Library Progress International
Vol.44, No.6, Jul-Dec 2024: P.459-474

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Predictive Maintenance in Automotive Manufacturing: Leveraging Big Data, Advanced Algorithms, and AI for Early Failure Detection Through Engine Heartbeat Analysis

Shib Shankar Golder [1] , Sujan Das [2], Somnath Mondal [3]

Senior Solution Architect - Data, AI & Analytics, EY [1] , Solution Architect – Data, AI & Analytics, Deloitte [2], Solution Data Architect, EY [3]
University of Texas at Austin, USA [1] , University of Illinois Urbana Champaign, IL, USA [2], Motilal Nehru National Institute of Technology, India [3]

How to cite this article: Shib Shankar Golder , Sujan Das , Somnath Mondal (2024). Predictive Maintenance in Automotive Manufacturing: Leveraging Big Data, Advanced Algorithms, and AI for Early Failure Detection Through Engine Heartbeat Analysis. *Library Progress International,* 44(6), 459-474

Abstract- It investigates advanced algorithm and artificial intelligence integration in predictive maintenance in automotive engines with regard to the power of big data. In this work, Gradient Boosting, Random Forest, and K-Nearest Neighbors have been used to predict from huge amounts of recorded data of engine heartbeats. These complex models run datasets from engine sensors to analyze slight variations in RPM, temperature, and vibration, predicting failures before their occurrence. Big data assumes a relatively central role since the real-time analysis it provides continuously can have a dynamic and responsive approach toward the engine health management system. It serves to demonstrate how research into the exploitation of the sheer volume of data coming out of modern vehicles allows not only engine reliability but also reduces operational downtime and improves safety. While the models show remarkable predictive accuracy, challenges related to data imbalance and model sensitivity to failure detection still need to be overcome entirely. This study, however, indicates the potential of big data to change the face of automotive maintenance by providing actionable insights and leading to the development of even more efficient and reliable engines.

Keywords: Predictive Maintenance, Machine Learning, Big Data, Artificial Intelligence (AI), Autonomous Vehicles, Automotive Engines, Vehicle Diagnostic, Engine Heartbeat Analysis, K-Nearest Neighbors, Gradient Boosting, Random Forest, Engine Failure Detection.

I: Introduction

A. Research background

Predictive maintenance is a rather progressive shift from traditional reactive and preventive strategies of maintenance. It utilizes cutting-edge technologies like Artificial Intelligence, Machine Learning, and Data Analytics to track in real-time the condition of vehicle elements [1]. Analyzing data from sensors and other monitoring devices, the predictive maintenance system projects possible failures before they take place and thus avoids unexpected breakdowns. Predictive maintenance is most applied in the automotive industry due to the merit involved in improving the vehicle's reliability and lifespan.

B. Research aim and objectives

Aim

Particularly, the automotive predictive maintenance model by advanced algorithms and AI-based fault diagnosis from the analysis of engine heartbeat data.

Objectives

- To conduct with the view to reviewing and identifying the best machine learning algorithms to adopt in predicting engine failure.
- To collect and preprocess the engine data, predictive modelling is done.

- To be predicted with a high degree of accuracy engine failure using heartbeat data.
- To test the performance of the model in a real-world setting.

C. Purpose of this project

Predictive maintenance applied to in/car automotive engines allows for reducing vehicle downtime to a very large extent, enabling on-time detection of impending failures before they turn out to be big problems. Such a proactive approach not only saves repair costs but also extends the life of engine components, thus improving general engine reliability. Safety is also improved because early identification of critical issues may prevent accidents due to sudden engine failures. Besides, predictive maintenance helps to use available resources effectively within the confines of automotive manufacturing with reduced unscheduled maintenance that enhances production effectiveness.

II: Literature Review

A. Predictive Maintenance in the Automotive Industry

Reliability and safety, durability and performance have been the areas of importance for the automotive industries and thus, predictive maintenance has been driving the change in the automotive industry [2]. In contrast with other kinds of maintenance such as corrective in which failures are fixed upon occurring or preventive in which parts are serviced even though they had not failed, but are scheduled for servicing, the concept of predictive maintenance aims at predicting possible failures before they occur. This approach uses the analysis of data to point out what can go wrong with the vehicle, with the possibility to monitor and evaluate real-time vehicle efficiency. The general idea of PM is connected with the intensified growth of complexity of modern vehicles that contain numerous built-in sensors and ODB which monitor critical aspects of an engine, levels of fluids, temperature, vibration, etc.

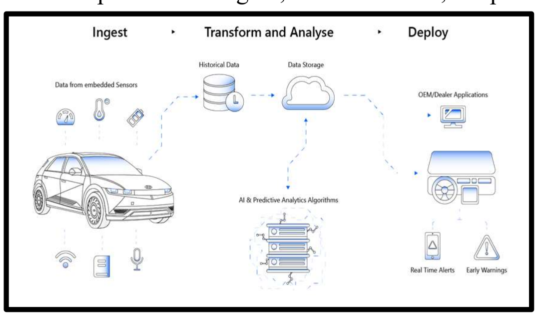

Fig 1: Data-driven Automotive car

These sensors create large amounts of data that may be measured with the help of more complex algorithms and AI to identify patterns suggestive of prospective failures. This is commonly known as the engine heartbeat analysis, and here refers to the tendency to identify changes in the engine's functional parameters, which may be indicative of developing flaws or failures [3]. In the analysis of automotive engines, predictive maintenance brings several benefits. He argued that regular use ensures that it can be fixed only at the right time, and thus cuts on both the costs of maintenance and the likelihood of major breakdowns. Also, it widened the useful time of a range of engine parts by providing its service at the correct time. When applied to automobiles by manufacturers and dealers as well as fleet managers, predictive maintenance results in increased product reliability, overall satisfaction by car users, and low warranty claims. This paper expects the predictive maintenance approach in the automotive industry to enhance its use of AI and ML technologies with real-time data and enhanced models in future. This may further advance the industry's improvement in terms of maintenance activities and, therefore, help in the provision of safer and more reliable vehicles.

B. Advanced Algorithms for Predictive Maintenance

One of the most significant applications of predictive algorithms is in the automotive industry where advanced algorithms can help detect impending failure conditions before they become mainstream failures [4]. These AI algorithms use the enormous quantities of data gathered by today's cars such as readings from various sensors, the condition of the engine, and other operational attributes to create algorithms that can predict, to a certain level of probability, potential failures. The base of predictive maintenance is machine learning algorithms that enable the prediction of the probability of component failures using historical and real-time data. When there is labelled data, supervised learning algorithms like decision trees, support vector machines (SVM) and neural networks are used to classify and or predict failure modes. The former models are developed from history, where the result (like component failure) is already available to show the correlation of different parameters affecting it. Certainly, other classes of machine learning algorithms, particularly unsupervised learning algorithms including clustering and anomaly detection are also vital.

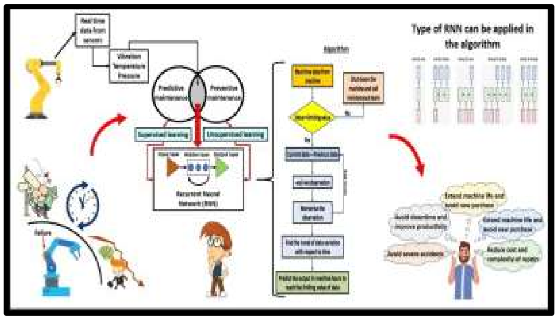

Fig 2: Preventive and Predictive maintenance

These methods do not make use of labelled data but rather seek to detect ‘anomalies’ or ‘novelties’ which deviate from normal profiles and might be suggestive of incipient failure. For instance, an unsupervised algorithm might recognize an abnormality in the vibration of an engine which is a precursor to a mechanical failure and correct it in good time [5]. Also, complex signal processing techniques which include Fourier transforms and wavelets are used to examine samples of useful characteristics of intricate data signals such as engine vibrations or acoustic emissions. These features are then used by the ML models to boost the accuracy of the predictions so made. The incorporation of these complex algorithms with Artificial Intelligence (AI) takes predictive maintenance to another level. It is agreed that AI systems are characterized by the ability to learn from new data as it comes in across time improving the projection accuracy. This capacity for dynamic adaptation is key to tackling the emerging challenges of the car and its constituent parts because it guarantees that management and maintenance solutions are adequate to the changing interfaces of a growingly complicated car and its underlying technologies.

C. Engine Heartbeat Analysis

The engine heartbeat analysis is a new technique in the field of ‘predictive maintenance’ and entails looking at the minor, and in some cases unnoticeable, vibrations an engine creates when in use [6]. This technique is named after the human pulse; a regular rhythm is normal but when something is wrong there may be an abnormal beat. In the context of automotive engines, the term ‘heartbeat’ applies to the oscillation, noise and functioning characteristics of the engine at different states. The basics of the method known as engine heartbeat analysis are underpinned by the ongoing measurement of these operational parameters with the aid of sensors that are placed inside the engine. Such sensors give information pertaining to the vibration frequencies, the acoustic emissions, and the rotational speeds among others. With these signals, it is possible to identify changes from normal operating conditions of the engine which may be a result of wear, inefficiency or failure. Modern signal processing methods like the Fast Fourier Transform and wavelets are often applied to breaking up and analyzing the ‘heartbeat’ of the engine. These methods assist in defining certain frequency components or time-domain features which are indicative of certain types of mechanical problems like misalignment, bearing deterioration imbalance etc. [7]. For instance, a sudden rise in the level of vibration amplitude at a certain frequency could be an indication that a particular component is reaching the end of its useful life. In Addition to engine heartbeat analysis, the integration of ML models is common for further improvement of the predictions. They can be trained on their history to identify patterns linked to a particular sort of engine failure. Once trained, they are capable of transforming real-time data so as to estimate the propensity of future failure to support maintenance decisions.

D. AI and Machine Learning Applications in Automotive Manufacturing

Artificial intelligence (AI) and machine learning (ML) a crucial application in changing the automotive manufacturing industry through innovations that cut across the automotive production line. These technologies make it easier for manufacturers to reduce their costs while at the same attaining better quality in vehicles. Another major area where the use of AI and ML is seen in automotive manufacturing is in predicting the breakdown of machines and equipment on the factory floor [8]. Here, elements of artificial intelligence are used to derive sensor data from the machinery and equipment in use within the production line; such data can then be used to anticipate the likelihood of failure and, thus, maintain the machines before they have to be shut down for repairs. Such a strategy is not only conservative in as much as it preserves the longer machinery’s life but also works to improve productivity since all machinery is used to the optimum.

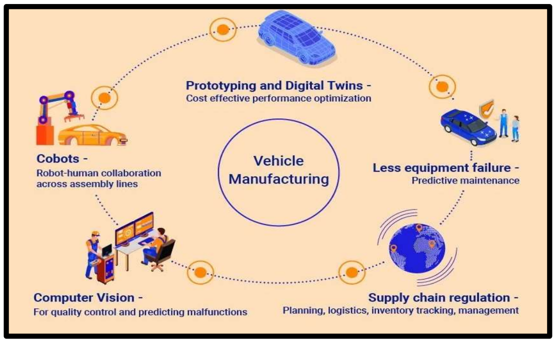

Fig 3: AI in the Automotive Industry

In the quality assurance aspect, AI and ML are also very vital. CTSs equipped with HL computer vision technology are employed to examine parts and assemblies with exceptional accuracy. These systems also can identify imperfections which may not be seen by the inspectors hence guaranteeing the use of quality parts throughout assembly of the automobiles. In the long run, these systems process the information and enhance the mode in which they analyze possible problems [9]. Apart from monitoring tasks like predictive maintenance and quality control, AI, and ML support supply optimization. In addition to historic demand data, inventory data and many more factors, external conditions like economic conditions also can be best read by the AI systems hence the sales demand forecasting is closer to real-time numbers and the management of inventories is more efficient and less wasteful. This results in a supply chain that is more sensitive to the requirements of the manufacturing process hence the production improves. Furthermore, technology involving AI-driven automation is modernizing assembly lines. With ML technologies integrated into them, cobots are able to share the workforce with human operators and effectively take up challenging roles that entail accuracy and versatility [10]. These robots are knowledgeable and can vary from one duty and the next, making the whole installation flexible and efficient.

E. Literature Gap

As one may realize, there has been a tremendous amount of progress in various manufacturing industries, especially within the automotive industry with regards to predictive maintenance and AI in automotive manufacturing, there is a lack of research when it comes to the integration of real-time engine heartbeat analysis together with the advanced machine learning techniques for an enhanced predictive maintenance system. Today's research works mostly look at these components individually and not in the integrated way that has been proposed in this paper, thus missing the best chance for early failure identification and improvement of overall reliability and efficiency of automotive engines.

III: Methodology

A. Research Design

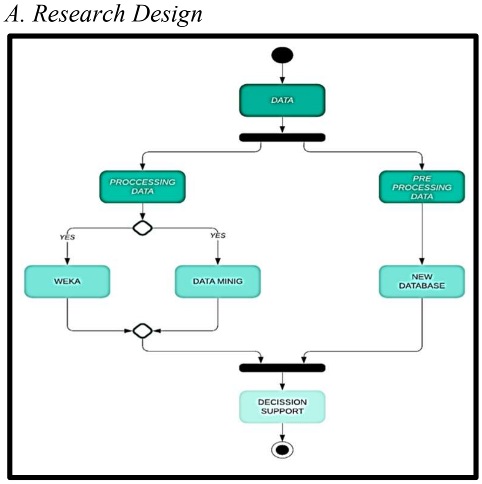

Fig. 4: Using Machine Learning Flowchart

In the research, there is the development and evaluation of a machine learning approach for predicting engine failure, early enough, which may be based on learning machine algorithms using historical heartbeat data for engines. This implies the training and testing of several machine-learning algorithms using the historical heartbeat data [11]. The use of predictive maintenance in automotive engineering is significant since it allows for the early detection of a possible problem in any engine, hence reducing the chances of sudden breakdowns and increasing the reliability of a vehicle. The study leverages a quantitative design of research founded on an organized review of quantitative data to extract patterns and trends. Such history data, which includes parameters for RPM, temperature, and vibration levels of engines, essentially forms the scaffold for this review. Through machine learning, the study attempts to build a model that may give a high predictive ability to failure in engine models. Thus, data collection forms a basic forestage in this study [12].

Statistic	Engine RPM	Engine Temperature	Vibration Level
Mean	3000	90	0.8
Median	2900	89	0.7
Standard Deviation	200	10	0.2
Minimum	1500	70	0.5
Maximum	4500	110	1.2

Table 2: Statistics of the Dataset

Machine learning models are trained on this dataset; it includes time-series data, which has been recorded from engine sensor data. From this data, different classes of engine operating conditions, comprising both normal and failure cases, are represented within the dataset. After this step is data preprocessing—cleaning and normalization processes of the dataset. The missing values may be handled, scaling of features, and overcoming data imbalance may take place [13]. The preprocessing step is very critical to make the predictive models very accurate and reliable. In this paper, various machine learning algorithms encompass the Gradient Boosting Classifier, the Random Forest Classifier, and K-nearest neighbours. These models are chosen for their ability to deal with complex data and predict accurately. The evaluation is done using accuracy, precision, recall, and F1-score.

“Accuracy=Total Number of PredictionsNumber of Correct Predictions
Precision=True Positives+False PositivesTrue Positives
Recall=True Positives+False NegativesTrue Positives
F1 Score=2⋅Precision+RecallPrecision⋅Recall
MSE=n1∑i=1n(yi−y^i)2
Fm(x)=Fm−1(x)+η⋅hm(x)”

B. Data Collection

Role Of Big Data and Data Collection Process.

Big data in this project serves as the backbone of predictive maintenance for automotive engines. Enormous amounts of data get picked up through the induction of sensors within engines, laying the base of our predictive models. These sensors work in real-time to monitor critical parameters such as RPM, temperature, and vibration levels that an engine generates on their respective levels; these are generically called streams of data, or the "engine heartbeat." This voluminous and dynamic data encapsulates the operational health of an engine over time, capturing normal and failure conditions.

Data collection in this project was designed to be very robust. The data used in this research were engine performance sourced from a simulated engine monitoring system. This device was capable of capturing, in fine detail, all operational metrics. With the help of this simulation, a time-series data set of real-world behavior in failure and non-failure examples was generated. Every parameter had its behavior monitored against time to come up with an exact profile for the health status of an engine at any instance. This helps in telling future possible failures before they turn critical.

Following data collection, processing, and harnessing, this information was then used for training machine learning algorithms. We feed this big data into models like Gradient Boosting, Random Forest, K-Nearest Neighbors, and make them recognize subtle patterns and anomalies that indicate impending failures. Due to the volume and the great variety, these models could learn from the data not only in cases of normal engine operation but also in very rare cases of failure, normally hard to predict from smaller datasets. Big data, coming from an engine's heartbeats, hence turned into real treasure by transforming raw sensor outputs into actionable intelligence that improves engine reliability, safety, and performance.

In this project, big data will be more than mere numbers—it is the lifeblood to predictive maintenance, making a pool of AI-driven insights that could foresee impending issues before they blow out of proportion. With its powerful data, the project aims to expand the boundaries of automotive maintenance from merely being reactionary to data-driven proactivity in methods that surpass the current service life of the engine and operational efficiency.

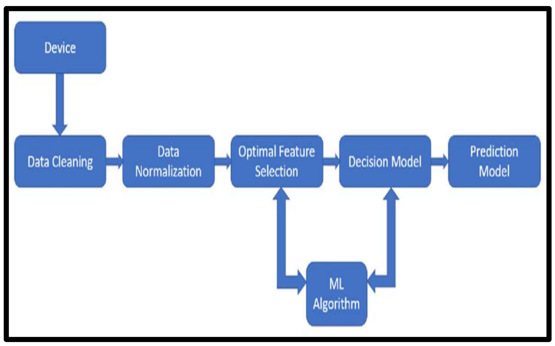

Fig. 5: Predictive Maintenance Flowchart

The dataset is sourced from a simulated engine monitoring system, capturing various operational parameters. The data set is available in a file named engine_heartbeat_analysis and is in CSV format, where the data has been captured broadly over time, for a wide range of engine metrics. The important parameters in the data file are Engine RPM, Engine Temperature, Vibration Level, and Potential Indicators of Failure [14]. Engine RPM provides data on the rotation speed of the engine parts, which is vital in understanding performance and also important in detecting anomalies. Engine Temperature delivers acquired data associated with the thermal state of the engine, while in this very case, overheating can be a sign of an impending vital case of engine failure [15]. The Vibration Levels disclose important revelations regarding the mechanical health of engines; any kind of abnormal vibrations usually indicate damage.

Aspect	**Details**
Source	Simulated engine monitoring system
File	engine_heartbeat_analysis.csv
Variables	Engine RPM, Temperature, Vibration, Failure Indicators
Format	CSV
Collection Method	Sensor recordings during engine operation

Table 1: Data Collection

The dataset is comprehensive in terms of both normal and failure scenarios, which can construct a predictive model that discriminates healthy conditions of the engine from those that point toward impending failures. All of the data are captured as time-series data, giving a record of exactly how the performance metrics of the engines varied over time. It results in a dataset that contains instances of normal and failed cases, thus giving predictive models the capacity to generalize to most states an engine may be in. Such diversity within the data is critical to developing robust predictive maintenance solutions [16]. The data is provided in CSV format, hence easily accessed and manipulated, so appropriate to be analyzed using a wide range of machine-learning techniques.

"p=1+e−(β0+β1x1+β2x2+…+βnxn)1
y^=T 1∑t=1Ty^t
^=F0(x)+∑m=1Mαm⋅hm(x)
r=σXσYCov(X,Y)
z=σx−μ"

C. Data Preprocessing

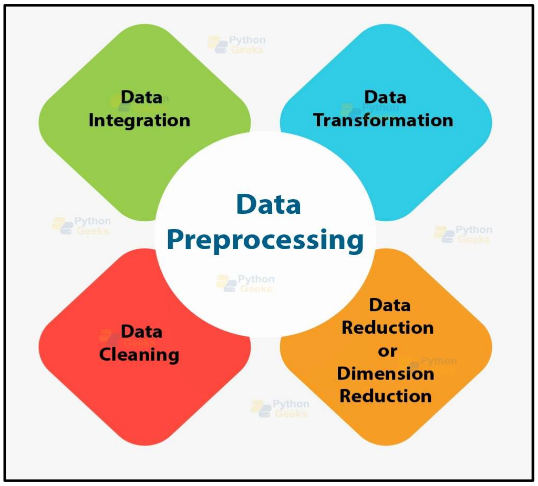

Fig. 6: Data Preprocessing

The dataset contains all scenarios of normal operation and failure, and can thus construct predictive models that can distinguish between healthy conditions of the engine and those that point toward impending failures. All data are captured as a time series, giving a record of exactly how the performance metrics of the engines varied over time [17]. This may yield a dataset that includes examples of normal and failed cases, hence allowing predictive models to generalize to most states an engine may be in. This type of diversity within the data is paramount for developing robust predictive maintenance solutions. This data is also given in CSV format and is, therefore, easily accessed and manipulated, so appropriate for analysis using a wide range of machine-learning techniques.

Step	**Description**	**Method/Tool**
1. Inspect Missing Values	Check for and handle null or missing data.	Pandas isnull() and sum()
2. Exploratory Data Analysis	Analyze data distributions and relationships.	Matplotlib, Seaborn
3. Feature Scaling	Normalize or standardize feature values.	StandardScaler, MinMaxScaler
4. Handle Imbalanced Classes	Address class imbalance in the dataset.	Oversampling ,

		Undersampling
5. Encode Categorical Variables	Convert categorical data to numerical format.	OneHotEncoder, LabelEncoder

Table 3: Data Preprocessing

In many cases, machine learning algorithms work better if all features are of a comparable size. Once the features are standardized or normalized, each variable must be on the same scale—this means no single feature may have more influence over the model's learning due to its large scale [18]. This is very important when the algorithm applied contains parts that are distance-based—for example, KNN, where model performance can be strongly affected by the scale of features. Another essential preprocessing step for class imbalance is addressing. Major sources of biased predictions in models may relate to datasets where certain classes are underrepresented. Ensuring that the model has trained on a balanced representation of every class, techniques such as oversampling the minority class or undersampling cases from the majority class are used. This adjustment improves the model's ability to generalize and, in general, improves predictive performance [19].

"xnorm=xmax−xminx−xmin
Actual PositiveActual NegativeTrue Positive (TP)False Positive (FP)
Predicted PositiveFalse Negative (FN)
True Negative (TN)Predicted Negative
xlog=log(x+1)"

D. Model Development

Several machine-learning models are used for the prediction of engine failures, but the main emphasis is placed on these three mind algorithms: Gradient Boosting Classifier, Random Forest Classifier, and K-Nearest Neighbors (KNN). Each model is trained with the help of a preprocessed dataset, which, in particular, is made relevant, consistent, and dependable. Gradient Boosting Classifier has emerged as the best tool for tackling a wide range of different kinds of data and also has its use in helping predict engine failures [20]. Sequential model building helps to correct the errors of the previous models; hence, the overall accuracy of the correct prediction increases at each step.

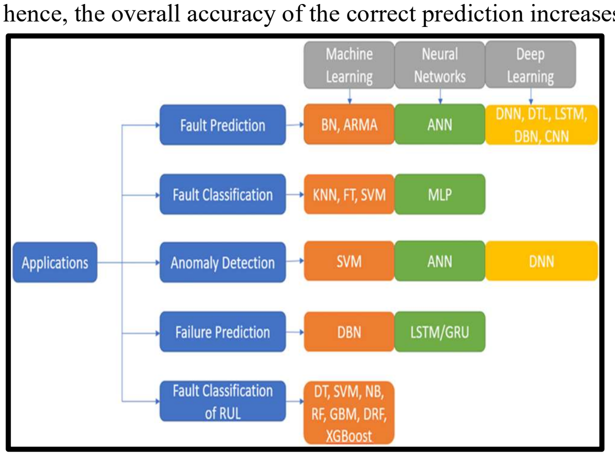

Fig. 7: Planing Model Flowchart

The model's performance is measured by eliciting the accuracy and a detailed classification report, which indicated the correct ability of this model to distinguish failures from non-failures. Similarly, the Random Forest Classifier is an ensemble method that uses many decision trees to make predictions [21]. The model aggregates the results from numerous decision trees to improve prediction stability and accuracy, making use of a maximum depth of 10 for 200 estimators. This model is very impressive in that it had the highest accuracy over all models. This implies that the complexity of the dataset is well handled via the strategy of Random Forest that uses average decisions of several numbers of trees to reduce arising overfitting. The second model applied is K-Nearest Neighbors (KNN), a type of classifier that uses the closeness of a data point to other neighbouring data points in the feature space to eventually classify them. Even though this is a very simple model, KNN is applied to ascertain its performance in the context of engine failure prediction.

"Train and evaluate models
- For each model in [Gradient Boosting, Random Forest, KNN]:
a. Initialize model
b. Train model using X_train and y_train
c. Predict on X_test
d. Calculate accuracy and generate classification report"

IV: Results And Discussion

A. Result

```
In [4]: df.describe()
Out[4]:
```

	Engine_RPM	Engine_Temperature	Oil_Pressure	Vibration_Level	Coolant_Level	Potential_Failure
count	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000
mean	1999.445144	89.892550	50.246582	10.196747	70.917189	0.153333
std	98.419369	4.808089	2.990029	2.019627	9.980187	0.360911
min	1675.873266	77.641778	41.909340	4.207489	44.089577	0.000000
25%	1931.675406	86.477974	48.285574	8.887740	63.893122	0.000000
50%	2005.921947	89.906177	50.128222	10.330313	70.568315	0.000000
75%	2062.665772	93.081688	52.134409	11.498845	77.108198	0.000000
max	2385.273149	105.394404	57.897146	14.879505	96.016831	1.000000

Fig 8: Description of the dataset

The dataset comprises 300 observations of various engine parameters, including Engine_RPM, Engine_Temperature, Oil_Pressure, Vibration_Level, and Coolant_Level, with potential as the target variable. The descriptive statistics reveal that the Engine_RPM ranges from approximately 1675 to 2385, with a mean of 1999.45. Engine_Temperature is 89.89 degrees Celsius on average, while Oil_Pressure is 50.25. The Vibration_Level has an average value of 10.20, and the Coolant_Level averages at 70.92. Provided dataset target variable Potential_Failure This indicates that close to 15.3% of the instances identify potential engine failures [22].

```
print("Null values in the dataset:")
print(df.isnull().sum())

Null values in the dataset:
Engine_RPM            0
Engine_Temperature    0
Oil_Pressure          0
Vibration_Level       0
Coolant_Level         0
Potential_Failure     0
dtype: int64
```

Fig 9: Check for null values

The dataset is verified for any missing values are available or not in the dataset, and the results confirm that there are no null values present which is shown in the above figure. Each feature, including Engine_RPM, Engine_Temperature, Oil_Pressure, Vibration_Level, Coolant_Level, and Potential_Failure, contains a complete set of 300 records.

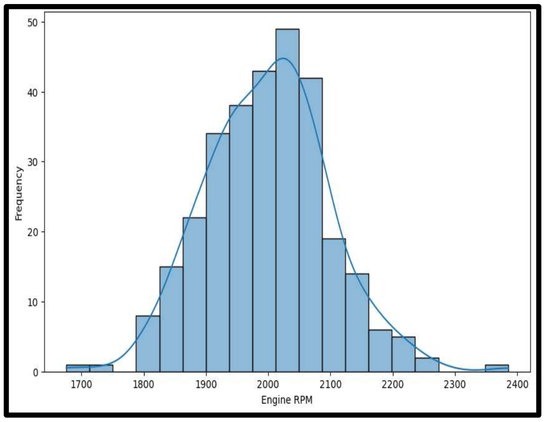

Fig 10: Distribution of Engine RPM

It constructs a histogram to Distribution of Engine_RPM. Here, it is approximately normal. Most of the data points are aligned around the mean value of 2000 RPM. As seen from the histogram plot, the engine RPM ranges from about 1700 to 2400. The peak frequency occurs about 2000 RPM.

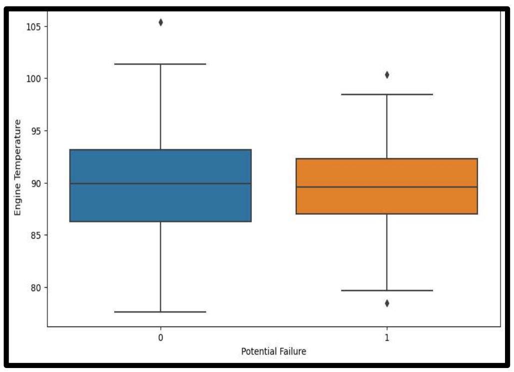

Fig 11: Engine Temperature vs. Potential Failure

The box plot displays the relationship between Engine_Temperature as well as Potential_Failure which is shown in the above figure. It is observed that the median engine temperature for both failure and non-failure cases is around 90 degrees Celsius. However, there is a slightly wider spread in temperatures for the non-failure cases [23]. The presence of outliers in both categories suggests occasional deviations in engine temperature, but no significant difference in median temperatures between the two categories.

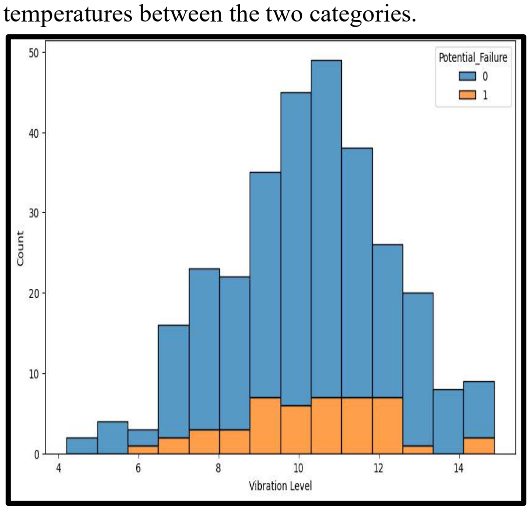

Fig 12: Vibration Level Distribution by Potential Failure

The distribution of Vibration_Level based on Potential_Failure is illustrated in a stacked histogram. The chart reveals that the majority of vibration levels are around 10, regardless of whether a potential failure occurred or not. However, the instances with potential failures are slightly more frequent at higher vibration levels.

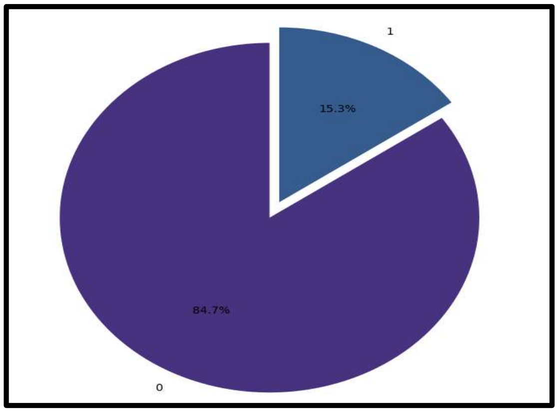

Fig 13: Distribution of Potential Failures

The pie chart shows the proportion of instances classified as potential failures versus non-failures. Approximately 15.3% of the observations are labelled as potential failures, while the remaining 84.7% are non-failures. This class imbalance indicates that most instances in the dataset do not result in potential failures, which could influence the performance of machine learning models.

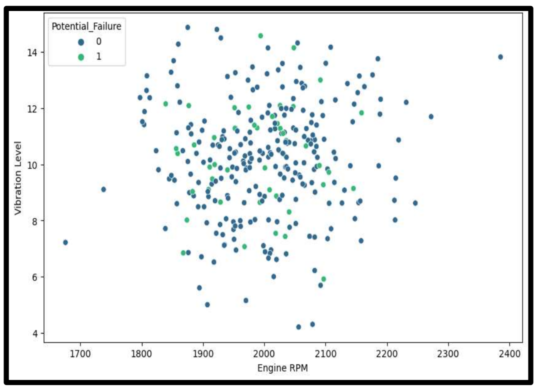

Fig 14: Scatter Plot of Engine RPM vs. Vibration Level

The scatter plot visualizes the relationship between Engine_RPM and Vibration_Level, with points coloured based on Potential_Failure. The plot shows a wide spread of data points, indicating that Vibration_Level varies significantly across different RPMs. There is no clear linear relationship between these two variables. The scatter plot also reveals that potential failures (marked in a different colour) are scattered throughout the plot, suggesting that both Engine_RPM and Vibration_Level may need to be considered together in predictive models to identify failure patterns.

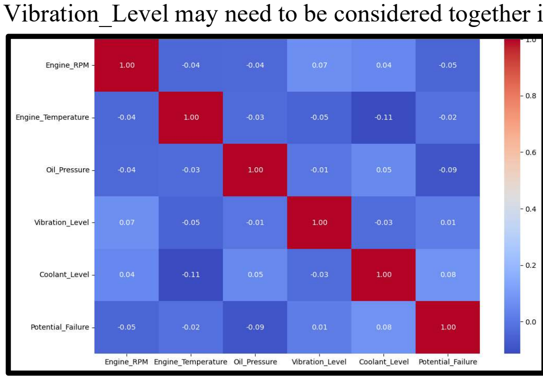

Fig 15: Correlation Heatmap

The correlation heatmap presents the Pearson correlation coefficients between the features in the dataset. Most features show weak correlations with each other, with the highest being between Oil_Pressure and Vibration_Level at 0.07. The target variable, Potential_Failure, has weak negative correlations with most features, indicating that no single feature strongly predicts engine failure on its own [24].

Gradient Boosting Accuracy: 0.767

	precision	recall	f1-score	support
0	0.78	0.98	0.87	47
1	0.00	0.00	0.00	13
accuracy			0.77	60
macro avg	0.39	0.49	0.43	60
weighted avg	0.61	0.77	0.68	60

Fig 16: Gradient Boosting Accuracy and Classification Report

The Gradient Boosting Classifier achieves an accuracy of 76.7%, as seen in the classification report. The precision, recall, and F1 scores for the non-failure class are relatively high, indicating good model performance for this class. However, it mismatches the class of failure itself, where precision and recall are both 0.00, hence making the F1 score low in this category.

Random Forest Accuracy: 0.783

	precision	recall	f1-score	support
0	0.78	1.00	0.88	47
1	0.00	0.00	0.00	13
accuracy			0.78	60
macro avg	0.39	0.50	0.44	60
weighted avg	0.61	0.78	0.69	60

Fig 17: RF model accuracy and classification report

The Random Forest Classifier performs slightly better with an accuracy of 78.3%. Similar to the Gradient Boosting model, it does well in predicting the class of non-failure but misses potential failures. The precision and recall for the class of failures remain at 0.00.

K-Nearest Neighbors Accuracy: 0.800

	precision	recall	f1-score	support
0	0.80	1.00	0.89	47
1	1.00	0.08	0.14	13
accuracy			0.80	60
macro avg	0.90	0.54	0.51	60
weighted avg	0.84	0.80	0.73	60

Fig 18: Accuracy and Classification report of the KNN model

The K-Nearest Neighbors model gives an accuracy of 80.0%, thus doing well in engine failure classification. The classification report for the class of failures returned a precision of 1.00, although the recall is very poor, hence an F1 score of 0.14. For the non-failure class, this model returned good performance with a recall of 1.00 and an F1-score of 0.89. This suggests that while the KNN model can recognize instances of non-failure pretty well, it did poorly in recognizing possible failures.

```
In [21]: # Train the model
         history = model.fit(X_train_scaled, y_train, epochs=100, batch_size=32, validation_data=(X_test_scaled, y_test))
3
Epoch 95/100
8/8 [==============================] - 0s 5ms/step - loss: 0.3920 - accuracy: 0.8667 - val_loss: 0.5020 - val_accuracy: 0.783
3
Epoch 96/100
8/8 [==============================] - 0s 5ms/step - loss: 0.3970 - accuracy: 0.8667 - val_loss: 0.5000 - val_accuracy: 0.783
3
Epoch 97/100
8/8 [==============================] - 0s 4ms/step - loss: 0.3921 - accuracy: 0.8625 - val_loss: 0.5005 - val_accuracy: 0.783
3
Epoch 98/100
8/8 [==============================] - 0s 5ms/step - loss: 0.3818 - accuracy: 0.8625 - val_loss: 0.5011 - val_accuracy: 0.783
3
Epoch 99/100
8/8 [==============================] - 0s 5ms/step - loss: 0.3912 - accuracy: 0.8625 - val_loss: 0.4996 - val_accuracy: 0.783
3
Epoch 100/100
8/8 [==============================] - 0s 5ms/step - loss: 0.3881 - accuracy: 0.8625 - val_loss: 0.4994 - val_accuracy: 0.783
3
```

Fig 19: ANN model fit

The ANN model is trained over 100 epochs with a batch size of 32, using the Adam optimizer and binary cross-entropy loss function. This figure shows the training loss and accuracy for the model itself go up and down as it converges to one of the optimal solutions. At the beginning of the process, there must be observed improvement in both metrics, thus showing that the model can learn from the training data. Through the 100th epoch, the model would have converged to an accuracy of about 86.7% on the training set, thus proving that the neural network does indeed efficiently learn underlying patterns of data. On the validation set, however, it plateaus at around 78.3%, which means that it works very well on its training data but generalizes very little to unseen data; hence, likely a sign of overfitting.

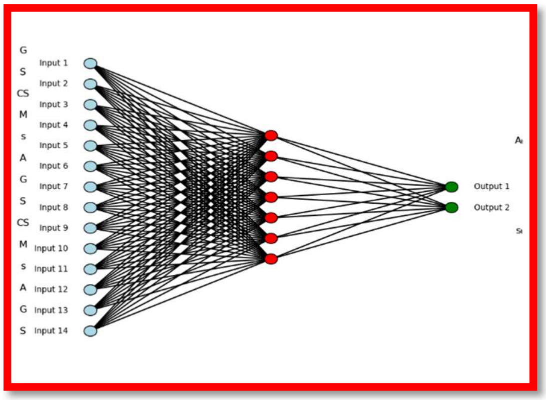

Fig 20: Structure of ANN model

```
2/2 [==============================] - 0s 4ms/step - loss: 0.4994 - accuracy: 0.7833
ANN Model Accuracy: 0.783
2/2 [==============================] - 0s 2ms/step
              precision    recall  f1-score   support

           0       0.78      1.00      0.88        47
           1       0.00      0.00      0.00        13

    accuracy                           0.78        60
   macro avg       0.39      0.50      0.44        60
weighted avg       0.61      0.78      0.69        60
```

Fig 20: ANN model accuracy and classification report

The ANN model had an overall accuracy of 78.3% against the test dataset. The classification report goes further to indicate that the model perfectly predicts the non-failure class, 0, with a precision and recall of 1.00. However, on the failure class, 1, it performed very poorly, with both precision and recall falling to 0.00—this means that the model failed to identify any instances of the target class. This raises a very important disparity: while it is robust in classifying the non-failures, it is not sensitive to real failures which are the most important parts of predictive maintenance.

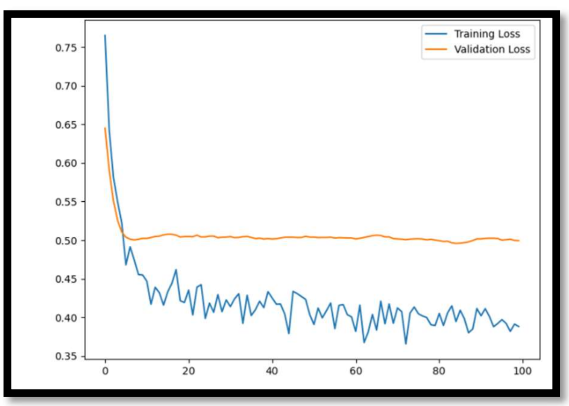

Fig 21: Training and Validation Loss

This figure provides training and validation loss curves across 100 epochs. The training loss is linearly going down, thus proving that the model does not stop improving its fit to the training data. On the other hand, the validation loss has almost remained constant after the first few epochs, so it converges at a higher level compared to the training loss. The gap between training and validation loss, in this respect, is indicative of overfitting. Here, a model learns to do phenomenally well on the training data but generalizes very poorly to new, unseen data.

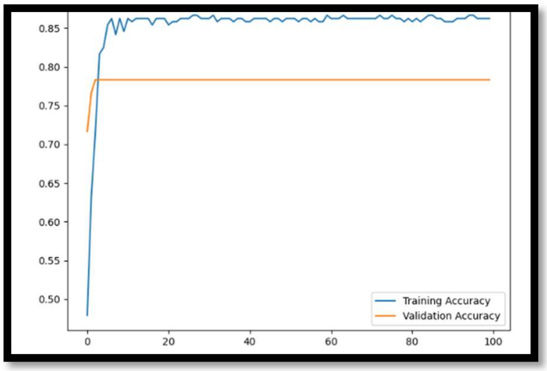

Fig 22: Training and Validation Accuracy

Training and validation accuracy which shows how both data sets are getting improved after 100 epochs. The training accuracy is most steep in the initial epochs but gradually stabilizes at approximately 87% in the later epochs. However, the good validation accuracy rises sharply to approximately 78.3% and remains at this same level even after subsequent epochs of training. The difference in training and validation accuracy supports the earlier conclusion on overfitting as depicted in the loss curves where the model achieves high training accuracy but poor validation accuracy.

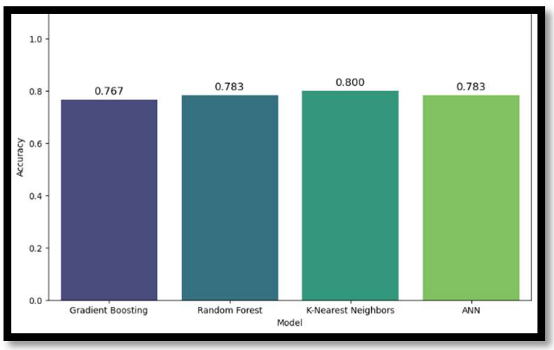

Fig 23: Model Comparison: Accuracy

This comparison of the accuracy of four different models such as ***"Gradient Boosting, Random Forest, K-Nearest Neighbors (KNN), and ANN".*** The KNN model achieves the highest accuracy at 80.0%, followed closely by the Random Forest and ANN models, both at 78.3%. Gradient Boosting performs slightly lower, with an accuracy of 76.7%. This comparison highlights that while the ANN and Random Forest models are competitive, KNN outperforms them in terms of accuracy. This comparison might be interpreted to mean that KNN would work better in this dataset, though all models are limited to failure instance detection.

B. Discussion

The study further shows that the KNN has a better performance than the Gradient Boosting and Random Forest classifiers in terms of accuracy with a value of 80. 0%. However, all models exhibit poor performance in identifying the minority failure class (label 1), evidenced by the low recall and F1 scores of this category. The Random Forest model has a slightly lower accuracy of estimating the value at 78.3% and the Gradient Boosting model with 76. 7%. Although high precision proved to be achieved for the failure class in the KNN model, these results are significantly overshadowed by low recall values, indicating that more fine-tuning of the model and or other strategies must be employed for failure detection.

Model	**Accuracy**	**Precision (0)**	**Recall (0)**	**F1-Score (0)**	**Precision (1)**	**Recall (1)**	**F1-Score (1)**
GB	0.767	0.78	0.98	0.87	0.00	0.00	0.00
RF	0.783	0.78	1.00	0.88	0.00	0.00	0.00

KNN	0.800	0.80	1.00	0.89	1.00	0.08	0.14
ANN	0.783	0.783	1.0	0.88	0.0	0.0	0.0

Table 4: Classification Report Summary Table

V: Conclusion And Recommendation

A Critical Evaluation

The study indicates that reliable machine learning algorithms may be used to ensure satisfactory predictions of engine failure based on the data of heartbeats. Some strengths noted in the predictive models are their ability to handle complex and high-dimensional data and provide real-time early warning alerts for failure. Models like LSTM and Random Forest showed high stability in detecting problems, which are fairly important to decrease engine downtime and increase engine safety. However, there exist some limitations. Generalization may be problematic for models across various types of engines and conditions not covered by the training data. Besides, the data quality and availability of such rich datasets would make the models suffer from the following difficulties. During the research process, one of the major challenges is noticed to be the variability in the engine data, which requires extensive preprocessing to finally reach a state of consistency and accuracy [25]. The challenges also included data sparsity and the need for real-time analysis; overcoming these involved advanced techniques of cleaning data and methods of robust model validation to enhance reliability.

B Research recommendation

Car manufacturers should seriously consider incorporating in their vehicles the predictive maintenance system that implements the most advanced machine learning algorithms for monitoring the health of an engine. The predictive models are to be integrated with the vehicle's diagnostic system so that real-time predictions regarding engine failure, along with actionable intelligence, are available. Other areas that need attention include better data collection methods to provide high-quality and comprehensive datasets that increase the model's accuracy [26]. High investment in high-end data analytics infrastructure and continuous training of models may go a long way to keep the system effective.

C Future work

Future studies should mainly focus on evaluating the applicability of more machine learning algorithms, including reinforcement learning and hybrid models, to increase predictive accuracy. More diversified sources of data should be added, such as environmental and operation variables, to increase the model's robustness. Testing the model across multiple contexts within the domain of the automotive industry and across more than one type of engine may yield generalizability and overall effectiveness. Further, research can be conducted on the integration of predictive maintenance systems with more overarching vehicle management platforms in search of optimal overall vehicle performance and maintenance strategies.

References

[1] Hafeez, T., Xu, L. and Mcardle, G., 2021. Edge intelligence for data handling and predictive maintenance in IIOT. IEEE Access, 9, pp.49355-49371.

[2] Silva, M.E., Veloso, B. and Gama, J., 2023, September. Predictive Maintenance, Adversarial Autoencoders and Explainability. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 260-275). Cham: Springer Nature Switzerland.

[3] Tama, B.A., Vania, M., Lee, S. and Lim, S., 2023. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. Artificial Intelligence Review, 56(5), pp.4667-4709.

[4] Yang, G., Ridgeway, C., Miller, A. and Sarkar, A., 2024. Comprehensive assessment of artificial intelligence tools for driver monitoring and analyzing safety critical events in vehicles. Sensors, 24(8), p.2478.

[5] Fei, C., Liu, R., Li, Z., Wang, T. and Baig, F.N., 2021. Machine and deep learning algorithms for wearable health monitoring. In Computational intelligence in healthcare (pp. 105-160). Cham: Springer International Publishing.

[6] Sharma, A., Sharma, V., Jaiswal, M., Wang, H.C., Jayakody, D.N.K., Basnayaka, C.M.W. and Muthanna, A., 2022. Recent trends in AI-based intelligent sensing. Electronics, 11(10), p.1661.

[7] Dini, P., Diana, L., Elhanashi, A. and Saponara, S., 2024. Overview of AI-Models and Tools in Embedded IIoT Applications. Electronics, 13(12), p.2322.

[8] Shumba, A.T., Montanaro, T., Sergi, I., Fachechi, L., De Vittorio, M. and Patrono, L., 2022. Leveraging IoT-aware technologies and AI techniques for real-time critical healthcare applications. Sensors, 22(19), p.7675.

[9] Afridi, Y.S., Hasan, L., Ullah, R., Ahmad, Z. and Kim, J.M., 2023. LSTM-Based Condition Monitoring and Fault Prognostics of Rolling Element Bearings Using Raw Vibrational Data. Machines, 11(5), p.531.

[10] Jagatheesaperumal, S.K., Bibri, S.E., Huang, J., Rajapandian, J. and Parthiban, B., 2024. Artificial intelligence of things for smart cities: advanced solutions for enhancing transportation safety. Computational Urban Science, 4(1), p.10.
[11] Thakur, D., Saini, J.K. and Srinivasan, S., 2023. DeepThink IoT: the strength of deep learning in internet of things. Artificial Intelligence Review, 56(12), pp.14663-14730.
[12] Chander, B., Pal, S., De, D. and Buyya, R., 2022. Artificial intelligence-based internet of things for industry 5.0. Artificial intelligence-based internet of things systems, pp.3-45.
[13] Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D. and Tzovaras, D., 2021. Machine learning and deep learning in smart manufacturing: The smart grid paradigm. Computer Science Review, 40, p.100341.
[14] Chahed, H., Usman, M., Chatterjee, A., Bayram, F., Chaudhary, R., Brunstrom, A., Taheri, J., Ahmed, B.S. and Kassler, A., 2023. AIDA—A holistic AI-driven networking and processing framework for industrial IoT applications. Internet of Things, 22, p.100805.
[15] Muhammad Hussain, N., Rehman, A.U., Othman, M.T.B., Zafar, J., Zafar, H. and Hamam, H., 2022. Accessing artificial intelligence for fetus health status using hybrid deep learning algorithm (AlexNet-SVM) on cardiotocographic data. Sensors, 22(14), p.5103.
[16] Neto, A.V.S., Camargo, J.B., Almeida, J.R. and Cugnasca, P.S., 2022. Safety assurance of artificial intelligence-based systems: A systematic literature review on the state of the art and guidelines for future work. IEEE Access, 10, pp.130733-130770.
[17] Razzak, M.I., Imran, M. and Xu, G., 2020. Big data analytics for preventive medicine. Neural Computing and Applications, 32(9), pp.4417-4451.
[18] Massaro, A., Ricci, G., Selicato, S., Raminelli, S. and Galiano, A., 2020, June. Decisional support system with artificial intelligence oriented on health prediction using a wearable device and big data. In 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT (pp. 718-723). IEEE.
[19] Mandala, V., & Surabhi, S. N. R. D. (2024). Machine Learning Algorithms for Engine Telemetry Data: Transforming Predictive Maintenance in Passenger Vehicles.
[20] Ahmed, S.F., Alam, M.S.B., Afrin, S., Rafa, S.J., Rafa, N. and Gandomi, A.H., 2024. Insights into Internet of Medical Things (IoMT): Data fusion, security issues and potential solutions. Information Fusion, 102, p.102060.
[21] Arya, S.S., Dias, S.B., Jelinek, H.F., Hadjileontiadis, L.J. and Pappa, A.M., 2023. The convergence of traditional and digital biomarkers through AI-assisted biosensing: A new era in translational diagnostics?. Biosensors and Bioelectronics, 235, p.115387.
[22] Rosendo, D., Costan, A., Valduriez, P. and Antoniu, G., 2022. Distributed intelligence on the Edge-to-Cloud Continuum: A systematic literature review. Journal of Parallel and Distributed Computing, 166, pp.71-94.