15th IFK
Proc. of the 15th International Fluid Power Conference <Symposium/Conference>
<DOI>

Machine Learning- Based Health Monitoring of High-Power Electro-Mechanical Actuator for Non-Road Mobile Machinery Using Experimental Data

Vinay Partap Singh[1,2,*], Tatiana Minav[1]

1 – *IHA lab, Faculty of Engineering and Natural Sciences, Tampere University, Finland.*

2 – *Danfoss Power Solutions, Nordborg, Denmark.*

* - vinaypartapsingh@tuni.fi , vinay.singh@danfoss.com

ORCID - https://orcid.org/0000-0002-0699-2712

Abstract.

This paper presents a deployable and interpretable framework for the health monitoring of high - power electro-mechanical actuators (EMAs) used in non-road mobile machinery (NRMM), turning standard sensor data into compact predictors, a health index, and finally a fused ensemble that tracks degradation for real-time maintenance.

Using experimental run-to-life data, four correlation-based coefficients identified in a previous study are computed to capture cross-sensor dynamics. An early-life baseline is established, and a bounded, monotone health index is derived using Mahalanobis distance-based deviation, persistence gating, and outlier-robust accumulation. Several machine learning regression models are trained and compared; three complementary models—Gaussian process regression, a bagged-tree ensemble, and a compact multilayer perceptron are selected for further process.

The models are then deployed in Simulink where raw sensor signals are transformed into correlation-based coefficients, which feed three trained regressors. The outputs of these individual regressors are fused via a performance-weighted scheme and smoothed with a rolling-window average to produce a bounded health index. In a separate 200-cycle dataset withheld from all training, the Simulink implementation tracks degradation accurately and suppresses transients. After an initial overshoot due to a shorter adaptive window, the HI remains close to 0.4 on a scale of 0 to 1, which is within the expected range for the withheld data from the HI map of the training data. The fused HI remains interpretable, monotonic in expectation, robust to brief operational disturbances, and consistent with training behaviour, and providing a practical basis for real-time maintenance decisions for high-power EMAs used in NRMM.

Keywords. Electro-mechanical actuators, non-road mobile machinery, machine learning regression, health monitoring.

ACRONYMS

- **EMA** – Electro-Mechanical Actuator
- **NRMM** – Non-Road Mobile Machinery
- **HI** – Health Index
- **MD** – Mahalanobis Distance
- **ML** – Machine Learning
- **MLP** – Multi-Layer Perceptron
- **GPR** – Gaussian Process Regression
- **RMSE** – Root Mean Square Error
- **FDI** – Fault Detection and Isolation
- **PHM** – Prognostics and Health Management
- **PRV** – Pressure Relief Valve
- **UI** – User Interface
- **EDA** – Exploratory Data Analysis
- **STD** – Standard Deviation
- **LOF** – Local Outlier Factor
- **MCSA** – Motor Current Signature Analysis

1. INTRODUCTION

Electro-mechanical actuators (EMAs) are rapidly being proposed for adoption in non-road mobile machinery (NRMM) due to their combination of high energy efficiency, precise motion control, and compatibility with the broader trend toward the electrification of heavy machinery. Unlike hydraulic systems, EMAs convert electrical energy directly into mechanical motion without the need for pressurised fluid, reducing energy losses and eliminating the complexity of pumps, valves, and hoses. This not only improves system level efficiency but also simplifies integration with modern, software-based control architectures that are increasingly prevalent in NRMM [1].

Despite these advantages, EMAs in NRMM operate under harsh and variable conditions that can accelerate component degradation. High mechanical loads, impact forces, and repetitive duty cycles impose wear on mechanical components, such as screw drives, bearings, and gears. Moreover, exposure to dust, moisture, and temperature extremes can degrade lubrication, increase friction, and cause corrosion. The electric motor and power electronics are also subject to thermal stress, insulation breakdown, and electrical component aging. Importantly, the majority of these degradation processes are not immediately apparent in the output of the actuator until a critical threshold is crossed, leading to sudden and unexpected failures. In contrast to hydraulic systems where issues, such as leaks or pressure drops, often provide an early warning, EMA faults can progress silently, making them more difficult to detect without continuous monitoring.

Research into health monitoring and prognostics for EMAs has grown steadily over the last few years for NRMM, with interest driven largely by aerospace applications and the broader adoption of more-electric systems. Early work in this area focused primarily on fault detection and isolation (FDI), aiming to identify actuator malfunctions as they occur rather than tracking gradual degradation. For example, a comprehensive prognostics and health management (PHM) framework for EMAs is developed by [2] by integrating model-based fault detection with a data-driven prognosis. Their system was validated on aerospace representative actuators, demonstrating accurate isolation of injected faults, such as motor winding failures and gear jams. However, as with much of the early PHM research, this work relied heavily on simulated fault scenarios and laboratory testbeds operating under simplified duty cycles, limiting its applicability to the variable, high-load operating environments found in NRMM. Subsequent studies extended EMA health monitoring by incorporating mechanical degradation mechanisms into the analysis. For example, the work on screw backlash and fatigue estimation examined the evolution of mechanical wear in transmission components, particularly in ball-screw assemblies [3]. The aforementioned study modelled wear progression and fatigue using analytical relationships and parameter estimation techniques, sometimes coupled with laboratory experiments under controlled loading. While they provided important insight into mechanical degradation pathways, they typically did not operate under the diverse and harsh conditions typical of NRMM service. A recurring theme across the literature is the use of synthetic or fault-injection data in place of genuine run – to - failure datasets.

A test rig has been developed for the experimental evaluation of a flight control EMA, with the ultimate goal of developing health monitoring algorithms in [4]. However, the actuator used in this study has a nominal load capacity of 300N. Furthermore, to generate the test data, the test rig has been subjected to axial loads up to 70% higher than nominal in order to accelerate degradation. The same research group proposes a general health monitoring approach for EMAs based on statistical process control methods using electric motor currents in [5]. In that study, the actuator does not reach failure, but testing was stopped at an event frequency of 60%. Another paper proposes a change detection algorithm for health monitoring of the same small EMA [6]. This algorithm compares the statistical features of the current experiment with past ones in batches, rather than providing an online estimation. Hence, the algorithm assumes that the external conditions on the EMA remain the same each time it is tested, which limits its application for real-time health monitoring. Two points stand out in the last three studies mentioned: first, the size of the EMA, with a nominal load capacity of 300 N, which is significantly low for any work function of NRMM; second, the assumption that external conditions (i.e., load, disturbances) remain the same during all test sessions, which is highly unlikely for any NRMM. The recent study by [7] notes that many approaches to actuator health assessment rely on artificially induced faults, such as deliberate increases in backlash or step changes in motor parameters, to generate degradation data for algorithm testing. While useful for algorithm development, these data often represent abrupt failures rather than the gradual degradation experienced in the real-life operating conditions. Furthermore, artificially induced faults can bias algorithms toward detecting the specific fault types injected, reducing generalizability.

In terms of sensing and feature extraction, multiple studies have explored multivariate feature spaces derived from actuator signals. These include time-domain statistics (i.e., mean, RMS, skewness, and kurtosis), frequency-domain parameters from vibration or current spectra, and residuals from physics-based models. For instance, [8] investigates condition monitoring through multiple sensor channels, applying statistical analysis to extract discriminative features. Similarly, the work in [9] discusses combining motor current signature analysis (MCSA) with vibration and temperature data to detect early anomalies. These approaches underscore the potential of feature fusion for more reliable health estimation. Feature extraction and multi-model fusion have been used for condition monitoring of an EMA in a train transmission in [10]. The study employs correlation-based feature selection and, through a Multimodal Fusion Improved Transformer Network, achieves better accuracy compared to traditional models. However, the study uses a 1:2 scaled-down experimental setup, and faults are induced manually rather than through natural degradation. That said, in most cases, the outcome is an anomaly score or classification label, not a continuous, interpretable health score.

The use of advanced machine learning (ML) methods for EMA health monitoring has also been investigated. Supervised learning techniques are applied in [11] to predict the actuator condition from sensor features. While these methods demonstrate strong offline performance, they are often computationally intensive, involving high-dimensional feature vectors and complex nonlinear models. Such complexity poses challenges for embedded real-time deployment, especially in NRMM applications where on-board computational resources may be limited. Moreover, many ML-based approaches in the literature are trained and validated on datasets that do not represent the full life cycle of the actuator, making them less robust to long-term drift and variability in operating conditions.

Another limitation is the focus on small to medium-sized EMAs, especially in aerospace and robotics contexts. In contrast, high-power EMA of the type used in heavy construction, mining, and agricultural machinery exhibit different thermal, mechanical, and electrical behaviours due to their larger components, higher load levels, and more severe environmental exposure. This gap is particularly significant given that high-power EMAs are now being deployed more frequently in NRMM applications, where predictive maintenance is essential.

This highlights a clear research gap where work on the health monitoring of high-power EMA working on duty cycles of NRMM is lacking so far.

This work directly addresses this gap by developing a health index for a high-power EMA using exclusively experimental degradation data from a long-term run-to-failure test. The method fuses multiple statistical features into a single normalised scale, enabling continuous tracking of actuator health. This work proposes a framework for constructing and predicting a health index for high-power EMAs operating under realistic duty cycles using experimental data. The starting point for this study is our previous research, in which a systematic feature selection procedure identified four statistical parameters from multi-sensor EMA data that exhibit consistent and monotonic degradation trends across the life cycle of the actuator [12]. Derived from load force, actuator velocity, and electric motor torque, these features form the basis for the HI construction.

A key challenge in this study is the absence of a clearly observed zero-health failure point. During the experiments, the actuator had an unexpected mechanical failure which is not directly related to the gradual degradation [12], which prevents the direct mapping of sensor data to a complete health trajectory from 1 (perfect condition) to 0 (end of life). To address this, the current study employs the Mahalanobis distance method [13, 14], which has been widely adopted in condition monitoring for combining multiple correlated features into a single metric that quantifies deviation from the healthy baseline. MD is advantageous in this context because it is scale-invariant, accounts for feature correlations, and produces a monotonic measure of degradation when applied to well-chosen features. The HI is then obtained by mapping MD values to a normalised scale where '1' corresponds to the healthy state and '0' to the end of life.

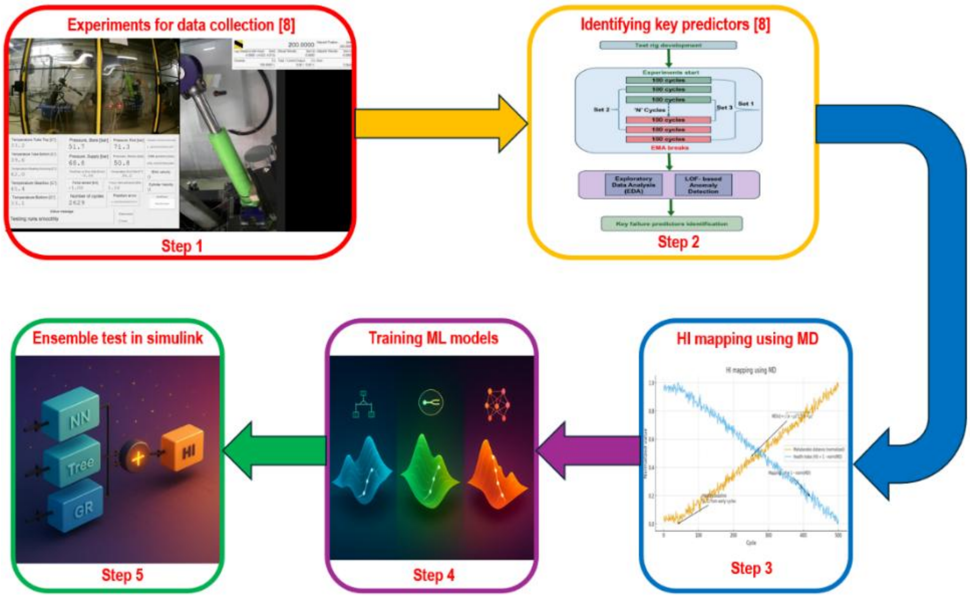

Figure 1. Overview of the study methodology

Figure 1 shows the overall process used for the study. In the above figure, Step 1 and Step 2 are covered in detail in our earlier study [12], which outlines the test rig, experiments, and the identification of key predictors of degradation. The current study carries forward the results from Step 2 and converts the key predictors from Step 2 into a score for HI using the MD. In Step 4, using MATLAB, several ML regression models are evaluated, and finally, the three best performing models in terms of the accuracy and size of the model are selected. These are a tri-layer feedforward neural network, a Gaussian process regression, and a bagged ensemble of decision trees. Each model is trained on the majority of the dataset, with a small, randomly selected segment of 200 cycles withheld for independent validation. Finally, in Step 5, the selected models are integrated into a Simulink-based ensemble, where their predictions are combined through a weighted moving average. The moving window length adapts over time (2, 10, 50, and 100 cycles), enabling early predictions with limited data and improving the stability as more observations become available. This adaptive windowing ensures that actionable health estimates can be generated immediately after startup, without waiting for extended operating histories. Finally, the ensembled model is tested in Simulink using the withheld 200-cycle experimental dataset, representing data that the models have never encountered during training.

The remainder of the paper is structured as follows. The next section of the paper introduces the test rig, the experimentation, and the key predictors of degradation identified in the previous study. Section 3 provides a detailed description of the data labelling process for the health index using MD. Section 4 then presents the training of machine learning regression models. Section 5 elaborates on the formation of the ensemble and its implementation in

Simulink as well as the testing of the framework on randomly excluded data. Finally, Section 6 concludes the paper with remarks and outlines potential future research directions.

2. Test Rig, Experiments, and Key Predictors of Degradation

The long run-to-failure test was conducted on a purpose-built, seesaw-type test bench at Tampere University. The EMA under test is a ball-screw linear actuator driven by a servomotor and loaded by a hydraulic cylinder whose chamber pressures are regulated by electronically controlled pressure-relief valves (PRVs). In combination with anti-cavitation valves, the PRVs let the hydraulic load act as a passive brake in both motion directions: as the EMA moves, pressure builds in the opposing chamber until it exceeds the PRV cracking pressure, producing a controllable, bi-directional resistive force. The test rig with control and monitoring UI is shown in Figure 2 [12].

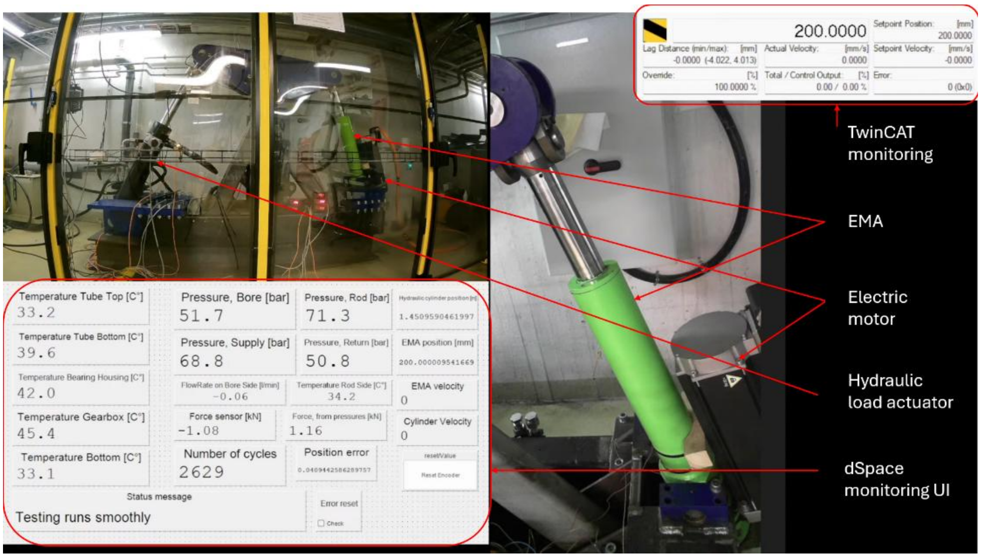

Figure 2. Test rig with monitoring UI [12]

A dedicated 20-bar PRV in the tank line maintains sufficient backpressure to overcome pressure drops across the anti-cavitation valves, and additional passive PRVs protect the circuit against fault-induced over-pressurisation which is important for safe unattended operation. The tested EMA has a 500mm stroke and is powered by an electric motor of 11.62 kW, and a nominal torque rating of 37 Nm at 3000 rpm. Whereas the hydraulic cylinder for loading is 150/80-550 mm and is controlled by an electronic PRV with a maximum operating pressure of 350 bar and maximum flow of 120 L/min.

The test rig has been fitted with numerous sensors on both sides, the hydraulic loading side and tested EMA side. On the tested EMA side, five temperature sensors are placed on different locations on the EMA, and the electric drive is used to measure the position and speed of the EMA as well as the torque and speed of the electric motor. On the hydraulic side, both the chamber pressures are measured, and a 250kN bi-directional load cell records the external force; the calculated cylinder force from these pressures is used to cross-validate the load-cell readings. The hydraulic cylinder position is monitored using a draw-wire position sensor and used as a safety interlock against drifting to stroke limits by comparing it with EMA position readings. Data collection and the control are split between a dSPACE MicroLab Box (hydraulic load control and centralised logging) and Beckhoff TwinCAT (electric motor control), with dSPACE recording all signals. The test rig was built to operate unsupervised and work 24/7 and hence fitted with numerous safety features against any accidents and failures.

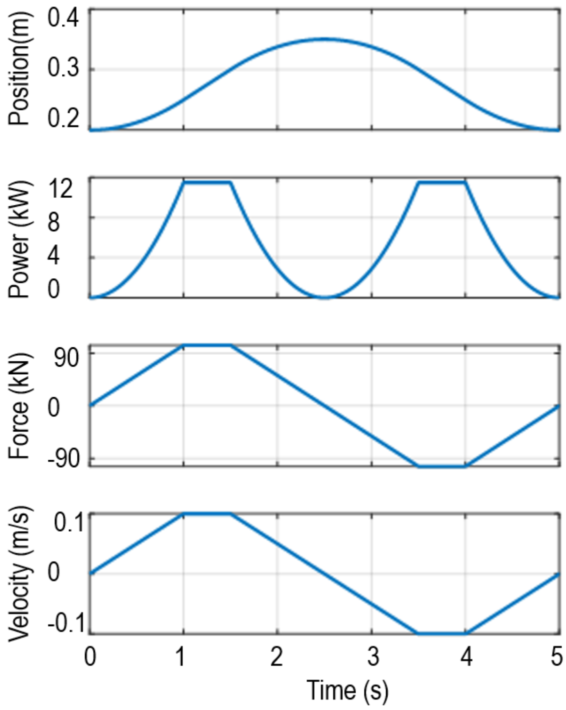

Figure 3. Utilized duty cycle for testing

One cycle for the tested duty cycles is shown in Figure 3. The duty cycle, designed with industrial partners, is an aggressive push-pull sequence: the actuator is driven at its maximum 100 mm/s within a mid-stroke window under a targeted load of 95kN. The bench is capable of regenerative energy via a proportional valve, but regeneration was not used in the current tests. As there is no active cooling, a 20-second pause follows each cycle to allow heat to conduct from internal interfaces (e.g., the screw nut) to the monitored surfaces; automated interlocks pause operation above safe temperature limits and resume once conditions normalise.

2.1. Key predictors of degradation

To quantify how degradation emerges in statistics rather than in any single raw channel, our previous study [12] compared early-life versus late-life operation using three matched pairs of datasets: (i) first 100 cycles versus last 100 cycles; (ii) second 100 cycles versus second-to-last 100 cycles; and (iii) third 100 cycles versus third-to-last 100 cycles. A sliding window was applied within each dataset to compute the summary features-mean, standard deviation

(STD), and skewness on the principal signals of load force, motor torque, and EMA velocity. Correlations among these windowed features were then evaluated, emphasising changes from the beginning to the end of life.

The exploratory data analysis (EDA) revealed consistent weakening of several cross-signal relationships as the actuator wore. Specifically, correlations that were strong and positive early on diminished markedly by the end of life. Complementing EDA, we performed local outlier factor (LOF) anomaly detection directly on the raw signal pairs (load - torque) and (load- velocity) which showed that late-life data occupies regions farther from the "healthy" cluster, with heavier distribution tails toward high anomaly scores [12].

From nine candidate correlation metrics (i.e., combinations of feature types across signals), four emerged as key predictors of degradation because they changed monotonically and consistently across all three dataset pairs:

1. **Correlation of the STD of load force with the STD of motor torque (C1)**
2. **Correlation of the STD of load force with the STD of actuator velocity (C2)**
3. **Correlation of the skewness of load force with the skewness of actuator velocity (C3)**
4. **Correlation of the mean of load force with the mean of actuator velocity (C4)**

In the present study, these four correlations are employed as key predictor signals to map the actuator's health index and to train the ML regression models.

A more detailed description of the test rig, experiments, and the procedure used to obtain the four signals lies outside the scope of the present study, as the primary objective here is to develop a health monitoring framework based on the results of our previous work. For comprehensive details regarding the test rig, experimental procedures, and signal acquisition process, the reader is kindly referred to our earlier study [12].

3. HEALTH INDEX AND DATA LABELLING

This section describes the construction of a scalar health index utilised to label data for supervised ML regression. The procedure operates on cycle-segmented signals and the four correlation-based predictors defined in Section 2, converting their windowed behaviour into (a) a multivariate deviation from a healthy baseline, (b) a smooth exceedance measure with persistence control, (c) a cumulative damage variable with outlier-robust scaling, and (d) a bounded HI suitable for training and evaluation.

The three signals of load force, actuator velocity, and electric motor torque are segmented into cycles using motion and reversal logic; cycles with insufficient samples are discarded to maintain statistical reliability (minimum sample count specified in Table 1). Low-activity cycles are filtered by an activity threshold defined as a fraction of a healthy reference quantile, ensuring that correlation estimates are not biased toward zero by quiescent operation. Within a rolling window of fixed length (see Table 1), it computes Spearman rank correlations for the predictor pairs defined in Section 2.

For distinct ranks, the coefficient ρ is

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \quad (1)$$

Here d_i is the rank difference and n is the number of observations in the window.

To quantify degradation, the Mahalanobis distance was employed to measure deviations from baseline correlations:

$$MD = \sqrt{(X - \mu)^T \textstyle\sum^{-1} (X - \mu)}, \quad (2)$$

Here X is the current correlation vector, μ is the baseline mean, and $\sum$ is the covariance matrix. MD was transformed into an exceedance score using a sigmoid function:

$$e = \frac{1}{1 + e^{-\frac{MD - z0}{k}}}, \quad (3)$$

Here $Z0$ sets the sensitivity threshold, and k controls the sigmoid steepness. Persistence gating was applied requiring exceedances in at least 4 of the last 7 cycles to reduce false alarms.

Damage D accumulation used an exponential relation:

$$D = \textstyle\sum e^{\beta(MD - z0)}, \quad (4)$$

Here β regulates the exponential rate, ensuring smooth, monotonic accumulation. Damage was capped at the 95th percentile to avoid extreme outliers, Finally Health Index is:

$$HI = 1 - \frac{D}{D_{95\%}} (1 - h_{min}), \quad (5)$$

Here, h_{min} denotes the minimum allowed HI.

Table 1 summarises the notation and hyperparameters used in this study. While the equation coefficients *(e.g., ρ and MD*) are computed dynamically from the data, the hyperparameters (e.g., *β, z0 and k)* were optimised through a grid search in MATLAB and cross-validation within literature-based ranges. The final selected values are reported in Table 1.

Table 1. Hyperparameters used to calculate HI

Parameter	Description	Value
MinSample	Minimum data samples per cycles to eliminate bad cycle data	500
$Z0$	Exceedance sensitivity threshold	3
k	Sigmoid Scaling	0.25
β	Damage exponential factor	1.2
h_{min}	Minimum HI	0.20

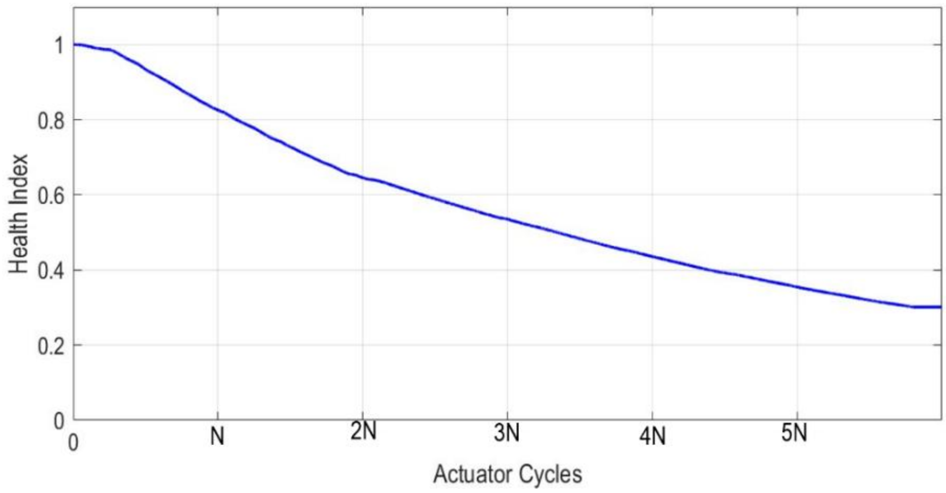

Figure 4. Health index mapping with actuator cycles

Applying the above pipeline to our full run-to-failure dataset gives a strictly non-increasing HI trajectory. The HI values remain near unity during early operation, decline gradually with time without collapsing to zero. This trajectory is presented in Figure 4, where the final HI is mapped with the number of cycles of the actuator. To respect the confidentiality constraints of the manufacturer of the tested EMA, the abscissa is de-identified: the numerical cycle count is replaced with a non-numeric value. The data of 200 cycles was withheld from the entire process to test the framework by deploying in Simulink at a later stage. The withheld data was from the dataset between the 4N and 5N points in Figure 4.

The monotonic decline of the HI demonstrates that the proposed methodology reliably captures progressive degradation of the actuator. Importantly, the absence of abrupt fluctuations indicates robustness against noise and transient effects in the sensor data. The gradual convergence towards a lower bound further confirms the stability introduced by the chosen hyperparameters and strategy. These characteristics collectively ensure that the HI provides an interpretable and reproducible representation of actuator health over its lifecycle.

It shall be noted that the proposed method, the selected hyperparameters, and the resulting HI trajectory are only for the present dataset and experimental conditions. We do not claim this representation to be a universal standard, as the actual health evolution of other actuators may differ. The results shown here are derived from a single run-to-failure experiment and therefore should be interpreted as an illustrative case study rather than a generalised benchmark. Neither this should be interpreted as the actual HI of the tested actuator, since the actuator failure was not directly linked to the gradual degradation, as explained earlier in Section 1.

4. TRAINING OF MACHINE LEARNING REGRESSION MODELS

The data and the HI prepared in Section 3 is used for the training of the ML regression models to estimate the HI directly from the four correlation-based predictors (Section 2).

The training data is divided into 90% for training in a 5-fold cross-validation and 10% for testing. Here, it is worth noting that this 10% testing data is different from the 200 cycles kept exclusively and randomly outside from the dataset for final testing in Simulink at later stages. Several model families were explored, and the three best performers were selected for complementary strengths are selected finally. These models are a Gaussian process regression, bagged decision tree ensemble, and feed-forward neural network.

Gaussian Process Regression (GPR): This provides flexible, non-parametric modelling with built-in uncertainty quantification which means it predicts behaviour while also estimating confidence in the prediction. The trained model has a zero-mean basis and an isotropic exponential kernel. Kernel scale and signal variance were determined by marginal-likelihood optimisation in the MATLAB regression learner app. The optimised sigma value is 1.7054e-4 which is obtained by the Bayesian optimiser of the MALTAB. The model achieved a test **RMSE** of **0.0254**, the lowest error among all candidates. The parity plot shown in Figure 5(A) shows tight adherence to the identity across the full health range with a minimal bias at the extremes, indicating an accurate and smoothly varying fit.

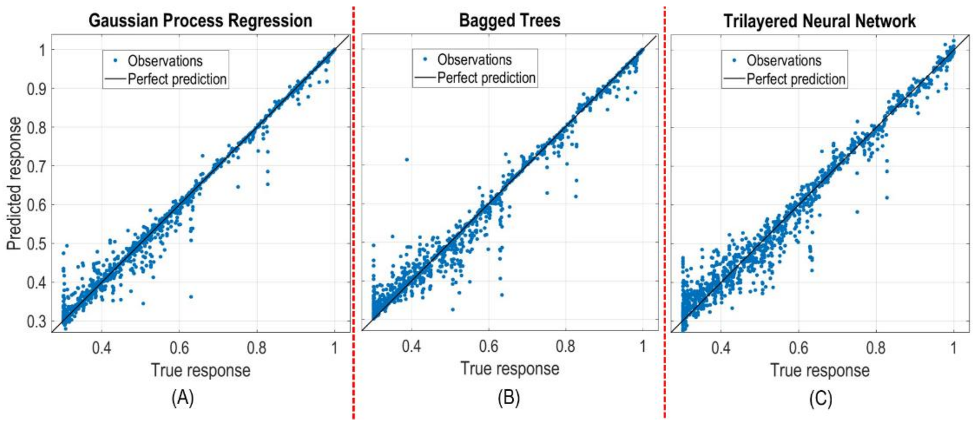

Figure 5. Regression models and predicted versus true response

Bagged Decision Tree Ensemble: An ensemble of 50 bootstrap-aggregated regression trees was fitted, with all predictors available at each split and a minimum leaf size of 1. No input standardisation was required. Bagging reduces the estimator variance by averaging high-variance trees, yielding a robust predictor that captures local nonlinearities without strong parametric assumptions. On the test set the ensemble delivered an **RMSE** of **0.0342**, the second best among the three used models. The parity plot shown in Figure 5(B) exhibits near-linear alignment with a modest dispersion in the mid-range.

Feed-Forward Neural Network (MLP): The last selected and trained model was a compact multilayer perceptron with three fully connected layers (sizes 60–300–50) and ReLU activations. Inputs for this were standardised, and a regularisation strength (Lemda) of 8e-5 was applied. Training used a maximum of 2000 iterations with cross-validation guiding early stopping and hyperparameter choice. The network provides a sufficient capacity to model smooth nonlinearities while remaining small enough to generalise from a modest feature set. On the test set, it achieved an **RMSE** of **0.0378**. The parity plot in Figure 5 (C) shows accurate tracking the near high and low ends of the HI with slightly conservative predictions in the lower-mid range. Owing to its compact size, the MLP offers very high inference throughput and a small model size, which are attractive for embedded deployment and closed-loop evaluation.

5. ENSEMBLE AND IMPLEMENTATION

The three regressors developed in Section 4 were deployed in Simulink as a real-time estimation chain that converts raw measurements into an operational HI. The architecture shown in Figure 6 implements three independently trained regression models as parallel branches in a Simulink environment. The same cycle segmentation, activity gating, and rolling-window settings used during labelling are embedded in this feature block to avoid a train–test mismatch. An upstream feature-extraction block processes the raw sensor signals of load force, actuator velocity and motor torque to compute the four correlation coefficients (C1–C4) described in Section 3. Once computed, the four features are fed concurrently into the three trained regression models explained in the previous section. Each branch is encapsulated as a standalone block with the exact hyperparameters learned during offline training; the blocks run synchronously on the same time steps, ensuring that all model outputs correspond to the same cycle. The modular design also simplifies diagnostic logging and makes it straightforward to replace or update individual regressors without affecting the rest of the system. Collectively, the three models provide redundant health estimates derived from complementary inductive biases, nonparametric kernel regression, variance-reduced ensemble partitioning, and parametric deep learning, thus meeting reliability requirements for critical machinery.

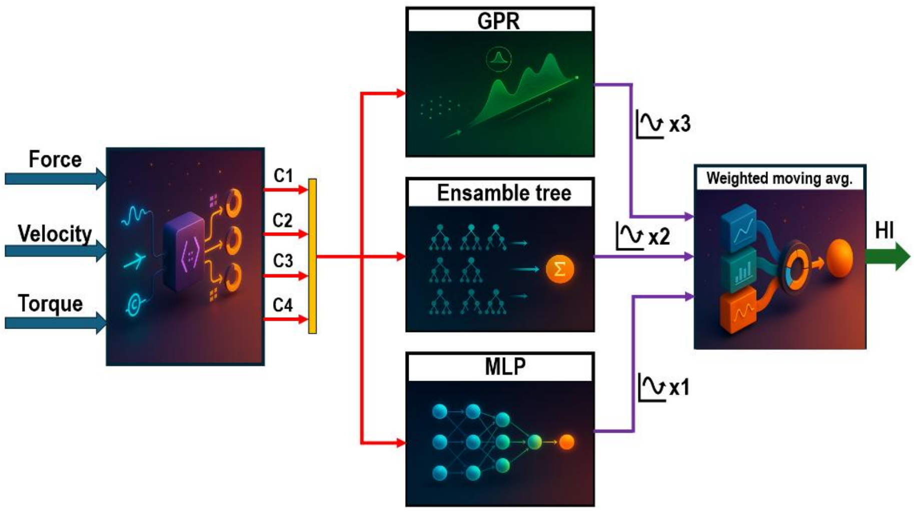

Figure 6. Simulink-implemented ensemble architecture

All three regressors are retained to exploit complementary inductive biases, i.e., kernel smoothing (GPR), partition-based nonlinearity (trees), and universal function approximation (MLP), so that the fused estimator is more robust to regime shifts and single-model failure modes than any constituent alone. Each model returns an instantaneous HI estimate at every cycle as shown in Figure 7. The GPR is the most accurate and stable across the range; the tree ensemble is competitive but slightly more variable in the mid-range; and the MLP reacts fastest but exhibits significant transients during the first few cycles, including a brief overshoot above the nominal upper bound of 1. As the operation proceeds, all three models converge toward consistent estimates in the mid-range. Although the MLP (neural network) exhibits transient overshoot at the beginning, its high sensitivity to rapid nonlinear changes increases ensemble diversity and early-change detectability, while the performance- and interpretability-weighted fusion prevents it from dominating, yielding a more resilient and stable estimate overall, as explained next in the section.

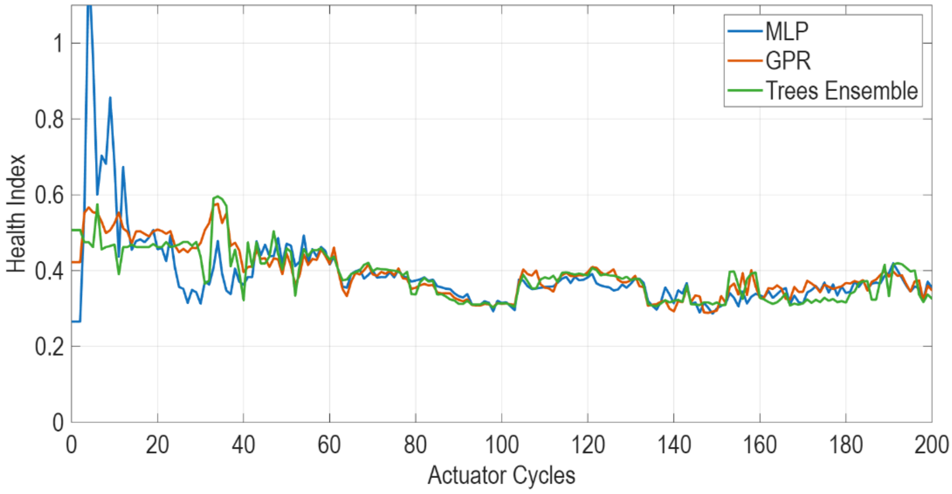

Figure 7. Individual output of instantaneous HI of regression models

The next step assigns fixed model weights to combine the three instantaneous estimates into a single operational HI on a rolling weighted average window. Weights were chosen using two principles. The first is predictive performance in the training stage in Section 4, where the GPR achieved the lowest RMSE, followed by the bagged trees, and finally the MLP. The second is model transparency at the global level, which is operationally valuable during deployment and auditing: GPR provides a principled predictive variance that can be surfaced as a confidence indicator; the tree ensembles support aggregated importance diagnostics and split-wise summaries; and the MLP, while fast, is less transparent as commonly called a black-box model [15]. Balancing these considerations, the fusion uses weights of 3:2:1 for GPR, trees, and MLP, respectively. This choice privileges the most accurate and interpretable models while retaining the responsiveness of the MLP as a minority voice that can nudge the fused estimate when rapid, genuine changes occur.

Since health evolves slowly while measurements and model outputs exhibit short-term variability, the fused weighted signal is smoothed by an adaptive moving-average window whose length increases as operational data accumulates. Immediately after start-up, only a handful of cycles are available; a 2-cycle window is therefore used to produce actionable estimates with minimal latency. This window maximises responsiveness but carries the largest variance, so it is used only long enough to cover the very early regime. Once sufficient data have accumulated, the window expands to 10 cycles, which reduces variance while still allowing the estimate to follow any real operating-regime transitions. After further operation, the window grows to 50 cycles, where the estimator begins to align with the slow time scale of mechanical wear rather than with short-term fluctuations. Finally, it settles at

100 cycles for steady operation, providing strong suppression of residual noise and model idiosyncrasies and enforcing temporal coherence consistent with gradual degradation. Transitions occur only when the next window can be fully populated with sufficient cycle count. Importantly, this adaptive smoothing acts after the instantaneous, weight-based fusion: the correlation windowing used for label construction in Section 3 is kept fixed and separate, ensuring methodological consistency while giving the online estimator the flexibility to manage its own bias–variance trade-off as evidence accumulates.

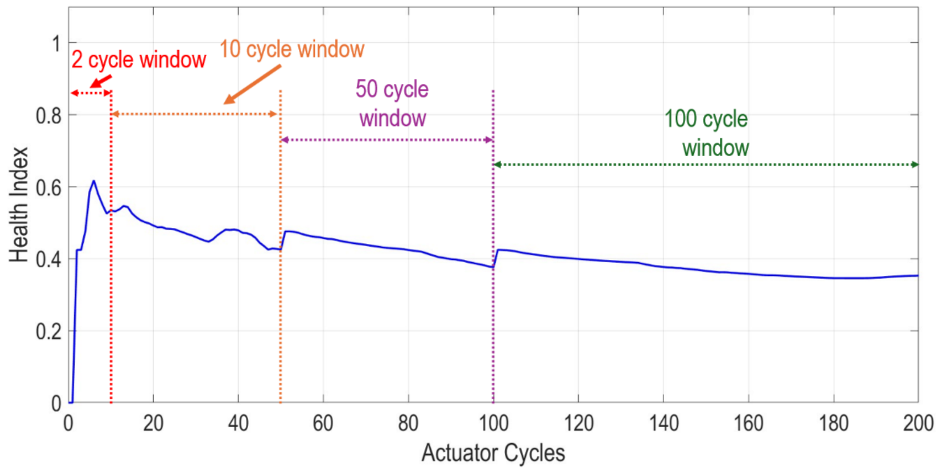

Figure 8. Final weighted average HI with varying window

Figure 8 presents the final HI after the fusion of three independent model estimations over the rolling window average. Vertical markers in the figure indicate the scheduled transitions of the adaptive window (i.e., 2→10, 10→50, and 50→100 cycles). Small, short-lived impulses appear at these markers due to a reweighting toward longer-horizon averages; they are not indicative of mechanical self-healing. Apart from these controlled adjustments, the fused curve is smooth, monotone in expectation, and bounded within the operational range. Here it should be noted that all results use a contiguous 200-cycle segment from the middle of the dataset that was randomly withheld from training and cross-validation; none of the models had prior exposure to these data. The response in Figure 8 remains within the interval expected from the held-out data shown in Figure 4. To protect the industry partner's confidentiality, the exact numerical range is intentionally omitted in Figure 4; however, the 200-cycle segment analysed here lies between the 4N and 5N markers in that figure, which corroborates the HI produced by the Simulink-deployed ensemble of Figure 8. At start-up, when the rolling window is short the fused HI is slightly elevated, driven by the initial peaks of the MLP visible in Figure 7. As the window length increases and more cycles are incorporated, the estimate stabilises and settles squarely within the 4N–5N interval, i.e., the region to which the test segment belongs.

Taken together, the Simulink ensemble provides a timely and stable HI estimate that generalises to previously unseen operating segments, reconciling early responsiveness with long-horizon coherence. These results validate the end-to-end pipeline—from feature extraction through weighted fusion with adaptive averaging, as a practical basis for real-time monitoring and maintenance decision support for high power EMAs.

6. CONCLUSION

This paper presented an end-to-end framework for the health monitoring of a high-power electro-mechanical actuator, from experiment design to real-time deployment. Building on correlation-based predictors established in our previous study, we defined a scalar health index that maps windowed feature relationships to a bounded health scale via an activity-gated, Mahalanobis-distance pipeline with persistence and damage accumulation.

Using the HI label and predictors, the ML regression models are trained, and the three best performing models are selected for their complimentary advantages. The selected regression models are a Gaussian process regression, a bagged-tree ensemble, and a tri-layered neural network. The three models were then imported into Simulink and fused by a weighted moving average that emphasises accuracy and transparency. An adaptive temporal window (i.e., 2→10→50→100 cycles) provided immediate estimates at start-up and increasing stability as more data accumulated. On a contiguous 200-cycle segment drawn randomly from the dataset and withheld from all training and validation, the fused output tracked the expected gradual degradation while suppressing early overshoot; brief impulses coincided with scheduled window transitions rather than physical recovery. Overall, the study provides a reproducible and deployable framework from raw measurements to robust health estimation.

7. ACKNOWLEDGEMENTS AND DECLARATION

This research is part of the Business Finland project EMMA-2 (2471/31/2021). The authors gratefully acknowledge this funding and support by Business Finland.

The authors gratefully acknowledge the contribution of Ville Raunio, Zakhar Afansev, Matias Alenius, and Toni Haapakoski from the Faculty of Engineering and Natural Sciences of Tampere University, for their contribution in building the test rig and carrying out experiments.

At the time of submission of this study, the first author, Dr. Vinay Partap Singh, was employed at Danfoss Power Solutions, Denmark; however, this study and all the experiments were solely carried out at the IHA lab of the Faculty of Engineering and Natural Sciences of Tampere University, Finland.

OpenAI’s ChatGPT was used to generate some graphics used in Figure 1 and Figure 6; this does not affect or reflect the methodology or results of the study. The authors take full responsibility for these graphics and for the content of the paper.

8. REFERENCES

[1] Huang, X., Yan, W., Cao, H., Chen, S., Tao, G., Zhang, J. (2024). Prospects for purely electric construction machinery: Mechanical components, control strategies and typical machines. Automation in Construction, 164, 105477.

[2] Balaban, E., Saxena, A., Narasimhan, S., Roychoudhury, I., Koopmans, M., Ott, C., Goebel, K. (2015). Prognostic health-management system development for electromechanical actuators. Journal of Aerospace Information Systems, 12(3), 329-344.

[3] Isturiz, A., Vinals, J., Abete, J. M., Iturrospe, A. (2012). Health monitoring strategy for electromechanical actuator systems and components, screw backlash and fatigue estimation. Recent Advances in Aerospace Actuation Systems and Components.

[4] Previdi, F., Maccarana, Y., Mazzoleni, M., Scandella, M., Pispola, G., Porzi, N. (2018, June). Development and experimental testing of a health monitoring system of electro-mechanical actuators for small airplanes. In 2018 26th mediterranean conference on control and automation (MED) (pp. 673-678). IEEE.

[5] Mazzoleni, M., Previdi, F., Scandella, M., Pispola, G. (2019). Experimental development of a health monitoring method for electro-mechanical actuators of flight control primary surfaces in more electric aircrafts. IEEE Access, 7, 153618-153634.

[6] Mazzoleni, M., Scandella, M., Maccarana, Y., Previdi, F., Pispola, G.,Porzi, N. (2018, August). Condition monitoring of electro-mechanical actuators for aerospace using batch change detection algorithms. In 2018 IEEE conference on control technology and applications (CCTA) (pp. 1747-1752). IEEE.

[7] Gómez de Mariscal, B., Villar Domínguez, A., Gallego, A., Fernández, J. A., Novillo, E., López de Calle, K., Fernández, A., Egaña, I. (2025, May 21–23). Health monitoring system for electromechanical actuation system: Pre-flight status assessment. In Recent Advances in Aerospace Actuation Systems and Components (pp. 119–126). Toulouse, France.

[8] Bajarunas, K., Baptista, M. L., Goebel, K., Chao, M. A. (2024). Health index estimation through integration of general knowledge with unsupervised learning. Reliability Engineering & System Safety, 251, 110352.

[9] Kim, S., Kim, N. H., Choi, J. H. (2020). Prediction of remaining useful life by data augmentation technique based on dynamic time warping. Mechanical Systems and Signal Processing, 136, 106486.

[10] Shi, C., Zhao, S., Chen, X., Wang, S., Liu, D. (2025). Condition monitoring of train transmission systems based on multimodal fusion improved transformer network. Sound & Vibration, 59(2), 2904-2904.

[11] Atamuradov, V., Medjaher, K., Camci, F., Zerhouni, N., Dersin, P., Lamoureux, B. (2020). Machine health indicator construction framework for failure diagnostics and prognostics. Journal of signal processing systems, 92(6), 591-609

[12] Raunio, V., Alenius, M., Minav, T. (2025, May). Identifying key predictors of failure in electro-mechanical actuator. In *International Conference in Aerospace Actuation Systems and Components* (pp. 127-134). INSA de Toulouse.

[13] Niu, G., Singh, S., Holland, S. W., Pecht, M. (2011). Health monitoring of electronic products based on Mahalanobis distance and Weibull decision metrics. Microelectronics Reliability, 51(2), 279-284.

[14] Jin, X., Wang, Y., Chow, T. W., Sun, Y. (2017). MD-based approaches for system health monitoring: a review. IET Science, Measurement & Technology, 11(4), 371-379.

[15] Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115

Biographies

Vinay Partap Singh received his master's degree in mining machinery engineering from IIT(ISM) Dhanbad, India in 2019, and Doctor of Science (Technology) from Tampere University, Finland in 2025. He is currently working as a Design Engineer in Danfoss Power Solutions in Denmark.

Tatiana Minav is an Associate Professor at Innovative Hydraulics and Automation lab of Tampere University. She has over 18 years of expertise on system electrification of non-road machinery and electro-hydraulic actuators solutions.