

IICOM – Interpretable Artificial Intelligence for Condition Monitoring

Public report



Project within: Vinnova FFI EMK
Author Abhishek Srinivasan, Scania CV AB
Date 10 Jan 2025



Fordonsstrategisk
Forskning och
Innovation

Content

1. Summary	3
2. Sammanfattning på svenska	4
3. Background	5
4. Purpose, Research Questions and Method	6
4.1. Method	7
4.1.1 Data Sources	7
4.1.2 Our Approaches	7
5. Objective	8
6. Results and deliverables	8
6.1 Results	8
6.1.1 Uncertainty Quantification	8
6.1.2 Causality	9
6.1.3 Interpretability	9
6.2 Deliverables	10
7. Dissemination and Publications	11
7.1 Dissemination	11
7.2 Publications	11
8. Conclusions and Future Research	12
9. Participating Parties and Contact Persons	13

FFI in short

FFI, Strategic Vehicle Research and Innovation, is a joint program between the state and the automotive industry running since 2009. FFI promotes and finances research and innovation to sustainable road transport. **For more information:** www.ffisweden.se

1. Summary

Predictive maintenance, an advanced maintenance approach, requires active condition monitoring models to estimate the health of the components, which in turn are used for maintenance planning. Current approaches include physics-based or domain knowledge-based methods, which require significant effort to develop, and data-driven methods, which often lack robustness, adaptability, and interpretability. To address this, our project has aimed to develop robust, adaptable and interpretable data-driven methods for condition monitoring (CM), and to evaluate those tools on selected systems in Scania trucks.

In this project, we have thus developed a set of complementary tools which addresses the challenges of machine learning (ML) for CM, specifically robustness, adaptability and interpretability. Some of the tools that were explored during the course of this project include: uncertainty quantification, causality, and counterfactual explanation approaches for ML models. Uncertainty quantification address robustness and adaptability, causality overlaps all the parts and counterfactual explanation for ML address interpretability.

Our result from different tools are as follows:

- Uncertainty quantification helps in disentangling knowledge uncertainty from process uncertainty, which clarifies the necessary action for the human users. Thus providing a measure of robustness for the model.
- Causality approaches like causal discovery can be used to discover underlying causal structure in engineered systems. Which can later be used for downstream modelling.
- Counterfactual explanation approaches for ML were improved for the application of CM.

The results from our project indicate significant potential to address the challenges. These developed tools have been demonstrated using Scania data, public datasets, or both.

2. Sammanfattning på svenska

Avancerade underhålls-strategier, som tillståndsbaserat och prediktivt underhåll, förlitar sig på modeller för tillståndsovervakning, som kan uppskatta olika komponenters hälsotillstånd, vilket i sin tur används för underhållsplanering. Tyvärr finns utmaningar med de modeller som används idag: De som är baserade på fysik-baserade simulatorer eller expertkunskap är väldigt arbetsintensiva att ta fram, medan de som är datadrivna inte är robusta, anpassningsbara, eller förklaringsbara. Projektet har därför syftat till att utveckla robusta, anpassningsbara, och förklarbara maskininlärnings-metoder för tillståndsovervakning för underhåll, samt att utvärdera dessa metoder på data från utvalda delsystem i Scania's lastbilar.

Vi har i projektet utvecklat ett antal verktyg för att hantera ovanstående utmaningar. Dessa verktyg inkluderar:

- Osäkerhets-kvantifiering, för att skilja osäkerhet i den underliggande processen från osäkerhet i maskininlärningsmodellen (orsakad av tex för få eller lågkvalitativa träningsdata). Detta ger ett mått på modellens robusthet, och därigenom hur pålitlig den är, vilket är viktigt att veta när man ska basera beslut på modellen.
- Metoder för att hitta kausala relationer i tids-serie-data från systemet som ska underhållas, och basera modellerna på dessa. Det kan ge robustare förutsägelser och bättre generalisering till nya situationer än traditionella maskininlärnings-metoder.
- En metod baserad på "kontrafaktiska förklaringar", som kan användas för att förstå den bakomliggande orsaken till en avvikelse i data. Detta är användbart när man ska identifiera och lokalisera fel som uppstår i systemet.

Såväl syntetiska data som data från Scania's system har använts för att testa och utvärdera metoderna. Resultaten visar på en stor potential för dessa metoder att möta utmaningarna kring maskininläring för tillståndsovervakning.

3. Background

Scania, a subsidiary of the Traton group and a manufacturer of heavy vehicles including trucks and buses. Scania has a vested interest in providing reliable products to its customers. One of the primary focuses is to enhance reliability through improved maintenance services, with predictive maintenance being a natural area of interest.

Maintenance services can be categorized into three types: reactive, preventive, and predictive. Reactive services respond to events or failures as they occur. Preventive services, on the other hand, are performed at regular intervals based on operational or calendar time. Predictive services actively monitor the current state of components to provide relevant maintenance suggestions. Predictive maintenance requires continuous monitoring through advanced condition monitoring models.

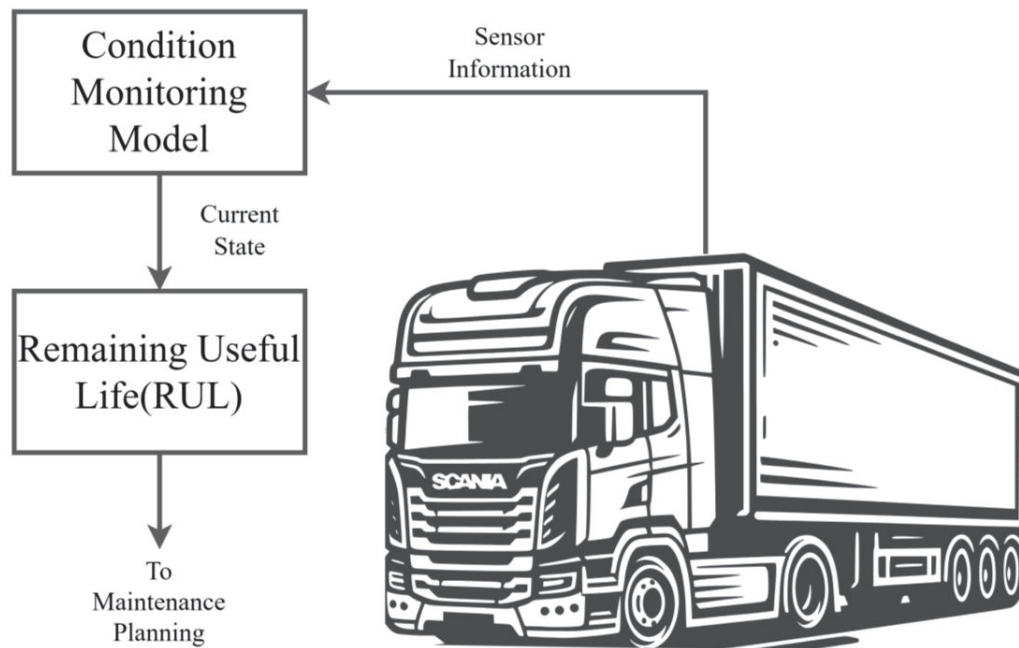


Figure 1, Illustration of modules involved in predictive maintenance process. Note: this illustration is one of the possible approaches to predictive maintenance. The sensor information is utilised by the condition monitoring model to determine the current health state of the component. This information is in turn taken used by RUL estimating model.

Predictive maintenance encompasses different elements (illustrated in the **Error! Reference source not found.**): the data source (sensor information), the condition monitoring model, the remaining useful life (RUL) model and the maintenance planner. The condition monitoring model uses the sensor information to determine the current state of the component or system that it is monitoring. This information is in turn utilized by the RUL model which uses the historic utilisation to predict the future utilization and estimates the remaining time to failure.

Currently, monitoring models could offer one of two types of health measures binary (healthy/unhealthy) or detailed health index (somewhere between 1 to 0). The monitoring models are primarily physics-based, requiring intensive modelling through domain knowledge and inputs from subject matter experts, while providing detailed health index. These approaches become less feasible on a larger scale due to the number of components and systems that need to be monitored. An other approach, rule-based models, works well as simple health monitoring tools that indicate whether the current state is healthy or not. However, they could only provide binary health measure.

Data-driven approaches for monitoring, which leverage statistics and machine learning, offer a solution by overcoming some of the limitations of physics-based and rule-based models. These data-driven approaches require less domain knowledge, less time from experts, and can provide binary and detailed health measures. However, machine learning models encounter challenges including sensitivity to spurious correlations, difficulties in interpretation, and limited adaptability, as they may not generalize effectively beyond the training data distribution.

This project aims to develop robust, interpretable and adaptable ML for condition monitoring. Robustness allows the models to have stable performance on different perturbation like attacks and noise modes, interpretability allowing the users of the model being able to understand the thought behind the models and finally adaptability allows learning new operating modes of the data which adapting to the new scenarios. We approached these problems with a set of tools including uncertainty quantification, counterfactual explanation and causality.

4. Purpose, Research Questions and Method

The primary objective of the project is to develop machine learning models for CM that are robust, adaptable, and interpretable. To achieve this objective, we have identified a series of research questions that address these three main themes. The research questions are as follows;

- RQ1. How does *Uncertainty Quantification* help condition monitoring? Can Uncertainty estimation be used for condition Monitoring (a binary health indexing case) and for RUL estimation?
- RQ2. *Causality* role in ML models, How can we obtain causal information from observable data? Can causality help in condition monitoring?
- RQ3. What are the limitations of current *counterfactual explanation* approaches to DL for anomaly detection (a simple binary health indexing case)? Can we improve the state of the art?

Addressing these research question has contributed to the three main themes. Corresponding contribution of these research question to the main themes are marked in different region illustrated in Figure 2.

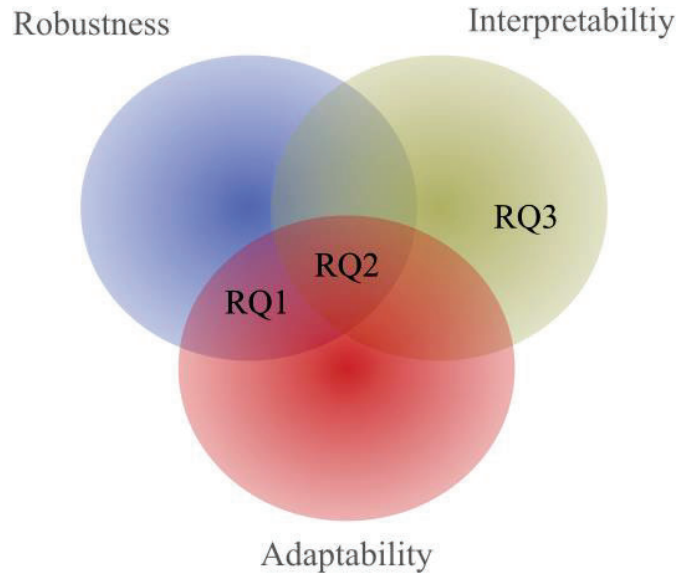


Figure 2 The Venn-diagram has three parts, one for robustness, adaptability and interpretability. The contribution of the research questions to the theme is marked in the illustration.

4.1. Method

4.1.1 Data Sources

This project utilized various data sources, both internal and external. Internal data were collected from field test trucks, focusing on systems such as the reductant pump, air pressure system, fuel system, and speed sensors. External sources included SKAB [1] and CMAPSS [2] datasets.

4.1.2 Our Approaches

Each of the research sub-question was addressed individually. The rest of this section describes the methods used to tackle these sub-questions.

Uncertainty Quantification focused on two problems: 1) anomaly detection, using both simulated data from models and real data from Scania's systems, and 2) Remaining Useful Life (RUL) prediction on the CMAPSS benchmark dataset.

Causal models concentrated on the problem of causal discovery over time, in engineered systems, utilizing both simulated models and data from Scania systems.

Counterfactual explanation of NN models focused on anomaly detection in a condition monitoring system, with experiments conducted on both Scania's data and SKAB public data-set.

5. Objective

The objectives of the project are as follows:

- Use casual models to build robust, reliable and interpretable condition monitoring leading to ML RUL model.
- Test the developed approaches on Scania systems such as NoX sensor, reductant pump and particulate filter.

The project was successful, with outcomes detailed in section 6 and the publication in section 7. Minor changes were made to the objectives; the updated version follows (changes italicised).

- *Build tools* for robust, *adaptable* and interpretable condition monitoring leading to ML RUL model.
- Test the developed approaches on Scania systems reductant pump, *air pressure system, fuel system and speed sensors*.

6. Results and deliverables

6.1 Results

As previously discussed, the goal was to develop robust, adaptable, and interpretable machine learning models for condition monitoring. This was achieved through the development of various tools, including uncertainty quantification, causality, and interpretability

6.1.1 Uncertainty Quantification

Uncertainty quantification methods measure the uncertainty in model's prediction. This tool address RQ1. This tool mainly contributes to the theme of robustness and partially to adaptability. Our publication (P1), focuses on quantifying the uncertainty bounds on model's prediction. Here, we developed single stage model to predict RUL directly from sensor data skipping the monitoring model (refer **Error! Reference source not found.**). We utilized ensemble probabilistic neural networks to model the uncertainties, as the ensemble agreed where the data existed and otherwise where they lacked knowledge on. Our ensemble approach was able to disentangle the different sources of uncertainties, from knowledge and process (i.e. epistemic and aleatoric). Here, our experiments on CMAPSS public data showed that this dis-entangling is vital in decision making processes. The disentangled uncertainty provides knowledge on what inputs the model was less confident

on, this information can be utilised to the data collection for model re-training. Later the developed method was also extended to anomaly detection case (i.e. a binary health index) for reductant pump use-case. Here, we additionally researched how the disentangled uncertainties could be utilized for fault diagnosis and retraining of models.

6.1.2 Causality

Causal models has been one of the tools that was not so straightforward in the application of condition monitoring. Causal models in theory should be able to help with all the main themes of the project goals. Some of the existing research has shown promises in this direction, Peters et. al [3] uses invariance to building adaptable models which can better the generalize outside the train space. Their modelling approach uses clear input and out feature space. However, the monitoring system requires total monitoring of the system rather than a specific feature. This brings us to understand the causal relationship between the feature space. Understanding these relationships from subject matter experts can be challenging as the knowledge about different parts exist in silos and never been studied in total.

To address this causal discovery approaches to determine the causal relationship between variable from observable data. Some of current state of the art approaches for causal discovery were tested on Scania data. The causal graphs generated by these algorithms were not satisfactory as they failed to realize obvious connections. We further developed an approach which supplements current causal discovery approaches and enables causal discovery on time-series data. This supplement framework uses mutual information rate to identify the skeletal graph which can further be used by normal causal discovery.

6.1.3 Interpretability

Interpretability approaches for machine learning models were another focus area for the project, addressing RQ3. We observed that most existing tools for explainable ML were designed for tabular data, whereas our data sources from the truck were time-series in nature.

To tackle RQ3, we selected a simple binary health index modelling method, an anomaly detection problem. Anomalies flagged by the model are usually hard to interpret due to several factors, including the necessity of knowing model's internal workings and having domain knowledge. To better understand why model flags a sample anomalous, we chose counterfactual explanations, which generate a "what if" scenario where the prediction is altered. Current state-of-the-art approaches generate counterfactuals by explaining the entire input space. However, anomalies in systems are usually caused by specific subsets of signals, making it impractical to explain the whole input space.

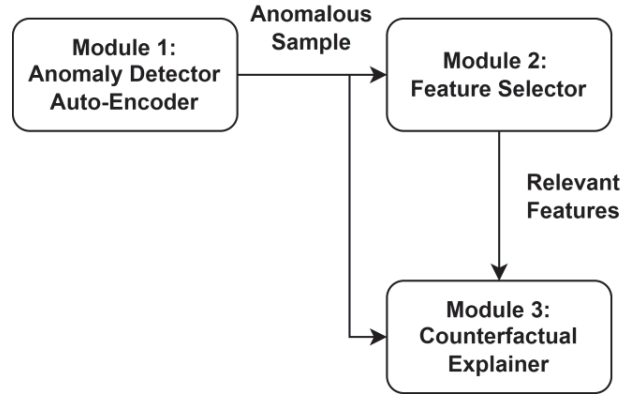


Figure 3 Illustration of developed counterfactual explanation framework.

Our work, detailed in publication (P2), focused on addressing this problem in counterfactual in explanation by selecting the right set of features for generating counterfactual explanations. Here we built a framework (illustrated in Figure 3) where the module two feature selector takes in the anomalous sample and evaluate each feature if they need to be explained or not. Based on the selection form module 2 our explainer in module 3 provides counterfactual explanations. Our example scenarios demonstrated that the explanations provided by our approach were much more meaningful compared to conventional ones.

6.2 Deliverables

List of deliverable as mentioned in the application,

- D1: Detailed review on current methods and models
- D2: Data collection of real-time data form test-rig
- D3: Data collection of vehicle real-time data
- D4: First internal progress report
- D5: First conference article submitted
- D6: Second internal progress report
- D7: Second conference article submitted

Almost all the deliverables were achieved as planned. Specifically, D1 and D3 to D7 were completed successfully. However, the deliverable D2 (data collection from the test rig) was not executed because we managed to collect the data directly from the vehicles, rendering D2 unnecessary.

7. Dissemination and Publications

7.1 Dissemination

How are the project results planned to be used and disseminated?	Mark with X	Comment
Increase knowledge in the field	X	This project has provided Scania with valuable insights into fault prognosis, adaptability, robustness, and interpretability. Additionally, it has advanced research by contributing to various scientific articles detailed in section 7.2.
Be passed on to other advanced technological development projects	X	The knowledge about the project will be carried over to the second half of the industrial PhD project, where the student will pursue on building models for prognosing failures.
Be passed on to product development projects		
Introduced on the market		
Used in investigations / regulatory / licensing / political decisions		

7.2 Publications

The contents of this research project has been published over several publications. List of publication over thesis project, past conference and planned future conference are listed below

Master thesis hosted during the same time on related topics.

1. Nouri, Ali. "Causal Discovery for Time Series: Based on Continuous Optimization." (2023).
2. Prasad, Deepthy, and Hampapura Sripada, Swathi. "Neural Network-based Anomaly Detection Models and Interpretability Methods for Multivariate Time Series Data." (2023).
3. Singapura Ravi, Varun, "Counterfactual Explanation for Auto-encoder based Anomaly detection." (2024). (Not published)

Past Conference Publications:

- P1. Srinivasan, Abhishek, et al. "Ensemble Neural Networks for Remaining Useful Life (RUL) Prediction." PHM Society Asia-Pacific Conference, vol. 4, no. 1, 4 Sept. 2023, <https://doi.org/10.36001/phmap.2023.v4i1.3611>. Accessed 6 Aug. 2024.
- P2. Srinivasan, Abhishek, et al. "Counterfactual Explanation for Auto-Encoder Based Time-Series Anomaly Detection." PHM Society European Conference, vol. 8, no. 1, 27 June 2024, pp. 9–9, <https://doi.org/10.36001/phme.2024.v8i1.4087>. Accessed 6 Aug. 2024.

Future Publications (Planned):

1. Srinivasan, Abhishek et al. Uncertainty Guided Diagnostics.
2. Srinivasan, Abhishek et al. Mutual Information Rate for Time-Series Causal Discovery.

8. Conclusions and Future Research

The aim of this project was to develop methods for building robust, interpretable and adaptable ML models for predictive maintenance in the scope of an Industrial PhD. This project covers the first half of the PhD.

Here we have developed an uncertainty estimation method for remaining useful life prediction (RUL) [P1], this method is based on an ensemble of neural networks, it decouples the aleatoric uncertainty from the epistemic uncertainty, making it possible to estimate the intrinsic uncertainty of the model (epistemic). This method helps to build more robust models, since we can pinpoint when a ML model is inferring from data out of the trained data distribution. This gives insights into knowing when to retrain a model to make it more robust and generic.



Additionally, a counterfactual explanation method was developed to make models more interpretable [P2], we show that this approach has advantages over previous, especially to diagnose classified anomalies.

Moreover, a systematic analysis of four systems and data collection has been done to evaluate the methods and further models in the second part of the PhD. First preliminary results using causal discovery methods have been achieved with the collected industrial data.

References:

- [1] Katset, I. D. and Kozitsin, V. O., 2020. Skoltech Anomaly Benchmark (SKAB). Kaggle, DOI: 10.34740/KAGGLE/DSV/1693952
- [2] Saxena, A., Goebel, K., Simon, D. and Eklund, N., 2008, October. Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (pp. 1-9). IEEE.
- [3] Peters, J., Bühlmann, P. and Meinshausen, N., 2016. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5), pp.947-1012.

9. Participating Parties and Contact Persons

<p>Scania CV AB</p> <p>Abhishek Srinivasan</p> <p>EECCL Connected Systems Scania CV AB 151 87 Södertälje</p> <p>Tel: +46700863712 E-mail: abhishek.srinivasan@scania.com</p> 	<p>RISE AB</p> <p>Anders Holst</p> <p>Data Analysis Unit RISE AB Box 1263 164 29 Kista</p> <p>Tel: +46 10 228 43 13 Email: anders.holst@ri.se</p> 
---	--