SC4020 Data Analytics and Mining

Academic Year 2025/2026

Semester 1

Group 18

LUNBERRY NOAH IWATA (N2503869H)

PIKERINGA ANTONINA DAILA (N2504101A)

RAHLFS FREDERIC MAURITZ (N2504096K)

SARAH EMILY ONG XIN WEI (U2440124G)

Abstract

This project focused on two graph-based neural networks, Graph Convolutional Network (GCN) AND Graph Attention Network (GAT). The models were then applied to our chosen datasets: the Wisconsin Breast Cancer dataset and the NASA Turbofan Jet Engine Degradation dataset. Selected to represent classification and regression, the chosen datasets reflect two distinct problem domains: medical diagnosis and mechanical prognostics. Both models were trained and evaluated using appropriate loss functions and metrics. Results showed that both architectures performed comparably on the breast cancer dataset (attaining an accuracy of 95.1%). On the NASA Turbofan dataset, both models achieved similar performance with MSE values differing by less than 0.1%.

Table of Contents

1 Introduction

Graphs are a type of data structure that maps objects (represented as nodes) and their relationships (represented as edges) [1]. Since graphs can be used to represent systems across a variety of areas. Consequently, Graph Neural Networks (GNNs), a machine learning method able to analyse such graphs, have emerged as an area for research and testing. GNNs are artificial neural networks that operate on the graphs. GNNs use a message-passing system to aggregate information from neighbouring nodes, allowing them to capture complex relationships in graphs [2].

This project focuses on two variants of GNN: Graph Attention Networks (GATs) and Graph Convolutional Networks (GCNs).

GATs are a neural network architecture that operate on graph-structured data [3]. GCNs use a semi-supervised learning approach on variant convolutional networks that can learn and predict based on the graph [3].

Our goal in this project was to compare the performance and appropriateness of GCN and GAT models on different datasets. We tested the two models on two datasets: the Breast Cancer Dataset and the NASA Turbofan Jet Engine Data Set. This report details how both models work, the performance on the two models on the chosen datasets, and comparative analyses on the performance by the two models on the chosen datasets.

2 Methods

In order to accurately assess the capability of two different graph networks, two diverse datasets were chosen. The two models were trained on each and evaluated.

2.1 Datasets

The two datasets used in this study were the NASA Turbofan Jet Engine Degradation dataset (C-MAPSS) and the UCI Wisconsin Breast Cancer dataset. These were selected to reflect two distinct real-world problems. Prognostics of mechanical systems and medical diagnosis. This ensured diversity in the training data while reducing domain-specific bias.

The Wisconsin Breast Cancer dataset consists of 32 columns. Two columns correspond to identifiers and diagnosis labels, respectively. The remaining 30 represent measurable tumor characteristics (e.g., radius, texture, area, smoothness, compactness, concavity, etc.). The target variable is the classification of the tumor as benign or malignant. Before training, all numerical features were normalized to zero mean and unit variance to ensure convergence during training. The dataset was split into training and test sets using an 80/20 ratio.

The NASA C-MAPSS turbofan jet engine degradation dataset provides multivariate time-series data simulating total life scenarios of jet engines. Each engine instance is characterized by 26 columns: one engine identifier, one time cycle index, three operational condition variables, and 21 sensor measurements. The target variable is the Remaining Useful Life (RUL), defined as the number of cycles left until engine failure. The dataset was provided with separate training and validation partitions: 100 engines for training and 100 engines for validation. For consistency, sensor features were normalized across engines.

The two datasets differ fundamentally, both in data modality (tabular biomedical features vs. multivariate time-series sensor data) and in target type (classification vs. regression). This diversity was intentional, aiming to assess the generalizability of the evaluated models across domains.

2.2 Models

Two graph networks with different layers were used. For the breast cancer dataset, nodes corresponded to patient samples, while the edges were chosen according to a k-NN graph. Node features corresponded to the 30 tumor attributes, and the output task was binary classification. For the turbofan dataset, nodes represented sensor states and operational conditions. Edges were defined through feature correlations. The GCN then predicted the Remaining Useful Life (RUL) as a regression task.

2.2.1 Graph Convolutional Network (GCN)

The Graph Convolutional Network (GCN) serves as a baseline model for learning from graph-structured data. A GCN updates node representations by aggregating feature information from neighboring nodes. The GCN architecture consisted of 2 convolutional layers with a RELU activation function, and dropout for regularization.

2.2.2 Graph Attention Network (GAT)

The Graph Attention Network (GAT) extends the GCN by introducing an attention mechanism over neighboring nodes. Instead of averaging features uniformly, GAT learns attention coefficients that weight the contribution of each neighbor during aggregation. This allows the model to focus on the most informative relationships within the graph, which, especially in datasets with high assumed covariance, can lead to significantly improved results. The GAT architecture consisted of 2 attention layers with 8 attention heads, followed by RELU. Again, dropout was used for regularization.

2.3 Training, Inference, and Evaluation

Both models were trained for 300 epochs on the NASA dataset, using MSE error due to the regressional nature of the target variable.

For the breast cancer dataset cross-entropy loss was employed due to the classification nature of the target, with 100 epochs. The trained models were then used for inference on the validation dataset and compared to the reference. For the breast cancer dataset, an F1 score was used for evaluation, while for the NASA dataset, the MSE is used to compare the performance.

3 Results

Section 3 will present the results of the models' predictions. Section 4 will analyse and draw conclusions from these results.

3.1 Breast cancer dataset

3.1.1 GAT model

In the GAT model, each patient was a node in the graph, and edges were made using the ten closest patients based on how similar their features were. The graph ended up with 569 nodes and 8,326 edges. This was then split it into 455 for training and 114 for testing.

The GAT had two attention layers with eight heads each, and the model was trained it for 100 rounds using cross-entropy loss. The loss started around 0.23 and went down to about 0.11 by the end, which means the model was learning steadily.

Table 1: GAT results on breast cancer dataset

Metric	**Score**
Accuracy	0.9561
Precision	0.9512
Recall	0.9286
F1-Score	0.9398
ROC-AUC	0.9944

3.1.2 GCN model

Table 2: GCN results on breast cancer dataset

Metric	**Score**
Accuracy	0.9561
Precision	0.9744
Recall	0.9048
F1-Score	0.9383
ROC-AUC	0.9924

3.2 NASA turbofan jet engine dataset

3.2.1 MSE error

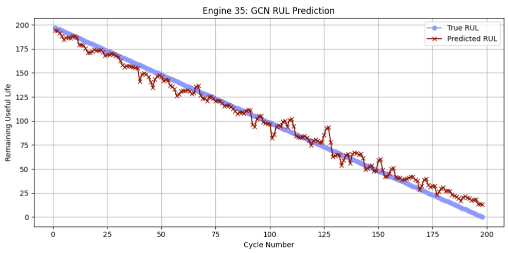
Table 3 shows the MSE error of both the GCN and GAT models. The GCN and GAT models have approximately the same MSE error, differing by less than 0.1%.

Table 3: NASA turbofan jet engine: GCN vs GAT comparison

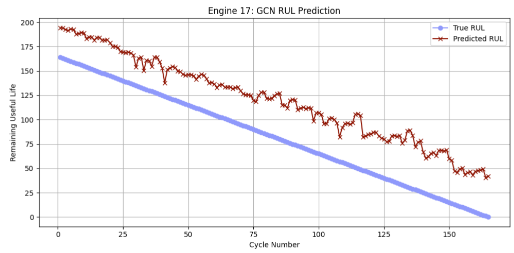
Model	**MSE error**
GCN	1365.4297
GAT	1363.8159

3.2.2 GCN model plots

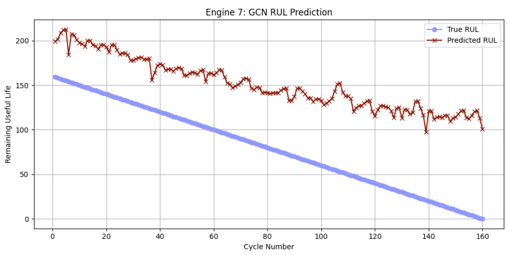
The results of the GCN model on the NASA turbofan jet engine dataset vary: Figure 1a is an example of a good fit of data, with the predictions well aligned with the true values. Other than some small spikes in the data, the overall downward slope is roughly accurate. Figure 1b shows an example of slight bias, with the slope of the predicted values being mostly correct but all predicted values being shifted upwards of the actual values by the same amount. Figure 1c and Figure 1d show moderate and high levels of bias respectively. Again, the slope of the predicted values is mostly accurate, but all predicted values are vertically shifted from true values. Interestingly, the example with the most bias (Figure 1d) has a much smoother slope than the examples with less bias.



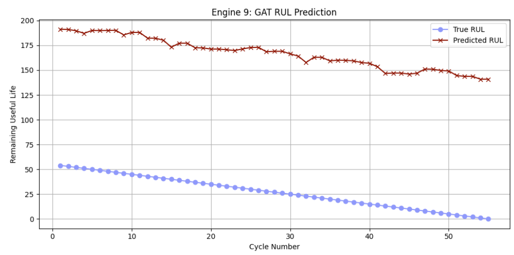
(a) Good fit



(b) Slight bias



(c) Moderate bias

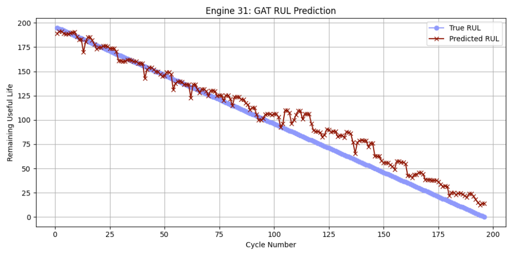


(d) High bias

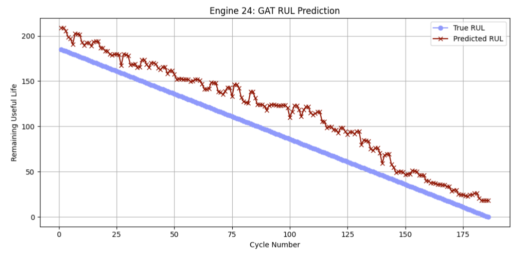
Figure 1: GCN model on NASA data

3.2.3 GAT model plots

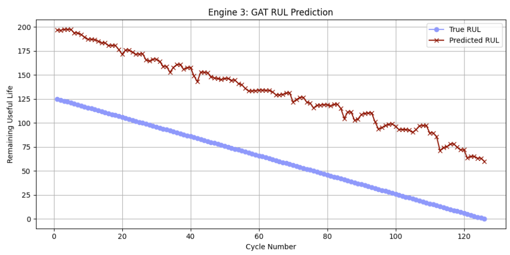
The results of the GAT model on the NASA turbofan jet engine dataset follow a similar trend those of the GCN model. Instances of good fits (Figure 2a), slight bias (Figure 2b), moderate bias (Figure 2c), and high bias (Figure 2d) can be found. The GAT model also has small spikes in the predicted values, but overall the slopes of the predicted values are accurate.



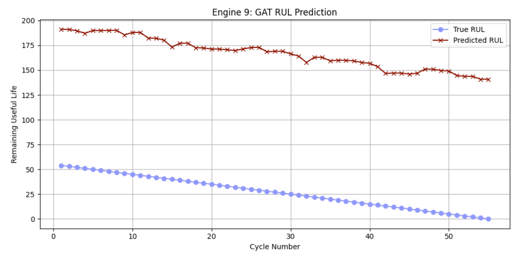
(a) Good fit



(b) Slight bias



(c) Moderate bias



(d) High bias

Figure 2: GAT model on NASA data

4 Analysis

4.1 Breast Cancer Dataset: GAT vs GCN

Head to head: both models hit the same accuracy (0.9561). GCN has higher precision (0.9744) but lower recall (0.9048). GAT flips that: lower precision (0.9512) but higher recall (0.9286). So GAT finds a few more malignant cases, while GCN is a bit stricter. GAT also has a slightly higher ROC-AUC (0.9944 vs 0.9924), which means it separates classes a touch cleaner across thresholds.

Table 4: Breast cancer: GAT vs GCN comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC-AUC**
GAT	0.9561	0.9512	0.9286	0.9398	0.9944
GCN	0.9561	0.9744	0.9048	0.9383	0.9924

4.2 NASA Turbofan Jet Engine Dataset: GAT vs GCN

The GCN and GAT models have similar MSE errors (see Table 3), indicating similar performance. This can be confirmed by analysing plots of the predicted RUL values against the true RUL values, as shown in Figure 1 and Figure 2. While both models have small oscillations throughout their predicted values, the models capture the general downward trend of the RUL data quite successfully.

Varying levels of bias can be seen in the predictions. In an attempt to explain this bias, the RUL values were normalised. This is because the range of RUL values in the datasets was large, with some engines having RULs in the range of 50, while others - 200. The cycle numbers can also range from ranges of 60 to ranges of 200. This makes it difficult for the models to predict precise values across all engines. Predictions display the least bias for the engines having RULs in the range of 200, and the most bias for those with RULs in the range 50. Unfortunately, the normalisation attempts the did not fix the bias, hence further investigation is required to solve this issue.

4.3 Comparison of GAT vs GCN across both datasets

In this report, two types of GNNs (GCNs AND GATs) were applied to two different tasks: classification (the breast cancer dataset) and prognosis (the NASA turbofan jet engine dataset). Interestingly, the the metrics and predictions and thus the performance of both models was highly similar for each dataset. This indicates that the GNN types have comparable results for classification and prognosis tasks. Further experimentation with more datasets is needed to clarify the extent of this conclusion.

5 Conclusion

This project compared the performance of Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) on two datasets of contrasting nature: a tabular biomedical classification problem and a multivariate time-series regression task.

Having achieved similar accuracy levels on the breast cancer dataset (95.6%) - with GAT exhibiting slightly higher discriminative power as reflected in the marginally larger ROC-AUC value - and captured general degradation trends on the NASA turbofan dataset, the results concretely show that both models performed comparably across both datasets. It is worth noting that on the NASA turbofan dataset, there were varying degrees of bias. This implies a vital role is played by the construction of the relational graph in performance stability.

Generally, albeit GATs can provide improved interpretability and fine-grained relational modelling through attention weighting, their advantages over GCNs depend on the dataset. With clear feature relationships, the simplicity and efficiency of GCNs often make it the more appropriate model. GATs on the other hand, offer more robust representations for data in complex domains with heterogeneous node connections or nonuniform influence.

Given that both models performed to nearly identical accuracy and effectivesness for the chosen datasets, future work could focus on larger, more heterogeneous graphs to determine the advantages provided by GAT over GCN for more complex graphs. Additionally, future work could work to solve the varying levels of bias found in the predictions on the NASA turbofan dataset.

List of References

[1] J. Zhou, G. Cui, S. Hu, *et al.*, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020, ISSN: 2666-6510. DOI: `10.1016/j.aiopen.2021.01.001`. (visited on 10/09/2025).

[2] L. Wu, P. Cui, J. Pei, and L. Zhao, Eds., *Graph Neural Networks: Foundations, Frontiers, and Applications*. Singapore: Springer Nature Singapore, 2022, ISBN: 978-981-16-6053-5 978-981-16-6054-2. DOI: `10.1007/978-981-16-6054-2`. (visited on 10/09/2025).

[3] J. Noble, *What is a Graph Neural Network — IBM*, https://www.ibm.com/think/topics/graph-neural-network, Feb. 2025. (visited on 10/09/2025).

Appendix A - Breakdown of Contributions

LUNBERRY NOAH IWATA (N2503869H)
GAT code for breast cancer dataset. Results and Analysis writeup.

PIKERINGA ANTONINA DAILA (N2504101A)
GCN code for NASA Turbofan dataset. Results and Analysis writeup.

RAHLFS FREDERIC MAURITZ (N2504096K)
GAT code for NASA Turbofan dataset. Methods writeup.

SARAH EMILY ONG XIN WEI (U2440124G)
GCN code for breast cancer dataset. Abstract, Introduction and Conclusion writeup.