Integrating Machine Learning and Real-Time Analytics for Engine Health Forecasting

Durga Sai Sriram Suggu  
Department of Computer Science and Engineering,Kalasalingam Academy of Research and Education,

Krishnakoil-626126.  
Tamil Nadu, India. saisriramsuggu@gmail.com

Sandeep Ravilla  
Department of Computer Science and Engineering,  
Kalasalingam Academy of Research and Education,

Krishnankoil-626126.  
Tamil Nadu, India.

sandeepravilla456@gmail.com

Amrutha Thottempudi  
Department of Computer Science and Engineering,

Kalasalingam Academy of Research

and Education*,*

Krishnankoil-626126

Tamil Nadu, India

amrutha.thottempudi@gmail.com

Pandiselvam Pandiyarajan  
Department of Computer Science and Engineering,  
Kalasalingam Academy of Research and Education,

Krishnankoil-626126.  
Tamil Nadu, India.

p.pandiselvam@klu.ac.in

Sk Tabasum Jaha

Department of Computer Science and Engineering,

Kalasalingam Academy of Research and Education,  
Krishnankoil-626126

Tamil Nadu, India.

sktabasumjaha@gmail.com

*Abstract*— In the domain of vehicle health management, predictive maintenance has emerged as a key innovation driven by improved machine learning and real-time data analysis capabilities. This research aims to create a system that uses live vehicle sensor data, like RPM and oil pressure/ temperature to monitor engine conditions. The predictive algorithm employs Gradient Boosting to infer the risk of failures, specify their probabilities, and forecast their maintenance weeks in advance. It is buttressed into a scalable and user-friendly application using Streamlit designed for usage by fleet operators or maintenance professionals. This solution focuses on predictive maintenance rather than reactive, which maximizes vehicle uptime, decreases Opex by 10-40%, and extends asset life span. It is tailored for use in multiple types of vehicles and allows rapid integration of IoT technologies that enable continuous monitoring. The challenges faced by the industry for data variability, and compliance with various security standards like GDPR and ISO/SAE 21434 are tackled through this project. The suggested framework and its participatory implementation transform fleet administration through the enhancement of resource distribution, the assurance of prompt maintenance, and the reduction of interruptions.

Keywords— Predictive Maintenance, Engine Health Monitoring, Real-Time Data Analytics, Machine Learning, Fleet Management, Vehicle Diagnostics, Performance optimization.

# Introduction

Predicting engine conditions is essential to the development of contemporary car health monitoring systems and serves as the basis for more intelligent, data-driven maintenance plans. By using real-time sensor data, predictive maintenance can identify irregularities and possible problems before they become serious failures. This method greatly improves the dependability and lifespan of automobiles while also lowering unplanned malfunctions and expensive repairs.

The focus of this study is on creating an intuitive web-based application that incorporates machine learning models to accurately predict engine conditions. In order to provide precise insights into the condition of the engine, the system analyzes a variety of sensor inputs, including engine RPM, lubricating oil pressure, fuel pressure, coolant pressure, and temperature fluctuations. The incorporation of sophisticated algorithms guarantees reliable forecasts, and the user-friendly interface makes analysis easier for maintenance experts.

A proactive approach to maintenance that guarantees operational reliability and cost effectiveness is made possible by intelligent vehicle management systems that accurately predict engine conditions. This study shows that sophisticated machine learning models, like Gradient Boosting Classifiers, can be used to efficiently analyze complex sensor data. To give useful information about engine health, important parameters like engine RPM, oil pressure, coolant levels, and temperature differentials are processed. In addition to making analysis easier, the interactive web-based application helps close the gap between unprocessed sensor data and insightful maintenance choices. With features like confidence scoring, real-time monitoring, and an intuitive interface, the system gives maintenance workers the resources they need to deal with problems quickly. This invention demonstrates how predictive maintenance can revolutionize fleet operations across multiple industries by decreasing downtime, increasing vehicle life, and improving fleet performance.

This solution improves overall resource utilization while giving stakeholders the tools they need to make informed maintenance decisions by fusing real-time data visualization with predictive analytics. The system's ability to process and display vital engine health data in an understandable way guarantees prompt detection of possible problems and permits preventative measures that lower the possibility of unplanned failures. This method minimizes downtime and maximizes productivity while optimizing resource allocation and greatly enhancing fleet-wide operational efficiency. According to the research, such sophisticated systems have the capacity to completely rethink conventional maintenance procedures by putting more emphasis on proactive management rather than reactive repairs. Through the use of state-of-the-art technology, this solution opens the door for affordable, dependable, and sustainable vehicle maintenance practices, offering significant advantages to a variety of sectors, including manufacturing, public services, logistics, and transportation.

# Related Works

S. Vasavi et al [1] In this paper, a fault prediction system based on edge computing is presented that uses both external and internal sensors to make real-time vehicle health predictions. Both a dashboard at the terminal and notifications via a mobile application are used to display risk information. A system like this lowers the latency between processing and sending vehicle data. The AK-NN ensemble of ANN and k-NN classifiers is used in the suggested system to enhance prediction performance. Shafi et al [2] In this paper outlines a method for predicting faults in the vehicle's four primary subsystems: the cooling system, exhaust system, ignition system, and fuel system. The strategy, which aims to increase vehicle uptime, was tested on seventy Toyota Corolla models. Shahbazi et al [3] This study introduces a hybrid framework designed to predict the condition of lithium-ion batteries used in electric vehicles (EVs), in response to the global transition towards sustainable energy. As EVs move away from fossil fuel reliance and towards battery power, the XGBoost model is employed to assess the state of health (SOH) of the batteries, owing to its exceptional predictive accuracy and generalization abilities. The experimental findings indicate that the model attains a high level of accuracy in forecasting SOH. Hari Pranesh et al [4] The research describes a machine learning approach to hydraulic brake system fault diagnosis. For the study, a real car brake system was taken into consideration. Using a piezoelectric accelerometer, the brake system's vibration signals were recorded under various fault scenarios. Shivansh Khurana et al [5] It is mandatory for all automakers to measure these engine emissions. This is typically accomplished by repeatedly testing the car, which is not a cost-effective method. The installation of costly test rigs is typically required for these procedures. However, predictive modeling can be used as a digital or virtual tool to test emissions accurately. The article has attempted to highlight a new trend in sophisticated machine learning algorithms that can be used to generate emission data in an easy-to-understand way. Al-refai et al [6] article propose a low-cost machine learning system that employs in-vehicle data to address three classification issues: driving style, road surface conditions, and traffic situations.

# Methodology

The system uses a supervised machine learning approach, where a model is trained based on historical engine data to classify whether the engine condition is normal or requires further investigation. The application is implemented using Streamlit, a Python-based web framework, to provide real-time interaction with the model. Input functions, such as motor speed, lubrication oil pressure, fuel pressure, cooling liquid pressure, oil temperature, and differences in cooling and temperature liquid, are standardized according to data set statistics to ensure reliable forecasts. These inputs are transformed to a pre-formed gradient model, loaded via a serialized picked-up file. The prediction function uses the model's predict and predict\_proba methods to determine the engine's state and confidence level. User-friendly sliders allow for real-time adjustment of feature values, providing an interactive experience for end users.

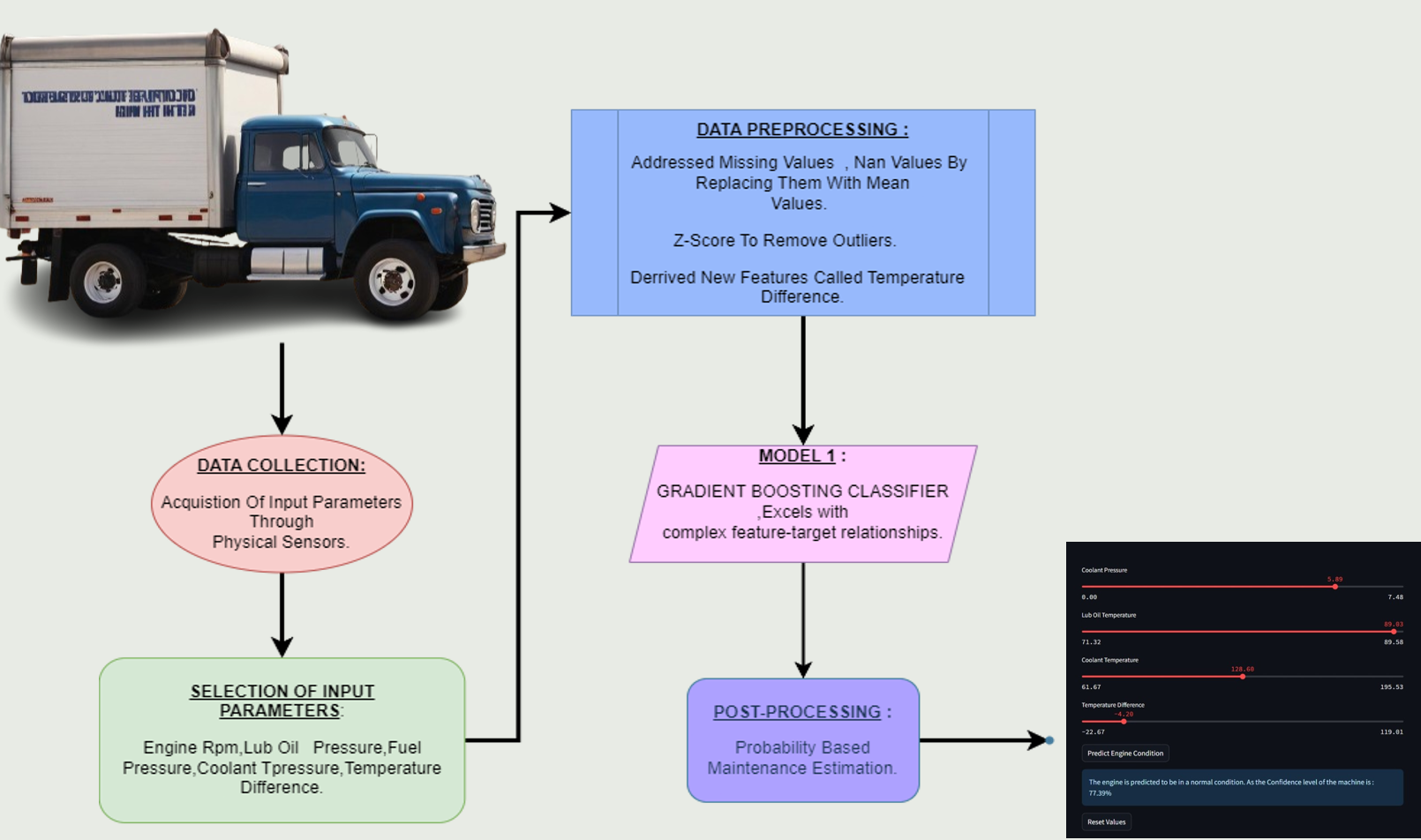


Fig.1. Workflow for engine prediction

Figure. 1 illustrates the workflow diagram outlining the methods used in the study. According to the flowchart, the preprocessing phase is crucial for adequately preparing the data for successful model training. This phase encompasses the following actions:

Gathering phishing and authentic data from reputable sources like Phish Tank, UCI Machine Learning Repository, and actual email records is the first step. To guarantee the datasets are impartial and clean: Incomplete, noisy, or redundant data entries are eliminated. Email headers, domain age, and URL length are among the pertinent information that are extracted. Class imbalances, which are frequent in phishing detection datasets, are addressed using data balancing approaches (oversampling or under-sampling).

***1. Data Pre-processing:***

*1.1* ***Addressing Missing Values:***

To maintain data quality, missing values are handled by substituting them with the mean of the corresponding feature. This method mitigates data loss while ensuring consistency, thereby preserving the integrity of the dataset for model training. By employing mean imputation, the central tendency of the data is retained, which helps to reduce the likelihood of introducing bias.

*1.2* ***Outlier Management:***

Outliers, which can distort predictions and adversely affect model performance, are addressed through the Z-Score method. Data points exhibiting a Z-Score greater than ±3 are classified as outliers and subsequently removed from the dataset. This process results in a more refined dataset, thereby improving the reliability of the predictive model.

*1.3* ***Feature Development:***

Feature development is essential for enhancing the model's performance. A new feature, Temperature Difference, is created by computing the difference between coolant temperature and ambient temperature. This feature captures significant patterns within the data, offering additional insights that contribute to improved predictive accuracy.

***2. Gradient Boost Classifier:***

The Gradient Boost Classifier improves predictive performance by reducing bias and variance through the sequential combination of models, which is based on the boosting principle. By building a sequence of weak learners, usually decision trees, this technique aims to minimize the loss function by having each model fix the mistakes of the one before it. The algorithm is especially well-suited for managing intricate datasets with nonlinear relationships because of its capacity to adaptively improve its predictions.

The ability of the Gradient Boost Classifier to withstand problems like noise and class imbalance, which frequently impair the performance of conventional models, is one of its main advantages. It uses regularization strategies, like subsampling and learning rate adjustments, to keep the model from overfitting and make sure it performs well when applied to new data. Additionally, the algorithm can be parallelized to maximize computational efficiency, which makes it a strong candidate for large-scale datasets due to its scalability.

The Gradient Boost Classifier can process a variety of input features, including temperature, pressure, and RPM sensor data, and successfully capture the interplay between these variables in the context of predictive maintenance. This makes it possible to accurately forecast system anomalies or failures, enabling maintenance teams to take preventative measures. Incorporating such a classifier into a framework for vehicle health monitoring not only improves prediction accuracy but also increases operational effectiveness and reduces costs.

***3. Post-Processing:***

The processing of nature in the predictive maintenance system, the outputs generated by the automatic learning model refer, which provides insights that can be used for an effective decision -making process. The model provides for maintenance opportunities based on sensor data, so that users can estimate the possibility of disorders in part and determine maintenance patterns with a 2-3 weeks delivery time. This possibility guarantees a proactive intervention, so that the risk of unexpected malfunctions is reduced to a minimum and maintenance teams can prioritize critical repairs. The system improves the predictive accuracy by integrating the random percentages for possible errors of the components, which offers a clear understanding of the engine health. In addition, he is responsible for different Ernst levels, so that maintenance professionals offer a more nuanced risk assessment. Additionally, it facilitates rapid maintenance planning, optimizes resource allocation, and reduces downtime, thereby improving overall operational efficiency and reliability. This post-processing framework ultimately enables fleet managers to make data-driven decisions that improve vehicle availability and reduce operating costs.

# Results And Discussion

The application uses the Gradient Boost Classifier (GBC) to predict engine conditions, and it outperforms other tested models in the study with a high accuracy of 72%. The efficacy of GBC in managing intricate patterns in the dataset was highlighted by the evaluation of other models, including Random Forest, Support Vector Machines (SVM), and Logistic Regression, which showed lower accuracy levels when compared to GBC. Maintenance experts can make well-informed decisions about engine health thanks to the system's predictions and confidence scores. It lowers the chance of operational failures and unscheduled downtime by effectively differentiating between routine operations and those that need urgent attention.

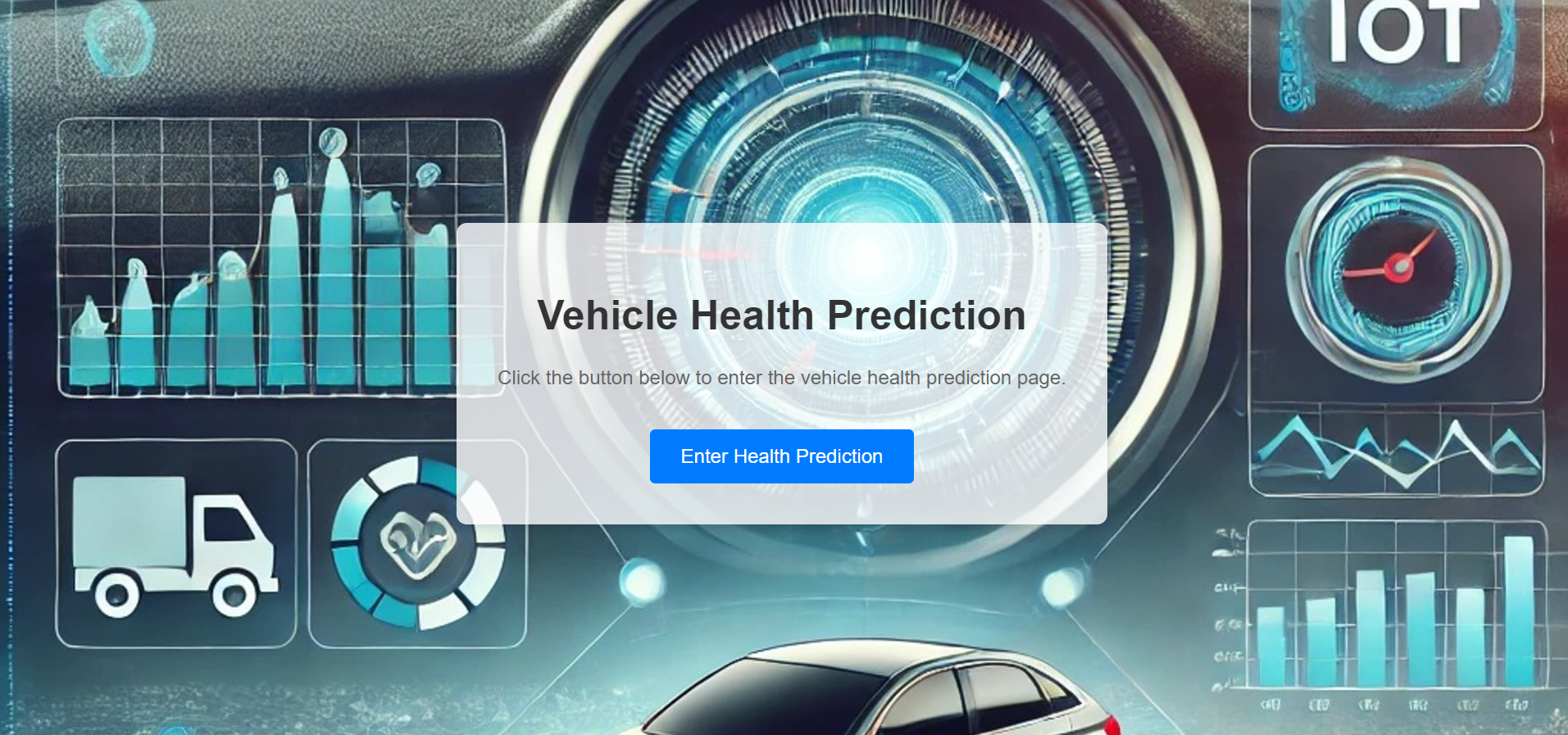


Fig.2. Web page for vehicle health prediction

The fig.2 illustrates an HTML web page crafted as the gateway to a vehicle health prediction system. This page boasts an aesthetically pleasing layout, incorporating various graphical components such as charts and gauges that represent data analysis and vehicle health assessment. A prominent button labeled "Enter Health Prediction" acts as the primary interactive feature, encouraging users to advance to the prediction module. The advanced design underscores the incorporation of IoT and analytics, highlighting the system's emphasis on predictive maintenance and real-time monitoring.

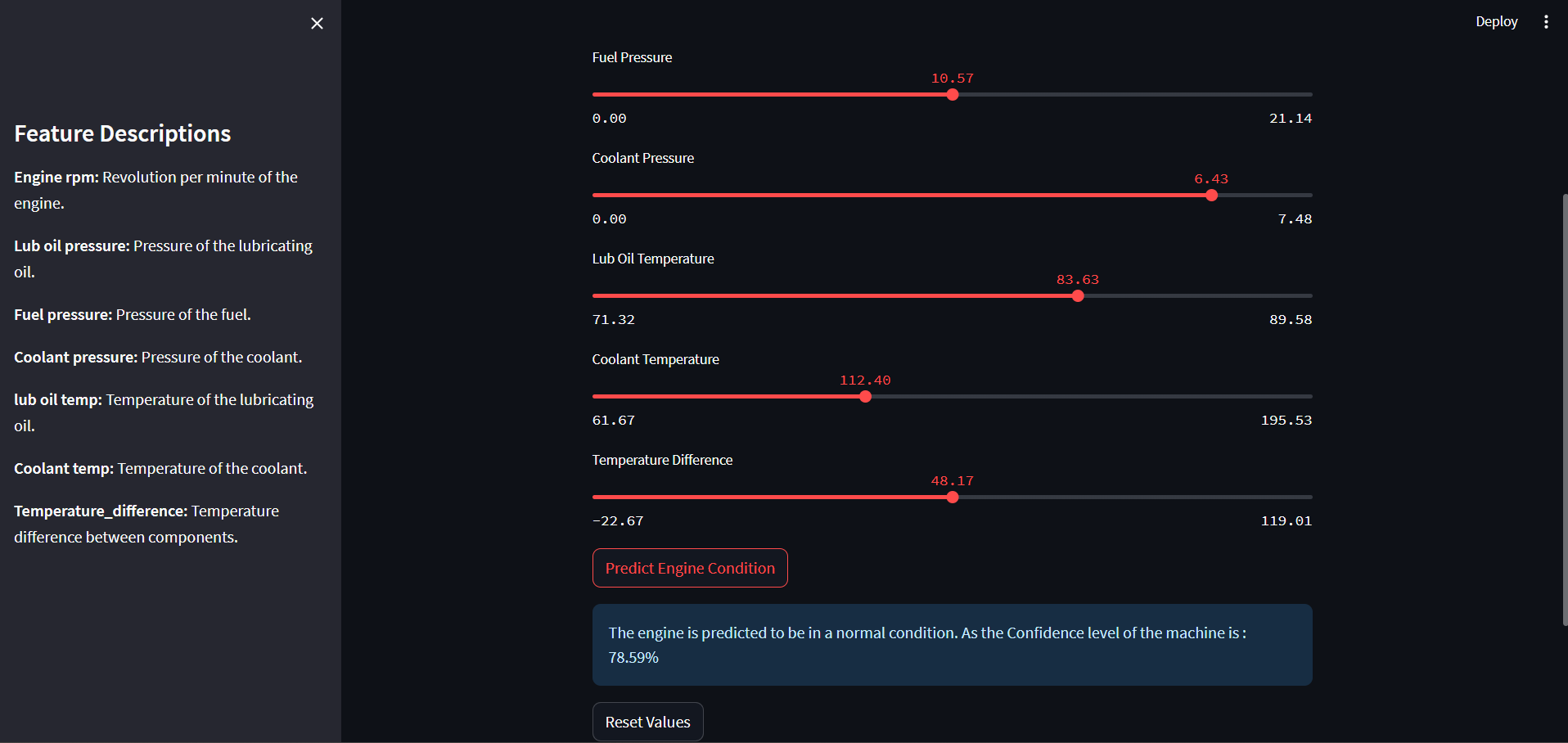


Fig.3. Confidence level: good

The fig.3 displays a predictive maintenance system's user interface for automobile engines.The engine health is analyzed by the system using a trained machine learning model based on the input data, and the prediction result and confidence level are displayed. At a 73.59% confidence level, the system forecasts that the engine is operating normally in this instance. This shows how well the system evaluates engine performance and gives users useful information for preventative maintenance scheduling. For improved user comprehension, the interface also has a feature description section.



Fig.4. Confidence level: Warning

The fig.4 depicts the Feature Input and Prediction Interface, which highlights the core functionality of the vehicle health prediction system. On the left, users will find descriptions of features such as engine RPM, lubricating oil pressure, fuel pressure, coolant pressure, and temperature differentials, offering clear insights into the input parameters. On the right, sliders enable users to enter real-time sensor values for these metrics. A button labeled "Predict Engine Condition" initiates the model's analysis of the inputs to forecast the engine's status. The prediction outcome is presented below, indicating the engine's condition and providing a confidence score of 45.85%.

|  |  |
| --- | --- |
| **Classification Models** | **Accuracy (%)** |
| Random Forest Classifier | 65.77% |
| XGBoost Classifier | 66.66% |
| Support Vector Machine | 64.69% |
| Logistic Regression | 65.80% |
| **Gradient Boost Classifier**  **(Proposed Model)** | 72.67% |

Table 1. Comparison with other state-of-art-models

Table 1 displays the results of several machine learning models that were used to forecast engine conditions. In analyzing the dataset, each model exhibits distinct strengths, with ensemble techniques such as Random Forest and Gradient Boost Classifiers demonstrating their capacity to manage intricate, non-linear relationships. In the meantime, algorithms that provide a baseline for comparison, like Support Vector Machine and Logistic Regression, simplify and make the predictive process easier to understand. XGBoost, which iteratively improves weak learners, is another example of how boosting techniques can increase prediction reliability. A formidable contender for predictive maintenance applications, the suggested Gradient Boost Classifier outperforms its competitors by utilizing sophisticated gradient optimization and reliable feature handling.  
  
**Accuracy Graph of a Proposed model**

The Gradient Boost Classifier (GBC) demonstrates superior accuracy, markedly surpassing the performance of alternative models. This underscores the efficacy of the GBC in yielding precise predictions for the specified dataset.

The system's practical utility is further enhanced by its user-friendly interface, which makes it highly adaptable for field applications by enabling real-time sensor data input. Furthermore, the application optimizes maintenance schedules and resource allocation by forecasting probable failures and visualizing confidence levels, thereby supporting proactive maintenance decisions. This model-driven method applies to a range of vehicle types and fleet sizes due to its notable scalability and adaptability. By adding more sensor parameters and sophisticated machine learning algorithms, future iterations could improve the system even more, possibly improving prediction accuracy and broadening its range of applications.

Expanding upon its good platform, the system's versatility makes it a useful instrument for both traditional and cutting-edge technologies, such as electric and driverless cars. It may be able to track complex systems specific to these next-generation cars, like autonomous driving subsystems, electric motor efficiency, and battery health, by utilizing advanced analytics. Moreover, integrating IoT (Internet of Things) capabilities may facilitate smooth data exchange between fleet management platforms, improving coordination and offering real-time insights on a bigger scale. The system might gain additional advantages from centralized data aggregation and machine learning model updates with the addition of cloud-based processing, guaranteeing steady advancements over time. To address recurrent maintenance issues, long-term planning may be made possible by combining predictive analytics with historical trends to find patterns that go beyond immediate failures. This change would greatly lessen the negative effects that vehicle downtime and inefficient maintenance have on the environment and the economy in addition to improving operational reliability.

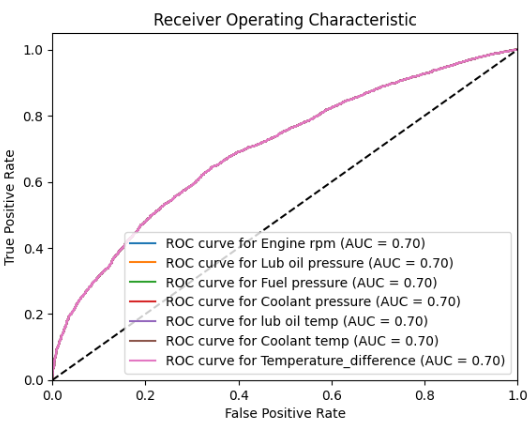


Fig.5. Receiver Operating Characteristic – Area Under Curve (ROC-AUC) Graph of Proposed model

Fig. 5 shows an examination of a Receiver Operating Characteristic (ROC) curve about various engine parameters utilized within a predictive maintenance framework. The plotted curves demonstrate the compromise between the True Positive Rate (sensitivity) and the False Positive Rate for each of the following parameters: engine RPM, fuel pressure, coolant pressure, lubrication oil temperature, coolant temperature, and temperature differential. The model exhibits moderate efficacy in distinguishing between engine states, as evidenced by an Area Under the Curve (AUC) value of 0.70 for each parameter. The curves that reside above the diagonal dashed line indicate that the model surpasses the performance of random guessing, whereas the diagonal dashed line represents the performance of a random classifier. This analysis underscores the significance of each input feature in enhancing the predictive capability of the model.

# Conclusion

In conclusion, Several classification models are evaluated to show how machine learning can be used to solve predictive maintenance problems. As the suggested method, the Gradient Boost Classifier continuously outperformed the other models among the tested models because of its capacity to manage intricate feature interactions and iterative optimization. The significance of choosing sophisticated algorithms that are suited to the complexities of the problem is highlighted by this, as this will guarantee more accurate predictions and better alignment with actual maintenance requirements. The outcomes highlight the potential for these models to completely transform engine health monitoring and maintenance practices.

The applications signifies a significant leap forward in the realm of transportation maintenance, proactively tackling potential challenges before they develop into operational interruptions. By leveraging artificial intelligence and real-time data analytics, the system delivers precise forecasts regarding engine health, facilitating timely and proactive measures. This methodology not only minimizes vehicle downtime but also substantially reduces maintenance expenses, enhances operational dependability, and prolongs the lifespan of essential vehicle components. By guaranteeing the smooth operation of vehicles, the system promotes uninterrupted activities, thereby boosting productivity, efficiency, and safety for all involved parties. Additionally, the incorporation of AI-driven insights into maintenance processes encourages more informed decision-making, enabling fleet managers and maintenance personnel to effectively optimize resources and scheduling. This innovation possesses considerable potential for scalability and adaptability, establishing itself as a transformative asset for the dynamic requirements of the automotive and transportation industries.

##### References

1. S. Vasavi, K. Aswarth, T. Sai Durga Pavan, A. Anu Gokhale, Predictive analytics as a service for vehicle health monitoring using edge computing and AK-NN algorithm, Materials Today: Proceedings, Volume 46, Part 17, 2021, Pages 8645-8654,ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.03.658.
2. Shafi, Uferah & Safi, Asad & Shahid, Ahmad & Ziauddin, Sheikh & Saleem, M.Q.. (2018). Vehicle Remote Health Monitoring and Prognostic Maintenance System. Journal of Advanced Transportation. 2018. 1-10. 10.1155/2018/8061514.
3. Jafari, S., Shahbazi, Z., & Byun, Y. -C. (2022). Lithium-Ion Battery Health Prediction on Hybrid Vehicles Using Machine Learning Approach. Energies, 15(13), 4753. https://doi.org/10.3390/en15134753
4. Hari Pranesh, Khiran Suresh, Swetha S. Manian, R. Jegadeeshwaran, G. Sakthivel, T.M. Alamelu Manghi, Vibration-based brake health prediction using statistical features – A machine learning framework, Materials Today: Proceedings, Volume 46, Part 2, 2021, Pages 1167-1173, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.02.060.
5. Shivansh Khurana, Shubham Saxena, Sanyam Jain, Ankur Dixit, Predictive modeling of engine emissions using machine learning: A review, Materials Today: Proceedings, Volume 38, Part 1, 2021, Pages 280-284, ISSN 2214-7853, <https://doi.org/10.1016/j.matpr.2020.07.204>.
6. Al-refai, G., Elmoaqet, H., & Ryalat, M. (2022). In-Vehicle Data for Predicting Road Conditions and Driving Style Using Machine Learning. Applied Sciences, 12(18), 8928. <https://doi.org/10.3390/app12188928>.
7. SKlearn. Available online: https://scikit-learn.org/stable/.
8. Numpy. Available online: https://numpy.org/.
9. Pandas. Available online: https://pandas.pydata.org/.
10. Matplot. Available online: https://matplotlib.org/.