# Engine Health State Index (EHSI)

Rohit Deo[1], Aman Yadav[2], Shruti Bharti[3], Nilesh Powar[4]

[1,2,3,4]*Cummins Technologies India Pvt. Ltd., Pune, MH, 411045, India*
*rohit.deo@cummins.com*
*aman.yadav@cummins.com*
*shruti.bharti@cummins.com*
*nilesh.powar@cummins.com*

## ABSTRACT

Assessing the health of diesel engines is challenging due to multiple coexisting failure modes, overlapping fault signatures, and highly imbalanced operational data. This paper proposes an Engine Health State Index (EHSI), a probabilistic health metric that aggregates risk estimates from multiple failure-code–specific models.

The framework employs a collection of binary classifiers, each trained to estimate the likelihood of a specific failure code from historical telemetry and diagnostic data. At each time step, the resulting failure risk vector provides a distributed representation of latent fault exposure rather than a single dominant failure mode. EHSI maps this risk distribution to a scalar health index using normalized uncertainty measures, enabling continuous tracking of health degradation without relying on explicit fault triggers.

Experiments on real-world diesel engine datasets show that EHSI produces smooth and interpretable health trajectories that correlate with impending failures while remaining sensitive to early-stage degradation. The proposed approach is model-agnostic, extensible to additional failure modes, and suitable for large-scale fleet monitoring applications.

**Keywords** – Remaining Useful Life, Prognostics, Predictive Maintenance, Engine Health Monitoring, Downtime reduction

## 1. INTRODUCTION

Cummins Inc. designs and manufactures diesel and alternative fuel engines, filtration systems, and power generation products. The company operates globally and serves customers across multiple industries and geographic regions. Over the past several years, Cummins has increasingly integrated data-driven methods into its engine platforms to support diagnostics and prognostics applications.

Prognostics on Cummins diesel engines has been practiced for more than seven years. Sensor-specific prognostics provides customers with an estimate of the remaining useful life (RUL) of components, helping them avoid unnecessary repairs and reduce downtime costs. The onboard modules record data from many sensors and construct datasets by aggregating sensor values for each trip. This irregular time-series data collection method is optimized and has proven effective in Cummins's prognostic applications. For in-warranty engines, replacement claims offer a reliable way to mark failures or anomalies in the dataset. Our previous work has demonstrated that this data supports accurate RUL estimation for several components.

The traditional component-focused prognostics is useful for reducing unplanned downtime for individual parts, but it has limitations when applied to a diesel engine. The engine consists of multiple interconnected subsystems, and failures often manifest as symptoms across several components. As a result, relying only on component-level outputs gives a fragmented view and does not provide a clear understanding of the overall engine health. It also becomes difficult to identify which issues are driving major degradation. These limitations create the need for system-level prognostics, where information from multiple components, subsystems, and failcodes is combined to form a single health measure.

Modeling the overall health of heavy-duty engines is constrained by the limited availability of publicly accessible fleet-scale datasets that contain verified failure annotations. Most prior studies rely on laboratory test benches or simulated degradation scenarios, which are useful for controlled experimentation but do not fully represent the variation, operating patterns, and usage diversity observed in real fleets. This gap makes it difficult to develop models that generalize well to actual field conditions. In this work, failure events are identified using warranty and service claim records, which provide operational markers of observed failures in the field. This grounding in real-world service data connects the degradation modeling to actual engine usage.

---

Many existing prognostic approaches focus on estimating the remaining useful life of individual components based on fault-specific degradation trajectories. While effective at the component level, such methods do not directly provide a unified view of how multiple degradation processes evolve simultaneously at the system level. In contrast, the proposed Engine Health Status Index (EHSI) provides a continuous system-level indicator that represents the changing likelihood of multiple component failures over time. This continuous measure allows the degradation trend to be observed earlier, giving technicians and fleet operators more context about emerging issues and enabling better planning for inspection or repair.

In this paper, we extend the traditional component-focused prognostics approach to one that captures the overall behavior of the engine. The Engine Health Status Index (EHSI) is developed to address this need by generating a single health indicator derived from patterns in sensor data. Instead of relying only on predictions built around individual components, we use failcodes as markers of failure and train separate classifiers for each failcode to estimate their respective probabilities. These probabilities are combined through defined mathematical operations to create a degradation profile. This profile forms the basis for the EHSI, which is interpretable and represents the collective health condition of the engine. The EHSI also identifies the most influential contributors to degradation, enabling technicians to focus on the issues that require attention on priority.

The paper is organized as follows. The rest of the paper is organized as follows. Section II presents the related work in prognostics and health indicator construction. Section III describes the proposed EHSI framework. Section IV explains the experimental setup and dataset used for evaluation. Section V presents the results and discussion. Section VI concludes the paper and highlights future research directions.

## 2. LITERATURE SURVEY

Many prognostic approaches in Prognostics and Health Management (PHM) focus on estimating the Remaining Useful Life (RUL) of individual components by analyzing fault-specific degradation trajectories. These approaches are effective when a component's degradation can be monitored independently. However, modern engineering systems often consist of multiple interacting components whose degradation processes evolve simultaneously. In such situations, component-level prognostics may not provide a complete understanding of the system's overall health condition. As a result, recent research has emphasized the development of system-level health representations capable of capturing the combined effects of multiple degradation processes over time.

A number of review studies have examined the evolution of PHM techniques and their application in complex systems. Cao, H., Yu, J., & Duan, F. (2025) review condition-based maintenance strategies in complex degradation systems and highlight the transition from single-component prognostics to multi-component system modeling. Their work discusses several analytical and probabilistic approaches used in maintenance decision making, including Physics-Informed Neural Networks (PINNs), Monte Carlo Simulation (MCS), Multi-State Fault Tree Analysis (MSFTA), Universal Generating Function (UGF), and Goal-Oriented (GO) methods. These methods help balance preventive and corrective maintenance actions while considering system interactions. However, the authors emphasize that accurately modeling the degradation behavior of interconnected components remains a major research challenge.

Similarly, Zhang, Y., Fang, L., Qi, Z., & Deng, H. (2023) provide a comprehensive review of RUL prediction approaches for mechanical equipment. They categorize PHM methods into traditional statistical approaches, machine learning models, and hybrid frameworks, and identify RUL prediction as a core technology in modern PHM systems. Their study highlights the advantages and limitations of different techniques in handling nonlinear degradation processes and uncertain operating conditions. In addition, Qiu, S., Cui, X., Ping, Z., Shan, N., Li, Z., Bao, X., & Xu, X. (2023) review deep learning techniques used for fault diagnosis and prognostics in industrial systems. Their work discusses architectures such as Autoencoders, Generative Adversarial Networks (GANs), Deep Belief Networks (DBNs), Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Transformer models, which have shown promising results in addressing issues such as data imbalance, high-dimensional sensor data, and multimodal data fusion.

Maintenance decision-making frameworks have also been integrated with PHM models to support optimal maintenance planning. For example, Yuan and Liu (2024) proposed a multi-attribute maintenance decision model for aircraft engines based on the entropy-weighted TOPSIS method. In their approach, different maintenance attributes are weighted to construct a composite evaluation index that assists in determining maintenance time and cost. By assuming Weibull distributions for system variables, the method integrates reliability and availability constraints into the maintenance decision process.

Another important research direction in PHM is the construction of Health Indicators (HIs), which are quantitative variables used to represent the degradation state of a system. Nguyen and Medjaher (2021) proposed a two-stage genetic programming-based framework for automated HI construction. In the first stage, relevant sensor signals and time-domain features are automatically selected, while in the second stage these features are combined into a mathematical HI expression. The optimization process considers multiple criteria including monotonicity, trendability, mutual information, and Spearman correlation to ensure that the constructed indicator accurately reflects the degradation process. Although the approach produces interpretable HIs suitable for prognostics, it assumes the availability of run-to-failure data and models degradation as a single trajectory, which limits its ability to represent multiple failure mechanisms simultaneously.

Unsupervised learning approaches have also been explored for HI estimation. Bajarunas, K., Baptista, M. L., Goebel, K., & Arias Chao, M. (2024) proposed a method in which an autoencoder is trained using a loss function derived from general degradation knowledge. The resulting health index is evaluated using metrics such as monotonicity, trendability, prognosticability, and correlation with degradation progression. Similarly, Qin, Y., Yang, J., Zhou, J., Pu, H., & Mao, Y. (2023) introduced a Supervised Multi-Head Self-Attention Autoencoder (SMSAE) for health indicator construction in rotating machinery. The model combines attention mechanisms with autoencoder structures to capture complex degradation patterns and improve similarity-based RUL prediction. Extending this concept, Yong and Brintrup (2022) proposed a Coalitional Bayesian Autoencoder framework, where separate autoencoders are trained for individual sensors and combined using a coalition-based learning strategy. This approach provides improved interpretability and robustness in condition monitoring applications.

A related line of research focuses on extracting meaningful features from high-dimensional sensor data. De Giorgi, M. G., Menga, N., & Ficarella, A. (2023) reviewed prognostic and diagnostic techniques used for jet engine health monitoring. Their study reports that dimensionality reduction techniques such as Local Linear Embedding (LLE), Principal Component Analysis (PCA), and Singular Value Decomposition (SVD) are effective in extracting compact representations of degradation patterns from large sensor datasets. In addition, model-based methods including gas path analysis, Kalman filtering, and genetic algorithms are shown to be effective for fault diagnosis and state estimation. Data-driven approaches such as Artificial Neural Networks (ANNs) and Bayesian Networks (BNs) are also highlighted for their ability to capture nonlinear relationships and uncertainty in engine degradation behavior. An interesting observation from the study is that fault detection accuracy is generally higher during the cruise phase of aircraft operation, where system conditions are relatively stable compared to the transient phases of take-off and climb.

For marine applications, Zhang, P., Cao, L., Dong, F., Gao, Z., Zou, Y., Wang, K., Zhang, Y., & Sun, P. (2022) proposed a hybrid prediction framework based on a Synthesized Health Indicator (SHI) constructed using Stacked Autoencoders (SAE) and Dynamic Kernel PCA (DKPCA). In this framework, multidimensional sensor data are transformed into a one-dimensional health variable normalized within the range (0,1), representing the system's degradation state. The SHI is then decomposed into trend and fluctuation components, and a hybrid prediction strategy is applied where statistical models capture the linear trend while machine learning models estimate the nonlinear residual component. Similar hybrid approaches have also been developed for cyber-physical systems, where Shcherbakov and Sai (2022) proposed a deep learning framework for predictive maintenance combining multiple neural network architectures.

Several classical methods have been proposed for RUL estimation of rotating machinery components such as bearings. Medjaher, K., Tobon-Mejia, D. A., & Zerhouni, N. (2012) developed a data-driven approach that identifies critical components and models their degradation using Mixture of Gaussian Hidden Markov Models (MoG-HMM) represented through Dynamic Bayesian Networks. The hidden states represent different degradation stages, enabling probabilistic RUL estimation. Mosallam, A., Medjaher, K., & Zerhouni, N. (2016) proposed a Bayesian-based data-driven prognostic method where health indicators are first extracted from sensor data through feature selection and transformation. The degradation trajectory of the monitored system is then compared with historical trajectories using k-nearest neighbors, and a discrete Bayesian filter is used to estimate the current degradation state and remaining useful life. Earlier work by Wang, T., Yu, J., Siegel, D., & Lee, J. (2008) introduced a similarity-based prognostics approach that estimates RUL by matching the degradation trajectory of a test unit with similar trajectories from a library of run-to-failure data.

With the advancement of machine learning techniques, deep learning methods have become increasingly popular for RUL prediction and HI construction. Liu, B., Gao, Z., Lu, B., Dong, H., & An, Z. (2022) proposed a deep learning-based RUL prediction model for bearings that converts vibration signals into time–frequency representations using Short-Time Fourier Transform (STFT). The resulting spectrograms are processed using a CNN-LSTM architecture combined with an attention module, enabling the model to capture both spatial and temporal degradation features. Experimental results on the PHM 2012 dataset demonstrate improved prediction accuracy compared to conventional methods.

Similarly, Chen, L., Xu, G., Zhang, S., Yan, W., & Wu, Q. (2020) developed an end-to-end Convolutional Recurrent

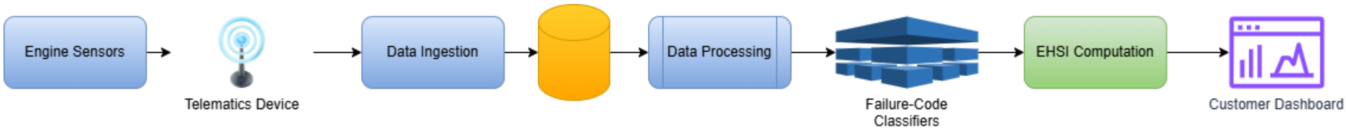

**Figure 1. System data pipeline for engine health monitoring and EHSI computation.**

Neural Network (CRNN) for constructing health indicators directly from vibration signals. In this model, CNN layers automatically extract local features while LSTM layers capture temporal dependencies in degradation signals. The approach eliminates the need for manually engineered features and demonstrates improved performance on the PRONOSTIA bearing dataset. Another related study by Guo, L., Lei, Y., Li, N., Yan, T., & Li, N. (2018) proposed a CNN-based health indicator construction method that automatically learns features from raw vibration signals. To address abnormal fluctuations in the constructed health indicators, the authors introduced an outlier correction technique based on the 3σ rule, which improves the monotonicity and reliability of the degradation trend.

Autoencoder-based approaches have also been explored for HI construction. Xu, F., Huang, Z., Yang, F., Wang, D., & Tsui, K. L. (2020) developed a Stacked Autoencoder (SAE) model that extracts degradation features from frequency-domain vibration signals of roller bearings. An exponential smoothing function is applied to reduce oscillations and improve the monotonic behavior of the health indicator. Experimental results show that the proposed approach outperforms several traditional indicators and clustering-based methods. In a related work, Kaji, M., Parvizian, J., & van de Venn, H. W. (2020) proposed a deep learning method that converts vibration signals into time–frequency images using Continuous Wavelet Transform (CWT) and processes them with a Convolutional Autoencoder (CAE). The Mahalanobis distance between healthy and degraded states is then used as a health indicator to track degradation, resulting in improved reliability compared with conventional feature-based methods. Overall, the existing literature demonstrates significant progress in data-driven prognostics, health indicator construction, and deep learning-based RUL prediction. Nevertheless, most current methods still focus on individual component degradation rather than capturing the interactions between multiple degrading components at the system level. Developing approaches that can effectively represent system-level health states and multiple concurrent degradation mechanisms remain an important research direction in PHM.

## 3. PROPOSED METHODOLOGY

This section describes the construction of the EHSI. The framework integrates failure-code risk estimation, transition-aware risk adjustment, and degradation regime identification to derive a system-level health indicator. Each stage of the framework is described sequentially, with the associated mathematical formulation introduced where required.

Figure 1 illustrates the overall data pipeline used in the study. Sensor measurements from the electronic control modules are aggregated at the trip level to form the engine state vectors used for modeling. These trip-wise feature vectors serve as inputs to the failure-code risk estimation models, which generate probabilities for the occurrence of different failure codes within the prediction horizon. The resulting risk estimates are further processed through the subsequent stages of the framework to produce the Engine Health Status Index, representing the overall health condition of the engine.

### 3.1. Failure-code risk estimation

Cummins diesel engines are equipped with an electronic control module that continuously gathers data from multiple onboard sensors and aggregates the measurements for each trip along with the associated fail codes. This results in an irregular time-series dataset where the aggregated trip records form the feature space and the instances where fail codes occur are marked as failures in the annotation labels. Since the remaining useful life (RUL) is divided into predefined ranges, range-wise class labels are assigned so that a classifier can predict the RUL class, which provides an estimate of degradation severity.

The structure of the proposed Engine Health Status Index framework is illustrated in Figure 2. The framework consists of four stages. In the first stage, classifiers estimate the probability of occurrence for each failure code using the engine state vector. In the second stage, these probabilities are adjusted using historical transition relationships between failure codes to account for common failure progressions. The third stage identifies whether the engine behavior

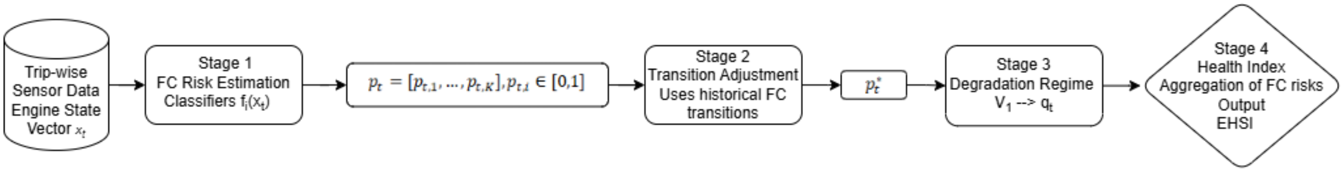

**Figure 2. Framework for Engine Health Status Index (EHSI) construction**

corresponds to a degradation regime and estimates the likely failure progression pattern. Finally, the adjusted failure risks and regime information are combined to compute the Engine Health Status Index, which summarizes the overall health condition of the engine.

The proposed framework estimates engine health by combining failure-code risk prediction, transition behavior between failures, and degradation regime identification. The input to the framework is the engine state vector $x_t \in \mathbb{R}^d$ , representing the trip-wise aggregated sensor measurements at time index t. For each failure code $FC_i$, a classifier estimates the probability associated with the predicted RUL class.

The collection of probabilities for all failure codes at time $t$ is represented as –

$$p_t = [p_{t,1}, \dots, p_{t,K}], p_{t,i} \in [0,1]$$

where $p_{t,i}$denotes the predicted risk associated with failure code $FC_i$ at time $t$. These probabilities are used in the subsequent stages of the framework for transition-aware correction.

### 3.2. Transition-aware risk adjustment

For each fail-code $FC_i$ , a supervised classifier $f(.)$ is trained to estimate the probability of obtaining that $FC_i$ for the input state vector $x_t$. The predicted probability for the $FC_i$ at time $t$ is given by –

$$p_{t,i} = f(x_t)$$

Here, $p_{t,i}$ denotes the probability that the failure code $FC_i$ will occur within the prediction horizon. The outputs of all failure-code classifiers are combined to form the failure-code probability vector $p_t$ . This vector represents the estimated risk of multiple potential failures at time $t$ Since several degradation processes may evolve simultaneously within the engine, these probabilities are evaluated jointly in the subsequent stages of the framework.

### 3.3. Transition-aware risk assessment

Failure events in diesel engines often do not occur independently. Certain failure codes tend to appear in sequence due to underlying degradation processes across related subsystems. To account for these dependencies, the predicted failure-code risks are adjusted using historical transition relationships observed in the data. Let $T$ denote a transition matrix that captures the likelihood of one failure code being followed by another based on historical observations. Using this transition structure, the initial probability vector $p_t \in \mathbb{R}^K$ is propagated to obtain a transition-informed risk estimate –

$$\tilde{p}_t = T(.)$$

Where, $T(.)$ represents a Markovian transition matrix. It's derived from historical failure sequences as follows –

$$T_{i,j} = P\left(FC_j \middle| FC_j\right)$$

Here,

$$T \in \mathbb{R}^{K\,X\,K}$$

However, the influence of transition propagation must be controlled to avoid over-adjustment when the classifier predictions are already confident. Further,

$$\mathrm{p}_{\mathrm{t+\Delta}} = \mathrm{p}_{\mathrm{t}} \,.\, T$$

Therefore, a confidence-based weighting mechanism is used to combine the original probability estimates with the transition-propagated risks.

$$p_t^* = \alpha_t \mathrm{p}_t + (1 - \alpha_t)\mathrm{p}_{\mathrm{t+\Delta}}$$

Where, $\alpha_t$ is a confidence weight that determines the relative influence of the original predictions and the transition-propagated risks. The confidence weight $\alpha_t$ is derived from an uncertainty measure computed from the entropy of the failure-code probability vector, allowing the influence of transition propagation to decrease when the classifier predictions are confident and increase when prediction uncertainty is high. The corrected probability vector $p_t^*$ represents the adjusted failure-code risk used in the subsequent stages of the framework.

### 3.4. Degradation regime identification

In addition to estimating individual failure-code risks, it is necessary to determine whether the engine behavior corresponds to normal operation or an evolving multi-component degradation pattern. For this purpose, a degradation regime classifier is introduced.

Let,

$$V_1 : \mathbb{R}^d \rightarrow [0,1]$$

denote a classifier that estimates the probability that the engine is operating in a degradation regime characterized by multiple interacting failure processes. For a given engine state $x_t$, the regime probability is defined as –

$$q_t = V_1(x_t) = P(\text{multi-FC degradation} \mid x_t)$$

Where, $q_t$ represents the likelihood that the engine has entered a degradation regime rather than normal operating behaviour. When the degradation regime probability indicates elevated risk, an additional model is used to characterize the expected progression of failures. Let,

$$V_2 : \mathbb{R}^d \rightarrow \mathcal{C}$$

denote a classifier that predicts the most likely failure-code progression pattern based on historical engine trajectories. The output –

$$\hat{c} = V_2(x_t)$$

represents the predicted failure-code sequence associated with the current engine state.

### 3.5. EHSI Construction

Let $p_t^* = [p_{t,1}^*, p_{t,2}^*, \dots, p_{t,K}^*]$ denote the transition-adjusted failure-code probability vector obtained from the previous stage. A degradation score is computed as a weighted aggregation of the failure-code risks:

$$D_t = \sum_{i=1}^{K} w_i p_{t,i}^*,$$

Where, $w_i$ represents the relative structural contribution of failure code $FC_i$ to the overall degradation state. These weights are derived from historical failure behavior observed in the engine population. The degradation score is then combined with the degradation regime probability $q_t$ to produce the Engine Health Status Index:

$$EHSI = 100(1 - h(q_t, D_t))$$

where $h(\cdot)$is a bounded monotonic mapping that converts the estimated degradation state into a normalized health index. Higher values of $EHSI$ correspond to healthier operating conditions, while lower values indicate increasing degradation risk. This formulation allows the health index to reflect both the aggregated failure-code risks and the broader degradation regime identified in the previous stage.

## 4. RESULTS AND DISCUSSIONS

This section presents the evaluation of the proposed framework using field data collected from operational engines. The analysis is conducted using fleet-scale telematics data collected from approximately 100,000 engines over a period of four years, where sensor measurements are aggregated at the trip level. Since publicly available datasets with detailed failure-code annotations for heavy-duty engines are limited, the analysis is conducted using internal fleet data collected from connected engine platforms. Failure codes are anonymized in the results to preserve confidentiality.

**Table 1. Failure-code wise performance of the classifiers**

| Fail code | Precision | Recall | F1 Score |
|---|---|---|---|
| FC1 | 0.56 | 0.55 | 0.56 |
| FC2 | 0.56 | 0.42 | 0.48 |
| FC3 | 0.69 | 0.32 | 0.44 |
| FC4 | 0.62 | 0.62 | 0.62 |
| FC5 | 0.65 | 0.49 | 0.56 |
| FC6 | 0.56 | 0.03 | 0.07 |
| FC7 | 0.63 | 0.53 | 0.58 |
| FC8 | 0.69 | 0.81 | 0.75 |
| FC9 | 0.61 | 0.72 | 0.66 |
| FC10 | 0.78 | 0.31 | 0.44 |
| FC11 | 0.63 | 0.53 | 0.58 |
| FC12 | 0.58 | 0.1 | 0.16 |
| FC13 | 0.68 | 0.6 | 0.64 |
| FC14 | 0.75 | 0.3 | 0.42 |
| FC15 | 0.69 | 0.1 | 0.18 |
| FC16 | 0.89 | 0.16 | 0.27 |
| FC17 | 0.57 | 0.18 | 0.27 |
| FC18 | 0.33 | 0.01 | 0.01 |
| FC19 | 0.39 | 0.16 | 0.23 |
| **Average** | **0.62** | **0.36** | **0.41** |

The transition-aware correction is regulated using an entropy-based uncertainty measure computed from the failure-code probability vector. Higher entropy indicates competing degradation hypotheses and lower confidence in the predicted failure progression, while lower entropy corresponds to more concentrated failure probabilities and greater confidence in the estimated degradation behavior.

**Table 2. Failure-code wise performance of the classifiers**

| **Model** | **Micro** | | **Macro** | |
|---|---|---|---|---|
| | Precision | Recall | Precision | Recall |
| V1 | 0.67 | 0.64 | 0.75 | 0.67 |
| V2 | 0.38 | 0.38 | 0.36 | 0.35 |

The evaluation focuses on three aspects: the performance of the failure-code risk models, the effectiveness of the degradation regime identification, and the behavior of the resulting Engine Health Status Index under real operational conditions.

### 4.1. Failure-Code Risk Prediction Performance

The first stage of the framework estimates the probability of individual failure-code occurrences within the defined prediction horizon. The distribution of failure-code occurrences is highly imbalanced across the fleet, with several failure modes occurring relatively infrequently compared to normal operating conditions.

Separate classifiers are trained for each failure code using the trip-wise engine state vectors. Table 1 reports the precision, recall, and F1-score achieved by these models. Failure codes are anonymized as FC-1, FC-2, …, FC-K to preserve confidentiality.

The results indicate that the classifiers can estimate the risk of multiple failure codes simultaneously, forming the basis for the subsequent stages of the framework. Precision values across most failure codes range between approximately 0.55 and 0.75, indicating that the models are generally able to identify relevant failure patterns in the sensor data.

However, recall varies significantly across failure codes, reflecting the imbalance and differing occurrence frequencies of failure events in the fleet data. Certain failure codes such as FC-8 and FC-9 exhibit relatively strong recall values, while others such as FC-6 and FC-18 show very low recall. The variation in recall across failure codes is expected in fleet-scale datasets, where some failures occur infrequently or show limited early indications in the available sensor measurements. In addition, the amount of training data available for each failure code varies significantly across the fleet, which further contributes to the observed differences in predictive performance across the classifiers.

Since the occurrence frequency of failure codes varies significantly across the fleet, both precision and recall are reported to provide a balanced view of classifier behavior. Recall is particularly important for identifying degradation-related failures prior to recorded maintenance events, while precision helps reduce false degradation indications during normal operation.

### 4.2. Degradation Regime Identification

The second stage of the framework identifies whether the engine behavior corresponds to a degradation regime involving multiple interacting failure processes. Two classifiers are used in this stage. The first model (V1) estimates the probability that the engine has entered a degradation regime. The second model (V2) predicts the likely failure progression pattern based on historical trajectories.

Table 2 summarizes the micro and macro precision and recall obtained for these models. The V1 classifier achieves micro precision and recall values of 0.67 and 0.64 respectively, indicating moderate capability in distinguishing degradation behavior from normal operation. The V2 classifier exhibits lower performance, with micro precision and recall values of approximately 0.38. This difference reflects the increased complexity of predicting structured failure sequences compared to detecting the presence of degradation. In addition, multiple degradation paths may lead to similar observed failure outcomes, increasing the ambiguity associated with sequence-level prediction.

In addition to the classifier outputs, the framework incorporates an entropy-based uncertainty measure computed from the failure-code probability vector. Higher entropy indicates that the predicted risks are distributed across multiple competing failure modes, reflecting lower confidence in the degradation estimate. Lower entropy corresponds to more concentrated failure probabilities and greater confidence in the predicted degradation behavior. This uncertainty information is used during the transition-aware correction stage to regulate the influence of historical failure progression patterns on the final risk estimates.

### 4.3. Engine Health Status Index Behavior

The final stage of the framework produces the Engine Health Status Index (EHSI), which summarizes the combined effect of multiple failure-code risks into a single health indicator.

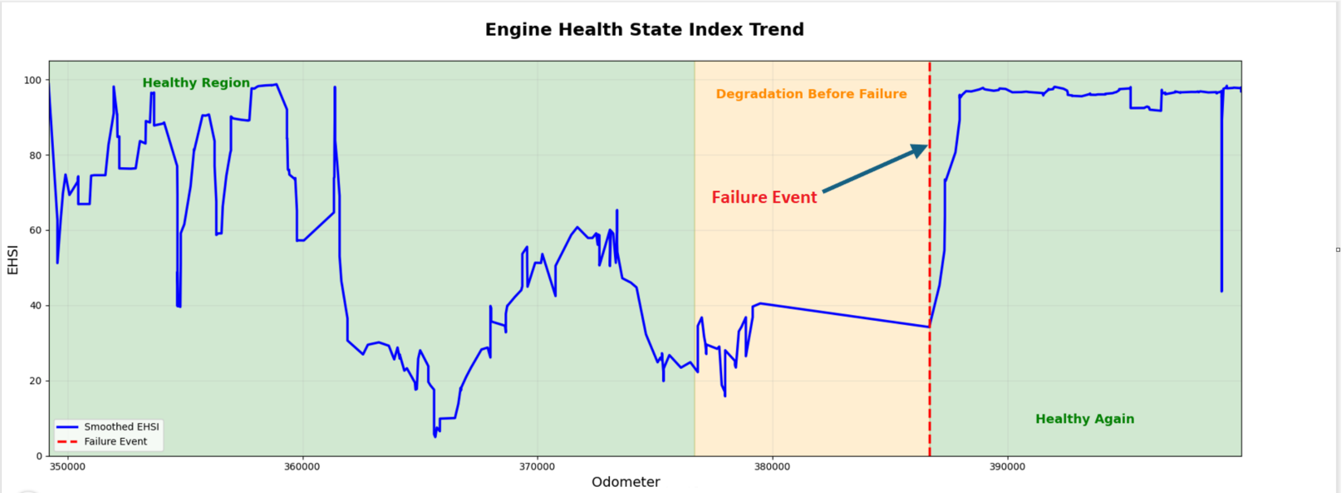

**Figure 3. Example trajectory of the Engine Health Status Index (EHSI) for a representative engine. The figure illustrates healthy operation, degradation prior to a recorded failure event, and recovery after maintenance**

To understand the behavior of the index under real operating conditions, the evolution of EHSI is analyzed over engine mileage for selected engine trajectories. The objective of this analysis is to examine whether the index reflects the gradual degradation of the engine prior to the occurrence of recorded failure events. Rather than representing the condition of a single component, the EHSI reflects the combined degradation tendency of multiple interacting subsystems observed through the failure-code risk estimates.

Figure 3 illustrates the evolution of the Engine Health Status Index (EHSI) across the operational timeline of a representative engine. During normal operation, the EHSI remains relatively high and stable, indicating a healthy operating condition. As degradation-related signals accumulate from the underlying failure-code classifiers, the EHSI begins to decline and enters a sustained low-health region prior to the recorded failure event. The red dashed vertical line denotes the failure event identified from the service records. Following the maintenance action associated with the failure event, the EHSI recovers and returns to a stable high-health region. This behavior suggests that the proposed index is capable of capturing degradation trends before the recorded maintenance event and reflecting the subsequent restoration of engine health.

Since failure events are identified from service and warranty claim records, the recorded event may occur after degradation has already become observable in the telematics data. Therefore, the EHSI trajectory should be interpreted as representing progressive degradation leading up to the maintenance event rather than the exact onset of physical failure. The observed recovery of the EHSI after the recorded failure event provides additional evidence that the index responds to changes in the underlying engine condition rather than remaining fixed to a historical degradation trend.

### 4.4. Limitations

Failure events in this study are identified using service and warranty claim records. In practice, the mileage recorded in service claims may not always align exactly with the telematics timeline because claims are typically reported after diagnostic confirmation and service completion. As a result, the recorded failure point may not correspond precisely to the moment when degradation first becomes observable in the sensor data. This temporal uncertainty introduces noise in the exact failure annotation and is an inherent limitation when using field maintenance records for prognostic analysis.

## 5. CONCLUSION AND FUTURE SCOPE

This paper presented a framework for constructing an Engine Health Status Index (EHSI) to represent the overall health condition of diesel engines using field sensor data. The proposed approach combines failure-code risk estimation, transition-aware risk adjustment, and degradation regime identification to aggregate multiple failure signals into a single system-level health indicator. By integrating probabilistic failure predictions with historical failure progression patterns, the framework provides a structured view of how degradation evolves across multiple engine components.

The experimental analysis using fleet data shows that the framework can capture degradation behavior prior to recorded failure events. Although the predictive performance of individual failure-code classifiers varies due to differences in failure frequency and data availability, their combined outputs enable the construction of a health index that reflects system-level degradation trends. The results demonstrate that the proposed formulation can provide a consistent system-level representation of degradation behavior using operational fleet data.

Although the framework was evaluated using heavy-duty diesel engine data, the overall formulation is not restricted to a specific engine platform or failure taxonomy. The approach may be extended to other connected industrial systems where degradation evolves through multiple interacting failure processes and operational telematics data is available.

Future work will focus on improving failure progression modeling and exploring approaches to better account for uncertainty in failure annotations arising from real-world service records. Additional investigation will also consider alternative formulations for integrating failure-code interactions and extending the framework to support different prediction horizons and engine platforms.

## REFERENCES

Cao, H., Yu, J., & Duan, F. (2025). Condition-based maintenance in complex degradation systems: A review of modeling evolution, multi-component systems, and maintenance strategies. Machines, 13(8), 714. https://doi.org/10.3390/machines13080714

Zhang, Y., Fang, L., Qi, Z., & Deng, H. (2023). A review of remaining useful life prediction approaches for mechanical equipment. IEEE Sensors Journal, 23(24), 29991–30006. https://doi.org/10.1109/JSEN.2023.3326487

Qiu, S., Cui, X., Ping, Z., Shan, N., Li, Z., Bao, X., & Xu, X. (2023). Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: A review. Sensors, 23(3), 1305. https://doi.org/10.3390/s23031305

Yuan, C., & Liu, S. (2024). Multi-attribute maintenance decision making for aircraft engines based on entropy weight TOPSIS approach. In Proceedings of the 2024 Global Reliability and Prognostics and Health Management Conference (PHM-Beijing) (pp. 1–6). IEEE. https://doi.org/10.1109/PHMBeijing63284.2024.10874564

Nguyen, K. T. P., & Medjaher, K. (2021). An automated health indicator construction methodology for prognostics based on multi-criteria optimization. ISA Transactions, 113, 81–96. https://doi.org/10.1016/j.isatra.2020.03.017

Bajarunas, K., Baptista, M. L., Goebel, K., & Arias Chao, M. (2024). Health index estimation through integration of general knowledge with unsupervised learning. Reliability Engineering & System Safety, 251, 110352. https://doi.org/10.1016/j.ress.2024.110352

Qin, Y., Yang, J., Zhou, J., Pu, H., & Mao, Y. (2023). A new supervised multi-head self-attention autoencoder for health indicator construction and similarity-based machinery RUL prediction. Advanced Engineering Informatics, 56, 101973. https://doi.org/10.1016/j.aei.2023.101973

Yong, B. X., & Brintrup, A. (2022). Coalitional Bayesian autoencoders: Towards explainable unsupervised deep learning with applications to condition monitoring under covariate shift. Applied Soft Computing, 123, 108912. https://doi.org/10.1016/j.asoc.2022.108912

De Giorgi, M. G., Menga, N., & Ficarella, A. (2023). Exploring prognostic and diagnostic techniques for jet engine health monitoring: A review of degradation mechanisms and advanced prediction strategies. Energies, 16(6), 2711. https://doi.org/10.3390/en16062711

Zhang, P., Cao, L., Dong, F., Gao, Z., Zou, Y., Wang, K., Zhang, Y., & Sun, P. (2022). A study of hybrid predictions based on the synthesized health indicator for marine systems and their equipment failure. Applied Sciences, 12(7), 3329. https://doi.org/10.3390/app12073329

Shcherbakov, M., & Sai, C. (2022). A hybrid deep learning framework for intelligent predictive maintenance of cyber-physical systems. ACM Transactions on Cyber-Physical Systems, 6(2), Article 17. https://doi.org/10.1145/3486252

Medjaher, K., Tobon-Mejia, D. A., & Zerhouni, N. (2012). Remaining useful life estimation of critical components with application to bearings. IEEE Transactions on Reliability, 61(2), 292–302. https://doi.org/10.1109/TR.2012.2194175

Mosallam, A., Medjaher, K., & Zerhouni, N. (2016). Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. Journal of Intelligent Manufacturing, 27, 1037–1048. https://doi.org/10.1007/s10845-014-0933-4

Wang, T., Yu, J., Siegel, D., & Lee, J. (2008). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In Proceedings of the International Conference on Prognostics and Health Management (pp. 1–6). IEEE. https://doi.org/10.1109/PHM.2008.4711421

Liu, B., Gao, Z., Lu, B., Dong, H., & An, Z. (2022). Deep learning-based remaining useful life estimation of bearings with time-frequency information. Sensors, 22(19), 7402. https://doi.org/10.3390/s22197402

Chen, L., Xu, G., Zhang, S., Yan, W., & Wu, Q. (2020). Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks. Journal of Manufacturing Systems, 54, 1–11. https://doi.org/10.1016/j.jmsy.2019.11.008

Guo, L., Lei, Y., Li, N., Yan, T., & Li, N. (2018). Machinery health indicator construction based on convolutional neural networks considering trend burr. Neurocomputing, 292, 142–150. https://doi.org/10.1016/j.neucom.2018.02.083

Xu, F., Huang, Z., Yang, F., Wang, D., & Tsui, K. L. (2020). Constructing a health indicator for roller bearings by using a stacked auto-encoder with an exponential function to eliminate concussion. Applied Soft Computing, 89, 106119. https://doi.org/10.1016/j.asoc.2020.106119

Kaji, M., Parvizian, J., & van de Venn, H. W. (2020). Constructing a reliable health indicator for bearings using convolutional autoencoder and continuous wavelet transform. Applied Sciences, 10(24), 8948. https://doi.org/10.3390/app10248948.

## BIOGRAPHIES

**Rohit S. Deo** received the bachelor's degree in Electronics and Telecommunication Engineering from Solapur University, Solapur, India, in 2011, and the master's degree in Signal Processing from University of Pune, Pune, India in 2014. He is currently a Senior Data Scientist with Cummins

Technologies India Pvt. Ltd., Pune, India, where he works on machine learning methods for predictive maintenance and engine health monitoring using fleet-scale telematics data. He has more than 11 years of experience in applied machine learning and artificial intelligence for industrial systems, with work spanning predictive modeling, time-series analysis, and large-scale data-driven engineering applications. His research interests include prognostics and health management, machine learning for engineering systems, time-series modeling, and interpretable artificial intelligence for industrial applications.

**Shruti Bharti** was born in Hajipur, Bihar, India. She received her B.Tech in Computer Science and Engineering from the National Institute of Technology Sikkim, India, in 2025. She completed a six-month internship at Cummins India as a Data Scientist and is currently working there in the same role, focusing on developing data-driven and scalable machine learning solutions for engineering applications.

**Aman Yadav** was born in Varanasi, India, on March 12, 1998. He earned a Bachelor of Technology in Computer Science and Engineering from the University Institute of Engineering and Technology (UIET), Kanpur, India, in 2021, and a Master of Technology in Artificial Intelligence from the Defence Institute of Advanced Technology (DIAT), Pune, India, in 2023. He worked as a Data Science Intern at Cummins Inc. from 2022 to 2023 for a duration of nine months. Since August 2023, he has been working as a full-time Data Scientist at Cummins Inc.

**Nilesh Powar** is the Advanced Analytics Director at Cummins. He holds bachelor in electronics engineering from University of Bombay, India, M.S in Computer Engineering from Wright State University and a doctoral degree in Electrical and Computer Engineering from University of Dayton, Ohio. He has over 20+ years' experience in field of image processing, machine learning, statistical pattern recognition and system integration. He had worked in the US as Distinguished Research Scientist for University of Dayton Research Institute, Dayton, OH, USA. Recent efforts involve data analytics for die casting, predictive analysis for supply chain management and video summarization using deep learning.