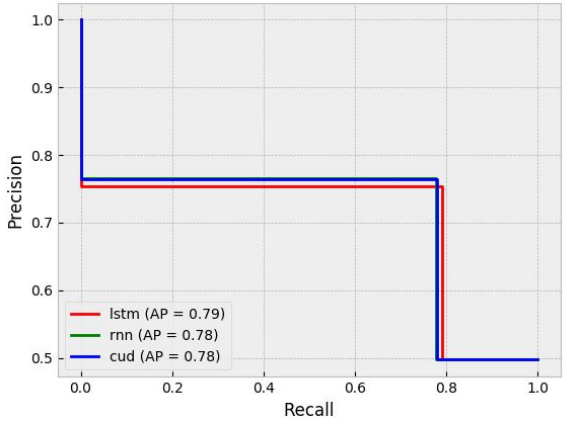
$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}. \qquad Precision = \frac{TP}{TP + FP}.$$

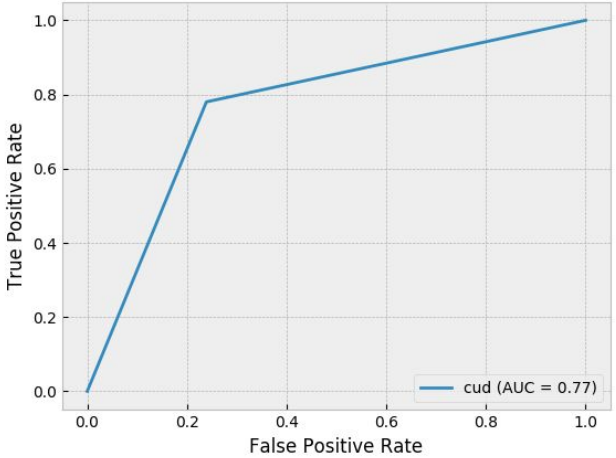
$$Recall = \frac{TP}{TP + FN}. \qquad F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall}.$$

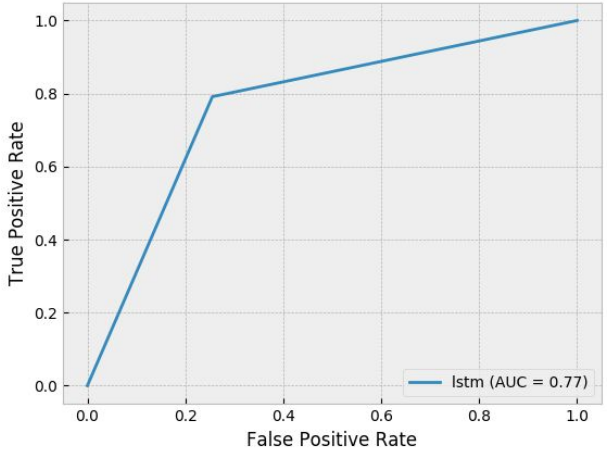
Training%	Precision	Recall	F-measure	Accuracy	Training size	Testing size	Obtained Accuracy
10%	68.49%	73.62%	70.96%	71.97%	6,335	21,952	74.45
20%	72.26%	72.30%	72.28%	72.29%	12,670	19,408	75.24
30%	69.90%	73.16%	71.49%	72.13%	19,005	17,044	76.10
40%	70.74%	72.98%	71.84%	72.27%	25,340	14,662	76.46
50%	70.19%	73.10%	71.62%	72.18%	31,675	12,147	76.89
60%	76.92%	71.53%	74.12%	73.15%	38,010	7,627	77.14
70%	68.34%	74.51%	71.29%	72.48%	44,345	6,105	77.25
80%	71.29%	73.90%	72.57%	73.06%	50,680	4,889	77.49

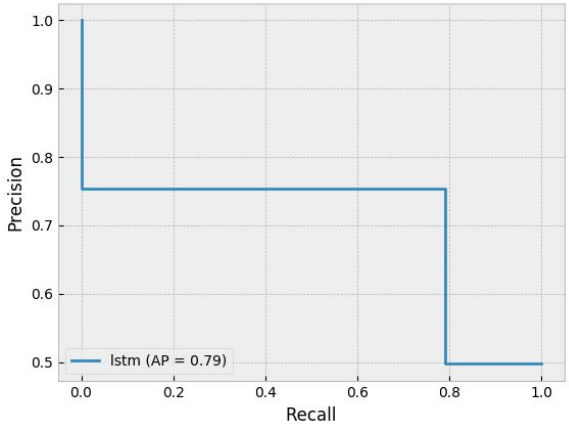

Results Comparison
Scores
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.758
0.741
0.769
0.77
0.7218
0.791
0.747
0.8
0.78
0.7019
0.764
0.746
0.776
lstm
mn
gru
cud
hmm
Accuracy
Precision
Recall

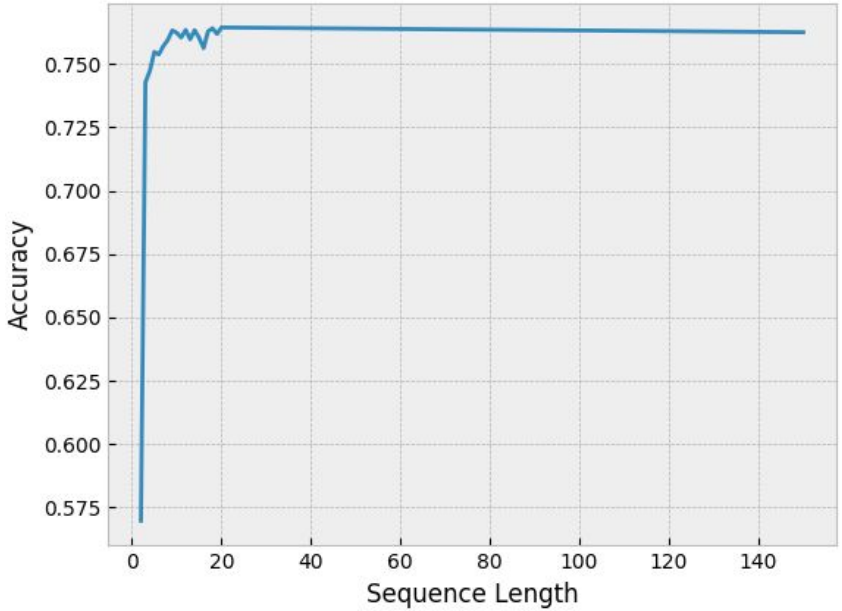

True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
lstm (AUC = 0.77)
rnn (AUC = 0.77)
gru (AUC = 0.77)
cud (AUC = 0.77)

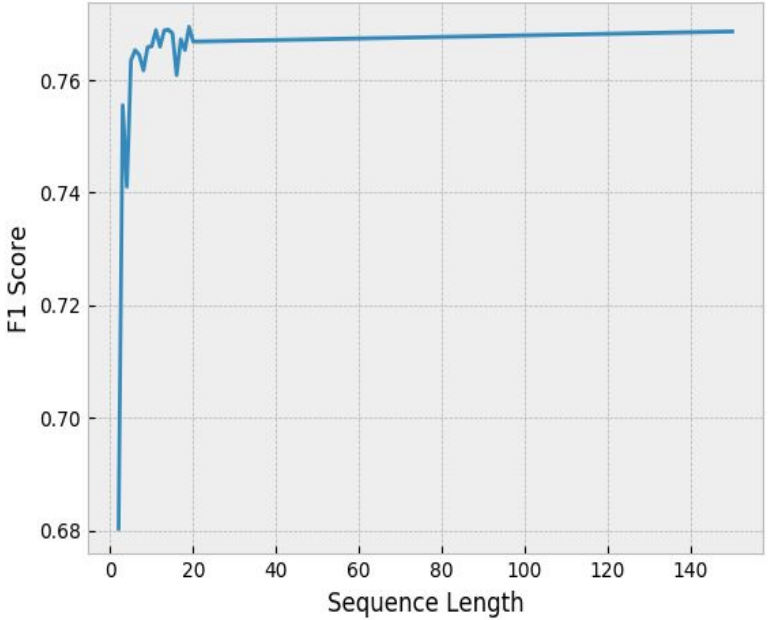

Precision
Recall
lstm (AP = 0.79)
rnn (AP = 0.78)
cud (AP = 0.78)


True Positive Rate
False Positive Rate
cud (AUC = 0.77)


True Positive Rate
False Positive Rate
lstm (AUC = 0.77)


Precision
Recall
lstm (AP = 0.79)


Accuracy
0.750
0.725
0.700
0.675
0.650
0.625
0.600
0.575
0
20
40
60
80
100
120
140
Sequence Length


F1 Score
0.76
0.74
0.72
0.70
0.68
0
20
40
60
80
100
120
140
Sequence Length

Thank you