Brandon Des Jardins
bdesjard

## Introduction

This report summarizes the methodology and results of predicting the remaining useful life (RUL) of engines given operational settings and sensor data. LSTM and GRU models were used and their results will be compared below.

## Methodology

Data preprocessing: Sensor data was cleaned and scaled/normalized. This was to keep all values between 0 and 1. This was done in order to put all values on the same scale so the LSTM and GRU would converge similarly. Furthermore, this prevents extremely large values from skewing the data.

Feature engineering: Upon a closer look at sensor data, it was noted that sensors 1, 5, 10, 16, 18, and 19 had no variance. These columns were dropped from the analyzed data frames.

Model training: LSTM and GRU models were trained with a focus on optimizing precision. The models were also analyzed using accuracy, precision, recall, and F1-score. The 16 and 8 units in the LSTM and GRU were used to keep the hidden states small, as not to increase computation time. Ideally, the GRU should require less units, but I felt I should test them with similar parameters. Two layers were used to capture higher-level, longer-term dependencies. I used sigmoid activation and binary cross entropy as this was a binary classification task. Overall, both models had around 3,000 parameters.

Evaluation: A threshold of 0.95 was used to fine tune precision and ensure recall was as high as possible. A lower threshold was identifying too many false positives, which would result in extra expenses and extra maintenance. I wanted to identify all true positives as it would be detrimental to miss one.

## Results

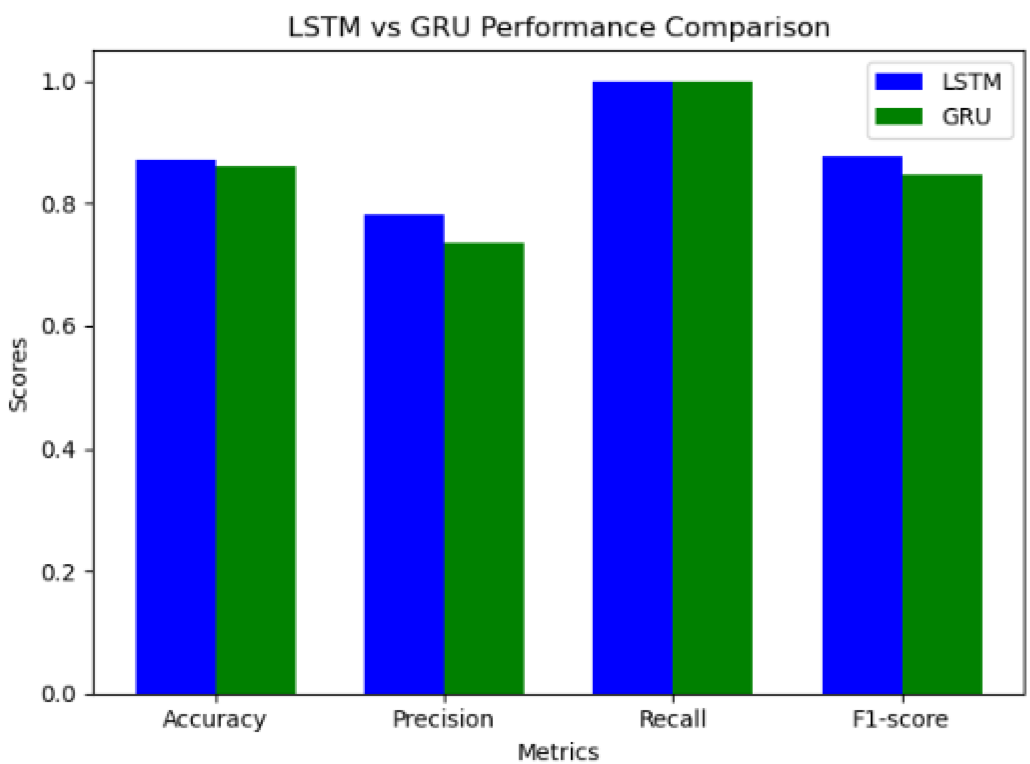
LSTM:

| | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| LSTM | 0.870968 | 0.781250 | 1.0 | 0.877193 |
| Template Best Model | 0.940000 | 0.952381 | 0.8 | 0.869565 |

GRU:

| | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| GRU | 0.860215 | 0.735294 | 1.0 | 0.847458 |
| Template Best Model | 0.940000 | 0.952381 | 0.8 | 0.869565 |

Comparison



## Conclusion

For training the LSTM had a training accuracy 0f 0.97 and a loss of 0.06. The GRU had a training accuracy 0f 0.97 and a loss of 0.05. They performed similarly in training. The SLTM converged in 17 Epochs while the GRU converged in 14 Epochs. This was expected as the GRU is faster.

For the test data, both models achieved perfect recall, meaning they classified all positive instances. I wanted to make sure both models accomplished this as we do not want to miss a faulty engine. The LSTM slightly outperforms the GRU in all of the other metrics by a couple percentage points. I felt that both models achieving over an 85% accuracy score was sufficient for initial tuning. The LSTM has 78% precision and the GRU has a 73% precision. Overall, the LSTM performs slightly better and predicts less false positives.

## Further Examination

A sequence length of 50 was used to model enough past information to help capture trends. We could try a model with different sequence lengths to see if this helps raise precision. Also, the models were only created with two layers and 16 and 8 units respectively. Ideally, the LSTM should have more units to achieve comparable performance as the GRU. I would recommend adding more layers to both models, more units to the LSTM, and then compare computational cost and number of parameters.