From Cloud to Control Cabinet: Edge AI for the Next Generation of Lift Monitoring

Tim Ebeling

Henning GmbH & Co. KG, Loher Str.4, 58332 Schwelm, Germany

Keywords: Predictive Maintenance, Edge AI, Condition Monitoring, Lift Monitoring, Neural Processing Unit, Hybrid Cloud-Edge Architecture, Anomaly Detection, Expert System, Sensor Fusion, Model Adaptation

Abstract. Predictive maintenance has become a key enabler for improving the reliability, availability, and lifecycle management of modern lift systems. Current solutions typically rely on connected sensor devices that transmit operational data to cloud-based platforms where advanced analytics and artificial intelligence (AI) models are executed. While this architecture enables powerful diagnostics and fleet-level insights, it introduces dependencies on connectivity, latency constraints, and additional data transfer requirements.

This paper presents the next development stage of a predictive maintenance system for lifts in which artificial intelligence models are executed directly on-site at the edge device installed within the lift system. The approach builds on an existing cloud-connected monitoring architecture in which substantial preprocessing and feature extraction already take place locally. However, until now, the final inference of the deep learning models has been performed in the cloud. By introducing an enhanced hardware platform equipped with dedicated Neural Processing Units (NPUs), it becomes possible to execute Deep Neural Network (DNN) models locally while maintaining acceptable processing time and energy consumption.

Running complex DNN models on embedded systems presents a significant challenge due to limited computing resources. Although NPUs can accelerate neural network inference compared to CPUs, this can lead to trade-offs between speed and model accuracy. Therefore, a balanced approach is necessary. Initial results from the presented architecture are promising and enable predictive real-time diagnostics directly within the lift system. This reduces reliance on cloud connections, improves response times, and paves the way for more autonomous and scalable predictive maintenance solutions in the lift industry.

1 INTRODUCTION

The requirements for modern lift monitoring have changed fundamentally in recent years. Building owners and operators no longer expect merely the reporting of faults that have already occurred; they increasingly require information about the future condition of their installations. The objective is to plan maintenance activities according to actual needs, prevent unplanned downtime, and optimise the economic utilisation of technical components.

Machine learning techniques, and particularly Deep Neural Networks (DNNs), offer significant potential in this context. To date, such methods have predominantly been deployed within cloud infrastructures, as both training and inference require substantial computational resources. However, the increasing availability of powerful embedded hardware now enables key elements of artificial intelligence to be executed directly on monitoring devices installed on-site.

The transition from cloud-based AI to Edge AI should not be regarded as a replacement of existing architectures, but rather as their logical evolution. The central question is therefore not whether cloud or edge computing is superior, but which tasks should be performed most effectively at which location.

2 FROM DATA COLLECTOR TO LOCAL DIAGNOSTIC NODE

Since 2004, the author's company has installed a five-figure number of lift monitoring systems worldwide. Initially, these systems represented an approach that had little connection to artificial intelligence. The foundation of the system is the continuous acquisition of operational data directly from the lift installation. Motion profiles, load and travel events, condition parameters, signal characteristics, and derived features are collected in such a way that meaningful technical conclusions regarding the condition of the system can be drawn.

This existing data infrastructure represents a significant strategic advantage. Many AI projects fail not because of a lack of algorithms, but due to insufficient historical data, inadequate data quality, or non-reproducible preprocessing methods. The monitoring platform already incorporates an architecture in which substantial preprocessing is performed locally. Raw data is filtered, condensed, and transformed into features suitable for diagnostics and trend analysis. While cloud-based evaluation enables the use of powerful models, it also requires reliable connectivity and introduces additional communication and operational requirements.

The transition towards Edge AI therefore represents a natural extension rather than a disruption of the existing architecture. The lift installation is no longer merely observed; it is equipped with a local diagnostic node. The device installed in the control cabinet evolves from a sensor gateway into an intelligent analysis platform capable of executing models, evaluating operating conditions, and prioritising relevant events. Existing expertise, historical data, and proven evaluation logic are leveraged to bring artificial intelligence closer to the point where technical conditions actually arise within the lift installation itself.

3 AI AS AN EXTENSION OF THE EXPERT SYSTEM – NOT A REPLACEMENT

The objective is to combine traditional expert systems with learning-based approaches. An expert system explicitly represents engineering knowledge through thresholds, plausibility checks, known fault patterns, condition classes, and experience-based decision trees. Such rules are transparent, auditable, and well established within safety-related environments.

Artificial intelligence complements this knowledge by identifying patterns that cannot easily be expressed through deterministic rules. Examples include subtle signal variations, combinations of multiple weak indicators, or installation-specific developments occurring over extended periods of operation.

A particularly interesting aspect for Edge AI systems is that the expert system can also contribute to the labelling process. Within real-world lift data, clearly defined and manually verified fault labels are relatively rare. Many events are ambiguous, occur with significant delay, or are only assessed during maintenance activities. Indirect labelling utilises existing rules, maintenance records, threshold violations, service reports, and recurring patterns to generate training data with sufficient explanatory value. Labels are therefore not created solely through manual annotation, but through the correlation of multiple technical indicators.

In this context, AI should not be viewed as a mysterious black box. Rather, it represents a second layer of analysis built upon an already well-established engineering foundation. The expert system provides structure, plausibility, and explainable intermediate states, while neural networks contribute

pattern recognition, generalisation, and probabilistic assessment. The result is a hybrid diagnostic concept that applies rule-based methods where rules are effective and learning-based methods where patterns become too complex for conventional approaches.

4 HYBRID EVALUATION THROUGH A PUSH/PULL ARCHITECTURE

Edge AI does not replace the cloud. Its strength lies in an effective distribution of responsibilities. Time-critical and repetitive analyses are performed locally on the monitoring device, including preprocessing, feature extraction, condition classification, anomaly detection, and event prioritisation.

The cloud continues to perform tasks that benefit from extensive computational resources, long-term data availability, and a fleet-wide perspective. These include model training, cross-installation comparison, validation, model deployment, and long-term trend analysis.

This results in a push/pull ecosystem. Devices do not continuously transmit large volumes of raw data. Instead, they selectively provide condensed information, unusual signal sequences, model metrics, and event correlations to the cloud. The cloud pushes improved models, configurations, or adaptation modules back to suitable device classes. Conversely, devices may pull updated models whenever improvements become available for a specific lift type, configuration, or detected operating pattern.

The objective is not merely data transfer but the creation of a continuously learning system that generalises knowledge across the fleet while remaining capable of local action.

From a practical perspective, four key benefits can be identified:

- Reduced dependency on permanent connectivity
- Faster response to local events
- Lower communication requirements through local data reduction
- Improved scalability, as not every evaluation must be performed centrally

Particularly within the lift industry, where installations often have long service lifecycles, varying modernisation levels, and highly heterogeneous operating environments, this hybrid architecture represents a practical and scalable solution.

5 MODEL ADAPTATION WITHOUT FULL EDGE TRAINING

A particularly promising building block is Low-Rank Adaptation (LoRA), a technique for parameter-efficient model customisation. The basic concept is straightforward: instead of retraining a large pre-trained model in its entirety, only small adapter structures or low-rank weight modifications are learned.

For on-site monitoring devices, this approach is especially attractive because full model retraining or complete model adaptation would require significant memory and computational resources.

Applied to lift monitoring, the concept assumes the existence of generalised foundation models trained across many installations and lift categories. Smaller adaptations can then be introduced for specific lift types, rope or belt configurations, modernisation states, or environmental conditions without redistributing the complete model.

Even with currently available monitoring hardware, controlled cloud-assisted adapter generation combined with local inference is already feasible.

Within this architecture, the expert system provides context and boundary conditions, the cloud trains and refines foundation models, and LoRA-based adapters form the bridge between generalisation and installation-specific specialisation. This also improves the efficiency of communication channels. Not all data needs to be transferred in raw form, while relevant anomalies and operational insights can still contribute to the continuous improvement of future model generations.

The result is an ecosystem that learns from every installation without requiring every installation to be trained independently.

6 SUMMARY

The combination of expert systems, Edge AI, and cloud-based fleet learning opens new opportunities for lift monitoring. In the short term, the primary focus lies on local inference, hybrid diagnostic approaches, and NPU-accelerated evaluation. In the medium term, installation-specific specialisation and adaptive learning methods will gain importance. In the long term, a distributed diagnostic ecosystem will emerge in which every installation contributes to improving overall models while simultaneously benefiting from the knowledge gained across the entire fleet.

The next generation of lift monitoring therefore shifts significant parts of artificial intelligence from the cloud into the control cabinet. By combining expert systems, Edge AI, on-site Neural Processing Units, and cloud-supported fleet learning, a system architecture is created that is faster, more resilient, and more scalable than purely cloud-based approaches.

The presented approach demonstrates that the future of predictive maintenance does not lie in moving intelligence entirely to a single location. Instead, the greatest benefit is achieved through an intelligent division of responsibilities between cloud and edge computing.

The cloud makes the system capable of learning; Edge AI makes it capable of acting locally.

BIOGRAPHICAL DETAILS



Tim Ebeling has been employed since 2003 as head of development with Henning GmbH & Co. KG. In this capacity, he has established the R&D centre in Braunschweig (Germany). A team of employees is now working there on the development and production of electronic and measurement components for lifts.

Since 2012, the author has also been Managing Director and a shareholder. One of his particular focal points is measurement technology. Especially in this area, the author looks back on many years of experience in the development of acceleration and rope load measuring systems. The author's professional goal is to enrich the lift market with innovative lift components.

In addition to his role as a board member of the German Elevator Association (VFA), he is active in working groups of the European Lift Association (ELA) and serves on the advisory boards of LiftJournal and the Centre for Elevator Technology in Roßwein (ZFA).