UNITED AIRLINE MAINTENANCE OPTIMIZATION

Haoyang Xu, Yanqian Pan, Yuying Wang

MSDS 498 2024, SEC 60

Northwestern University, Capstone

18/08/2024

Abstract

This research project aims to enhance the reliability and accuracy of United Airlines' aircraft data systems by identifying wrong data records. With thousands of aircraft real-time data, we need to ensure that data quality is accurate for safe and efficient operations. Our approach involves developing an unsupervised model for error detection, which will be achieved by comparing aircraft left ratio and right ratio under the same aircraft over time. Our final deliverables includes a comprehensive analysis, featuring visualizations and statistical summaries to demonstrate the system's effectiveness. Our project will help United Airline to enhance data accuracy and quality so as to improve the operational efficiency and ensure safety.

Introduction

In the aviation industry, it is significant to ensure the aircraft's accuracy and reliability in order to maintain safe and efficient operations. United Airline operates numerous data systems that continuously transmit real-time information from aircraft. If United Airline fails to identify an error in the data in time, it may cause operational inefficiencies, safety risks, and false decision-making problems. Therefore, in this project, we will address the critical need for data quality assurance within United Airlines' aircraft systems.

The primary objective of this research is to develop an unsupervised model to effectively detect data errors. By integrating this model, we expected to help United Airline better monitor the existing data infrastructure in real-time and better identify the error in the dataset. By doing so, we aim to enhance the overall quality of the data, thereby supporting more accurate data-driven decision-making processes.

The reason why this research is crucial is that our analysis and models can help United Airline improve operation performance and safety. By ensuring that data is accurate and reliable, we aim to mitigate the risks associated with erroneous data and contribute to more efficient and informed decision-making. Finally, we will demonstrate the effectiveness of the implemented system and its benefits for the airline through visualizations and statistical summaries from our model.

Literature Review

1. *Importance of Ensuring Aircraft Data Accuracy*

Aircraft data accuracy is fundamental to ensuring safety, optimizing operations, and adhering to regulatory standards in the aviation industry. Llopis (2023) emphasizes that poor data quality can severely impact maintenance efficiency and safety, leading to operational inefficiencies and potential safety risks. Ensuring high-quality, accurate data is essential for effective predictive maintenance, which not only enhances safety by identifying potential mechanical failures before they occur but also reduces maintenance costs through optimized scheduling. Wang and Shen (2013) further discuss the importance of data quality in reducing uncertainty in complex system analyses, which is crucial for maintaining the reliability and airworthiness of aircraft. These insights are aligned with the findings of Kwakye, Jennions, and Ezhilarasu (2024), who highlight that platform health management systems heavily rely on accurate data to function effectively, thereby supporting safe and efficient aircraft operations.

2. *Autoencoders for Anomaly Detection*

Autoencoders have emerged as a powerful tool in detecting anomalies within time-series data, particularly in scenarios where labeled data is unavailable. Korba et al. (2023) describe the application of an algorithmic approach to optimize aviation maintenance, highlighting the role of autoencoders in identifying abnormal data records. Their approach demonstrates the utility of autoencoders in improving data quality and maintenance processes by detecting subtle anomalies that may otherwise go unnoticed. Similarly, Githinji and Wa Maina (2023) explore the use of deep LSTM Autoencoders for anomaly detection in IoT sensor data, emphasizing the importance of lower-dimensional embeddings and reconstruction error in reducing false positives and negatives. The work of Khanmohammadi and Azmi (2024) extends this discussion by applying a D-CNN-LSTM Autoencoder framework to connected and automated vehicles, showcasing the versatility and efficacy of autoencoders in various transportation-related contexts, including aviation.

3. *Predictive Maintenance and Prognostics in Aviation*

Predictive maintenance, driven by advanced machine learning models, is becoming increasingly important in the aviation industry. Wang et al. (2024) discuss how predictive maintenance scheduling based on remaining useful life (RUL) predictions can significantly improve the efficiency and reliability of aircraft engines. This approach is complemented by the work of Upadhyay and Amhia (2023), who propose an LSTM-based model for predicting the RUL of aircraft engines, demonstrating its potential to reduce unscheduled maintenance and extend asset lifespan. Baptista, Henriques, and Prendinger (2021) also contribute to this conversation by reviewing classification prognostics approaches in aviation, which leverage data-driven techniques to forecast component failures and optimize maintenance schedules. These studies collectively underscore the critical role of predictive maintenance in enhancing the operational efficiency, safety, and cost-effectiveness of aviation maintenance practices.

Methods

To detect anomalies within the aircraft system data provided by United Airlines, we first take a rough overview of the dataset. The dataset consists of 61,879 rows and 27 variables, focusing primarily on numerical variables related to the left and right engine parameters, which are expected to behave symmetrically under normal conditions. While we took a closer look at the data, we found that we did not need many categorical variables. However, there were two exceptions: 'Timestamp' (the time recorded at takeoff) and 'ACID' (aircraft type ID), which were essential for our analysis.

To better understand our dataset, we conducted an Exploratory Data Analysis (EDA). Initially, we plotted time series for each variable to observe trends over time (Figure 1). However, due to significant noise, it is hard and difficult for us to generate any actionable insights from these plots. Therefore, we grouped the data by aircraft (ACID) and plotted comparisons between the left and right engine variables.

Through this approach, we find that while some variables performed similarly for both engines, others exhibited significant discrepancies. However, the discrepancies do not mean the abnormality of the dataset. As long as the difference is constant over time, the left and right variables will still be considered as normal values in the dataset. Hence, to further investigate these discrepancies, we plotted the differences between the left and right engine variables for each aircraft over time. If the difference remained consistent, the data was considered normal. However, if the difference has significant variability, then it indicates potential anomalies, particularly those showing sudden drops.

In addition, during the EDA, we also observed that some variables exhibited higher values in the first half of the year compared to the second half, suggesting a potential seasonality effect. To test this, we plotted both the original time series and a deseasonalized version. The minimal difference between these plots suggested there’s no seasonality effect in our dataset. However, this conclusion remains tentative since we only have a one-year data span. For the current analysis, we chose to ignore the seasonality effect.

Following the EDA, we employed an autoencoder model due to its ability to handle high-dimensional data and capture complex, nonlinear relationships that are prevalent in our dataset. Traditional models might struggle to accurately identify subtle anomalies within such data, but autoencoders are specifically designed to learn compressed representations of input data, making them ideal for our needs.

For the modeling process, the data was split into two halves based on the date. The first half of the year, up to June 30, 2023, was used for training and validation. Specifically, 80% of this half-year data was allocated for training, and the remaining 20% was used for validation. The second half of the year served as the test data. This approach allowed the autoencoder to learn the normal behavior patterns from the training data and validate its performance before applying the model to detect anomalies in the test data. The autoencoder was trained to minimize the reconstruction error between the input and output data, with high reconstruction errors in the test set indicating potential anomalies.

Model

The model used in this project is Autoencoder. It is one kind of neural network architecture to compress input data into its essential features, then reconstruct the original input from the compressed representation. It is a powerful tool for anomaly detection due to the ability of learning complex data and identifying deviations from normal patterns.

Data-preprocess

There are a total of 44 different aircrafts in this dataset, and we only extract ACID 9679 as our target since it is the most representative. Before starting the splitting data, we convert some unexpectedly labeled categorical data variable to numeric data (WF_R, WFRQ_L, WFRQ_R). The date 6-30-2023 is selected as the cutoff point since it is the middle data of a year. Then top 80% data of the first year will be set as training data and the remaining is the validation data. Then both of these two datasets will be used to predict the second half year by using AutoEncoder. Dividing data by half year will help to maintain sequential data's consistency and seasonality. Although it is a one-year data, this model hopes to help United Airline deal with more data than just one year. Time series data often contain seasonal variations or long-term trends. In many practical applications, such as aviation data, sales data, meteorological data, etc., the half-year time span is usually long enough to capture these seasonal or cyclical changes. This partitioning ensures that the model learns the seasonal pattern of the data during training, which allows for more accurate detection of anomalies during validation and testing. Second reason is to prevent data leakage. If the data set is divided directly at random, it may result in the validation set or test set containing data at the same point in time as the training set, which can lead to data leakage and affect the generalization ability of the model. Breaking it down by time avoids this problem by ensuring that the data in both the validation set and the test set are points in time that the model has not seen before. The last not least reason is it will better evaluate a model's generalization ability. Breaking down by time provides a more realistic assessment of the model's performance on future data because the data for both the validation and test sets are subsequent time periods. This division tests whether the model has actually learned the laws of time series, rather than just a "memory" of past data. Then the process of normalizing will be performed to minimize the data into the range from 0 to 1, it is an encoding process to help model faster convergence and improve its stability. In order to help the model better learn from data's features, a window function is also performed and the window size is set as 5.

Model Structure

The Autoencoder model is relatively simple in structure, using encoder and decoder parts symmetrically. The encoder progressively compresses the input data to a low-dimensional feature representation (32 dimensions), while the decoder re-expands it back to the dimensions of the original input data. By minimizing reconstruction errors (the difference between input and output), the model can learn the main features of the input data and identify abnormal data points by detecting high reconstruction errors when tested.

We do not add many complicated structures like regularization techniques or residual connections because this sequential data is easy to interpret, adding too many techniques might lead the model to be overfitting.

After 50 epochs, training loss is 0.0015 and the validation loss is 0.000921(Figure 5).they are pretty close and during the training, both of these two losses are not increasing; this indicates the model does not have any risks of overfitting. It can also be proved in the Training and Validation Loss plot (Figure 6). At the beginning of the plot, the loss decreases extremely and then stays constant. In the tail of these two lines, they are very close to each other to prove there is no overfitting problem.

After ensuring the model performs well, we need to dive into which time period has abnormal values. We construct the plot of reconstruction error over time for ACID 9679 (Figure 7), and this plot can clearly point out the abnormal periods. In the Autoencoder model, a high reconstruction error usually means an exception. Autoencoder compress and reconstruct data by learning the patterns of normal data. When the model encounters normal data similar to the training data pattern, it can reconstruct accurately and the reconstruction error is low. However, when we encounter unprecedented abnormal data, the model cannot be effectively restored due to the large deviation between these data and the normal model, resulting in a significant increase in reconstruction errors. Therefore, the reconstruction error is used as the main indicator of anomaly detection. The higher the reconstruction error, the more likely the data is to deviate from the normal mode, indicating that there may be an anomaly in the period. In this plot, the blue line (Train Reconstruction Error), has a horizontal line around month May, this means during that time the plane is not flying so we do not have to consider this situation. The reconstruction error of the green test set has a significant fluctuation, especially around September 2023 and November 2023, the reconstruction error increases significantly. In September 2023, this is the most significant anomaly in the entire time series, and the reconstruction error increases dramatically at this time. This means that during this time period, the model detected a situation that did not match the pattern of the training data, that is, a data exception. The volatility in November 2023 is not as striking as September's, but there was still a period when the margin of error continued to rise. This suggests that there may have been more anomalies during this time as well. In order to prove, the prediction is made correct, we will dive in to find the specific period of the most significant anomaly. After zooming in, this fluctuation happened from 9-24 to 9-26(Figure 8).

The purpose of this model is to find outliers and determine the time period over which they occur. Although there are two dramatic fluctuations, we will only focus on the most striking one. In order to better observe the difference between the reconstructed data and the original data, I drew the comparison diagram of each variable (only the most representative ones were selected) (Figure 9 & 10) In Figure 9, this is the plot for EPRCM_L(Engine Pressure Ratio Corrected for Left Engine), the blue line represents original data and orange line represents reconstructed data. Looking at the real data, we can see that the EPRCM_L has not transmitted data back since September 24 and the autoencoder has stimulated the data that should be returned every day based on the previous data. When we go back to the rare data, we find that during this period, aircraft 9679 is indeed empty. So we can assume that EPRCM_L is indeed an outlier during this time period. In the plot Figure 10, it is the same format as Figure 9, this is the plot for TRA_L(Thrust Angle for Left Engine). The actual data line has been transmitting very stable data back unlike the last plot, but the difference between these two lines are quite big. Such a big gap is also worth the attention of the staff, but at present we are only concerned about finding suspected outliers, and more sophisticated algorithms are needed to determine whether they are outliers

We also plot the reconstruction error of every variable to observe (Figure 11 & 12). Plotting differences helps to show the error between the predicted and actual values more visually. This kind of graph is able to clearly quantify the magnitude of the error, especially in the abnormal time period, where sharp changes in the difference can be more clearly identified. Compared to directly looking at the two curves of the original value and the predicted value, the difference chart can simplify the understanding of the error, help analyze the trend of the error over time, and quickly find anomalies. In addition, such graphs provide a focused, clear perspective when analyzing error distributions over multiple time periods, helping to judge model performance more effectively.

Results

The autoencoder model successfully captured the patterns in the training and validation data, enabling effective reconstruction of the test data. After training on data up to June 2023 and validating the model's performance, the test data from July 2023 onward revealed key insights into potential anomalies. The model's training and validation losses showed smooth convergence, indicating no overfitting and solid learning of essential patterns. Upon evaluating the reconstruction error, two significant spikes were observed in September and November 2023, with the most prominent anomaly detected between September 24 and 26, where the reconstruction error remained consistently high. Further analysis of specific variables, such as EPRCM_L and TRA_L, revealed discrepancies where the original data either dropped to near-zero (like EPRCM_L) or showed substantial differences (like TRA_L) compared to predicted values, flagging potential data transmission issues. Visualizing the differences between original and reconstructed data clarified these errors, with sharp fluctuations clearly indicating when anomalies occurred. Overall, the results confirm that the autoencoder effectively identified significant data deviations, demonstrating its capability to enhance data monitoring and flag critical anomalies in real-time operations.

Conclusion

This project successfully illustrates the capability of leveraging an autoencoder for anomaly detection in complex aircraft data systems. By focusing on critical engine performance parameters and utilizing reconstruction error as the primary indicator, we identified specific periods where data deviated significantly from expected patterns, flagging potential anomalies. The autoencoder model, trained on historical data, effectively learned the inherent patterns and relationships within the dataset, enabling it to highlight discrepancies when tested on unseen data. Notably, the model detected a significant anomaly in September 2023, where the reconstruction error spiked, indicating a potential data issue that requires further investigation.

The comprehensive analysis provided by the model, including visual comparisons between actual and predicted values, as well as the quantification of differences, demonstrated its practical application in real-time data monitoring. The ability to isolate these periods of abnormal data flow is critical for enhancing operational safety, ensuring timely maintenance, and preventing false decision-making based on erroneous data.

Looking ahead, while the current approach has proven effective, there remains potential for enhancement. Incorporating more advanced techniques, such as hybrid models or regularization strategies, could further improve detection accuracy and robustness. Additionally, expanding the dataset beyond a single year would allow the model to better capture seasonal effects and long-term trends, making it more adaptable to a broader range of scenarios.

In summary, the autoencoder-based model provides a scalable, efficient, and accurate method for detecting data anomalies in United Airlines' operations. By integrating this system into their existing data infrastructure, United Airlines stands to significantly enhance their data quality assurance processes, ultimately supporting safer, more efficient, and reliable flight operations.

References

Ai, Yi, Yuanji Wang, Weijun Pan, and Dingjie Wu. “A Deep Learning Framework Based on Multisensor Fusion Information to Identify the Airplane Wake Vortex.” *Journal of Sensors* 2021 (2021): 1–10. https://doi.org/10.1155/2021/4819254.

Baptista, Marcia L, Elsa M.P Henriques, and Helmut Prendinger. “Classification Prognostics Approaches in Aviation.” *Measurement : Journal of the International Measurement Confederation* 182 (2021): 109756-. https://doi.org/10.1016/j.measurement.2021.109756.

Chen, Haihua, Jiangping Chen, and Junhua Ding. “Data Evaluation and Enhancement for Quality Improvement of Machine Learning.” *IEEE Transactions on Reliability* 70, no. 2 (2021): 831–47. https://doi.org/10.1109/TR.2021.3070863.

Githinji, Stephen, and Ciira Wa Maina. “Anomaly Detection on Time Series Sensor Data Using Deep LSTM-Autoencoder.” In *2023 IEEE AFRICON*, 1–6. IEEE, 2023. https://doi.org/10.1109/AFRICON55910.2023.10293676.

Khanmohammadi, Fatemeh, and Reza Azmi. “Time-Series Anomaly Detection in Automated Vehicles Using D-CNN-LSTM Autoencoder.” *IEEE Transactions on Intelligent Transportation Systems*, 2024, 1–12. https://doi.org/10.1109/TITS.2024.3380263.

Kwakye AD, Jennions IK, Ezhilarasu CM. Platform health management for aircraft maintenance – a review. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering. 2024;238(3):267-283. doi:10.1177/09544100231219736

Korba, Peter, Patrik Šváb, Michal Vereš, and Ján Lukáč. "Optimizing Aviation Maintenance through Algorithmic Approach of Real-Life Data." Applied Sciences 13, no. 6 (2023): 3824. https://doi.org/10.3390/app13063824.

Llopis, Carlos. "The Impact of Data Quality on Aircraft Safety." *EXSYN Aviation Solutions*, July 3, 2023. https://www.exsyn.com/post/the-impact-of-data-quality-on-aircraft-safety.

Upadhyay, Rupali, and Hemant Amhia. "LSTM-Based Approach for Remaining Useful Life Prediction of Air Craft Engines." *ITM Web of Conferences* 57 (2023): 3004-. https://doi.org/10.1051/itmconf/20235703004.

Wang, L., Y. Chen, X. Zhao, and J. Xiang. 2024. "Predictive Maintenance Scheduling for Aircraft Engines Based On Remaining Useful Life Prediction." IEEE Internet of Things Journal, January, 1-1–1. doi:10.1109/JIOT.2024.3376715.

Wang, Endong, and Zhigang Shen. "A Hybrid Data Quality Indicator and Statistical Method for Improving Uncertainty Analysis in LCA of Complex System – Application to the Whole-Building Embodied Energy Analysis." *Journal of Cleaner Production* 43 (2013): 166–73. https://doi.org/10.1016/j.jclepro.2012.12.010.

Yang, Kaifei, Yakun Wang, Xiaodong Han, Yuehua Cheng, Lifang Guo, and Jianglei Gong. "Unsupervised Anomaly Detection for Time Series Data of Spacecraft Using Multi-Task Learning." *Applied Sciences* 12, no. 13 (2022): 6296-. https://doi.org/10.3390/app12136296.

Appendix

Table1

B777 E24 Report

- **timestamp**: The specific date and time when the data was recorded.
- **ACID**: Aircraft Identification.
- **DPT**: Departure airport code.
- **DST**: Destination airport code.
- **DATE**: The date of the flight.
- **GMT**: Greenwich Mean Time when the data was recorded.
- **FM**: Flight Mode (e.g., Takeoff, Landing, Cruise).
- **SER_NUM_L**: Serial Number of the Left Engine.
- **SER_NUM_R**: Serial Number of the Right Engine.
- **EPR_L**: Engine Pressure Ratio for Left Engine.
- **EPR_R**: Engine Pressure Ratio for Right Engine.
- **EPRCM_L**: Engine Pressure Ratio Corrected for Left Engine.
- **EPRCM_R**: Engine Pressure Ratio Corrected for Right Engine.
- **N1_L**: Fan Speed as a percentage of maximum (Left Engine).
- **N1_R**: Fan Speed as a percentage of maximum (Right Engine).
- **EGT_L**: Exhaust Gas Temperature for Left Engine.
- **EGT_R**: Exhaust Gas Temperature for Right Engine.
- **N2_L**: Core Engine Speed as a percentage of maximum (Left Engine).
- **N2_R**: Core Engine Speed as a percentage of maximum (Right Engine).
- **WF_L**: Fuel Flow to Left Engine.
- **WF_R**: Fuel Flow to Right Engine.
- **WFRQ_L**: Fuel Flow Rate for Left Engine.
- **WFRQ_R**: Fuel Flow Rate for Right Engine.
- **TRA_L**: Thrust Angle for Left Engine.
- **TRA_R**: Thrust Angle for Right Engine.
- **P2_L**: Pressure at station 2 in Left Engine.
- **P2_R**: Pressure at station 2 in Right Engine.
- **P2_5_L**: Pressure at station 2.5 in Left Engine.
- **P2_5_R**: Pressure at station 2.5 in Right Engine.
- **P5_L**: Pressure at station 5 in Left Engine.
- **P5_R**: Pressure at station 5 in Right Engine.

Figure 1 Engine Pressure Ratio (EPR) Left and Right (perform similarly over time)

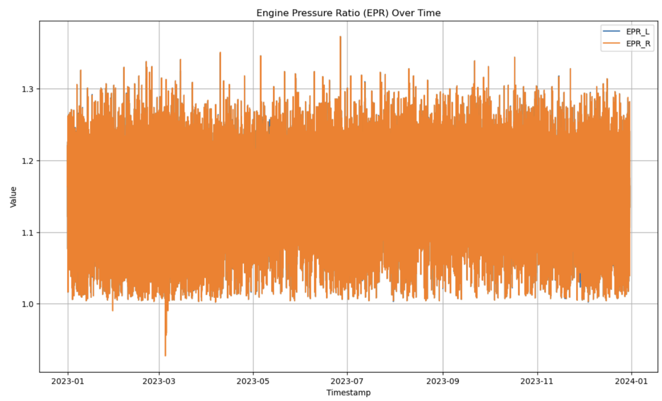

Figure 2 Engine Pressure Ratio (EPR) Left and Right (perform similarly over time) by aircraft 968

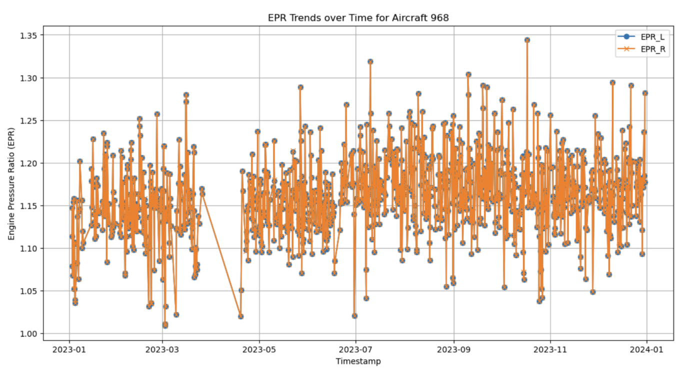

Figure 3 Exhaust Gas Temperature (EGT) Left and Right (Significant discrepancies)

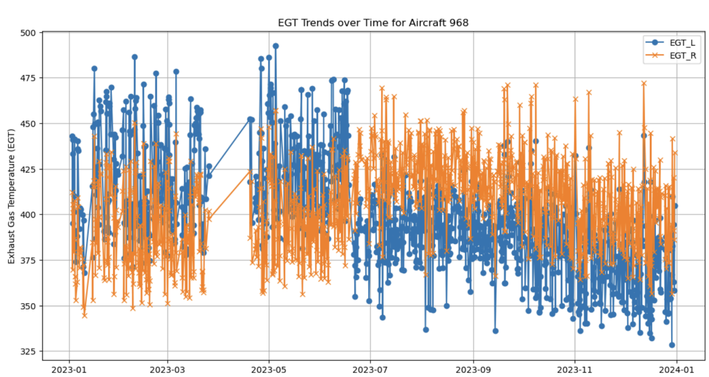

Figure 4 Exhaust Gas Temperature (EGT) Difference

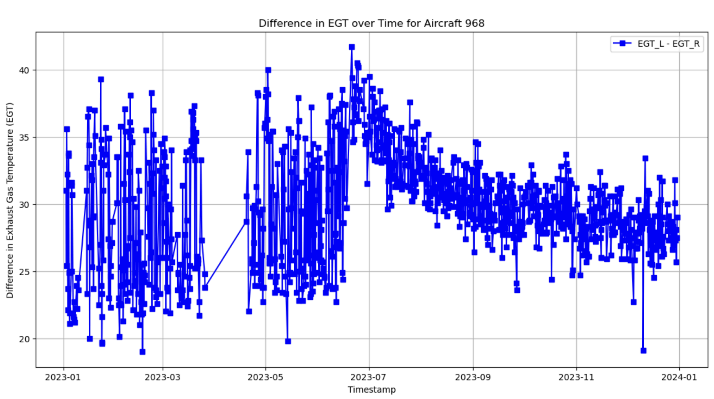

Figure 5 Engine Pressure Ratio (EPR) Left Original vs Deseasonalized

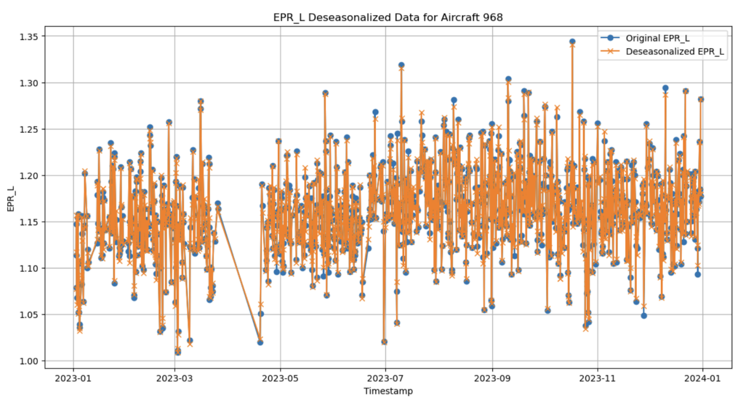

Figure 6

```
Epoch 9/50
16/16 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 0.0138 - val_loss: 0.0115
Epoch 10/50
16/16 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 0.0129 - val_loss: 0.0102
Epoch 11/50
16/16 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 0.0114 - val_loss: 0.0089
Epoch 12/50
16/16 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 0.0093 - val_loss: 0.0079
Epoch 13/50
...
Epoch 49/50
16/16 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 0.0014 - val_loss: 9.8544e-04
Epoch 50/50
16/16 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 0.0015 - val_loss: 9.2182e-04
```

Figure 6

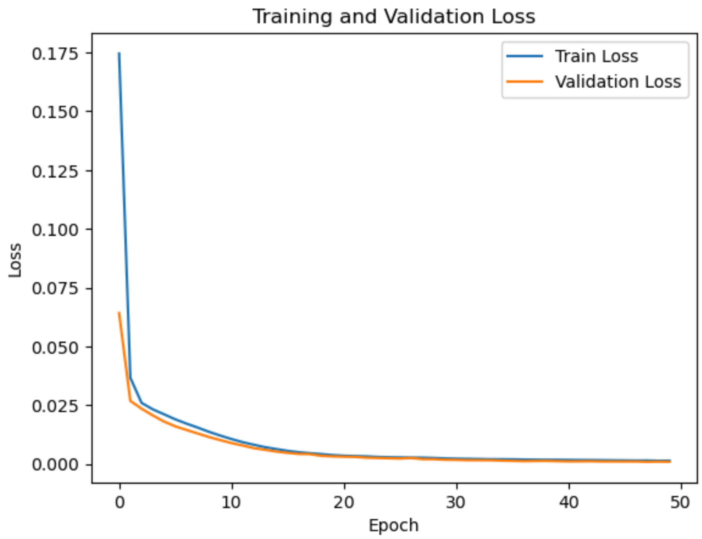

Figure 7

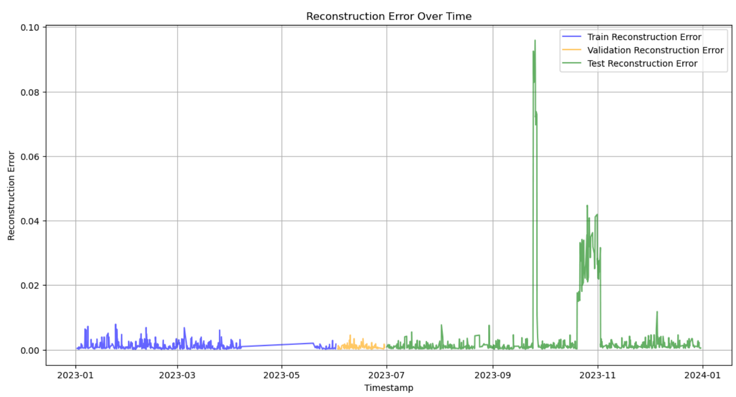

Figure 8

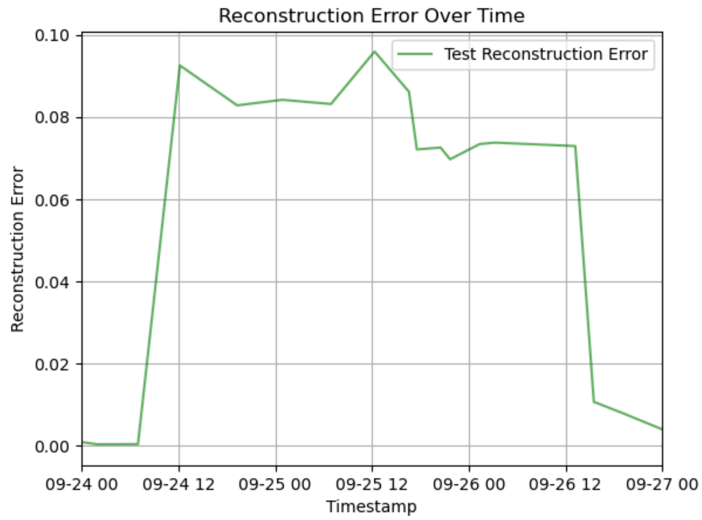

Figure 9

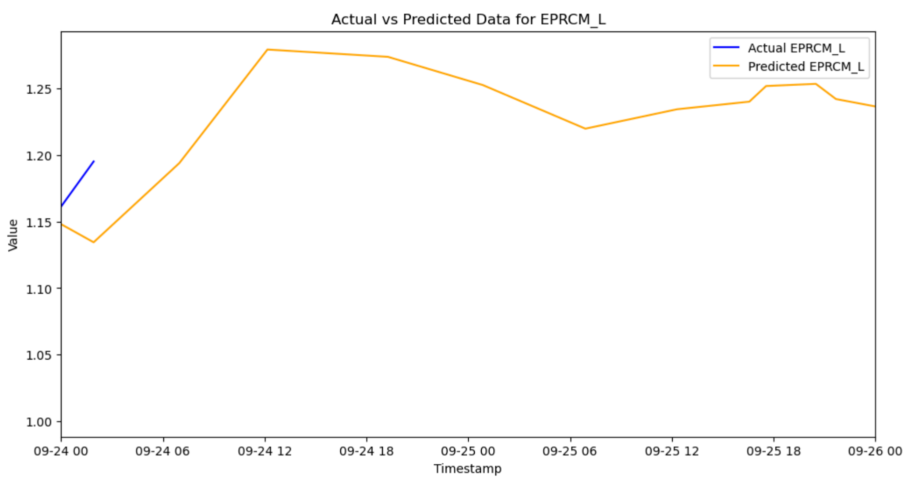

Figure 10

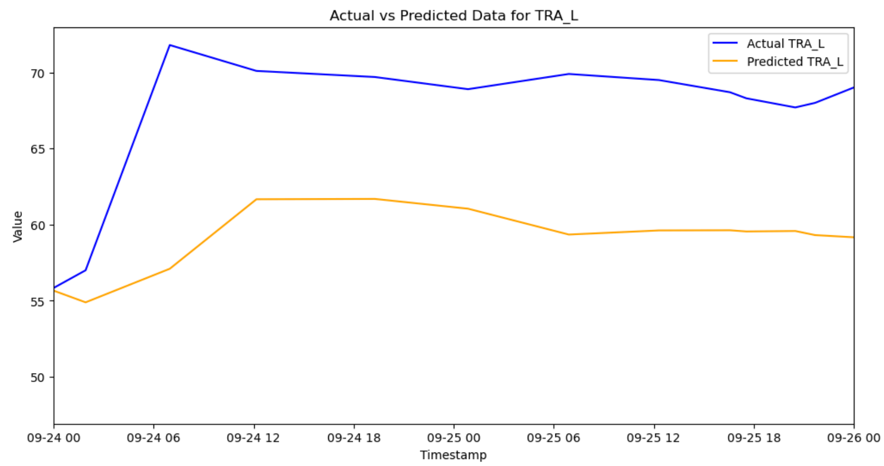

Figure 11

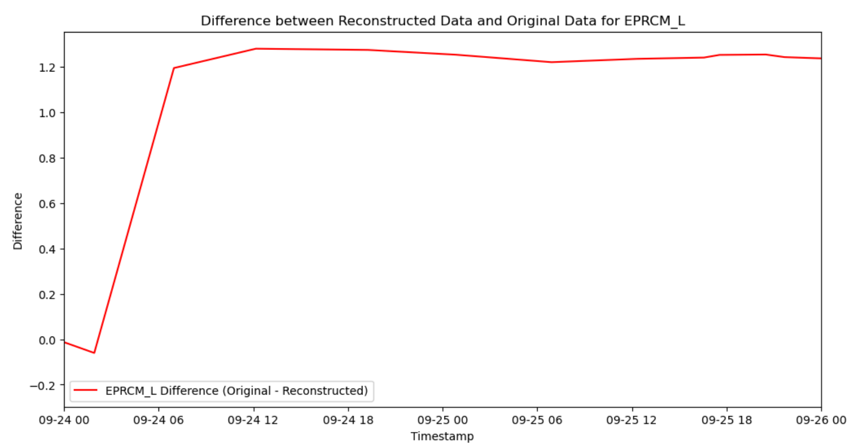

Figure 12