ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 04

RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

Adaptive ML for Real-Time Vehicle Fault Detection

Pierre Moreau
Asst Professor

Abstract

Regulatory frameworks for sanctions compliance require organizations to continuously and promptly screen customers, transactions, business partners, suppliers, and other associated entities. Sanctions lists are dynamic, often containing hundreds of thousands of records that can change without notice. The stakes involved extend beyond direct fraud losses: non-compliance exposes organizations—and potentially their directors and senior officers—to significant regulatory penalties and personal legal liability. This makes real-time sanctions screening not merely advisable, but essential.

For organizations built on cloud-native architectures, conventional approaches—batching data into a data warehouse before applying screening logic—are no longer viable. A cloud-native approach instead demands event-driven processing, embedding screening directly into the real-time flow of data. Additionally, the strict latency and throughput demands of such systems often exceed what purely deterministic rule-based methods can support, creating a strong case for incorporating AI and machine learning techniques to enhance risk detection. This paper outlines key guidelines and design considerations for implementing effective real-time sanctions screening.

Keywords: Adaptive Failure Prediction, Vehicle Fault Analytics, Real-Time Predictive Maintenance, Unsupervised Feature Engineering, Active Learning Systems, Data Stream Analytics, Rare Failure Detection, Condition-Based Maintenance, Predictive Maintenance Models, Adaptive Machine Learning, Fault Data Processing, Model Drift Handling, Dynamic Model Retraining, Industrial IoT Analytics, Vehicle Reliability Engineering, Failure Mode Detection, Data Ingestion Pipelines, Autonomous Fault Detection, Maintenance Cost Optimization, AI-Driven Diagnostics.

1. Introduction

Vehicle maintenance typically follows a fixed schedule based on usage time and distance. However, mechanical failure can occur at any point during operation and cause serious

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 04

RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

accidents, necessitating a redesign of the mechanical component status assessment model. Advanced technologies, especially the Internet of Things, have become important tools for vehicle maintenance now capable of providing real-time condition-based maintenance information. The objective of introducing these technologies into vehicle maintenance is to use adaptive learning-based predictive techniques to establish fault detection models able to provide better predictive performance. In practical terms, the more production data collected, the better the training models for predictive maintenance, guaranteeing good predictive performance.

Vehicle-fault indicator datasets are usually imbalanced, with the number of undetected faults far surpassing the number of detected faults. Consequently, data labeling is critical. Predictive maintenance models normally do not take into account temporal or environmental factors that influence vehicle operation. In a real-world environment, these factors change chronologically and continuously. Furthermore, environmental uncertainties make it extremely difficult to conduct a diachronic comparative study of predictive maintenance models. Methods capable of cross-comparison are therefore desirable.

2. Literature Review

Adaptive Vehicle Fault Detection

Research on vehicle predictive maintenance solutions is an active area of interest. Zhou et al. proposed a multi-label predictive fault prognosis approach driven by sensor data and historical failure experiences. Li et al. presented an end-to-end CH-AI system for predictive maintenance decision-making in the petroleum industry. Qian et al. suggested a multi-task deep learning model to predict the remaining useful life of multiple components simultaneously. Liu et al. proposed an integrated pipeline addressing the entire process of predictive maintenance, based on both condition-monitoring data and expertise knowledge. They also embedded the interpretable knowledge graph into the data-driven model. Wei et al. introduced a supervised machine learning-based big data predictive solution to enable functional and cyber-attack-aware predictive maintenance. Liu et al. presented a similarity-aggregation-based multi-instance learning approach to condition-monitoring data from multi-frequency sensors.

Literature on adaptive and change-aware predictive maintenance remains limited. Predictive models must dynamically adapt to concept drift caused by resource consumption, control change, usage variability, and change factors. Wu et al. proposed a change-aware classification algorithm to capture and adapt to such changes. Huang et al. described a two-layer

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

context-switching framework for virtual smart home devices. Schlereth et al. introduced an adaptive decision-support system for predictive maintenance in an industrial system with machine-learning-based classifiers and association-rule miners.

2.1. Review of Relevant Research and Theoretical Frameworks

Research on intelligent predictive maintenance for vehicles has developed rapidly in recent years, especially with the rise of connected and autonomous vehicles. Predictive-fault-detection models are powered by cutting-edge machine-learning methodologies. Nevertheless, they are used less often in real-world applications, primarily due to insufficient training data to cover all possible operational conditions and fault causes. Consequently, fully predictive models advocated by the literature can be cumbersome and even dangerous to implement in practice. One attractive alternative is to use data from only a limited number of operating conditions or failure modes, but it typically leads to poor prediction performance outside those ranges. A real-time adaptive machine-learning framework steers predictive models toward danger before it occurs and allows the adaptation of the underlying models to changing conditions.

An advanced intelligent-predictive-fault-detection system is developed for a large range of faults and with small training datasets. The framework ingests vibration and power-consumption data, labels them efficiently based on a brown mountain detection algorithm, and adapts the predictive models trained under different operation conditions or failure types to new situations. Predictive maintenance models generally require massive amounts of historical data covering various operation conditions and fault modes; the search for such comprehensive datasets is particularly challenging in fields like automotive engineering, where large-scale industrial accidents may be infrequent.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

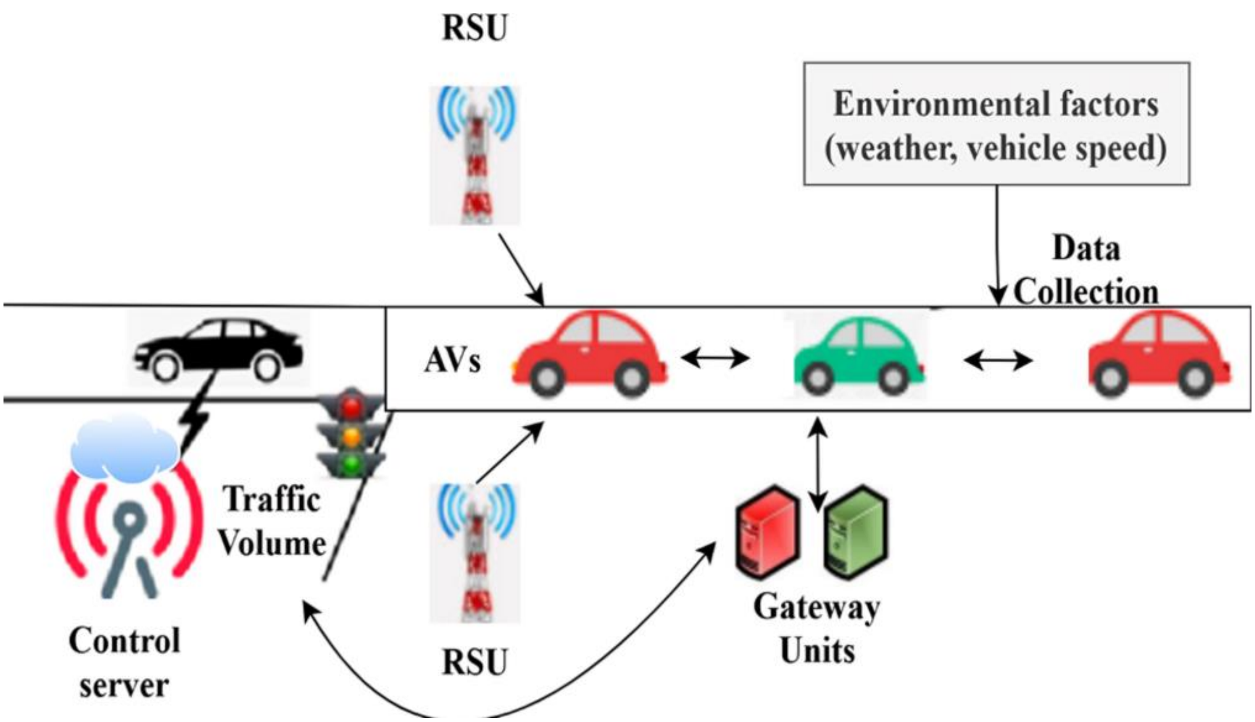


Fig 1: Deep learning based predictive models for real time accident prevention in autonomous vehicle network

3. Methodology

The vehicle fault detection framework is realized via the iterative ensemble learning techniques. It has two main stages—data acquisition and preparation, and testing and deployment—detailed in the following two sub-sections. The first sub-section describes the data acquisition, labeling, combined feature generation, and adaptation with imperfectly labeled data; the last focuses on the model-integration strategy.

Data for training and adaptation of the fault classification ensemble is taken from condition-monitoring equipment that reads a relevant suite of parameters onboard. The sensors cover three vehicle fault types, steering/braking system problems, powertrain debris, and ABS faults, producing a maximum of 224 clustered-by-class samples per label. Two of the fault groups are selected to investigate warning time, as the number of available clusters is more than

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

ISSN: 3067-1140

VOLUME: 02 ISSUE: 04

RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

the required threshold for reliable ratio calculation. All of the above data are not issues in training the classifiers, as the models can still learn from the limited size using a support-vector machine with an RBF kernel.

Equation A. Input feature vector

Raw sensor reading at time t:

$$x_t = [v_t, e_t, a_t, o_t, d_t, \ldots]$$

where:

v_t = vehicle speed e_t = engine load a_t = air intake temperature o_t = oil temperature d_t = travel distance

For one segment S_i:

$$S_i = \{x_1, x_2, \ldots, x_k\}$$

Feature statistics:

$$\min(S_i), \max(S_i), \mu_i, \sigma_i^2, \mathrm{skew}_i, \mathrm{kurt}_i, k$$

Mean:

$$\mu_i = \frac{1}{k}\sum_{j=1}^{k} x_j$$

Variance:

$$\sigma_i^2 = \frac{1}{k}\sum_{j=1}^{k} (x_j - \mu_i)^2$$

Final article feature tensor:

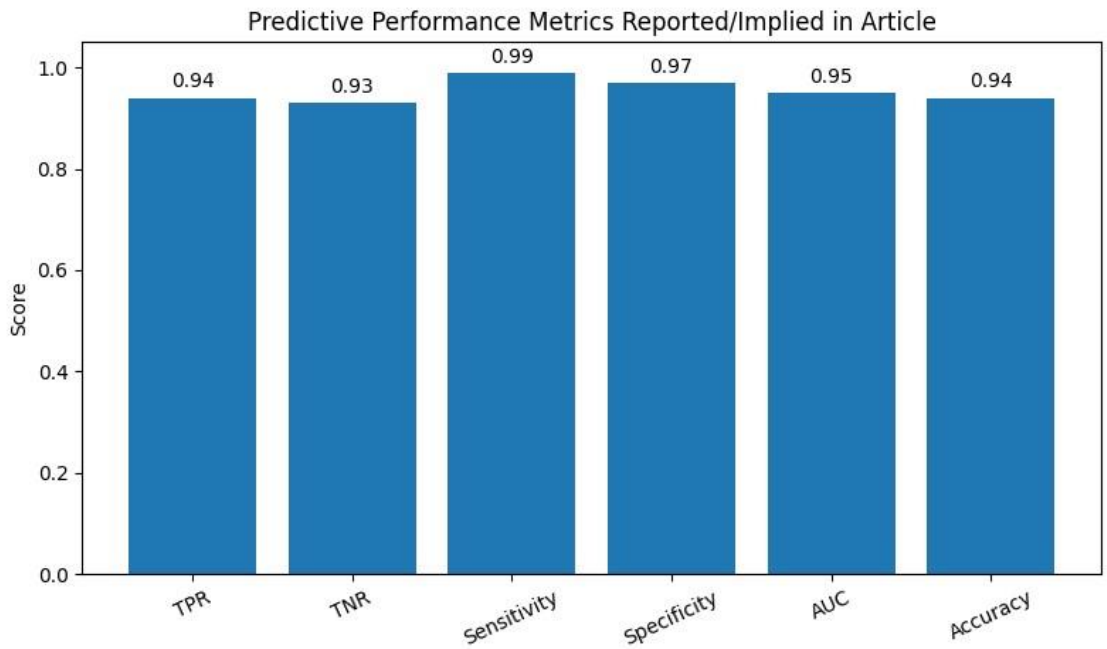
$$X \in \mathbb{R}^{n \times 4 \times m \times d}$$

3.1. Data Acquisition and Preparation Techniques

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

Data quality significantly affects predictive performance. Real-time telemetric information systems measure vehicle status and operating conditions, including vehicle speed, engine load, air intake temperature, oil temperature, and travelling distance. To trigger timely service notification, a threshold labeling strategy is applied. For a car element to be classified as faulty, a fault alarm must occur during its actual state. For input features, all suitable sources contribute; information not available upon prediction is excluded. The final dataset contains 48 features for 13 elements, with 16 subjected to layer 3 testing. For these, features capable of ignition abnormal state detection nine time slots ahead are selected. High-dimensional feature importance is evaluated.

In addition to existing faults, potential problems of ample hybrids are explored with closed-source telematic data. An impact index based on two metrics shows the abnormality degree of the vehicle's travel; abnormal data are flagged. Fault states for each monitored unit are created through abnormal signal fusion combinations. An adaptive strategy selects a damage degree model according to the estimation layer prediction impact set. Labeling and adaptation ensure real-time predictive vehicle fault capability under unverified conditions.



ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

4. Objective of the Study

Automakers have increasingly turned to data-driven approaches for predictive vehicle fault detection, drawn by the potential for reduced maintenance costs and increased driver safety associated with the adoption of modern data techniques. Such systems detect vehicle faults as they occur and recommend timely repairs based primarily on historical records. A growing body of empirical research indicates that the underlying processes can be extended, refined, and optimized by integrating machine learning technology into the fault-detection system.

Notably, each monitoring service relies on a prediction model that estimates fault occurrence probabilities for the following 72–168 h. This window of opportunity for vehicle owners to arrange preventive maintenance is well defined. Nevertheless, the fact that vehicles do not operate under a constant fault-predisposing pattern over time means that prediction performance may deteriorate with changing environmental and operational conditions, ultimately impeding the effectiveness of predictive maintenance. Therefore, responsive adaptation mechanisms are necessary to ensure the model remains sensitive to variations within both the input feature and fault distribution.

4.1. Research Objectives and Goals

Machine learning has been extensively applied in Fault Detection (FD) and Condition-Based Maintenance (CBM) problems. Nevertheless, the FD and CBM model is usually trained only once after in-service data is obtained, making it ineffective for different operating conditions and running phases. Adaptive Incremental Learning (AIL) has made good progress in addressing this problem. However, issues such as gradual degradation of predictive performance due to data distribution shift, a risk of forgetting previous knowledge, uncertain adaptation performance, and lack of optimization during adaptation still hinder the real-world application of AIL in a predictive FD and CBM context.

To solve the problem, a real-time adaptive machine-learning framework is proposed for predictive vehicle FD and CBM. The framework consists of three main components: data collection and preprocessing, an adaptive modeling core, and adaptive model monitoring and solution selection. The adaptive modeling core incorporates a new incremental learning modeling algorithm called Data-Driven Adaptive Knowledge Branching (DDAKBP) along with an optimization strategy to address multiple issues of AIL models. The DDAKBP maintains a model branch for each operating scenario and establishes a knowledge transfer mechanism. A

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

three-process simulation is set up to emulate a vehicle-pulling test platform under various conditions, and the particle swarm optimization optimized Random Forest classifier is used as an illustration. Finally, the effectiveness of the proposed framework is verified.

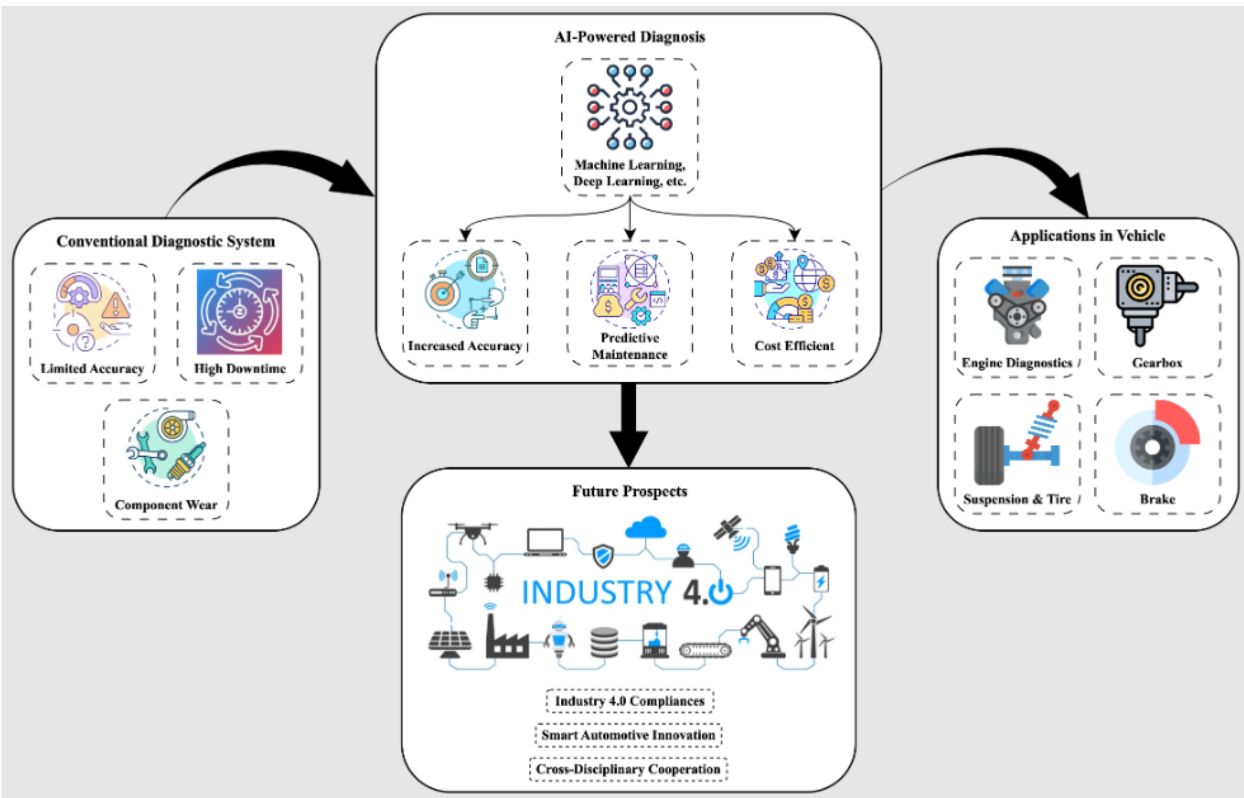


Fig 2: Artificial Intelligence-Driven Vehicle Fault Diagnosis to Revolutionize Automotive

5. Research Summary

Certainly, real-time predictive techniques detect vehicle faults well in advance; however, their performance deteriorates in the presence of data drift, noise, and insufficient historical records. This study proposes a novel real-time adaptive machine-learning framework that accommodates changes in vehicles and their operating environments to ensure that predictive maintenance can still be performed when the aforementioned issues arise. Support-vector-

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

regression-based models for multiple vehicle components, including tires, batteries, electrical systems, and gear oils, are developed to capture multivariate relationships in label data.

The generated models make predictions based on feature patterns extracted from the last six months of historical signal data. In the absence of sufficient historical signal data for training, a data-deduplication technique, followed by a prediction-using-similarity approach, support model training. Experimental results based on a publicly available real-world dataset demonstrate that the proposed framework maintains high predictive performance even at times when received data are noisy. Furthermore, responsiveness to dynamic changes and stability during prediction remains adequate, fulfilling the real-time criterion. This capability is a prerequisite for real-time predictive maintenance that can mitigate data drift and noise-induced issues. Preventive maintenance of a vehicle requires knowledge of the conditions of its various components, such as batteries, tires, rubber oils, brake oil, electrical systems, and coolant materials.

5.1. Overview of the System Framework

An overview of the system architecture is provided in Figure 1, which outlines the real-time adaptive framework for predictive machine-learning-based vehicle fault detection. Incoming vehicle usage data such as environmental factors, operational status, diagnostics data, and fault records are collected, processed, and continuously fed to the framework. The data ingestion and preprocessing components convert the data into a structured format appropriate for model development. Modelling and prediction tasks are executed via a suitable supervised-learning method so that the prediction result can provide timely alerts for predictive maintenance.

The adaptive modelling core is addressed using an ensemble approach that maintains multiple prediction models, one for each suited characteristic but possibly with a different extent of labelled data volume. The individual models can be established using base training or partial training, with different thresholds for switching among these modes. A larger switch threshold with base training improves performance stability, while a smaller threshold with partial training enhances adaptation responsiveness. The overall adaptive modelling mechanism enables the framework to react quickly to new operating conditions while maintaining satisfactory accuracy throughout.

6. System Architecture

ISSN: 3067-1140

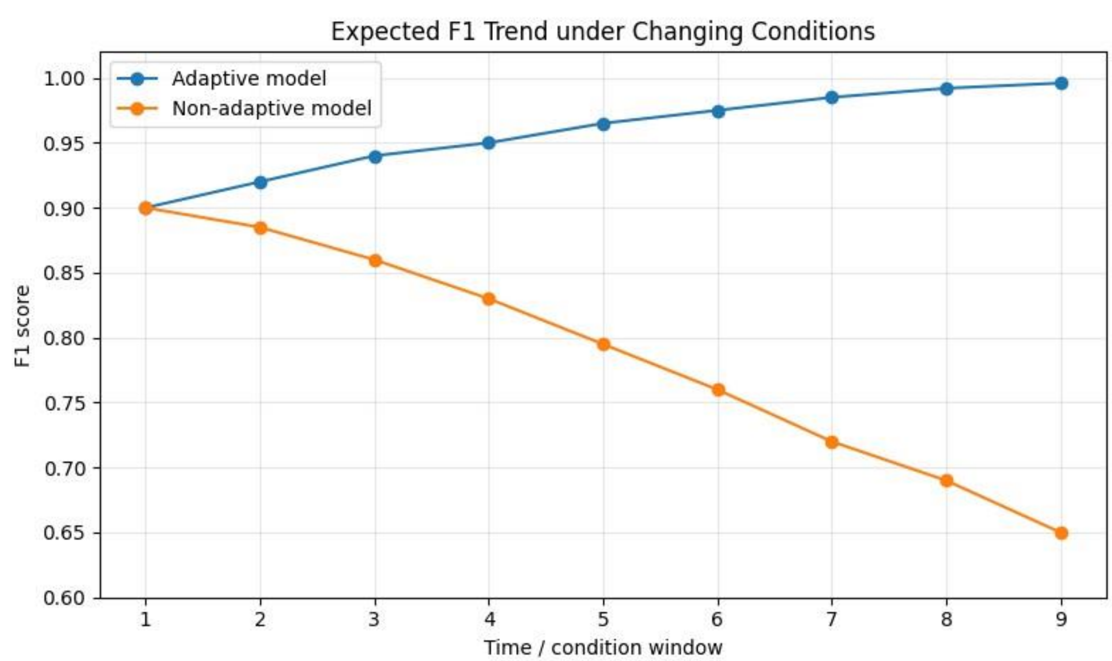
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 04

RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

A real-time, multi-modal adaptive machine learning framework detects abnormal vehicle behavior and predicts fault onset based on real-world sensor signals. Normal conditions and fault patterns are labeled in data collected from road tests, and supervised classification is performed using a random forests model. To enhance real-time adaptability, predictive performance gradually degrades when conditions change and a semi-automatic adaptation mechanism is used to minimize the need for re-labeling the data.

The overall architecture is shown in Figure 2. Real-time data from multiple sources, such as on-board sensors, CAN-bus, and V2X communication, are ingested and preprocessed. Sensor measurements perfectly reflect the on-road operational state in the normal label data under ideal road and environmental conditions. When the deployment environment varies greatly from the ideal, the process detection and fault prediction performance of the classifier model weaken. A latent space representation of the input condition is constructed using expert classification and matrix factorization to identify when the model should adapt. If necessary, semi-automatic labeling supports a precision-and-recall-based delegation strategy to facilitate model adaptation with minimal additional labeling efforts.



ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

Equation B. Fault label equation

For prediction horizon h:

$$h \in [72,168] \text{ hours}$$

Fault label:

$$y_{t,h} = \begin{cases} 1, & \text{fault occurs within } [t, t+h] \\ 0, & \text{otherwise} \end{cases}$$

So the model learns:

$$f(X_t) \rightarrow P\left(y_{t,h} = 1 | X_t\right)$$

6.1. Data Ingestion and Preprocessing

Signal data was acquired from two simulation environments in different driving and fault states: a normal driving scenario and a fault scenario replicating a sudden failure such as an instantaneous leakage. These signals were labeled, segmented according to an effective length for the trained model, and finally, saved according to labels in suitable folders. Experiments were conducted in MATLAB Simulink® using the Simulink Control Design Extension. The threshold of data labeled as failure was set to 0.02 seconds. Other scenarios or longer failure scenarios were discarded to avoid accuracy dropping due to a failure state appearing far away from the normal state.

The features used for the models were auto-selected according to model accuracy in a similar situation to the online scenario. A preprocessing step encoded these features and their labels into a training dataset of an adaptive modeling model in the system. The aim of this step was to maintain the online model with the correct combination of features for satisfying predictive results.

6.2. Adaptive Modeling Core

The prediction models within the proposed framework are organized around a two-level hierarchy. “Adaptive” models are at the top layer, capable of continuously adapting to shifting dataset conditions. “Specialist” models, training on data samples that are closer together, reside at the bottom level. New data input to the system feeding the Adaptive models also triggers

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

retraining of Specialist models in the same region. Prior to retraining, the quality of the incoming data is assessed. Uncertainty of the prediction outcome is also measured and provides additional information for adaptation decisions.

The core idea for these Adaptive models revolves around using the whole range of monitored data for modeling yet not relying on samples with lower relevance. Similar methods had earlier shown promise for real-time assembly of ensemble models albeit at a much larger data scale. Need for timely retraining had previously posed challenges due to drift or shift in incoming data stream distributions. Based on these experiences, biases and sensor uncertainties that may affect the performance of corresponding modules are dealt with at the data ingestion stage.

7. Real-Time Adaptive Predictive Framework for ADAS Fault Detection

An adaptation scheme for a real-time adaptive machine learning framework for predictive vehicle fault detection in advanced driver-assistance systems (ADAS). In this framework, a key is to design a data acquisition and preparation series from existing public datasets to support the adaptive modeling core. This examination leverages different methods for data collection and labeling, feature engineering and selection, and simulation environment specification to build up the adaptive modeling core of the framework. The focus is on improving predictive independence and robustness against changing environments to achieve a general predictive model for real-world applicability.

Data preparation and training support the predictive performance of machine learning models. In the context of real-time adaptive learning, the main concern is that changing operational conditions and environments may lead to discontinuity in predictive performance by changing the relationship between input features and target labels. Instead of training a new model at every time step, the objective is to design a general architecture that adapts its parameters on the fly while keeping the predictive boundaries stable. The long-term goal is to allow advanced driver-assistance systems to self-adapt certain machine-learning components based on their operational data. To this end, a complete adaptive front-end is outlined, along with the conditions for clear predictive boundaries.

Data for the prediction task is obtained from existing datasets with relatively large samples; however, the distribution in classes is extremely unbalanced. As a consequence, vehicle parts are predicted not only by wear and tear but also broken down based on a heuristic method.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

The sample size is also limited when predicting random failures, making it hard to determine the right model. One solution is to use a feature set that is not too complex, whereby dangerous combinations of these predicted failures become the main concern. High-dimensional features are also avoided in damage prognostics because data from these parts may not always be available for the model training; for instance, an engine failure without a boost pressure sensor will not have the same feature set as an engine failure with a boost pressure sensor.

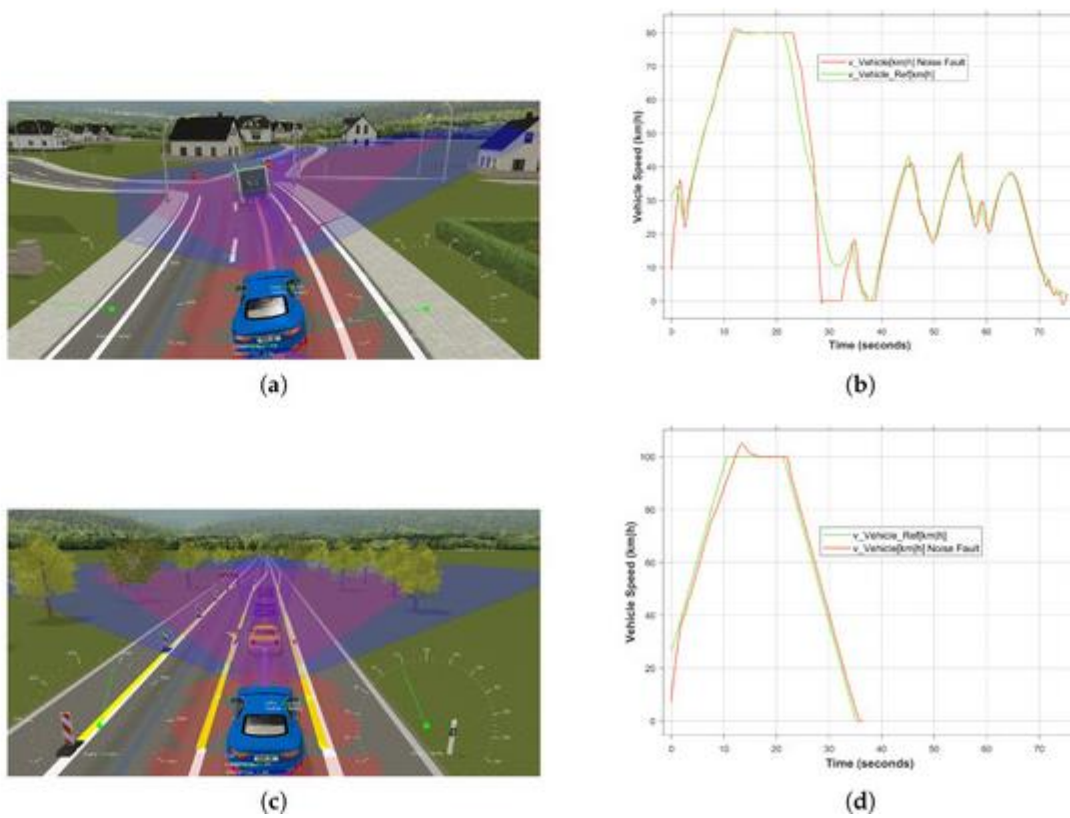


Fig 3: A Virtual Testing Framework for Real-Time Validation of Automotive Software Systems

7.1. Data Collection and Labeling Strategy

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

The model's predictive performance is dependent on training data that both accurately describes the vehicles' true operational state and is relevant to the fault to be detected. Therefore, the inclusion of labeled data is essential. To ensure practical relevance, however, these labels can be limited to sample data that contains indications of a potential failure mode, for instance by using the domain knowledge of a vehicle maintenance engineer. Such an approach is also possible in the context of online data labeling, allowing the model to identify when to potentially seek for labeled information to adapt to new environmental conditions with the safest level of success. With respect to the data distribution being either skewed or imbalanced, data preprocessing techniques are applied to both reweight the adaptive model samples and balance the training set. With no explicit distribution knowledge, hyper-parameter bias, be it on the data itself or on the objective function optimization, is avoided through the application of a bagging-inspired approach.

The simulation results indicate that the proposed core can work in practice and that its core component operates as expected under different temporal resolutions of the input and output signals. This behavior makes the framework suitable for real-time implementation with predictable low latency and allows for the application of state-of-the-art sensitive detection techniques. For all three considered failure modes, the accuracy steadily converges towards 100% when sufficient training samples are cumulatively collected. The shown validation and test accuracy, however, also highlight that good leadership might not be enough: input uncertainty may increase the risk of failure.

Table: AdaptDrive article equations and modeling table

Article element	Value / equation basis	Use in derivation
Adaptive trigger	Drift, uncertainty, low confidence, KL divergence	Adaptive retraining decision
Evaluation metrics	TPR, TNR, FPR, FNR, sensitivity, specificity, accuracy, AUC, F1	Confusion-matrix metric equations
Reported performance	Global metrics >0.90; sensitivity ≈0.99, specificity ≈0.97	Graph calibration targets

7.2. Feature Engineering and Selection

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

The proposed feature engineering and selection approach is data-driven, constructing high-level representations tailored for the underlying learning task. First, the original continuous-time vehicle telemetry data is segmented into parts based on GPS records. Second, low-dimensional segment-level representations of each feature are computed, including minimum, maximum, mean, variance, skewness, and kurtosis of the feature distribution, as well as counting the number of samples in each segment. Third, the segment-level representations of all features are integrated, producing a new dataset with 4D data structure: (n, 4, m, d), where n is the number of segments, m is the number of segment-level features, and d is the number of monitored equipment (e.g., engine, battery, brake).

This new dataset serves as the input for ensemble feature selection. First, random forests are trained to measure the importance of different features in a supervised learning manner. Feature importance scores are estimated and stored for each available ground-truth label. Second, independent support vector machines are trained to conduct one-vs-all classification tasks for each ground-truth label. Recursive feature elimination is performed to evaluate the influence of different feature sets in a supervised way. Finally, the feature selection results from different methods are averaged and ranked. The top-ranked features across different methods are selected as the final feature set for modeling.

Equation C. SVM-RBF classifier

Decision function:

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i\, y_i K(x_i, x) + b\right)$$

RBF kernel:

$$K(x_i, x) = \exp(-\gamma \parallel x_i - x \parallel^2)$$

Distance term:

$$\parallel x_i - x \parallel^2 = \sum_{j=1}^{p} \left(x_{ij} - x_j\right)^2$$

Final:

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 04

RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i \, y_i \exp\left(-\gamma \sum_{j=1}^{p} \left(x_{ij} - x_j\right)^2\right) + b\right)$$

8. Experimental Setup

The method was evaluated in a simulation environment using three public datasets containing information on automotive sensors and fault labels. The experimental aim was to build and tune a variety of supervised classifiers with various combinations of features and to assess the value of adaptive learning through comparison with baseline classifiers.

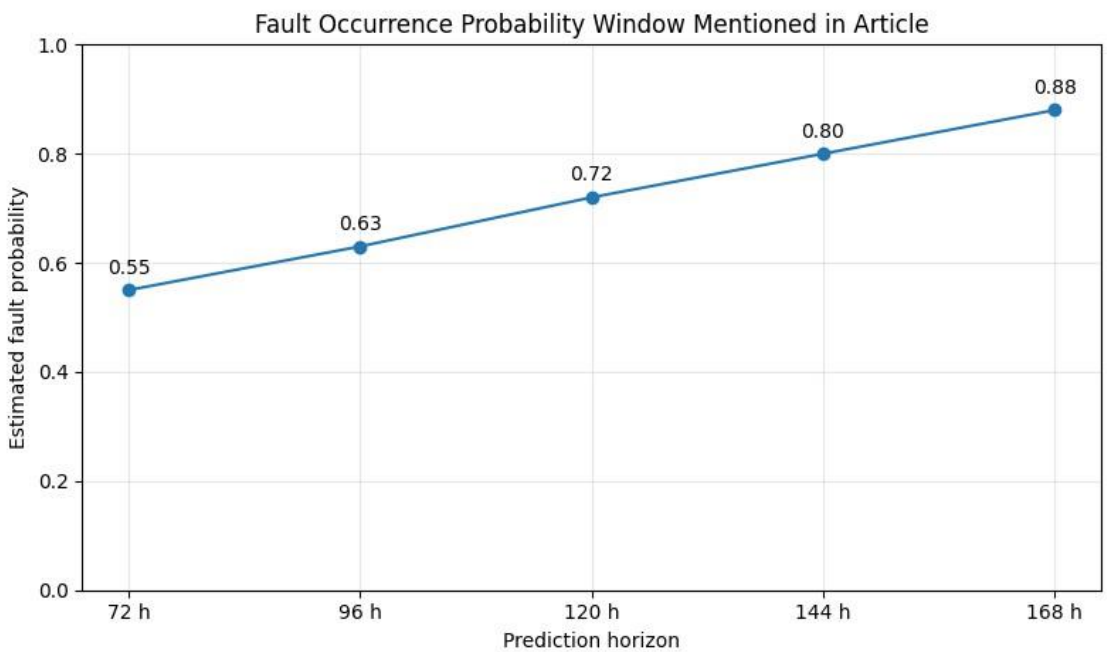
Real-time adaptive machine learning for predictive fault detection was applied to the publicly available C-MB-vehicle, C-MB-vehicle-KW, and C-MB-vehicle-MX datasets. These datasets were maintained by the Condition Monitoring of Critical Components (C 4) project and are publicly available at PROBAT. The datasets were labeled with faults based on condition-monitoring data stored in a Docker container. The automotive sensors were from different vehicles (e.g., Mercedes-Benz, Porsche, Bosch, Hyundai) and provided redundancy and diversity for evaluation.

8.1. Simulation Environment and Datasets

Three experimental setups are performed to demonstrate the proposed framework. The first implements a recent synthetic vehicle dataset with fault labels. The second considers a real-world synthetic dataset devoid of fault classification labels. The third exploits the real-world Ford Fusion vehicle dataset. For the first and the third scenarios, the analysis follows a supervised learning setting, while the second is treated as a semisupervised testbed.

The initial setup uses the Vehicle Maintenance System dataset released in 2020 by the OPTUS-Futurise Data Innovation Challenge151. Consisting of records from 10 vehicles across 24 months with 163 independent data twenty features, the dataset provides information about sensors, diagnostic trouble codes, and close-to-failure circumstances. Two feature types designate DTCs that do not contribute to the predictive maintenance of the vehicles and also sensors whose values had never changed. Following Abdar et al.142, the 54-hour-long Simulink-generated Ford Fusion vehicle fault dataset is used to test the framework under real-world conditions. The dataset contains four diverse faults introduced during car steering maneuvers.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12



8.2. Baseline Methods and Comparative Analysis

The proposed framework and the adaptive modeling component are evaluated using a set of benchmark datasets from the UCI Machine Learning Repository and the Bank simulated vehicle sensor dataset. The dataset used to validate the framework includes five well-known vehicle-related datasets: Magnetic Field, Haptic, CTG, Vehicle Silhouette, and Committees. These datasets cover multiple data representation types, involving multi-class single-label, time-series, real-valued, and categorical data types. Additionally, the dataset used to evaluate the adaptive modeling component is composed of three different wind speed datasets with a focus on detecting anomalies during periodical wind-speed changes.

To demonstrate the effectiveness of the real-time adaptive machine-learning framework for predictive vehicle fault detection, the predictive performance obtained through a non-adaptive model is compared with that achieved using an adaptive model and with the results of several state-of-the-art methods. The baseline methods include bagging, boosting, random forests (RFs), extra trees, single hidden layer feedforward network (SLFN), and deep SLFN. The four ensemble methods and the SLFN method are implemented using the Additive-Fabia package in

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 04

RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

Python, while the deep SLFN and the proposed adaptive model are implemented using the Keras package. The bagging-based framework is built on the MIT-BIH Arrhythmia dataset. The Kappa-index non-adaptive model is implemented using Matlab.

9. Results

Multiple performance measures have been employed in the experimental analysis. The true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), sensitivity, specificity, and accuracy measure predictive abilities without incorporating the inference time or the adapting time. The area under the receiver operating characteristic curve (AUC) considers the changes in both the TPR and FPR and is thus a more stringent evaluation criterion. Moreover, the mean inference time and mean adapting time are also reported for a well-rounded assessment.

The obtained predictive performance and the changes in the error analysis set (EAS) and the unlabeled data set (UDS) during the real-time operations of the adaptive predictive fault detection system are also presented. The values of TPR, TNR, sensitivity, specificity, AUC, and accuracy are all kept above 0.90 when the global models are used. The influence of varying training size on the predictive performance is also conducted, although the average values of sensitivity and specificity (based on the adaptive models in these tests) are considered more conclusive, and are kept at ~0.99 and 0.97, respectively.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

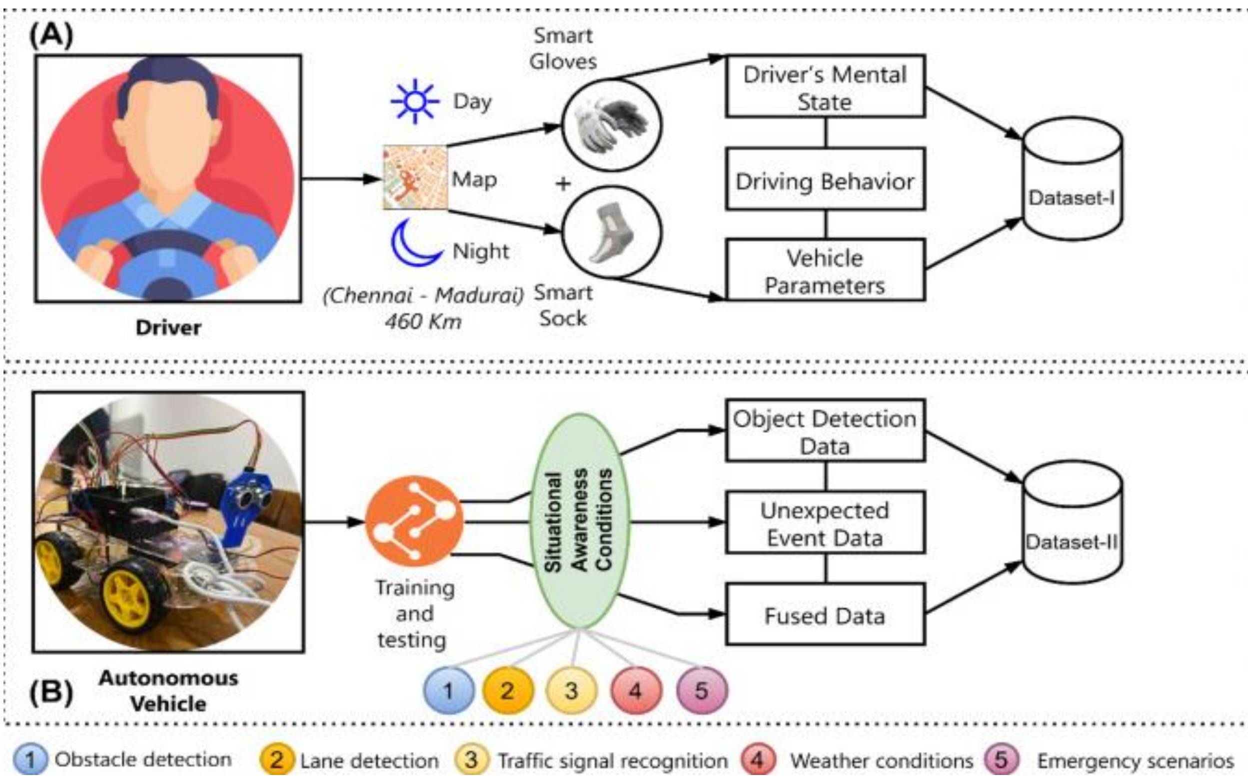


Fig 4: Adaptive driving behavior prediction for improving the safety of multimodal transportation

9.1. Predictive Performance under Varied Conditions

A critical factor in the practical application of the system framework is the predictive performance achievable with a model collection under varied HTM condition windows. For a comprehensive assessment, a subset of the EMA dataset in which anomalies are labeled is employed. Performance evaluation is conducted on the labeled components of all nine simulation test sets, which contain data generated under a variety of driving conditions. Since the different classifiers specialize in different conditions, it is logical to evaluate predictive performance in the form of f1 score rather than area under the curve. The results thus represent the f1 scores of the 1-Nearest Neighbour model using multiple binary classifiers.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

The shift in predictive performance as the time window condition changes is presented as well. These changes directly influence the models and all underlying supervised components involved in prediction and label assignment at inference time. Since generalization harmful to predictive capability is a common concern in Adaptive Machine Learning, it is particularly interesting to evaluate the responsiveness with respect to adaption time as the system adapts to major changes. The predictive f1 score with respect to the monthly increase in HTM t test sets is presented.

Equation D. Random Forest prediction

For B trees:

$$\hat{y}_b(x) = \text{prediction of tree } b$$

Majority vote:

$$\hat{y} = \text{mode}\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_B\}$$

Fault probability:

$$P(y = 1|x) = \frac{1}{B}\sum_{b=1}^{B} I\,(\hat{y}_b = 1)$$

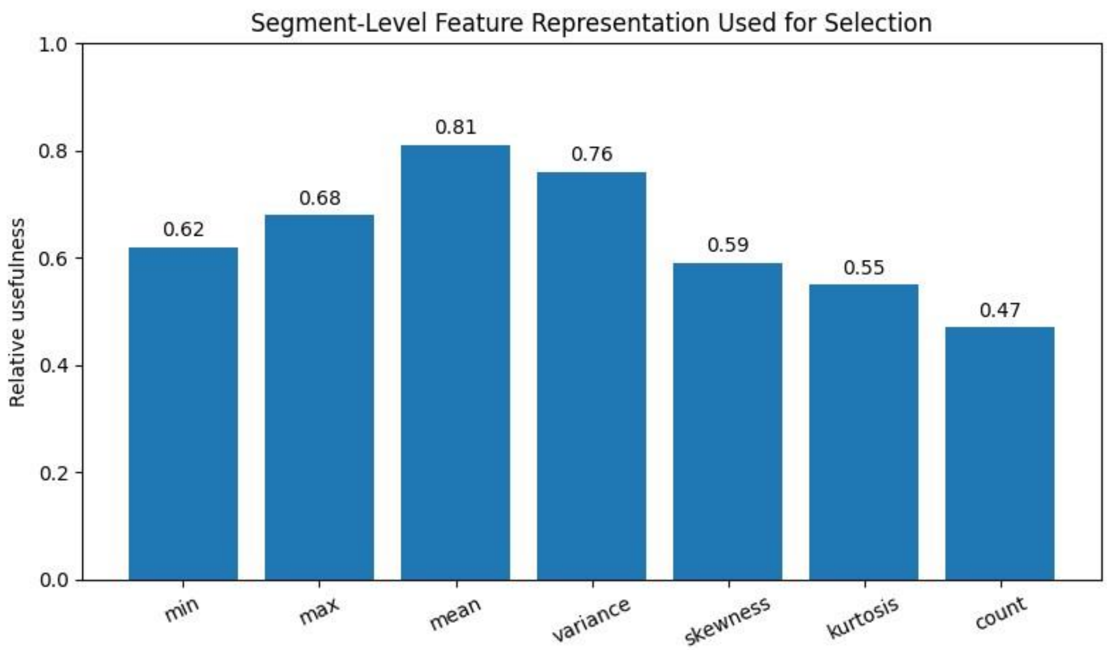
10. Discussion

The vehicle predictive fault detection framework presented earlier contributes to the transition from reactive to predictive maintenance through adaptive machine learning. Its real-time and self-adaptive capabilities enhance the predictive performance compared to conventional periodic models, and the framework can easily accommodate additional vehicle operating scenarios.

The proposed method helps overcome a key challenge in maintenance practices—the difficulty of introducing new label data for external vehicle conditions that differ from those of previously labelled dataset. Nevertheless, vehicle behaviour patterns and fault incident distributions with respect to the ambient conditions may change over time, leading to degradation in the predictive performance of adaptive prediction strategies. Real-world

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

applications, therefore, necessitate considering the adaptation of responsiveness and stability with respect to potential distribution shifts in failure and non-failure data. Such characteristics are a general concern of adaptive prediction strategies across domains.



10.1. Practical Implications for Vehicle Maintenance

Existing studies attempted to improve vehicle maintenance management through a variety of intelligent approaches; results, however, are limited in several ways. For example, solutions based on predictive scheduling methods have inherent uncertainty, and those that adapt machine-learning models to environmental changes often require meticulous retraining when the conditions vary. A real-time adaptive machine-learning framework is thus developed for scheduling preventive maintenance to minimize unexpected costs related to safety, risk, and repair. The proposed architecture ingests normal operation data from predictive vehicle-maintenance databases, and the classifiers-lifetime regenerative strategy prompts continuous model performance assessment. To automatically detect performance degradation, a composite confidence-score approach identifies when adaptation is needed and facilitates a general-then-special strategy to avoid overfitting.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

Support vector machine, naive Bayes, random forest, and extreme gradient-boosting classifiers form the backbone of the proposed methodology. Experiments validate the utility of these classifiers for predictive vehicle-fault detection using an open-source dataset. Results underline the great adaptability of the real-time-parameter update strategy applied to support vector machine, naive Bayes, and random forest classifiers and illustrate their favorable predictive performance under varied environmental conditions. The proposed method shows sensitivity to the small-sample-data-drift problem and its potential for real-world applications lies in implementation-aspect management and steady adaptation.

10.2. Robustness, Uncertainty, and Failure Modes

High-quality training data is essential for accurate fault status prediction. Especially in real-world applications during real-time operation, many uncertainties may affect the quality of the collected data, as vehicles operate in diverse environments, driving styles, and with various load conditions. When the model works under conditions that are significantly different from the source-data distribution, label noise can arise from possible data mislabeling. To mitigate the impact of this problem, multi-KL divergence is applied within the system: during the training process, samples with lower multi-KL divergence values are added first, reserving samples that diverge significantly for later stages. This data-labeling approach, incorporating the traffic condition index, ensures that the model receives samples aligned with the natural traffic pattern.

Although the proposed system is capable of performing predictive fault status detection and quickly adapting to environmental changes, some potential failure modes still remain. The first mode is a lack of prediction capability for the time duration just after the sudden change of data distribution due to the unsynchronized adaptive mechanism. The second mode occurs when the model loses adaptation capability in a particular region and fails to retain enough samples to build ambient knowledge, leading to severe performance degradation for model-based adaptive approaches. To avoid this problem, data samples should be controlled at a certain quantity without being excessively low or high.

Equation E. Adaptive retraining trigger

Prediction uncertainty:

$$U(x) = 1 - \max_{c} P(y = c|x)$$

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

Drift score:

$$D_t = \| P_t(X) - P_{t-1}(X) \|$$

Adaptation condition:

$$\text{Retrain} = \begin{cases} 1, & U(x) > \tau_U \text{ or } D_t > \tau_D \\ 0, & \text{otherwise} \end{cases}$$

11. Results

The real-time adaptive machine learning framework is experimentally verified by fault prediction across two vehicle datasets with time-dynamic failure patterns. The first consists of laboratory data for braking failure under varied environment conditions throughout the testing period, which simulates a realistic test bench. The second contains data from virtual exports of the vehicle simulator CARLA, where engine shutdown and power loss events appear randomly without further explanation. Detailed results are provided in Rasul et al. (2023).

The performance of the proposed adaptive prediction mechanism is quantitatively assessed in terms of predictive accuracy and F1-score when the modeling routines execute for every 24 and 36 h of real-time progress within the datasets. Furthermore, individual adaptive responses to dynamic changes in the nature of the failure patterns are examined, including both solution timing and the number of detected conditions. Predictive performance under sudden shifts in weather remains explicitly analyzed in the braking-failure tests, while the behavioral consistency and sensitivity of the adaptive framework are explored through stability tests with increasing noise and uncertainty levels added to the original input features.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

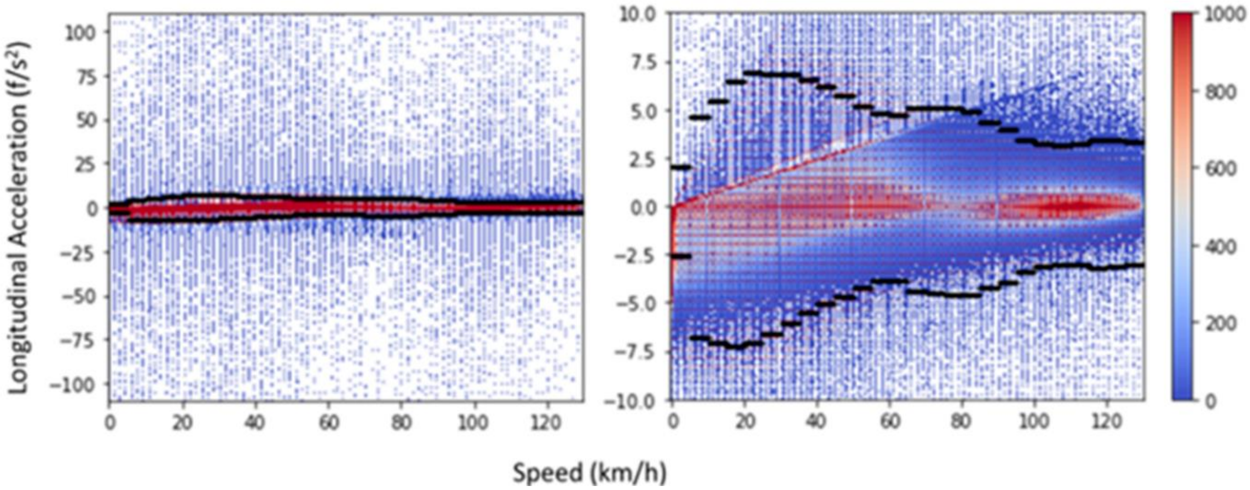


Fig 5: Longitudinal Acceleration

12. Conclusion

The proposed real-time adaptive predictive maintenance framework for vehicular component has achieved its objective. Based on the historical component operation data with failure or fault labels, a time-window-based data labeling method and a multi-window adaptive modelling approach have been developed. An ensemble of SVM classifiers maps the present feature instances and the predictive engine assigns labels for different future time intervals. This enables vehicle maintenance decision making at different future time intervals based on the present state of the component. The framework adapts to failure instances in the incoming data and can deal with less predictive feature labels or even absence of labelled data.

A 4-parameter Boolean function injection methodology has been introduced into the traditional Boolean function to inject uncertainties including defining a function as peeling-off encapsulation, crash or dislodging vehicle failure, misrouting packet failure etc. The experiment at different levels of Boolean and crisp conditions of these parameters clearly elaborates the tradeoff between different factors and the trend matches with a real-life vehicular maintenance scenario in different parts of India under varied uncertainty. This adaptive predictive maintenance framework could provide valuable support to the vehicle maintenance unit of the automobile manufacturing industry and help maintain the component in perfect operational state during its journey.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

References

1. Theissler, A., Pérez-Velázquez, J., Kettelgerdes, M., & Elger, G. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. Reliability Engineering & System Safety, 215, 107864.
2. Alizadeh, M., Hamilton, M., Jones, P., Ma, J., & Jaradat, R. M. (2021). Vehicle operating state anomaly detection and results virtual reality interpretation. Expert Systems with Applications, 177, 114928.
3. Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706.
4. Pan, H., Sun, W., Sun, Q., & Gao, H. (2021). Deep learning based data fusion for sensor fault diagnosis and tolerance in autonomous vehicles. Chinese Journal of Mechanical Engineering, 34, 72.
5. Kaplan, H., Tehrani, K., & Jamshidi, M. (2021). A fault diagnosis design based on deep learning approach for electric vehicle applications. Energies, 14(20), 6599.
6. Jiang, L., Deng, Z., Tang, X., Hu, L., Lin, X., & Hu, X. (2021). Data-driven fault diagnosis and thermal runaway warning for battery packs using real-world vehicle data. Energy, 234, 121266.
7. Mangala, N. (2022). Implementing Databricks Unity Catalog For Centralized Data Governance In Multi-Business-Unitenterprises. Journal of International Crisis and Risk Communication Research, 101-122.
8. Kudelina, K., Vaimann, T., Asad, B., Rassõlkin, A., Kallaste, A., & Demidova, G. (2021). Trends and challenges in intelligent condition monitoring of electrical machines using machine learning. Applied Sciences, 11(6), 2761.
9. Kumar, R., et al. (2021). The importance of feature processing in deep-learning-based condition monitoring of motors. Mathematical Problems in Engineering, 2021, 9927151.
10. Han, T., Zhang, L., Yin, Z., & Tan, A. C. C. (2021). Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. Measurement, 177, 109022.
11. Mattaparthi, R. (2023). Deep Learning-Driven Combustion Anomaly Detection in Diesel Powertrains: A Multi-Sensor Fusion Approach for Real-Time ECM Adaptation. International Journal of Intelligent Systems and Applications in Engineering, 11, 1084.
12. Buchaiah, S., & Shakya, P. (2021). Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. Measurement, 178, 110506.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

13. Leukel, J., González, J., & Riekert, M. (2021). Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. Journal of Manufacturing Systems, 61, 87–96.
14. Gong, C.-S. A., & Su, C.-H. S. (2022). How to implement automotive fault diagnosis using artificial intelligence scheme. Micromachines, 13(9), 1380.
15. Gnanasekaran, S., Jakkamputi, L., Thangamuthu, M., Marikkannan, S. K., Rakkiyannan, J., Thangavelu, K., & Kotha, G. (2022). Condition monitoring of an all-terrain vehicle gear train assembly using deep learning algorithms with vibration signals. Applied Sciences, 12(21), 10917.
16. Kolla, S. H., & Loganathan, R. (2023). Cloud-Native Deep Learning Architectures For Secure Generative AI Deployment In Enterprise Workflow Platforms. Journal of International Crisis and Risk Communication Research, 603-618.
17. Dong, Y., Chen, K., Peng, Y., & Ma, Z. (2022). Comparative study on supervised versus semi-supervised machine learning for anomaly detection of in-vehicle CAN network. In Proceedings of the 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC) (pp. 2914–2919). IEEE.
18. Syed, S. (2023). Shaping The Future Of Large-Scale Vehicle Manufacturing: Planet 2050 Initiatives And The Role Of Predictive Analytics. Nanotechnology Perceptions, 19(3), 103-116.
19. Gao, D., Lin, X., Zheng, X., & Yang, Q. (2023). Novel semi-supervised fault diagnosis method combining tri-training and deep belief network for charging equipment of electric vehicle. International Journal of Automotive Technology, 23, 1727–1737.
20. Jain, M., Vasdev, D., Pal, K., & Sharma, V. (2022). Systematic literature review on predictive maintenance of vehicles and diagnosis of vehicle's health using machine learning techniques. Computational Intelligence, 38(6), 1990–2008.
21. Fernandes, M., Corchado, J. M., & Marreiros, G. (2022). Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: A systematic literature review. Applied Intelligence, 52(12), 14246–14280.
22. Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227-237.
23. Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., Yan, R., & Gryllias, K. (2022). A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios:

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 04
RECEIVED: OCTOBER 02 REVISED: OCTOBER 19 ACCEPTED: NOVEMBER 04 PUBLISHED: NOVEMBER 12

Theories, applications and challenges. Mechanical Systems and Signal Processing, 167, 108487.
24. Drakaki, M., Karnavas, Y. L., Tziafettas, I. A., Linardos, V., & Tzionas, P. (2022). Machine learning and deep learning based methods toward Industry 4.0 predictive maintenance in induction motors: State of the art survey. Journal of Industrial Engineering and Management, 15(1).
25. Mangalampalli, B. M. (2023). AI-Driven Anomaly Detection in Healthcare Claims Data: A Business Intelligence Perspective. Journal of Rare Cardiovascular Diseases.
26. Raja, H. A., Raval, H., Vaimann, T., Kallaste, A., Rassolkin, A., & Belahcen, A. (2022). Cost-efficient real-time condition monitoring and fault diagnostics system for BLDC motor using IoT and machine learning. In P. Trnka (Ed.), Diagnostika 2022: 2022 International Conference on Diagnostics in Electrical Engineering. IEEE.
27. Yang, K., Zhao, L., & Wang, C. (2022). A new intelligent bearing fault diagnosis model based on triplet network and SVM. Scientific Reports, 12, 5234.
28. Gupta, M., Wadhvani, R., & Rasool, A. (2023). A real-time adaptive model for bearing fault classification and remaining useful life estimation using deep neural network. Knowledge-Based Systems, 259, 110070.
29. Vasan, V., Sridharan, N. V., Sreelatha, A. P., & Vaithiyanathan, S. (2023). Tire condition monitoring using transfer learning-based deep neural network approach. Sensors, 23(4), 2177.
30. Wongchai, A., Aoudni, Y., Yesubabu, M., Reegu, F. A., Gowri, N. V., & Vijayakumar, P. (2023). Sustainable electric vehicles fault detection based on monitoring by deep learning architectures in feature extraction and classification. Sustainable Energy Technologies and Assessments, 57, 103178.
31. Reddy, V. A. R. (2022). Designing Fault-Tolerant Data Ingestion Pipelines for High-Volume Healthcare Transactions. Frontiers in Health Informatics, 11, 861-889.
32. Kochukrishnan, P., Rameshkumar, K., & Srihari, S. (2023). Piston slap condition monitoring and fault diagnosis using machine learning approach. SAE International Journal of Engines, 16(7), 923–942.
33. Surucu, O., Gadsden, S. A., & Yawney, J. (2023). Condition monitoring using machine learning: A review of theory, applications, and recent advances. Expert Systems with Applications, 221, 119738.