

# Car Engine Health Monitoring System Using Machine Learning

Akanksha S. Pawar<sup>1</sup>, Gauri N. Rahane<sup>2</sup>, Anuja A. Chate<sup>3</sup>, Dr. Avinash S. Kapse<sup>4</sup>

<sup>1</sup>Computer Science & Engineering, Anuradha College of Engineering & Technology, Chikhli.

<sup>2</sup>Computer Science & Engineering, Anuradha College of Engineering & Technology, Chikhli.

<sup>3</sup>Computer Science & Engineering, Anuradha College of Engineering & Technology, Chikhli.

<sup>4</sup>Computer Science & Engineering, Anuradha College of Engineering & Technology, Chikhli.

\*\*\*

**Abstract** - This review paper explores the advancements in intelligent engine health monitoring systems that integrate machine learning, explainable artificial intelligence (XAI), and predictive maintenance strategies. The study reviews various approaches for analyzing critical engine parameters such as engine RPM, lubrication oil pressure, fuel pressure, coolant pressure, oil temperature, and coolant temperature to assess engine health and detect anomalies.

The paper evaluates the performance of multiple machine learning algorithms, including Random Forest, Gradient Boosting, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression, in classifying engine conditions as healthy or faulty. Special emphasis is given to the use of feature engineering techniques, such as temperature difference and pressure ratio, to enhance model accuracy.

Furthermore, the review highlights the importance of Explainable AI methods, particularly SHAP (SHapley Additive exPlanations), in providing transparency by identifying the contribution of individual features toward fault prediction. This improves trust and interpretability in AI-driven diagnostic systems.

The integration of web-based frameworks like Flask for real-time user interaction and RESTful APIs for mobile application support is also examined. Additionally, the study discusses database-driven systems for storing vehicle-specific parameters and enabling predictive maintenance through automated scheduling and reminders.

The paper concludes that combining machine learning with explainable AI and intelligent safety recommendation systems significantly enhances fault detection accuracy, reduces unexpected engine failures, and supports proactive maintenance. These systems have strong potential for real-world applications in automotive diagnostics, fleet management, and smart transportation systems.

**Key words:** Car Engine Monitoring, Machine Learning, Random Forest, Data Preprocessing, Low-pass Filter, Algorithm Comparison Introduction.

## 1. INTRODUCTION

The automotive industry is undergoing a significant transformation driven by advancements in artificial intelligence, data analytics, and smart technologies. Traditional vehicle maintenance strategies, such as reactive and scheduled maintenance, often lead to unexpected failures, increased operational costs, and reduced system reliability. In contrast, predictive maintenance, powered by machine learning, enables early detection of potential faults and ensures timely intervention, thereby improving efficiency and safety.

Modern vehicles are equipped with numerous sensors that continuously generate data related to engine performance, including parameters such as engine speed (RPM), oil pressure, fuel pressure, coolant pressure, and temperature levels. However, extracting meaningful insights from this data remains a challenge without the application of intelligent analytical models. Machine learning algorithms have emerged as powerful tools for analyzing such complex datasets and identifying hidden patterns associated with system failures.

In recent years, various machine learning techniques, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Gradient Boosting, and Random Forest, have been explored for fault detection and classification tasks. Among these, ensemble methods such as Random Forest have gained prominence due to their robustness, high accuracy, and ability to handle non-linear relationships in data.

Despite the effectiveness of these models, a major limitation is the lack of interpretability, often referred to as the "black-box" problem. In safety-critical domains such as automotive systems, understanding the reasoning behind model predictions is essential. To overcome this

limitation, Explainable Artificial Intelligence (XAI) techniques, such as SHAP (SHapley Additive exPlanations), have been introduced to provide insights into feature contributions and enhance model transparency.

This paper presents a review and implementation of an AI-based engine health monitoring system that integrates machine learning, explainable AI, and web-based deployment. The system not only predicts whether an engine is healthy or faulty but also provides detailed diagnostic insights, safety recommendations, and maintenance actions. Additionally, it incorporates a predictive maintenance scheduler that automatically generates reminders, ensuring proactive system management.

The proposed system is implemented using the Flask web framework, enabling real-time interaction through both web interfaces and mobile APIs. It also supports vehicle-specific configurations, allowing customized threshold values for different engine types. This makes the system adaptable and scalable for real-world applications.

The primary objectives of this study are to:

- Analyze the effectiveness of machine learning models in engine fault detection
- Enhance model interpretability using XAI techniques
- Develop a real-time, user-friendly diagnostic platform
- Enable predictive maintenance through automated scheduling

The remainder of this paper is organized as follows: the next section reviews related work in engine diagnostics and predictive maintenance, followed by methodology, system design, results and discussion, and finally, conclusions and future scope.

### Problem Statement:

In modern automotive systems, engine performance and reliability are critical factors that directly impact vehicle safety, efficiency, and maintenance costs. Traditional maintenance approaches, such as reactive maintenance (repair after failure) and preventive maintenance (scheduled servicing), are often inefficient as they fail to detect hidden faults in real time. These methods can lead to unexpected engine breakdowns, increased downtime, and higher repair expenses.

With the increasing availability of sensor data in vehicles, large volumes of engine performance parameters—such as engine RPM, oil pressure, fuel pressure, coolant pressure, and temperature—are generated continuously. However, without intelligent analysis, this data remains underutilized and fails to provide actionable insights for early fault detection.

Moreover, existing diagnostic systems often lack:

- Real-time fault prediction capabilities
- Interpretability of machine learning models (black-box issue)
- Integration of safety recommendations and maintenance planning
- User-friendly platforms for both web and mobile access

As a result, there is a need for an intelligent, scalable, and interpretable system that can analyze engine data, predict faults accurately, explain the causes of failures, and provide proactive maintenance and safety recommendations.

### Objectives:

- To develop a machine learning-based model
- To enhance prediction accuracy using feature engineering
- To compare multiple machine learning algorithms
- To integrate Explainable Artificial Intelligence (XAI)
- To design a real-time diagnostic system
- To provide safety recommendations and fault diagnostics
- To implement a predictive maintenance scheduler.
- To enable vehicle-specific customization
- To improve overall vehicle reliability and reduce maintenance costs
- Develop ML model for engine fault detection.
- Improve accuracy using feature engineering.
- Use SHAP for explainability.
- Build web/API-based diagnostic system.
- Provide safety and maintenance recommendations.

## 2. LITERATURE SURVEY

Recent research underscores the effectiveness of machine learning and ensemble methods in engine health and fault diagnosis:

Chukwudi et al. proposed an ensemble deep learning model for vehicular engine health prediction, combining Random Forest, Decision Trees, and K-Nearest

Neighbours to achieve robust classification performance [1].

Yang et al. reviewed predictive maintenance in transportation systems, analyzing various ML models applied in railway and vehicle contexts [2].

Chen et al. presented a deep learning framework for semiconductor manufacturing, illustrating how advanced models can improve fault detection in production systems [3].

Zhang et al. used ensemble deep learning for gearbox fault diagnosis in wind turbines, demonstrating transferability of such approaches to mechanical systems [4].

Kumar et al. applied ML to sensor data from power transformers, illustrating how condition monitoring can preempt component failure [5].

Liu et al. surveyed deep learning methods for anomaly detection in industrial processes, highlighting challenges such as data imbalance and interpretability [6].

In engine-specific contexts, Random Forest Classifier has been applied successfully to fault classification and health monitoring. For example, Ricordeau et al. applied Random Forest Classifier to engine health monitoring tasks, showing that it handles complex sensor data and enables interpretability in aviation engine systems [7]. Bastard et al. demonstrated that Random Forest Classifier enhances robustness in engine health management systems [8].

Together, these works validate the use of ensemble methods for mechanical system diagnostics and support the design choices in this project.

Author/Year	Title	Method Used	Findings
R. Sharma (2021)	Prediction for vehicle failures	ANN	Showed high accuracy but required large dataset
P. Singh (2022)	Smart car health monitoring	CNN + Sensor Data	Real-time engine parameter monitoring possible

Gurulakshmi S S (2023)	Engine Health Prediction Using Ensemble Deep Learning	Random Forest, SVM	Achieved good accuracy in engine fault classification
Prashant pillai(2024)	An Ensemble Deep Learning Model for Vehicular Engine Health Prediction	Support Vector Machine (SVM), and Naive Bayes	Early detection of failures, reduced breakdowns
G.Nagappa (2025)	An Ensemble Deep Learning Model for Vehicular Engine Health Prediction	ML Models(SVM, KNN, etc.)	Simple, accurate prediction with comparison of models

**Table no. 1: Literature Survey**

**Related Works**

Several studies have applied machine learning and deep learning techniques to predictive maintenance and fault detection in mechanical systems. Techniques such as SVM, ANN, and Random Forest have shown success in various domains, but limited research exists specifically for car engine health monitoring with multiple algorithm comparisons and signal preprocessing. This work bridges that gap by introducing comprehensive preprocessing along with algorithm evaluation.

[1] N. Balakrishnan et al. In (2021) authors designed a novel approach to aero-engine health monitoring by utilizing the Whale Optimization Algorithm-based Artificial Neural Network, WOANN. The proposed method comprises Analysis of performance data from a military aircraft engine using HUMS data.

[2] Ağbulut et al. (2021) used several deep learning and machine learning methods, such as ANN, to lower CO and NOx emissions

[3] Liu et al. (2022) The research work deals with a micro gas turbine engine with microalgae biofuel blends A20

and jet-A fuel. Performance and Emission characteristics are predicted with the help of LSTM networks.

[4] Matharasi, D. P. (2023) It discusses the use of biodiesel blended with hydrogen obtained from inedible rubber seed oil in an unaltered diesel engine, In the same year to correlate with this the author proposed another research.

[5] Mariani, V. C. et al. (2023) presented an enriched version of a Random Forest model incorporating feature engineering for the prediction of NOx along with CO2 emission by a dual fuel engine, utilized in Compressed Natural Gas blending with diesel fuel is used to enhance the excellent operation of the engine. To prevent engine degradation during its performance.

[6] Wang et al. (2024) stated that the data-driven dynamic health index of the EBA system, as used for diagnostics and predicting its performance, based on machine learning model usages in order to prevent the degradation of the engine. In the same year to improve the performance of the models.

[7] Chukwudi, I. J et al. (2024) applied the deep learning voting ensemble vehicular engine health prediction model which is being suggested for real-time detection of potential problems within the engine and, in that regard, termed proactive maintenance.

### 3. PROPOSED SYSTEM ARCHITECTURE

The proposed system architecture for the AI-based engine health monitoring and predictive maintenance platform is designed to provide real-time diagnostics, explainable insights, and proactive maintenance recommendations. The architecture integrates multiple layers including data acquisition, preprocessing, machine learning, interpretability, and user interaction via web and mobile interfaces.

#### System Components

##### 1.1 Data Acquisition Layer

Sensors: Collect real-time engine parameters such as:

- Engine RPM
- Lubrication Oil Pressure
- Fuel Pressure
- Coolant Pressure
- Oil Temperature
- Coolant Temperature

Vehicle Database: Stores historical vehicle-specific parameters and thresholds for personalized diagnostics.

##### 1.2 Data Processing and Feature Engineering Layer

- Unit Conversion: Converts raw sensor readings into standard units required by the machine learning model (e.g., PSI → Bar for pressures).
- Feature Engineering: Computes derived features such as:
  - Temperature Difference (Oil Temp – Coolant Temp)
  - Pressure Ratio (Fuel Pressure / Oil Pressure)
- Data Scaling: Applies a pre-trained scaler to normalize the input features for consistent model performance.

##### 1.3 Machine Learning and Fault Prediction Layer

- Model Selection: Uses an ensemble-based Random Forest classifier for fault detection, selected based on high accuracy, precision, and recall.
- Prediction: Classifies engine state as Healthy or Faulty.
- Probability Scores: Provides confidence levels for both healthy and faulty predictions.

##### 1.4 Explainable AI Layer

- SHAP Integration: Implements SHAP (SHapley Additive explanations) to interpret model predictions.
- Fault Diagnosis: Identifies critical engine parameters contributing to anomalies.
- Fault Registry Mapping: Maps the critical parameters to predefined safety warnings, recommended actions, and maintenance guidelines.

##### 1.5 Safety Recommendation & Predictive Maintenance Layer

- Safety Recommendations: Generates actionable guidance such as:
  - Precautionary measures
  - Driving limits
  - Emergency instructions
- Predictive Maintenance Scheduler: Automatically generates maintenance reminders based on predicted faults.
- Database Storage: Stores reminders, status, and historical diagnostic reports.

##### 1.6 User Interaction Layer

- Web Interface (Flask): Provides real-time data input, prediction results, fault diagnostics, and maintenance tasks.

- REST API: Supports mobile integration for remote monitoring and control.

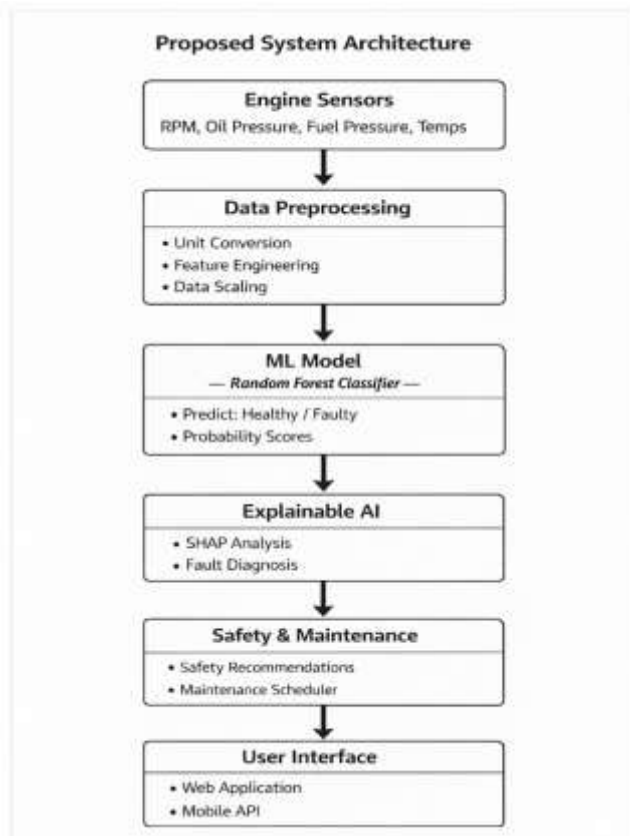


Fig.1 Proposed System Architecture Diagram

## 4. METHODOLOGY

### A. Overview

The proposed system presents an intelligent engine health monitoring and predictive maintenance framework that integrates machine learning, explainable artificial intelligence, and web-based technologies. The primary objective of the system is to detect engine faults, analyze their root causes, and provide actionable safety and maintenance recommendations in real time.

The architecture follows a layered design approach, ensuring modularity, scalability, and efficient data processing. It consists of four primary layers: the Presentation Layer, Application Layer, Machine Learning Layer, and Data Layer. Additionally, Explainable AI and Predictive Maintenance modules are incorporated to enhance system transparency and usability.

### B. Architectural Layers

#### 1) Presentation Layer

The Presentation Layer serves as the user interaction interface of the system. It is responsible for collecting input data and displaying the results in an intuitive format.

This layer is implemented using HTML, CSS, and JavaScript, integrated with Flask templates for dynamic content rendering. It provides functionalities such as user authentication, input forms for engine parameters, visualization of prediction results, and access to maintenance scheduling features. Users input key engine parameters, including engine speed (RPM), oil pressure, fuel pressure, coolant pressure, oil temperature, and coolant temperature. The interface ensures ease of use and supports real-time interaction with the backend system.

#### 2) Application Layer

The Application Layer acts as the **core processing unit** of the system and is implemented using the Flask web framework. It manages communication between the frontend, machine learning components, and the database. The key responsibilities of this layer include request handling, session management, input validation, and execution of business logic. It also exposes RESTful APIs to enable integration with external systems such as mobile applications.

This layer consists of several functional modules:

- **Authentication Module:** Handles user registration and login using secure password hashing mechanisms.
- **Prediction Module:** Processes user inputs, performs preprocessing, and communicates with the machine learning model to generate predictions.
- **Vehicle Management Module:** Stores and retrieves vehicle-specific operational thresholds to improve prediction accuracy.
- **Maintenance Scheduler Module:** Manages maintenance tasks, including creation, updating, and deletion of reminders.
- **Service Locator Module:** Assists users in identifying nearby service facilities.

#### 3) Machine Learning Layer

The Machine Learning Layer forms the analytical backbone of the system. It is responsible for predicting the health status of the engine based on input parameters. Multiple machine learning models are evaluated, including Random Forest, Gradient Boosting, Support Vector Machines, K-Nearest Neighbors, and Logistic Regression. Among these, the Random Forest classifier is selected due to its superior accuracy, robustness, and ability to handle nonlinear relationships.

The data processing pipeline includes:

- **Data Preprocessing:** Conversion of input units (e.g., pressure from PSI to Bar) and handling of inconsistencies.
- **Feature Engineering:** Derivation of additional features such as temperature difference and pressure ratio to enhance predictive performance.
- **Feature Scaling:** Standardization of features using scaling techniques to ensure uniformity.
- **Prediction:** Classification of engine condition into healthy or faulty categories along with probability scores.

#### 4) Explainable AI Module

To enhance interpretability, the system incorporates an Explainable AI module based on SHAP (SHapley Additive Explanations). This module identifies the contribution of each input feature to the final prediction and provides a ranked list of parameters influencing the decision. It enables users to understand the root cause of faults, thereby increasing trust in the system.

#### 5) Fault Diagnosis and Safety Recommendation Module

The system includes a rule-based fault diagnosis module that maps abnormal parameter values to predefined safety guidelines. A fault registry is maintained containing information such as safe operating ranges, risk descriptions, precautionary measures, driving limitations, and maintenance recommendations. Based on the identified faults, the system generates actionable insights to assist users in making informed decisions.

#### 6) Predictive Maintenance Module

The Predictive Maintenance Module provides proactive maintenance support by automatically generating service reminders when a fault is detected. The system schedules maintenance tasks based on the severity of detected issues and stores them in the database. Users can view, update, and manage these reminders through the interface, thereby reducing the risk of unexpected failures.

#### 7) Data Layer

The Data Layer is responsible for persistent data storage and retrieval. It is implemented using an SQLite database.

The database consists of the following tables:

- **Users Table:** Stores authentication credentials.
- **Vehicles Table:** Contains vehicle-specific operational parameters.

- **Reminders Table:** Maintains maintenance schedules and task statuses.

This layer ensures data integrity, efficient querying, and reliable storage of system information.

#### C. System Workflow

The overall workflow of the system is as follows:

1. The user inputs engine parameters through the interface or API.
2. The Application Layer validates and preprocesses the input data.
3. The processed data is passed to the Machine Learning Layer.
4. The model predicts the engine condition and generates probability scores.
5. The Explainable AI module analyzes feature contributions.
6. The Fault Diagnosis module generates safety recommendations.
7. The Predictive Maintenance module schedules necessary maintenance tasks.
8. The results are displayed to the user through the Presentation Layer.

#### D. Deployment Architecture

The system can be deployed on both local and cloud environments. The backend server is implemented using Flask, while the trained machine learning model is stored as serialized files. The architecture supports integration with mobile applications and IoT devices for real-time data acquisition.

#### E. Advantages of the Proposed Architecture

The proposed architecture offers several advantages:

- High accuracy in fault detection using machine learning
- Real-time diagnostics and decision support
- Enhanced transparency through Explainable AI
- Proactive maintenance scheduling
- Modular and scalable system design
- Compatibility with web and mobile platforms

#### 5. CONCLUSION

This paper presented a comprehensive review and design of an intelligent engine health monitoring and predictive maintenance system based on machine learning techniques. The proposed framework integrates data pre-processing, feature engineering, model training, and

explainable artificial intelligence to deliver accurate and reliable engine fault diagnosis.

The use of advanced machine learning algorithms, particularly the Random Forest classifier, enables the system to achieve high prediction accuracy and robustness in identifying engine conditions. Furthermore, the incorporation of an Explainable AI module enhances transparency by providing clear insights into the factors influencing the model's decisions, thereby improving user trust and system interpretability.

In addition to fault detection, the system extends its functionality by incorporating a fault diagnosis and safety recommendation module, which translates model outputs into actionable guidance for users. The integration of a predictive maintenance module further strengthens the system by enabling proactive maintenance scheduling, reducing the likelihood of unexpected failures and minimizing operational downtime.

The modular and layered architecture of the system ensures scalability, flexibility, and ease of integration with modern technologies such as mobile applications and IoT-based sensor networks. This makes the proposed approach suitable for real-world deployment in automotive and industrial domains.

Overall, the proposed system demonstrates the potential of combining machine learning and explainable AI for developing **smart, reliable, and user-centric diagnostic solutions**. Future work may focus on incorporating real-time sensor data, deploying the system on cloud platforms, and exploring deep learning models to further enhance performance and scalability.

### Acknowledgement

I sincerely thank all who contributed to this project's success. I am especially grateful to my guide, Dr. A. S. Kapse, for his guidance, support, and encouragement. I also thank the faculty of Anuradha College of Engineering and Technology for their assistance and resources.

### REFERENCES

- [1.] I. J. Chukwudi, N. Zaman, M. A. Rahim, M. A. Rahman, M. J. Alenazi, and P. Pillai, "An Ensemble Deep Learning Model for Vehicular Engine Health Prediction," *IEEE Access*, 2024.
- [2.] Z. Yang, L. Wang, and X. Liu, "Predictive Maintenance for Railway Systems Using Machine Learning Algorithms: A Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 3, pp. 1276–1287, Mar. 2023.

- [3.] H. Chen, K. Wang, and S. Li, "Deep Learning-Based Predictive Maintenance Framework for Semiconductor Manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 37, no. 1, pp. 78–87, Feb. 2023.

- [4.] Y. Zhang, J. Li, and Q. Zhao, "An Ensemble Deep Learning Approach for Fault Diagnosis of Wind Turbine Gearboxes," *IEEE Trans. Ind. Electron.*, vol. 69, no. 5, pp. 4143–4152, May 2022.

- [5.] Kumar, S. Singh, and R. Gupta, "Machine Learning-Based Predictive Maintenance for Power Transformers Using Sensor Data," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1056–1065, Mar. 2023.

- [6.] X. Liu, Y. Chen, and Z. Wang, "Deep Learning for Anomaly Detection in Complex Industrial Processes: A Survey," *IEEE Trans. Ind. Inform.*, vol. 18, no. 1, pp. 13–25, Jan. 2024.

- [7.] J. Ricordeau and J. Lacaille, "Application of Random Forests to Engine Health Monitoring," in *ICAS Conference*, 2010.

- [8.] G. Bastard et al., "Engine Health Management in Safran Aircraft Engines," PHM Society Conf., 2016.

- [9.] G. Nagappa et al., "An Ensemble Deep Learning Model for Vehicular Engine Health Prediction," *IJESI*, Vol. 14, Issue 4, April 2025