

From the graph above, we can see while the given linear RNN starts off with better results, it eventually is over taken by lstm and gru. Theoretically this makes sense. The reason why the linear RNN has a better start is because it needs less parameters and hence less data than lstms and grus to improve itself. But since we had sufficient data, the gru and lstms were able to improve themselves and eventually be more accurate than the linear RNN. Furthermore we see that LSTM gives slightly better results that GRU and that is fine as well as these two methods vary according to the situation and it seems that in this situation LSTM won.

The **code (in .pynb format)** for this question has been included in the 2a folder, along with a copy of the image of the graph. Python3 is used.