Enhancing Automotive Testing Efficiency with AI-Driven Monitoring and Assessment Solutions

Geevarghese George, Georgios Koutroulis, Gerhard Schagerl, Milan Zivadinovic

{geevarghese.george, georgios.koutroulis, gerhard.schagerl, milan.zivadinovic}@avl.com

Abstract: Due to the ever increasing worldwide competition between premium automotive manufacturers and the stringent regulation standards, novel methods from the field of AI are required to substantially decrease developments costs without sacrificing the end quality of the product. However, it is challenging, first, to identify the right use case within the automotive product development and testing phases, and second, to choose the appropriate AI algorithms that will address the problem. In this article, three innovative solutions are discussed that aims at enhancing the testing efficiency, and ultimately decrease development times and costs.

First, an anomaly detection framework is presented for testing of newly developed internal combustion engines (ICE). Next, we present a data-driven methodology to perform End-of-Line (EoL) quality check to detect abnormalities in a semi-automated manner. Finally, a transfer learning approach for accurately estimating the State of Health (SoH) of e-vehicles is discussed. We demonstrate the feasibility and benefits of utilizing such statistical approaches for handling large amounts of telemetry data. The proposed solutions have the potential to automate and accelerate the testing process while minimizing the risk of faulty units.

1 Introduction

The demand for and transition to electric vehicles (EVs) has introduced new challenges in the automotive industry, with leading manufacturers reporting a loss of $20-30k per vehicle sold relative to the sales volume [Ltd]. Moreover, various economic factors in the post-pandemic era have driven up costs throughout the product lifecycle, even as efforts to further reduce greenhouse gas emissions and improve safety are intensifying. To this end, performing any kind of vehicle testing during the development phase, ranging from measuring emissions to validating the durability of the engine or an electric component is of paramount importance.

Such testing procedures are complex and need to be carefully designed, executed and ultimately monitored. Testbeds provide a controlled environment where various components, systems, and the entire vehicle can be rigorously tested and validated. These testing procedures typically generate large amounts of data which may be further analysed by modern data-driven techniques to extract valuable insights. Automated and efficient processing of this data supplies fast and accurate feedback to the development teams, enabling them to make informed decisions and making effective decisions in the upstream developmental process. In this paper, we focus on developing methods to address these challenges in the different phases of vehicle development and manufacturing.

2 Methodology

In the following sections, we describe three data-driven solutions, focusing on their methodology and impact on the automotive industry during the product development and testing phases. In Section 2.1 we describe the application of a graph neural network-based anomaly detection framework for automotive engine testbeds. In Section 2.2 we discuss a statistically derived methodology for classification of EoL testing data of engines. In Section 2.3 we present the methodology for the accurate estimation of battery SoH leveraging transfer learning.

2.1 Anomaly detection in automotive engine test-beds

In the context of vehicle development, anomaly detection [Agg] can be used to identify issues in the development process, such as faulty components or suboptimal configurations. The endurance test is a typical validation and verification (V&V) cycle conducted on a testbed to verify the reliability of an internal combustion engine (ICE). This performed by running the engine under a specific load and speed profile for a predetermined period of time. Data are sampled at a relatively high frequency (1Hz) by sensors placed on the engine and the testbed to generate multivariate time series (MTS) data, which may include channels such as torque, rotational speed, temperature, and pressure.

2.1.1 Testbed data description

The dataset consists of multiple test runs with varying test scenarios and operating conditions for a unit under test (UUT). It includes 864 measured channels (sensor readings), approximately an hour long, spread across 575 measurement files. Channels are manually categorized into three classes: *output channels* (customer-defined), *important channels* (as identified by AVL domain experts), and *unimportant channels*. Channels irrelevant to the UUT, not present in all files, constant, erroneous, and redundant channels were removed based on expert review. The raw dataset does not contain any labeled anomalies. However, the engineers maintain a testbed diary which contains the information about the test runs and the anomalies detected during the test runs. This information is used as a ground truth to validate the detected anomalies by the model.

Characteristic patterns in the testbed signals, including level shifts at the beginning and high fluctuations at the end, were observed. The model should learn these patterns as normal behavior and detect anomalies such as sudden noisy spikes or drops in the signals. The anomaly detection framework is applied to such high frequency MTS data in an unsupervised [HTF] manner to extract channel interdependencies and forecast UUT failures.

2.1.2 Anomaly detection framework

The anomaly detection model that we employed is based on the recent developments in the domain of geometric deep learning. More specifically, we used the Graph Deviation Network (GDN) [DH] which is a graph-based model that uses Graph Attention Networks (GATs) [VCC+] to forecast the future values of a MTS.

The main benefit of a graph-based MTS forecasting model is that the model is able to learn non-linear, temporal dependencies along with structural- relationships between different channels. In order to achieve this, LASSO feature selection [HTF] is applied to the *important* and *output* channels of the MTS. The selected features are used to create an adjacency matrix, which is then transformed into embedding vectors to learn the graph structure. The GATs model takes these embedding vectors and self-lags [Box] of the channels, where a masked attention operation [KW] is performed over the edges of the graph.

The model performs a rolling forecast with a horizon of one timestep. The forecasting error is measured by common regression metrics such as mean squared error (MSE) and coefficient of determination (R^2). For quantifying a value as an anomaly, a model specific metric, known as the Graph Deviation Score (GDS) [DH] was also used. If the forecasted value is outside a threshold δ, set to the maximum GDS value per file, the model flags the value as an anomaly. The full workflow is depicted in Fig A.2.

2.2 Classification for End-Of-Line testing of engines

While the previous section focused on anomaly detection methodologies for testbed data during endurance testing of ICE engines, the following section describes a series of cold test (non-operating engine) measurements conducted sequentially as part of the EoL quality check. A typical EoL test sequence involves comprehensive evaluations, including initial inspections, cold start tests, step-wise RPM changes under varying load conditions, and emissions assessments. Each test sequence focuses on specific engine parts or behaviors and is associated with a specific set of channels.

The current practice during analysis of the recorded data is to use global extrema values (minimum and maximum) for each channel to detect possible faults. For this, a short window is selected and comparison is made with the predefined bands to check if the extrema value lies within the allowed range. If the value falls outside the predefined bands, a fault is reported. However, this methodology has drawbacks as the thresholds are set empirically and may not be optimal for all cases.

2.2.1 EoL data description

Each sensor ideally produces a time series with same length and frequency attributes during the measurement. We selected the single largest group of engine type, tested with the same execution sequence as the source data for the proposed EoL solution. Even though the data from such a test sequence is of high precision, the time series can have noise and time-misalignment. In those cases, the original time series may also need filtering and time scale modification to align the time series.

2.2.2 EoL fault detection framework

Instead of predefined boundaries within time windows, we utilize a mathematical model that captures normal behavior of a time series. We employ a widely known sampling technique in statistics called *bootstrapping*, to construct a reference curve. To perform this, we select random samples of time series from a population of recorded time series. A mean signal is constructed by point-wise averaging of the selected sample. This process is repeated several times (hundreds to thousands) to create a new population of averaged time series'. The new population, according to Central Limit Theorem [Wil], will approach a normal distribution for a large sample size. The mean of this new population is the best estimate of the mean of the initial population of recorded time series. Based on the new population, the reference curve is constructed as the mean of the new population[1]. For each data point of the reference curve, a point-wise standard deviation (σ) can also be measured. The confidence intervals can typically vary for several standard deviations. Since the reference curve data points follow the Normal distribution, approximately 99.7% of the observations should fall within $\pm 3\sigma$ range. By relaxing the range to $\pm 4\sigma$, we should practically cover all *normal* observations. During thee classification phase, any data point outside this range can therefore be characterized as an *outlier*, using any statistical hypothesis testing such as a z-test.

2.3 Estimation of battery State-of-Health of electric vehicles

The current market for accurate SoH estimation is on the rise, with equal demand for cloud-based SoH estimators and predictive maintenance solutions. In this section, we demonstrate a data-driven machine-learning based approach as an alternative to the traditional physical state observer-based SoH estimation. Typically before the production of battery packs starts, in the design phase, the characteristics of cells are measured in a controlled laboratory environment. On the otherhand, recorded SoH data from the field is known to be noisy and inaccurate. To combat this issue with fleet data, a solution based on transfer learning [PY] was developed at AVL for accurate SoH estimation.

2.3.1 SoH data description

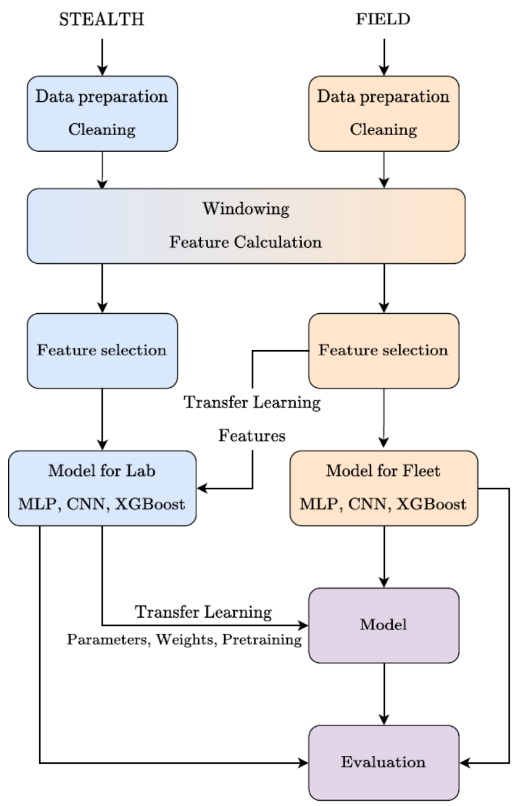
The data used for the SoH estimation usecase was collected from two different sources: laboratory data (STEALTH) and fleet data (FIELD). We collected data from two sources: laboratory (STEALTH) and fleet (FIELD). For the laboratory data, we cycled 8 lithium-ion cells on two testbeds. The relevant channels were current, voltage, and temperature. The data included aging cycles with periodic capacity measurements, tested over 100-200 days at a sampling rate of 4Hz. For the fleet data, we gathered telemetry data from 208 vehicles operated between 2019 and 2022. The relevant channels included current,

[1] Please note that such a reference curve can only be constructed for channels that show similar behavior on set of historical reference engines.

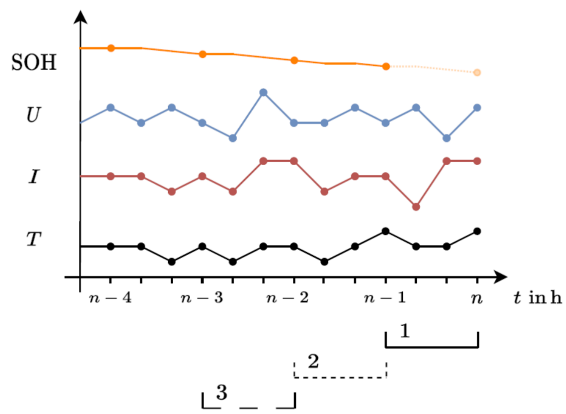
voltage, temperature, and SOH from the onboard BMS. The data was measured at a lower sampling rate of $1/30$Hz compared to STEALTH.

2.3.2 Framework for SoH Estimation

The proposed framework for SoH estimation leveraging transfer learning is depicted in Fig. 1a. The first step is to separately perform data cleaning, plausibilization and standardization for the two datasets. The preparation stage also includes the calculation of SoH from capacity, followed by adding a resting (indicating charge, drive or storage mode) channel. Several common statistical, energy-based, SoC change related and heatmap features were calculated additionally.



(a) Proposed framework for SoH estimation using transfer learning.



(b) Representative figure demonstrating the window used as feature and the last point of SoH as target.

Figure 1: A data driven SoH estimation framework and training methodology.

We use a window-based approach with 1-hour and 120-hour windows to compute input features. The last SoH value in the window is used as the target value (Fig. 1b). We additionally performed feature selection using LASSO coefficients [HTF] and recursive backward elimination.

Three classes of machine learning algorithms, namely multilayer perceptron (MLP), convolutional neural network (CNN), and extreme gradient boosted trees (XGBoost) were compared for SoH estimation. By making sure that the features were generalizable across the two datasets (Fig. 1a), we were able to train the models using transfer learning. Transfer learning involves using knowledge from one domain to improve performance in another. In this case, the models were first trained on the STEALTH dataset and those model weights were fine-tuned for the FIELD dataset. This is appropriate since the area of application and modelling tasks for the two datasets are be the same. Please note that for XGBoost, there are several ways to perform transfer learning, such as transferring hyperparameters, continued training, and continued update, all of which were experimented with.

3 Results and Discussion

As described in Sec 2.1, the GDN model aims to learn the multivariate dependencies of the input channels during training. The graph structure that the model learned is shown in Fig A.1. The validation of the graphs was performed by domain experts, who described the graphs as excellent visual indicators of UUT behavior and a valuable tool for the detection of meta-trends. The generated graph had a total of 57 nodes (channels) of which 53 had edges (relationships) that reflected the real interdependencies to the UUT.

Another result from our study is that GDN excels in meta-trend analysis, providing clear and low-noise trend detection (Fig 2). It outperforms benchmark models[2] by identifying not only known trends but also discovering additional anomalous sequences related to erroneous behavior in the Unit Under Test (UUT). In terms of the step-ahead forecasting task, GDN performs satisfactorily, with mean squared error (MSE) of 0.0052 ($\sigma = 0.0064$) and R^2 of 0.7727 ($\sigma = 0.0052$), however it is not as performant as benchmark models.

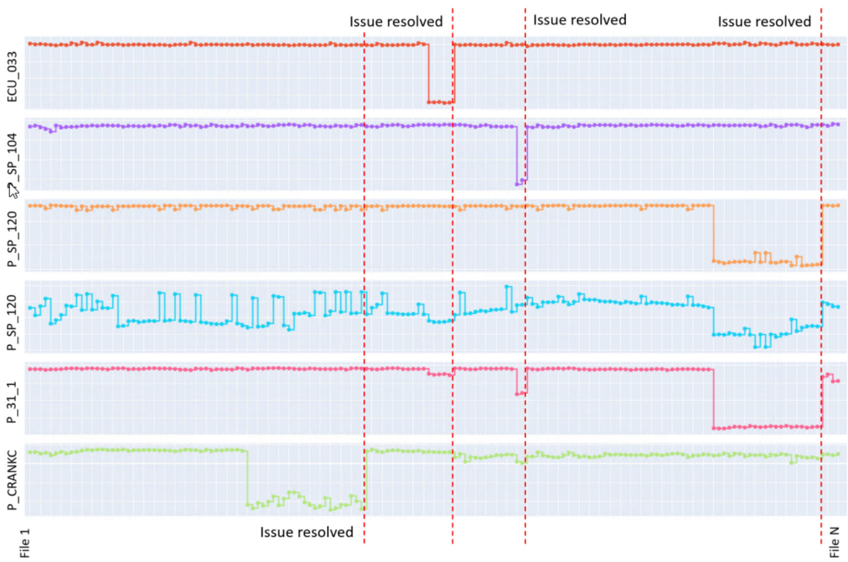
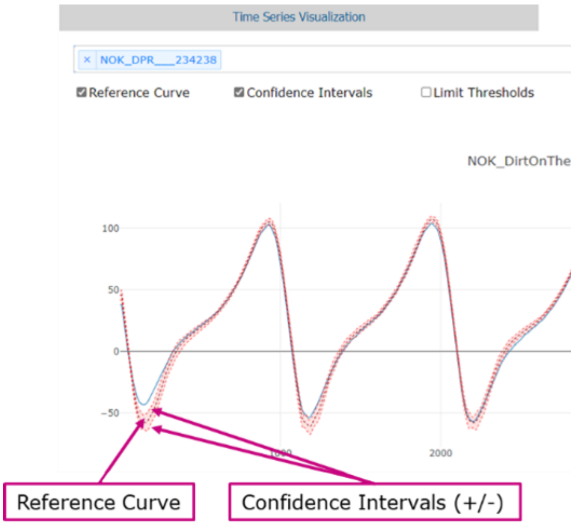


Figure 2: A section of six representative channels for unseen test files. The vertical axes show the R^2 metric per channel and the horizontal axis indicates the respective test files.

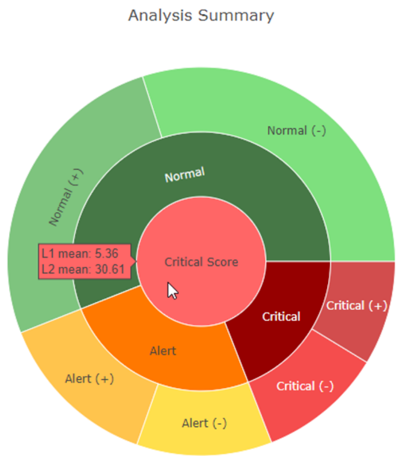
In Sec 2.2, an enhanced methodology for classification of EoL testing data of engines was

[2]MLP and LSTM were benchmarked on the same task.

presented. The main requirement for the OEM is to correctly classify possible erroneous engines during cold test in order to direct them to the second phase of a hot test. The presented methodology has helped OEMs to classify engine faults accurately and to reduce the number of false positive and false negative detections. A dashboard integrated with the AVL Data Analytics™ platform allows for interactive, real-time visualization for the operator along with the possibility to automate the process. A snapshot of the dashboard indicating the reference curve, adjustable confidence intervals and observed signal (Torque) is shown in Fig. 3a.



(a) Dashboard indicating the reference curve, confidence intervals and observed signal using the proposed methodology. The confidence intervals at the point level can be adjusted by the operator based on the boundary conditions.



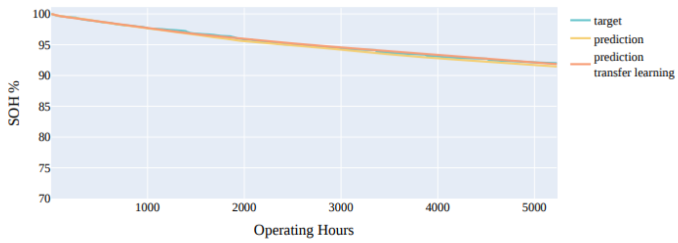
(b) Summary statistics for each severity band regarding the ratio of points found in each severity band (confidence interval). Additionally, the middle point in the sunburst figure represents the overall fault severity measured by mean L_1 and L_2 norm.

Figure 3: Artifacts from the point-wise severity analysis of EoL testing data.

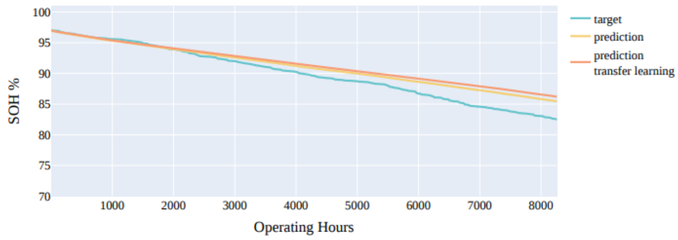
The fault severity is determined by the distance between the observed signal and the reference curve. This classification allows for a more granular distinction of the measured data points based on confidence intervals. This information can be visualized as part of the time series, or more conveniently as a sunburst chart as shown in Fig. 3b. Finally, the reference curve is dynamically updated to guarantee drift event amortization.

In Sec. 2.3, we described the methodology for the accurate estimation of battery SoH leveraging transfer learning. The important features for estimation included min, max, median SoC, mean current, temperature-based, and heatmap features. Additionally, energy throughput and resting time were relevant. Heatmap features were only useful for the FIELD dataset as the STEALTH dataset does not reach extreme temperatures. Lower SoH values and higher energy throughput were observed in the STEALTH data due to long age cycle tests.

With transfer learning, significant improvements were observed in comparison to training from scratch. For example, MLP showed a 35% reduction in the loss from start, and XGBoost showed a 36% reduction. The SoH estimation made with transfer learning of the XGBoost model is shown in Fig. 4. Pretraining on the STEALTH dataset helped the models converge faster and achieve better performance. XGBoost, with the transfer of hyperparameters strategy, resulted in the best performance, but other strategies mentioned in Sec. 2.3.2 were also competitive.



(a)



(b)

Figure 4: SoH estimation using XGBoost: The two plots show the (a) best, and (b) median prediction in terms of loss from start in the test set, for the model trained from scratch (yellow) and the model with transfer learning (orange) from STEALTH.

4 Conclusion

The automotive industry is undergoing a significant transformation with the advent of electric vehicles and the increasing demand for sustainable transportation solutions. Leveraging the vast amount of data generated during the vehicle lifecycle is crucial to minimize costs and improve the efficiency of the industry as a whole. In the near future, EU's data and AI strategy through the Digtial Europe Programme for shared data spaces would ease the barrier to data sharing across sectors and industries, including the automotive domain, to foster innovation and improve the quality of products and services.

In this context, we have explored how AVL can continue to support OEMs in addressing challenges related to testbed validation, end-of-line testing and state-of-health estimation for preventive and proactive action. Automation and digitalization tools such as AVL's Data Analytics™ is integrating these methodologies to provide customers with real-time feedback from various phases of the vehicle lifecycle.

The methodology we have developed for anomaly detection provides a novel approach for detecting meta-trends in time series data. Recent work carried out at AVL in the context of EV fleet monitoring has shown that GDNs are performant in detecting point anomalies [TMC+] as well. The technology presented for EoL testing has already been offered to OEMs to help them circumvent downtrends in engine production quality. The integration of the proposed methodology into existing workflows has the potential to significantly impact the testing phase by improving the quality of detections, reducing the time spent to analyze faults, and consequenty reducing the cost for engine testing. We also demonstrated that estimating fleet-wide SoH using transfer learning of models trained on laboratory data is a viable approach. Integrating such models into the BMS or on a cloud-based environment for real-time estimation and predictive maintenance is the logical next step.

5 Acknowledgements

This work was partly supported by the SM4RTENANCE (European Deployment of Smart Manufacturing Asset 4.0 MultilateRal DaTa Sharing SpacEs for an AutoNomous Operation of CollAborative MainteNance and Circular SErvices) consortium, which received funding from the Digital Europe Programme of the European Union Under Grant Agreement No 101123423.

References

[Agg] Charu C. Aggarwal. *Outlier Analysis*. Springer International Publishing.

[Box] George Box. *Box and Jenkins: Time Series Analysis, Forecasting and Control*, pages 161–215. Palgrave Macmillan UK.

[DH] Ailin Deng and Bryan Hooi. Graph Neural Network-Based Anomaly Detection in Multivariate Time Series.

[HTF] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer.

[KW] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks.

[Ltd] The Editorial Team at Plunkett Research Ltd. The Global Automotive Industry: Key Challenges and Opportunities.

[PY] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. 22(10):1345–1359.

[TMC+] Thomas Alexander Kristan, Milan Živadinović, Christian Rupert Rehrl, Roman Kern, and Alwin Tuschkan. AI-Based Digital Twin - Anomaly Detection and Diagnostics for HV Battery Behavior and Performance. 2024 JSAE Annual Congress (Spring).

[VCC+] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks.

[Wil] Rand R. Wilcox. *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*. Springer.

A Supplementary Figures

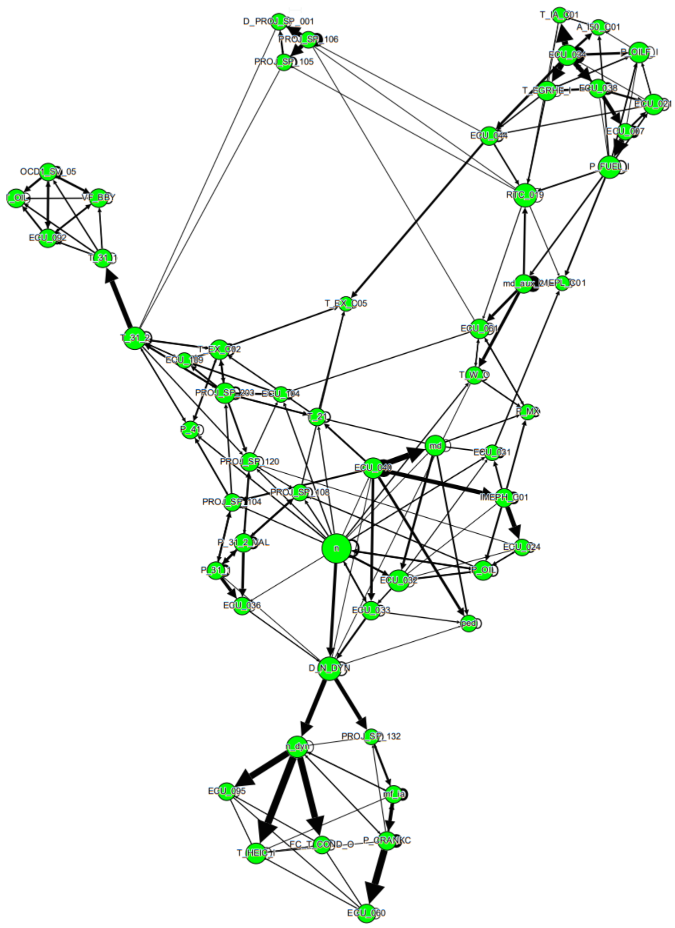


Figure A.1: The learned graph structure by the Graph Deviation Network (GDN). Each node is a channel and edge corresponds to relationships between the channels.

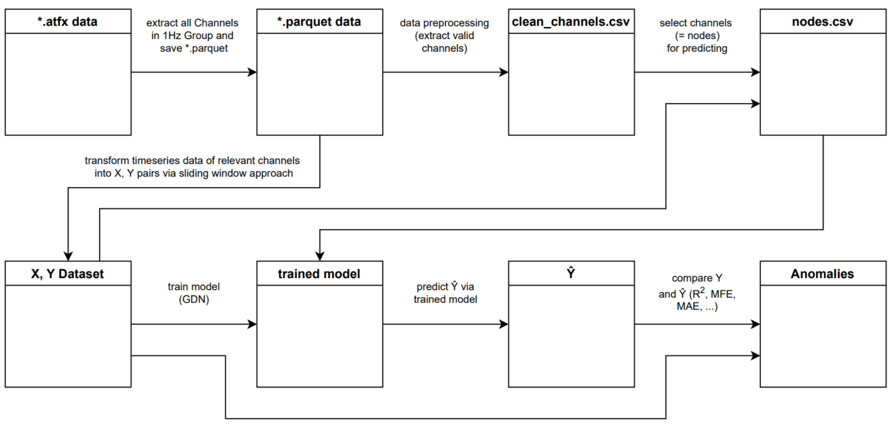


Figure A.2: Workflow for the anomaly detection task in automotive engine test-beds.