IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

An Ensemble Machine Learning Model for Vehicular Engine Health Prediction

V. Chandra Sekhar Reddy[1], B. Sree Kavya Sudha[2] , CH. Deepikamani [3], B. Paramesh[4] , R. Sahithi[5]
Associate Professor, Computer Science and Engineering[1]
Student, Computer Science and Engineering[2,3,4,5]
ACE Engineering College, Ghatkesar, India

Abstract: *In the automotive industry, maintaining optimal engine health is vital for vehicle performance and longevity. Traditional diagnostics often lack real-time capabilities, creating the need for predictive systems. This project proposes an ensemble deep learning approach combining Random Forest, Decision Tree, and K-Nearest Neighbors (KNN) algorithms. Random Forest achieved 84% accuracy, leveraging its strength in managing complex data. Decision Tree offered 76% accuracy with strong interpretability, while KNN reached 80%, using instance-based classification. The ensemble model enhances overall accuracy and robustness, supporting real-time engine monitoring and proactive maintenance. This intelligent framework offers a scalable solution for predictive automotive diagnostics and health management. By combining these individual models into an ensemble framework, the system significantly improves prediction accuracy, robustness, and overall reliability. The ensemble method addresses the limitations of each algorithm while leveraging their strengths to produce more consistent and precise outcomes. The model is trained on a comprehensive dataset containing engine parameters such as temperature, pressure, vibration, fuel usage, and emissions. As a result, it facilitates early anomaly detection, supports proactive maintenance, and reduces repair costs and vehicle downtime. This ensemble deep learning solution offers scalability and intelligence for next-generation automotive health diagnostics*

Keywords: Engine Health Monitoring, Ensemble Learning, Random Forest, Decision Tree, KNN, Predictive Maintenance, Automotive Diagnostics, Machine Learning.

I. INTRODUCTION

In the age of rapid technological advancement, agriculture is undergoing a significant transformation through the integration of Artificial Intelligence (AI) and Machine Learning (ML). These technologies are being harnessed to address critical aspects such as soil fertility prediction, crop recommendation, and plant disease detection each playing a pivotal role in increasing agricultural productivity and sustainability. This transition is especially vital for a country like India, where agriculture remains one of the primary economic pillars. With over two- thirds of the population directly or indirectly dependent on agriculture and the sector contributing nearly 20% to the national GDP, the wellbeing of farmers is intrinsically linked to the nation's development. This document is a template. An electronic copy can be downloaded from the website. For questions on paper guidelines, please contact the International Journal committee as indicated on the Journal website. Information about final paper submission is available from the website Despite advancements, maintaining optimal engine health remains a major challenge. Engine parameters such as RPM, lubricating oil pressure, fuel pressure, coolant pressure, and various temperature readings must be continuously monitored to detect anomalies. Traditional reactive maintenance methods often result in unexpected breakdowns and high repair costs. Additionally, real-time analysis of these multidimensional datasets is difficult without automated systems. Smaller fleet operators and individual vehicle owners often lack access to sophisticated diagnostic tools, making early issue detection even more difficult. As with other industries, the lack of accessible, intelligent systems limits the ability to transition from reactive to predictive maintenance. Traditional reactive maintenance methods often

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 5, June 2025 Impact Factor: 7.67

result in unexpected breakdowns and high repair costs. Additionally, real- time analysis of these multidimensional datasets is difficult without automated systems. This project proposes a machine learning-based engine health prediction system, integrating an ensemble model that analyzes real-time sensor data to assess engine condition. By leveraging a combination of Random Forest, Decision Tree, and K-Nearest Neighbors (KNN) algorithms, the model captures complex, nonlinear interactions between parameters. This hybrid model improves accuracy and generalization, offering reliable insights into engine performance. It supports proactive maintenance planning and early anomaly detection, significantly reducing operational risks and unplanned downtime. The engine health model is trained using a dataset comprising historical engine condition labels and corresponding sensor measurements. These include continuous signals from various subsystems, such as fuel and lubricant pressures, coolant temperatures, and engine speed (RPM).Through feature engineering and preprocessing, the data is structured for efficient learning and optimized prediction. The ensemble approach allows the model to detect subtle variations and patterns, even in noisy datasets, enabling it to diagnose performance degradation or emerging faults with high accuracy.Moreover, this work aims to build a web-based interface for real-time monitoring, allowing integration with existing vehicle diagnostics systems. Fleet managers and vehicle owners can access condition predictions, maintenance alerts, and historical trends through an intuitive dashboard. The application of AI in automotive maintenance also raises concerns around data privacy, scalability, and ethical implementation. While predictive models offer improved reliability, misuse or overdependence on technology without proper oversight can lead to misdiagnoses or neglect of mechanical evaluations.Therefore, the project emphasizes user-friendly design, responsible model deployment, and cost-effective scalability to ensure widespread adoption across commercial and private sectors alike. Therefore, the project prioritizes an intuitive and accessible design to accommodate users with varying technical expertise. It ensures responsible and transparent deployment of predictive models, maintaining ethical standards and reliability. Additionally, the solution is developed with cost-effective scalability in mind, making it feasible for both individual vehicle owners and commercial fleet operators. By focusing on affordability, usability, and secure implementation, the project aims to drive widespread adoption across the automotive sector, supporting a shift towards proactive maintenance practices.

II. LITERATURE REVIEW

A literature survey on ensemble deep learning for vehicular engine health prediction examines predictive maintenance using neural networks, logistic regression, and reinforcement learning. Ensemble models enhance fault detection and remaining useful life estimation. Recent studies emphasize real-time monitoring, multi-model fusion, and interpretable AI, aiming to improve decision-making and accuracy in predictive maintenance systems, ensuring reliability and efficiency in vehicular engine performance management.

Isinka et al. [1] propose an ensemble deep learning model tailored for real-time vehicular engine health monitoring. Their framework emphasizes the detection of anomalies and performance degradation, enabling proactive maintenance and reducing system downtime. Similarly, Sruthi et al. [2] explore ensemble deep learning techniques, comparing multiple models across various datasets to evaluate prediction precision and robustness in engine health diagnostics.Jammal and Srour [3] employ machine learning techniques— specifically Multi-Layer Perceptron (MLP) and Logistic Regression—to develop a predictive model for engine failure. Their approach integrates data preprocessing, feature engineering, and validation to optimize maintenance scheduling and minimize failure risks.Mun et al. [4] focus on military ground vehicles and implement an ensemble learning-based predictive model using multivariate Long Short-Term Memory (LSTM) networks. Their model leverages real-time operational data to monitor component health, outperforming traditional LSTM approaches in accuracy and reliability classification. The application of edge computing for real-time vehicle fault detection is introduced by Vasavi et al. [5]. Their system uses the AK-NN algorithm to process sensor data from both internal and external sources, providing predictive analytics as a service. This architecture supports immediate fault detection and reduces latency in vehicle diagnostics. Rahim et al. [6] present a hybrid deep learning-based Vehicular Engine Health Monitoring System (VEHMS) that combines Convolutional Neural Networks (CNNs) and Bi-Directional Gated Recurrent Units (BiGRUs). The model classifies engine health status into categories such as good, critical, moderate, and minor by integrating real-time sensor data with infrastructure vulnerability assessments. In a related work, Rahim, Rahman, and Razzak [7] design an intelligent risk

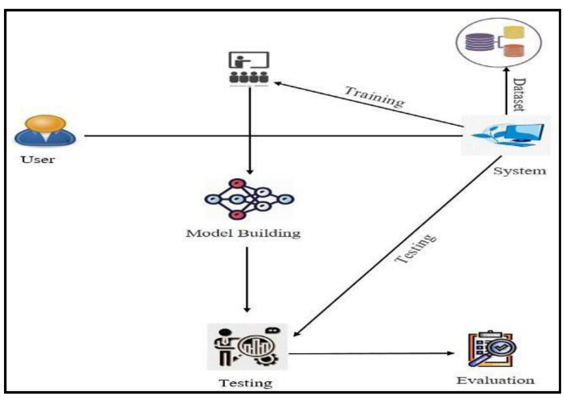
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 5, June 2025 Impact Factor: 7.67

management framework to monitor vehicular engine health. Their system utilizes sensor-actuator data and vulnerability identification to classify engine conditions, applying machine and deep learning algorithms to enhance decision-making in maintenance planning. Rahman and Rahman [8] propose a secure and intelligent big data analytics framework for vehicle health monitoring. Their approach leverages the Internet of Everything (IoE) to connect stakeholders, data, and vehicle systems, addressing the lack of smart features in current vehicles. The framework targets increased safety, better maintenance, and more efficient transport systems. Altogether, these studies highlight the transformative potential of AI-powered systems in vehicular engine health monitoring. By leveraging deep learning, ensemble models, edge computing, and big data analytics, these systems enhance prediction accuracy, reduce maintenance costs, and improve overall vehicle performance and safety.

Altogether, these studies emphasize the transformative impact of AI-powered systems in vehicular engine health monitoring. By integrating deep learning, ensemble techniques, edge computing, and big data analytics, these models provide real-time diagnostics, early fault detection, and proactive maintenance strategies. Such innovations not only improve the accuracy of engine condition predictions but also help reduce maintenance costs and prevent unexpected failures. sustainable transportation management.

III. PROPOSED WORK

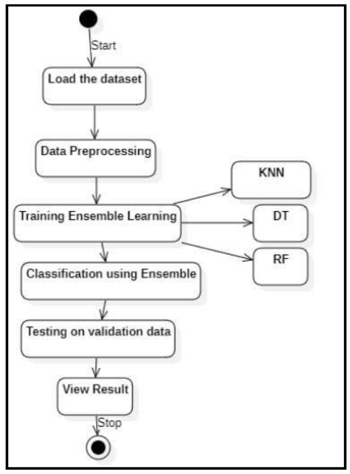
The architecture of the proposed system leverages ensemble deep learning techniques to support advanced predictive maintenance in the automotive industry. It integrates multiple machine learning algorithms Random Forest, Decision Tree, and K-Nearest Neighbors (KNN)—to enhance engine health prediction accuracy. The system begins by collecting input data such as engine RPM, lubricating oil pressure and temperature, fuel pressure, coolant pressure, and coolant temperature. This data is transmitted to the backend, which interfaces with a centralized repository containing historical performance records and training datasets.



The collected data is processed through stages of model training, validation, and evaluation. Trained ensemble models are applied to detect anomalies and predict potential failures. The final output, indicating the engine's health status or recommended maintenance actions, is displayed to the user in an intuitive format. By continuously learning from both historical and real-time data, the architecture enables a proactive and adaptive maintenance strategy. This structured flow—from data acquisition to intelligent diagnostics—ensures reduced downtime, cost-effective repairs, and extended engine lifespan, ultimately improving reliability in vehicle maintenance operations.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

Workflow of Proposed System :



The flowchart illustrates the step-by-step process of an ensemble- based predictive maintenance system. It begins with loading the dataset, followed by data preprocessing to clean and prepare the input. The system then trains an ensemble learning model using three algorithms: K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF). These models work collectively to improve prediction accuracy. After training, classification is performed using the ensemble model. The system then tests the model on validation data to evaluate performance. Finally, the results are presented to the user, completing the process with a comprehensive analysis of the engine's health status. The flowchart presents the process flow of a machine learning- based ensemble system for engine health prediction. It begins with loading the dataset, which includes critical engine parameters such as RPM, temperature, and pressure readings. The data then undergoes preprocessing to handle missing values, normalize scales, and prepare it for analysis. Next, ensemble learning is employed by training three different classifiers K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF)—on the processed data. These models are combined to create a more robust predictive system, leveraging the strengths of each algorithm. Finally, the prediction results are displayed to the user, supporting informed decision-making for proactive vehicle maintenance.

User Module

The **User Module** provides an interactive front-end interface that facilitates smooth user interaction with the system. Its key features include:

- **Index Page**: Landing page that introduces the application and provides navigation to login or registration.
- **Register Page**: Allows new users to create an account by submitting their details.
- **Login Page**: Authenticates users and grants access to the system.
- **User Home Page**: Personalized dashboard displaying tools and content based on user roles.
- **Prediction Page**: Interface for inputting data to receive engine health predictions.
- **Graph Page**: Visualizes system performance and prediction trends using charts.
- **Logout**: Ends the current session and redirects the user to the homepage.

System Module

The System Module is responsible for processing sensor data from vehicles to predict engine health and ensure effective maintenance.

It includes the following submodules:

Data Collection and Preprocessing Submodule :

- **Data Collection**: Gathers sensor data such as engine RPM, oil pressure, fuel pressure, coolant pressure, oil temperature, and coolant temperature.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

- **Preprocessing**: Cleans and normalizes the data, handles missing values, and applies feature scaling to prepare it for model input.

Ensemble Model Overview Submodule :

- **Model Integration**: Introduces the ensemble of Random Forest, Decision Tree, and KNN models.
- **Ensembling Logic**: Combines predictions from multiple models to improve robustness and accuracy.

Model Training Submodule :

- **Dataset Split**: Divides the collected data into training and validation sets.
- **Training**: Trains the ensemble models (Random Forest, Decision Tree, KNN) using the preprocessed sensor data.
- **Hyperparameter Tuning**: Optimizes model performance by adjusting key parameters for each algorithm to achieve higher accuracy.

Prediction Submodule :

- **Data Input**: Accepts real-time sensor data for analysis.
- **Prediction Generation**: Applies the trained ensemble models to classify engine health status.

Performance Evaluation Submodule :

- **Metrics Calculation**: Computes accuracy, precision, recall, and F1-score to evaluate model performance.
- **Model Validation**: Compares model outputs against known validation data to assess effectiveness.

Model-Specific Functionality Submodule :

- **Random Forest**: Uses multiple decision trees to perform majority-vote-based classification.
- **Decision Tree**: Makes predictions using a single tree structure for interpretability.
- **K-Nearest Neighbors (KNN)**: Classifies engine condition by comparing input to the nearest training samples.

IV. METHODOLOGY

This section outlines the machine learning models used for engine health prediction, Ensemble Learning for Enhanced Accuracy, and proactive maintenance in smart automotive systems. Random Forest, Decision Tree, and K-Nearest Neighbors analyze engine sensor data to enhance diagnostic accuracy and support efficient, data-driven vehicle maintenance decisions.

Engine Health Prediction

Models analyzed include Random Forest, Decision Tree, and K- Nearest Neighbors (KNN). These models were selected for their complementary strengths in classification and regression tasks to improve predictive maintenance in the automotive industry.

Random Forest (RF):

An ensemble learning method that builds multiple decision trees during training. It aggregates the predictions of individual trees using majority voting, leading to improved accuracy (84%) and reduced overfitting. It is effective at capturing complex interactions in engine sensor data such as RPM, oil pressure, and coolant temperature.

Decision Tree (DT):

A supervised learning algorithm that splits data based on feature values to create interpretable rules. With an accuracy of 76%, it helps identify nonlinear patterns and is valuable for understanding decision paths in engine diagnostics.

K-Nearest Neighbors (KNN):
An instance-based learning method that classifies engine health conditions based on the similarity to neighboring data points. It achieved an 80% accuracy and complements other models by making predictions based on real-time proximity in feature space.

Machine Learning Models for Engine Health Prediction
The proposed system utilizes three supervised machine learning algorithms—Random Forest, Decision Tree, and K-Nearest Neighbors (KNN)—to predict engine health by analyzing sensor data such as RPM, oil pressure, fuel pressure, and coolant temperature.
Random Forest:
Captures complex relationships in engine sensor data and reduces overfitting by aggregating predictions from multiple decision trees.
Decision Tree:
Offers interpretable rules and identifies nonlinear patterns in parameters like oil pressure, RPM, and coolant temperature.
K-Nearest Neighbors:
Uses instance-based learning to classify engine health based on similarity to nearby data points in the feature space.
Ensemble Advantage:
Combining all three models improves prediction accuracy, enhances reliability, and ensures robust engine health diagnostics for proactive automotive maintenance.

Proactive Maintenance and Operational Efficiency
The proposed system enables real-time monitoring and predictive maintenance by effectively learning from both historical and live engine sensor data. It continuously analyzes key parameters such as RPM, oil pressure, fuel pressure, and temperature readings to detect anomalies and predict potential failures.

V. CONCLUSION AND FUTURE WORK

This project presents a robust ensemble-based machine learning system that integrates Random Forest, Decision Tree, and K-Nearest Neighbors (KNN) algorithms to predict engine health. By analyzing critical parameters such as RPM, oil pressure, coolant temperature, and fuel pressure, the system enables early detection of faults and supports proactive maintenance strategies. The user-friendly interface and real-time monitoring capabilities ensure efficient interaction and high accuracy, helping reduce repair costs, extend engine lifespan, and enhance overall vehicle reliability.Future enhancements aim to incorporate advanced deep learning models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) to better capture temporal and spatial patterns in engine performance data. The integration of IoT-based real-time data collection will facilitate continuous monitoring and enable remote diagnostics. The system's functionality can be expanded to classify specific engine faults, thus improving its diagnostic precision. Additionally, implementing online learning techniques will allow the model to adapt dynamically to new patterns without requiring complete retraining. To ensure broader accessibility and ease of use, especially for large-scale fleet operators, the solution will be deployed as a scalable cloud-based or mobile application. These enhancements aim to transform the system into a comprehensive smart automotive maintenance platform. Furthermore, incorporating predictive analytics and anomaly detection techniques will help in forecasting potential issues before they escalate, ensuring minimal downtime. Integration with vehicle telematics systems can offer holistic insights into driving behavior and environmental impact on engine performance. Collaboration with automotive manufacturers and service providers could lead to customized solutions tailored for different vehicle types. Ensuring cybersecurity and data privacy in connected environments will also be a critical area of focus. Continuous updates based on user feedback and real-world data will refine the system's accuracy, making it a reliable and intelligent solution for next-generation automotive maintenance and engine health management.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

REFERENCES

[1]. Isinka Joseph Chukwudi, "An Ensemble Deep Learning Model for Vehicular Engine HealthPrediction",IEEE,2024.

[2]. Geetha Priya and S. Gurulakshmi, “Engine Health Prediction Using Ensemble Deep Learning”,ICOFE,2024

[3]. Hussein Jammal and Farah Srour, “Prediction of Automotive Vehicles Engine Health Using MLP andLR”,ICR,2024.

[4]. Md. Arafatur Rahman and Md. Mustafizur Rahman, “A Secure and Intelligent Framework for Vehicle Health Monitoring Exploiting Big-Data Analytics”, IEEE, 2022.

[5]. Sungkwang Mun and Joonsik Hwang, “Ensemble Learning for Fault Condition Prediction and Health Status Monitoring in Military Ground Vehicles",IISE,2023.

[6]. S. Vasavi and K. Aswarth, “Predictive analytics as a service for vehicle health monitoring using edge computing and AK-NN algorithm”, ScienceDirect, 2021.

[7]. Md. Abdur Rahim and Md. Mustafizur Rahman, “Deep learning-based vehicular engine health monitoring system utilising a hybrid convolutional neural network”, ScienceDirect, 2024.

[8]. Md. Abdur Rahim and Md. Arafatur Rahman and Imran Razzak “An Intelligent Risk Management Framework for Monitoring Vehicular Engine Health”, IEEE, 2022.