IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 October 2023, accepted 2 November 2023, date of publication 8 November 2023,
date of current version 14 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3331316

RESEARCH ARTICLE

EngineFaultDB: A Novel Dataset for Automotive Engine Fault Classification and Baseline Results

MARY VERGARA[1], LEO RAMOS[2,3], (Student Member, IEEE), NÉSTOR DIEGO RIVERA-CAMPOVERDE[4], (Member, IEEE), AND FRANCKLIN RIVAS-ECHEVERRÍA[3,5], (Senior Member, IEEE)

[1]Higher School of Engineering, Science and Technology, Valencian International University, 46002 Valencia, Spain
[2]Computer Vision Center, Universitat Autónoma de Barcelona, Bellaterra, 08193 Barcelona, Spain
[3]Kauel Inc., Houston, TX 77027, USA
[4]Grupo de Investigación en Ingeniería del Transporte (GIIT), Universidad Politécnica Salesiana, Cuenca 010105, Ecuador
[5]Escuela de Ingeniería, Pontificia Universidad Católica del Ecuador Sede Ibarra, Ibarra 100112, Ecuador

Corresponding authors: Mary Vergara (maryjosefina.vergara@professor.universidadviu.com) and Leo Ramos (leo.ramos@kauel.com)

This work was supported by the Grupo de Investigación en Ingeniería del Transporte (GIIT) at the Universidad Politécnica Salesiana, Ecuador.

ABSTRACT This paper introduces EngineFaultDB, a novel dataset capturing the intricacies of automotive engine diagnostics. Centered around the widely represented C14NE spark ignition engine, data was collected under controlled laboratory conditions, simulating various operational states, including normal and specific fault scenarios. Utilizing tools such as an NGA 6000 gas analyzer and a USB 6008 data acquisition card from National Instruments, we were able to monitor and capture a comprehensive range of engine parameters, from throttle position and fuel consumption to exhaust gas emissions. Our dataset, comprising 55,999 meticulously curated entries across 14 distinct variables, provides a holistic picture of engine behavior, making it an invaluable resource for automotive researchers and practitioners. For evaluation, several classifiers, including logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, and a feed-forward neural network, were trained on this dataset. Their performance, under standard configurations and a simple neural network architecture, offers foundational benchmarks for future explorations. Results underscore the dataset's potential in fostering advanced diagnostic algorithms. As a testament to our commitment to open research, EngineFaultDB is freely available for academic use. Future work involves expanding the dataset's diversity, exploring deeper neural architectures, and integrating real-world automotive conditions.

INDEX TERMS Engine fault, automotive engine, spark ignition engine, fault classification, machine learning, deep learning.

I. INTRODUCTION

Engine malfunction in automobiles is not only a pressing safety concern but also bears significant environmental consequences. As motor vehicles remain a dominant mode of transportation globally, their role in air pollution is undeniable [1], [2], [3]. Particularly, engine faults can lead to escalated emissions of harmful pollutants such as carbon monoxide, hydrocarbons, and carbon dioxide [4], [5].

These amplified emissions further intensify climate change and pose health risks. Alongside the environmental concerns, engine malfunctions can incur high repair costs, unexpected downtime, and in dire scenarios, severe accidents [6]. Therefore, the early and precise detection of motor faults stands paramount.

Traditional techniques for diagnosing motor faults have predominantly revolved around manual inspections, complemented by specialized diagnostic tools [7]. While effective, these methods are often labor-intensive, expensive, and occasionally prone to human errors [8]. The surge of advancements in machine learning (ML) and deep learning (DL) has ushered in a new era, with automated solutions for fault detection and classification emerging as potent alternatives. These cutting-edge techniques promise

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

swift, economically efficient, and exceptionally accurate diagnosis, even for intricate systems with a multitude of variables [9], [10].

However, the lack of comprehensive and well-structured datasets hampers the optimization of ML and DL applications in fault detection [11], [12]. Addressing this pivotal gap, this paper unveils EngineFaultDB, a meticulously crafted dataset, amalgamating data from laboratory-controlled experiments and real-world diagnostics, honed for automotive engine fault classification. This dataset, rich in its composition, consists of 55,999 entries spanning across 14 distinct variables, capturing both standard operational metrics and specific fault conditions.

Coupled with the introduction of EngineFaultDB, this work embarks on a detailed baseline analysis employing a spectrum of ML and DL algorithms, spotlighting the dataset's versatility. This rigorous evaluation not only establishes preliminary benchmarks for the dataset but also illuminates its prowess across various analytical paradigms. Our assessment, which includes a simple yet effective three-layer neural network configuration, further accentuates the depth and utility of EngineFaultDB.

Amidst an ongoing digital transformation, the automotive industry stands at the crossroads of change. In this milieu, EngineFaultDB serves as a beacon for the research community, manufacturers, service technicians, and end-users alike. This dataset fosters a seamless transition from conventional diagnostic paradigms to the realm of advanced analytics, reinforcing its tangible real-world implications.

EngineFaultDB emerges as an invaluable asset for the scientific cohort delving into automotive diagnostics. With its extensive and diverse dataset, it facilitates profound insights, catalyzing innovations in fault detection techniques. This foundational groundwork beckons further exploration, paving the way for real-time diagnostic systems and anticipatory maintenance solutions.

II. DATASET DESCRIPTION

A. DATASET ACQUISITION

1) TOOLS AND EQUIPMENT

The cornerstone of our data collection efforts is the utilization of a C14NE spark ignition engine, which is housed within the vehicle shown in Fig. 1, chosen for its representation of a typical automotive powerplant. This engine, as outlined in Table 1, offers key specifications that align with real-world automotive spark ignition engines, ensuring the relevance and applicability of our dataset. This test engine serves as the heart of our experimental setup, allowing us to simulate various operating conditions and capture a broad range of engine fault scenarios.

FIGURE 1. Car and test engine used for fault data acquisition.

To assess the concentration of harmful emissions in the exhaust gases of the vehicle under study, we employed an NGA 6000 gas analyzer. This analyzer, illustrated in Fig. 2, boasts a valid calibration certificate and employs the non-dispersive infrared absorption (NDIR) method for measuring carbon monoxide (CO), hydrocarbons (HC), and carbon dioxide (CO2). Additionally, it utilizes electrochemical cells for the measurement of oxygen (O2) and nitrogen oxides (NOx). The specifications of the gas analyzer are meticulously detailed in Table 2.

FIGURE 2. Gas analyzer device.

TABLE 1. Test engine specifications.

Specification	Detail
Maximum power	83.7 HP @ 6000 RPM
Torque	113.56 N.m @ 3000 RPM
Displacement	1388 cc
Injection system	Multipoint
Fuel consumption	6.8 l/100 km
Valve configuration	SOHC

TABLE 2. Gas analyzer specifications.

Characteristic	Range	Sensitivity	Unit
Carbon monoxide	0.00 – 0.99	0.01	%
Hydrocarbons	0 – 9999	1	ppm
Carbon dioxide	0.0 – 20	0.10	%
Oxygen	0.0 – 25	0.01	%
Nitrogen oxides	0 – 5000	1	ppm
Lambda	0 – 2.000	0.001	–
Air/fuel ratio	0.0 – 99.0	0.1	–

Facilitating seamless data collection is the USB 6008 data acquisition card (DAQ) from National Instruments, portrayed in Fig. 3. This DAQ card, outlined in Table 3, offers a robust platform for interfacing with various sensors and capturing data with exceptional accuracy. The USB 6008 DAQ card acts as the bridge between our experimental setup and data processing system, ensuring the precision and integrity of the collected data.

FIGURE 3. USB 6008 device.

TABLE 3. USB 6008 data acquisition card specifications.

Specification	Detail
Measurement type	Voltage
Analog input channels	8
Analog input and output resolution	12 bits
Maximum voltage range for analog input	-10V to 10V
Maximum voltage accuracy for analog I/O	7mV
Analog output channels	2
Maximum voltage range for analog output	0V to 5V
Total current capacity	10mA

FIGURE 4. Laboratory setup used for engine testing and data collection.

TABLE 4. Types of fault addressed and their specific conditions.

Fault type	Fault name	Conditions
Fault 1	Rich mixture	Incorrect sensor performance High fuel pressure Defective injector Faulty pressure regulator Clogged air filter Clogged fuel return line
Fault 2	Lean mixture	Incorrect sensor performance Low fuel pressure Defective injector Faulty pressure regulator
Fault 3	Low voltage	Worn spark plugs Faulty ignition cables Defective coil Faulty sensor wiring

2) EXPERIMENTAL DESIGN AND METHODOLOGY

Data acquisition in this study was carried out using a methodology based on the design of experiments and the response surface technique. The primary objective was to assess and record the behavior of the car engine under various operating conditions, including normal conditions and specific faults. All the setup and experimental procedures were meticulously assembled and conducted in a university laboratory setting, shown in Fig. 4. This controlled environment ensured precision in data collection and allowed for the replication of specific conditions, thus enhancing the reliability and validity of the findings.

The faults include rich mixture, lean mixture, and low voltage, each characterized by distinct variations in parameters as outlined in Table 4. It is essential to note that measurements were taken when the sensors were in the appropriate state for accurate data acquisition. In instances where inconsistent or erroneous measurements occurred, they were rigorously screened and excluded from the dataset.

To induce the mentioned faults, controlled adjustments were made to the relevant parameters as outlined above. During each experiment, relevant variables were comprehensively monitored and recorded to capture the intricate nuances of the engine's response to these induced faults.

Once the experimental design was defined, the response surface technique was applied to explore the intricate relationships between the input factors and the engine's responses, including performance parameters and emissions. This approach provided a holistic understanding of how changes in these factors affect the engine's behavior under different fault conditions.

In addition to studying the engine's response to fault conditions, we conducted experiments under standard operating conditions. This allowed us to make a detailed comparison between how the engine performed under typical conditions and how it behaved during the induced fault scenarios. The methodology for the dataset formation is graphically depicted in Fig. 5.

This methodology, founded on the principles of experimental design and the use of the response surface technique, facilitated precise and systematic data collection. It played a crucial role in constructing a comprehensive dataset. This dataset, comprised of meticulously gathered information,

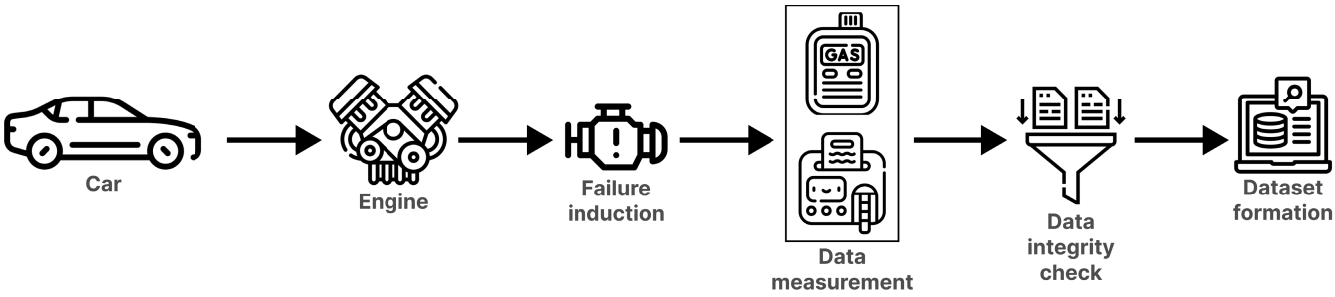


FIGURE 5. Methodology used for the EngineFaultDB formation.

offered valuable insights into the characterization of car engine behavior under different conditions, especially in the context of identifying and documenting engine failures.

B. ATTRIBUTES & VARIABLES

EngineFaultDB stands out as a comprehensive assembly of automotive engine diagnostics, featuring 55,999 meticulously curated entries. The essence of this dataset is captured through its 14 distinct variables, each pinpointing specific metrics and measurements that are imperative for interpreting engine performance and potential malfunctions. Herein, we detail the specific attributes that constitute this dataset:

- **Manifold Absolute Pressure (MAP):** A measure of the pressure within the intake manifold, crucial for the engine's electronic control system to regulate fuel injection and ignition timing, measured in kilopascals (kPa).
- **Throttle Position Sensor (TPS):** Provides information about the position of the throttle, influencing fuel injection, ignition timing, and other engine parameters. It is measured as a percentage.
- **Force:** Represents the engine's torque or rotational force. It is measured in newtons (N).
- **Power:** Quantifies the rate at which work is done or energy is transferred in the engine. It is measured in kilowatts (kW).
- **Revolutions Per Minute (RPM):** Indicates the engine's speed, detailing how many times the engine's crankshaft rotates per minute.
- **Fuel consumption l/h:** Illustrates the engine's fuel consumption rate.
- **Fuel consumption l/100km:** Relays the engine's fuel efficiency over a given distance.
- **Speed (km/h):** The vehicle's travel speed.
- **Carbon monoxide (CO):** CO concentration in the exhaust gases. It is measured as a percentage.
- **Hydrocarbons (HC):** Concentration of unburnt hydrocarbons in the exhaust. It is measured in parts per million (ppm).
- **Carbon dioxide (CO2):** CO2 concentration in the exhaust, indicative of combustion efficiency. It is measured as a percentage.
- **Oxygen (O2):** Oxygen amount in the exhaust, offering insights into the combustion process. It is measured as a percentage.
- **Lambda:** The air-fuel equivalence ratio.
- **Air-Fuel Ratio (AFR):** Ratio of air to fuel in the combustion chambers.

All variables within the dataset are numerical. Metrics such as RPM and fuel consumption are continuous, while certain variables like the engine fault classifications fall into discrete categories: 0 (no fault), 1 (fault type 1), 2 (fault type 2), and 3 (fault type 3). Notably, the latter serves as the target variable. The units adopted for these metrics align with standard automotive diagnostic conventions, with RPM denoting engine speed, L/H and L/100KM measuring fuel consumption, and other variables reflecting standard units for their respective gas concentrations.

III. EXPLORATORY DATA ANALYSIS

A. DATASET DISTRIBUTION

The dataset contains a total of 55,999 entries, classified into four categories representing different fault types. A closer look at the distribution of data points across these fault types reveals nuanced patterns, as shown in Fig. 6. Specifically:

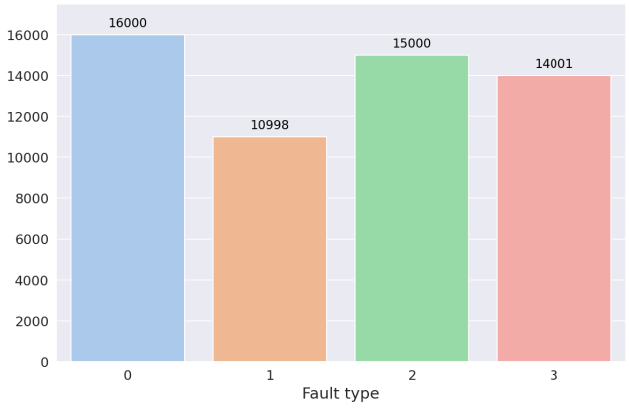


FIGURE 6. Distribution of fault types in EngineFaultDB.

- **No fault (0)**: 16,000 entries, which account for approximately 28.57% of the dataset. This suggests that just

over a quarter of the data represents normal engine operation without any discernible malfunctions.

- **Fault type 1**: 10,998 entries, representing about 19.64% of the total dataset. This is the least prevalent fault type, suggesting its rarity or specific conditions for occurrence.
- **Fault type 2**: 15,000 entries, making up 26.79% of the dataset. It holds a prevalence slightly above fault type 1, but less than the no-fault category.
- **Fault type 3**: 14,001 entries, equating to approximately 25% of the dataset. This suggests that it's relatively common, albeit marginally rarer than fault type 2.

The dataset exhibits a reasonably balanced distribution, with no single fault type overwhelmingly dominating. Such balance is critical, as it ensures that machine learning or deep learning models trained on this data don't develop a bias towards a particular class due to over-representation. Nevertheless, there exists a mild disparity between fault type 1 and the other categories, which researchers should remain cognizant of. While this might pose a minor imbalance, it's also a reflection of real-world scenarios where some faults may be rarer than others.

Given the near-balanced distribution, the dataset stands well-poised for robust model development. Algorithms are likely to have enough data points across categories to discern intricate patterns specific to each fault type. However, for those aiming to achieve the highest classification accuracy, techniques such as oversampling the underrepresented fault type 1 or employing weighted loss functions can be considered to further fine-tune the models.

In essence, the EngineFaultDB dataset, with its balanced fault types, furnishes an optimal foundation for researchers and engineers to sculpt, assess, and hone diagnostic algorithms, ensuring adaptability across diverse engine fault scenarios.

B. DESCRIPTIVE STATISTICS

The descriptive statistics, shown in Table 5, of the EngineFaultDB dataset offers a comprehensive picture of engine behavior and emission characteristics.

The MAP shows considerable variability. The mean of 1.83 kPa, in juxtaposition with the maximum of 4.55 kPa, suggests that the dataset captures not only typical engine operating conditions but also scenarios of heightened load or possible engine stress.

The variation TPS readings is noteworthy. A higher standard deviation relative to its mean emphasizes that throttle positions varied widely during data collection. This variation can hint at the inclusion of diverse driving behaviors, from idling to aggressive accelerations, encompassing a broad spectrum of engine demands.

Force and Power, two interrelated metrics, offer a peek into the engine's performance spectrum. The vast difference between the minimum and maximum values in both parameters indicates that the dataset comprises instances from low-load conditions, perhaps idling or running at steady low speeds, to peak performance scenarios.

RPM metrics underscore the inclusion of different gear ratios and engine speeds, which is crucial for a dataset meant for diagnostics. Engines exhibit different behaviors across the RPM range, so this spread ensures the dataset's versatility.

The fuel consumption metrics, especially the l/100Km readings, which have a wide range, reinforce the dataset's comprehensiveness. Lower readings indicate highway-like efficient driving conditions, while higher values suggest urban or city driving with frequent stops and starts, which are less fuel-efficient.

Emissions data, including CO, HC, and CO2, provides insight into the combustion efficiency and potential faults in the engine or exhaust system. The high variability in HC, for instance, suggests that the dataset has captured conditions from efficient combustion to potential misfires or incomplete burns.

The Lambda and AFR metrics are pivotal in understanding combustion dynamics. Lambda values close to 1 or AFR values near the stoichiometric point (around 14.7 for gasoline engines) indicate optimal combustion. The variability in these readings implies that the dataset includes a mix of optimal and sub-optimal combustion scenarios, essential for training diagnostic models to recognize both standard and faulty conditions.

Overall, the statistics reveal a dataset that captures the depth and breadth of engine behaviors, from diverse driving conditions to potential malfunction scenarios. This richness ensures that models trained on this data will be robust and versatile, equipped to handle real-world automotive challenges.

C. CORRELATION ANALYSIS

Moving forward, Fig. 7 provides a heatmap detailing the correlation coefficients among the various attributes of our dataset. The heatmap was created using Python and the libraries Matplotlib and Seaborn to visualize the Pearson correlation coefficients between the attributes in the dataset. The shades of red signify strong positive correlations, while the deeper blues indicate strong negative relationships. The closer the shade is to white, the weaker or nonexistent the correlation between the paired attributes.

Starting with MAP, it shows strong positive correlations with TPS and Force, with coefficients around 0.88. This consistent relation implies that as the throttle opens, both the intake manifold pressure and the force exerted by the engine increase concurrently.

Force and Power also display a pronounced positive relationship. Their correlation coefficient of approximately 0.62 suggests that as the engine delivers more force, there's a corresponding increase in its power output, albeit the correlation isn't as strong as one might expect.

Diving into combustion metrics, Lambda and AFR show an almost perfect correlation of 1.00. This perfectly linear relationship reflects their intertwined nature in combustion

TABLE 5. Descriptive statistics of the dataset variables.

	MAP (kPa)	TPS (%)	Force (N)	Power (kW)	RPM	C. (l/h)	C. (l/100km)	Speed (km/h)	CO (%)	HC (ppm)	CO2 (%)	O2 (%)	Lambda	AFR
mean	1.83	1.40	286.69	5.66	2398.05	4.50	8.94	51.69	1.93	188.45	13.04	0.59	0.96	14.17
std	0.84	0.91	378.77	7.68	932.01	2.22	3.15	20.14	1.99	111.05	1.05	0.22	0.07	0.97
min	0.45	0.38	2.58	0.47	1066.45	1.92	5.19	22.76	0.42	1.79	8.65	0.20	0.69	10.21
25%	1.22	0.90	76.85	0.99	1830.20	2.98	6.57	39.41	0.64	158.80	12.84	0.41	0.94	13.78
50%	1.54	1.01	92.50	2.40	2105.59	3.82	8.07	45.39	1.13	178.26	13.24	0.53	0.98	14.37
75%	1.94	1.26	257.99	4.70	2761.37	5.14	9.92	59.51	2.46	203.68	13.64	0.79	1.01	14.82
max	4.55	4.05	1537.12	33.95	5013.40	14.81	20.04	107.54	10.13	975.66	15.13	1.15	1.15	16.89

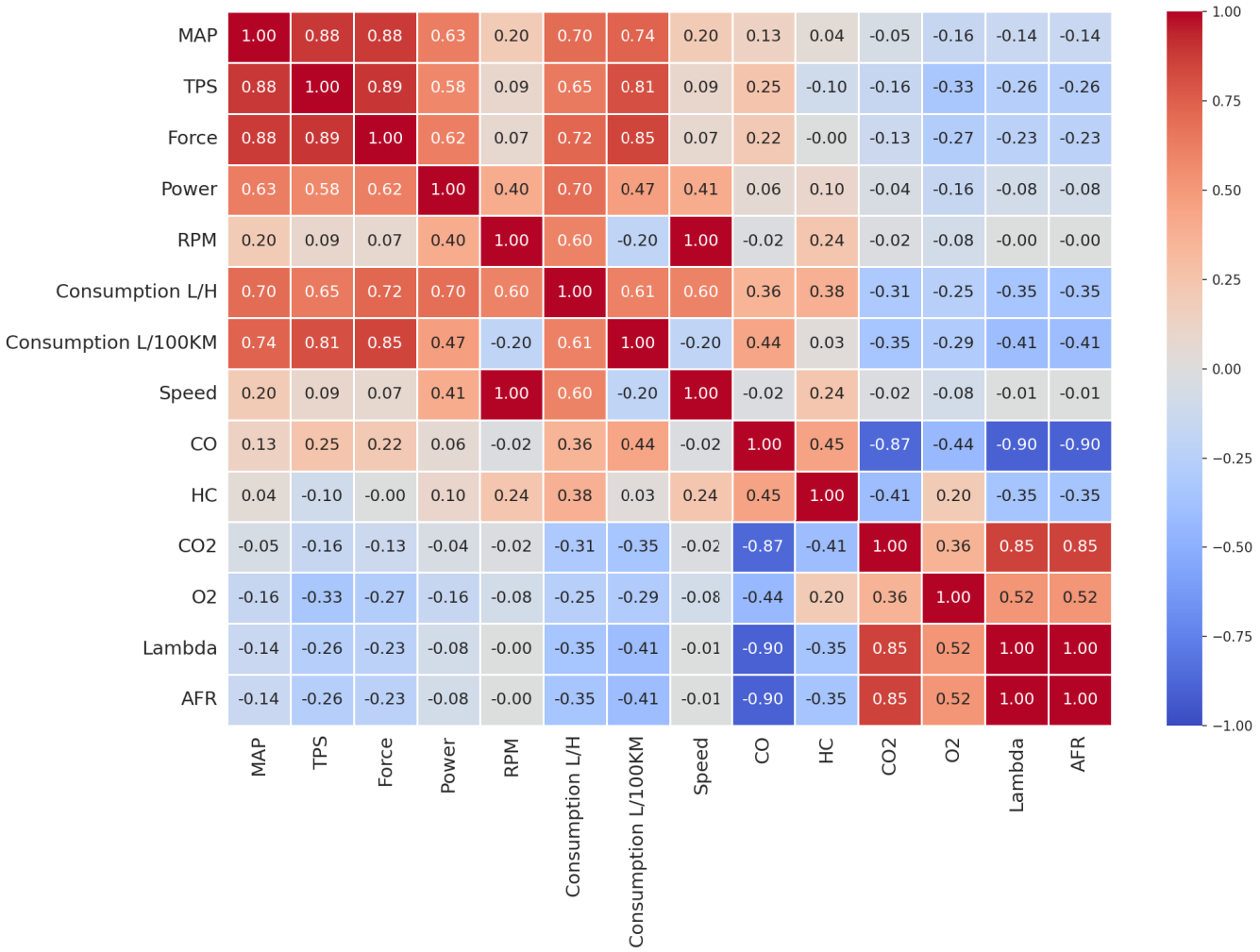


FIGURE 7. Heat map of the correlation of EngineFaultDb variables.

dynamics. Lambda's values directly influence AFR readings, making them crucial in understanding and diagnosing combustion efficiency and potential engine irregularities.

The RPM metric demonstrates moderate correlations with several other parameters. Its association with Power stands out with a coefficient of around 0.40, emphasizing that engine speed has a role in influencing its power delivery. Additionally, the speed of the vehicle has a 1.00 correlation with RPM, suggesting a direct proportional relationship between the two.

Fuel consumption metrics present interesting correlations. While consumption l/h maintains a correlation of approximately 0.70 with MAP, its counterpart, consumption l/100Km, has a correlation of 0.74 with the same variable, suggesting that both metrics, though related, might be influenced differently by intake pressures.

In the domain of emissions, the relationship between CO2 and O2 remains robust, with a correlation coefficient of about 0.85. This emphasizes their reciprocal nature: higher O2 levels might result in increased CO2 emissions due to efficient combustion. HC, on the other hand, don't show strong correlations with most other parameters, pointing to its unpredictable nature in the emissions.

In summary, the correlation heatmap presents a nuanced perspective on the interplay among various engine parameters. This comprehensive view aids in deepening our understanding of engine performance, emissions interactions, and the potential implications of faults, thereby equipping researchers and technicians with valuable insights to address engine anomalies effectively.

D. BOX PLOT ANALYSIS

Box plots, as depicted in Fig. 8, offer a visually informative method to explore data distribution and can be particularly useful in understanding the behavior of various engine parameters across different fault types. They efficiently depict the central tendency, spread, and potential outliers in the data.

First, we have the MAP (Mean Arterial Pressure) in the four types of failures. Fault type 0 (no fault) exhibits a wide dispersion with a median close to 1.5 kPa, demonstrating higher variability in MAP readings under normal conditions. For fault type 1, the distribution of MAP is more concentrated around 1.5 kPa, indicating consistency in the readings.

Fault type 2 presents a median above 1.5 kPa, with occasions where unusually low and high MAP readings are recorded, as indicated by the outliers. Lastly, fault type 3 shows a median similar to fault type 2, with occasional increases in the readings. Faults 2 and 3 exhibit less dispersion in their measurements compared to faults 0 and 1.

Transitioning to the TPS, fault type 0 (no fault) has a slightly wider interquartile range than the others, suggesting varied throttle positions during normal operation. Fault type 1 is depicted with a lower median TPS value, potentially indicating a throttle position pattern associated with this specific fault. Both fault types 2 and 3 show a compact interquartile range. It is important to note that all fault types present outliers, with type 0 having the furthest outliers, suggesting occasionally anomalous throttle readings even when normal conditions are present.

Shifting our attention to the Force, interestingly, in all faults, the median is particularly low, close to 100N, suggesting that under all studied conditions the engines operate without exerting excessive force. Faults 0, 1, and 3 exhibit little dispersion in the measurements, as the spread of their boxes is compact, especially faults 2 and 3. Fault 1 shows greater variability in the upper data due to the extent of the whisker. Additionally, faults 0, 2, and 3 display outliers at the upper end.

Transitioning to Power, a similar pattern to that observed with Force emerges. Fault type 1 exhibits greater dispersion in its measurements, fault 0 appears more compact than fault type 1, and faults 2 and 3 display the most compact interquartile ranges. Additionally, fault type 1 shows more variability in its upper end. Outliers are visible in fault 0, and there is a higher presence of these in faults 2 and 3, extending up to almost 35 kW. The medians of faults 2 and 3 are close to 5 kW, while those of faults 0 and 1 are below 2.5 kW.

Turning our attention to RPM, the fault-free scenario (0) presents a centralized RPM distribution with a median hovering around 2000 RPM, much like fault type 1. Faults 2 and 3 exhibit medians close to 2500 RPM. The distribution of measurements for fault type 2 displays the highest degree of dispersion among all, while the distribution for fault type 1 appears notably more compact. Moreover, this fault exhibits the lowest variability of all, as indicated by its interquartile ranges, particularly the upper whisker, being relatively short. Additionally, there are notably distant outliers observed in faults 0 and 1.

Transitioning to the consumption l/h, the fault-free scenario (0) demonstrates contained consumption, with the majority of values hovering within the range of 3 to 5 l/h, representing typical consumption rates under normal conditions. The median in this fault scenario is close to 4 l/h. Fault type 1 exhibits a median close to 3 l/h, indicating that this particular fault might lead to the engine consuming less fuel per hour, possibly due to inefficiencies or disruptions in the combustion process. Additionally, this fault exhibits the highest degree of dispersion in its measurements, as well as variability and outliers in the upper quartile.

Fault type 2 presents a median above 4 l/h, very similar to what is observed in fault type 3. In fact, fault types 2 and 3 display nearly identical graphs, differing only in the median value. Both faults exhibit compact interquartile ranges and outliers above the upper whisker. This suggests that fault types 2 and 3 have a very similar pattern of fuel consumption in l/h, which presents a challenge for precise classification.

Delving into fuel consumption in l/100Km, the non-fault scenario (0) establishes a benchmark, where most vehicles average fuel consumption between 7 and 12 l/100Km. In fact, the graph for this scenario appears to be perfectly symmetrical. Fault type 1 exhibits a wider interquartile range than the previous fault and a median close to 8 l/100Km. Furthermore, this fault displays the highest variability in the upper whisker.

Faults 2 and 3 prove to be very similar once again, as seen in fuel consumption in l/h. Both have a median just below 8 l/100Km, they have the most compact interquartile range, and the highest number of outliers over the upper whisker. Additionally, it's clear that there is no symmetry as the medians are very close to the third quartile.

Moving on to Speed, in the no-fault scenario (0), most vehicles appear to travel at an average speed ranging between 35 and 60 km/h, indicating regular, everyday driving. For fault type 1, the median is close to the no-fault scenario, around 40 km/h. Furthermore, this fault exhibits a more compact interquartile range. One commonality between these

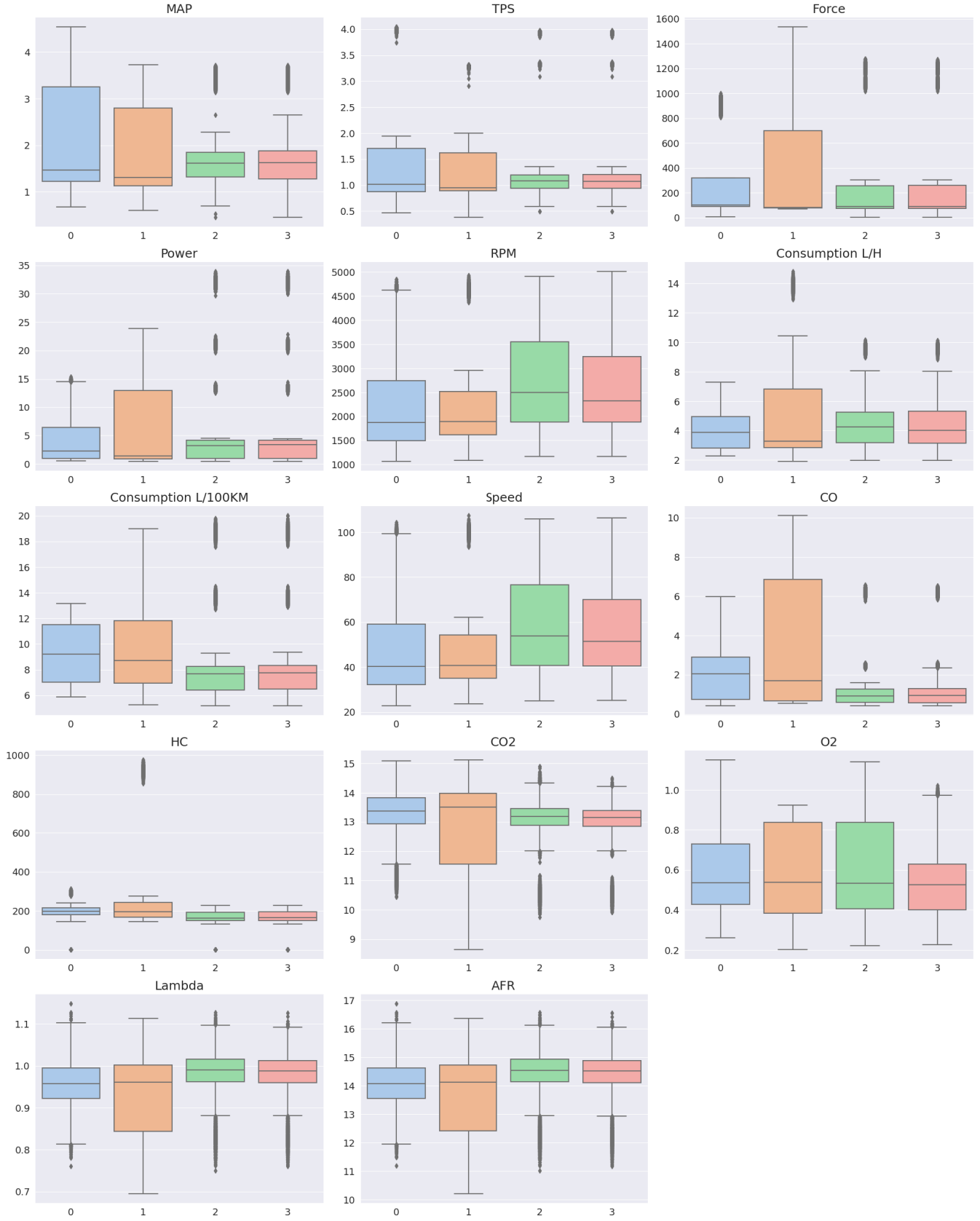


FIGURE 8. Comparative box plots of variables for each fault type.

two faults is the presence of outliers above the upper whisker, with fault type 1 having the highest number of them.

As for fault type 2, it displays a median close to 60 km/h. Moreover, it has the widest interquartile range among all faults. This wide distribution signifies the varied impact of the fault type on speed, possibly indicating sporadic issues that occasionally restrict the vehicle's speed. Fault type 3 also shows a median close to 60 km/h, but the interquartile range is more compact. Additionally, faults 2 and 3 appear to exhibit a similar trend in terms of data variability, as evidenced by their whiskers.

Transitioning to the CO emissions, for vehicles with no detected faults (0), CO emissions remain relatively low, typically below 3%. This represents a standard or reference scenario in which vehicles emit CO within expected levels, ensuring environmental compliance. Fault type 1 exhibits significantly elevated CO emission levels, with an interquartile range between 1 and 7%, approximately. This marked increase suggests that vehicles with this particular fault type may have malfunctioning exhaust systems or issues with fuel combustion, leading to higher carbon monoxide emissions.

Contrastingly, fault type 2 displays considerably low emissions, implying that this fault type may not be directly related to exhaust or combustion inefficiencies. Fault type 3 shows the same behavior. Once again, it is observed that faults 2 and 3 have nearly identical interquartile ranges and medians. Additionally, they are the only faults that exhibit outliers above the upper whisker. This indicates that, once more, these two faults are similar and present a challenge for classification.

Turning our attention to the HC emissions, vehicles with no detected faults (0) emit relatively low levels of HC, with most values clustering below 250 ppm, and a narrow spread suggesting consistent performance among these vehicles. Fault type 1 shows slightly elevated HC emissions, with a broader interquartile range. However, it is the fault that exhibits the farthest outliers, nearly reaching 1000 ppm, indicating a severe malfunction in that particular instance.

Moving on to the next faults, once again, it is observed that fault types 2 and 3 have very similar measurements. Both have a median below 200 ppm and a compact interquartile range. Additionally, they have outliers below the lower whisker, technically within the same range and location. It is important to note that, overall, all four fault types have compact interquartile ranges, and their whiskers are not excessively long.

Shifting our focus to the CO2 emissions, for vehicles with no detected faults (0), CO2 emissions are relatively consistent, with the majority of values centered around 13.5%, and a noticeable degree of symmetry. However, there are outliers below the lower whisker. Fault type 1 exhibits a wide interquartile range without symmetry, along with similarly long whiskers. The median for this fault type is also above 13.5%.

For faults 2 and 3, once again, they display a compact and symmetrical interquartile range, nearly identical, as well as similar medians and whiskers. However, in this case, differences are observed in the outliers. Fault type 2 appears to have a higher number of outliers, both at the lower and upper ends, than any other category.

Observing the O2 emissions, for vehicles with no detected faults (0), the interquartile range extends from approximately 0.4 to 0.8%. The median for this fault, as well as for all fault types, is around 0.5%. The data dispersion for fault type 1 is the second widest of all and lacks symmetry. In other words, there is a wide interquartile range, indicating that this particular fault might be associated with a higher percentage of oxygen in the exhaust.

Fault type 2 displays a slightly more compact interquartile range compared to fault type 1. The most notable difference is observed in the upper whisker, which reaches much higher values than that of fault type 1. Fault type 3 has the most compact interquartile range of all, but it also exhibits outliers above the upper whisker. Interestingly, unlike what was observed with CO and CO2 emissions, fault types 2 and 3 are not identical.

Regarding the Lambda values, fault type 0 demonstrates a median close to 0.95. This group exhibits a relatively consistent dataset, as evidenced by a symmetrical interquartile range. However, there are some outliers both below and above the whiskers, suggesting occasional deviations. For vehicles categorized under fault type 1, the Lambda median is slightly above fault type 0. This group has a somewhat broader data spread compared to fault type 0, indicating greater variability. Interestingly, there are no significant outliers for this category.

On the other hand, vehicles in fault type 2 have their Lambda value median close to 1. The data distribution for this category is quite symmetrical above and below this median, suggesting a balanced dataset for this fault type. Additionally, it has a significant number of outliers, especially below the lower whisker. As for fault type 3, once again the graph is almost identical to fault type 2, only slightly shifted downwards. Furthermore, it appears to have a slightly smaller number of outliers than fault type 2 but still more than faults 0 and 1.

Shifting our focus to the AFR values, vehicles with fault type 0 exhibit an AFR distribution with a median slightly above 14. The spread indicates a relatively consistent AFR around this median, although there are outliers suggesting that some vehicles may occasionally operate outside the typical range. For vehicles with fault type 1, the AFR median appears to be slightly above 14. The distribution is much more dispersed, the most dispersed of all, given the wide interquartile range and lack of symmetry.

Fault types 2 and 3 once again present similar patterns, as both the interquartile range, median, and whisker lengths are similar. The medians are close to 14.5, and both fault types exhibit a significant number of outliers at both the lower and

upper extremes. It is also important to note that these two fault types have the most compact distributions of all.

The analysis of box plots for various performance metrics in vehicles with different fault types provides valuable insights into their behavior. Firstly, it is evident that each fault type exhibits unique characteristics, leading to diverse impacts on vehicle performance. This highlights the importance of accurate fault diagnosis to address specific issues effectively.

Interestingly, some fault types, such as 2 and 3, demonstrate strikingly similar patterns across different metrics. This suggests potential correlations between these fault types or similar underlying causes that affect vehicle behavior. Further investigation is warranted to explore these relationships and their implications.

In conclusion, the box plot analysis provides comprehensive insights into the behavior of different fault types in vehicles, offering valuable information for fault diagnosis, performance optimization, and emissions control in the automotive industry. Further research and targeted strategies are essential to address the unique characteristics and challenges posed by each fault type. The analysis of box plots for various performance metrics in vehicles with different fault types provides valuable insights into their behavior. Firstly, it is evident that each fault type exhibits unique characteristics, leading to diverse impacts on vehicle performance. This underscores the importance of precise fault classification to effectively address specific issues. Notably, fault types 1 and 2, while distinct in their effects, did not exhibit any unusual or unexpected behaviors. Their distinct differences were well-defined and easily distinguishable.

Interestingly, fault types 2 and 3 demonstrate strikingly similar patterns across different metrics. This suggests potential correlations between these fault types or similar underlying causes that affect vehicle behavior. Further investigation is warranted to explore these relationships and their implications.

In conclusion, the box plot analysis provides comprehensive insights into the behavior of different fault types in vehicles, offering valuable information for fault classification, performance optimization, and emissions control in the automotive industry. Further research and targeted strategies are essential to address the unique characteristics and challenges posed by each fault type.

IV. CLASSIFICATION

A. DATA PREPROCESSING

An integral part of any data-driven research, especially when employing machine learning or deep learning techniques, is data preprocessing. The quality and manner in which the data is prepared can significantly impact the performance of classification algorithms. For our dataset, the preprocessing steps were straightforward yet crucial for the subsequent stages of our analysis.

Considering that our dataset is free from missing values, the primary focus during preprocessing was on scaling the data. Features with different scales can influence the model performance, particularly for algorithms that rely on distances or gradients. To ensure uniformity and to prevent any feature from disproportionately influencing the classifiers, we employed the Min-Max Scaler. This technique rescales each feature to a specified range, typically [0,1], maintaining the structure and relationships within the data. The formula for Min-Max scaling is:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where X_{scaled} is the rescaled feature, X is the original feature value, and X_{min} and X_{max} are the minimum and maximum values of the feature, respectively.

By utilizing the Min-Max Scaler, we ensured that all features in our dataset had equivalent influence on the classification models, paving the way for robust and unbiased baseline results.

B. CLASSIFIERS

1) LOGISTIC REGRESSION

Logistic regression (LR) is one of the most straightforward yet powerful algorithms for binary and multi-class classification problems. The model estimates the probability that a given instance belongs to a particular class by applying a logistic function to a linear combination of features [13]. The logistic function ensures that the output falls between 0 and 1, making it interpretable as a probability.

The simplicity of LR makes it highly interpretable and easy to implement, thus serving as an excellent baseline model. However, it might underperform when dealing with non-linearly separable data or complex relationships between features.

2) DECISION TREE CLASSIFIER

The Decision tree (DT) classifier is another supervised learning algorithm used primarily for classification tasks. It works by recursively splitting the dataset based on the value of the selected features, thus constructing a tree-like model of decisions [14]. The algorithm uses metrics such as Gini impurity or information gain to select the feature that produces the most significant separation of classes at each node [15].

The strength of DTs lies in their simplicity and interpretability. The final model can be visualized as a tree, making it easy to understand and explain. However, DTs are highly susceptible to overfitting, especially when the tree is deep, capturing noise in the training data [16].

3) RANDOM FOREST

Random forest (RF) is an ensemble learning method, offering a more robust model by combining multiple DTs. The fundamental idea is to build numerous trees during training and to take the mode of their outputs for classification [13].

RF is particularly effective because it tends to correct the overfitting problem of DTs. The algorithm's inherent randomness makes it robust and flexible, capable of handling both categorical and numerical features efficiently [17].

4) SUPPORT VECTOR CLASSIFIER

Support Vector Classifier (SVC) is a supervised machine learning algorithm that belongs to the family of Support Vector Machines (SVMs). The primary objective of SVC is to find a hyperplane that effectively segregates the data points of different classes in a high-dimensional feature space [18]. The hyperplane is chosen such that it maximizes the margin between the classes. The margin is the distance between the closest data points (or support vectors) of the different classes [19].

Advantages of using SVC include its effectiveness in high-dimensional spaces and its robustness against overfitting, especially in cases where the number of dimensions exceeds the number of samples [20]. However, the algorithm may require careful hyperparameter tuning and can be computationally intensive for larger datasets.

5) K-NEAREST NEIGHBORS

K-Nearest Neighbors (KNN) is a non-parametric, lazy learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point. To classify the new point, it looks at the k nearest labeled data points (neighbors) and assigns the label by a majority vote [13]. Since KNN stores the entire dataset, it's called a lazy learner, it doesn't learn an explicit mapping from the input space to the output space during the training phase but instead waits until classification is required to compute the output [21].

One of the primary advantages of KNN is its simplicity and its ability to adapt easily to changes since there's no explicit training phase. It often performs surprisingly well with sufficient data. However, its performance deteriorates with an increase in the dimensionality of data [22]. Moreover, as the dataset grows, both the space and time complexity can become prohibitive, making preprocessing and dimensionality reduction steps essential.

6) NAIVE BAYES

Naive Bayes (NB) is a probabilistic machine learning classifier based on Bayes' theorem, with an assumption of independence among predictors. In simple terms, it calculates the probability of a particular event based on prior knowledge of conditions related to the event [23]. It's called 'naive' because it assumes that each input feature is independent of the others, which is a naive assumption in real-world scenarios [23], [24].

Naive Bayes is simple, efficient, and scalable, making it ideal for high-dimensional data. While great for text tasks like spam detection, its performance can falter with irrelevant features, as it assumes independent feature contribution.

7) FEED-FORWARD NEURAL NETWORK

Feed-forward neural network is a type of artificial neural networks that consist of an input layer, hidden layers, and an output layer [25]. One of the primary advantages of using feed-forward neural networks is their ability to approximate any continuous function, given a sufficient number of neurons in the hidden layers [25], [26]. This property, known as the universal approximation theorem, makes them highly versatile for a wide range of tasks.

On the downside, the training process involves back-propagation and gradient descent algorithms, which can sometimes be computationally expensive [27]. Moreover, the choice of activation functions, learning rate, and other hyperparameters can significantly impact the network's performance, requiring rigorous experimentation for optimal results. The feed-forward neural network is highly flexible and capable of learning complex mappings from inputs to outputs [28].

C. PERFORMANCE METRICS

1) ACCURACY

It is a commonly used metric for classification problems, offering a straightforward way to measure the overall effectiveness of a model [29]. Mathematically, accuracy is defined as shown in Equation 1, where TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Higher values of accuracy are generally desirable as they indicate a higher proportion of correct classifications. However, it's essential to note that accuracy can be misleading in cases of imbalanced datasets.

2) PRECISION

It is a popular metric often employed to ensure the reliability of positive classifications. Mathematically, precision is defined as per Equation 2, where TP is the number of true positives and FP is the number of false positives. Higher values of precision indicate fewer false positives, making it a crucial metric when the focus is on reducing incorrect positive classifications [30].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

3) RECALL

It is especially crucial in contexts where missing out on a true positive is considerably more problematic than getting a false positive. Mathematically, recall is defined as shown in Equation 3. Here, TP represents the number of true positives, which are the instances correctly identified as the positive class. FN stands for false negatives, which are the instances that belong to the positive class but are incorrectly classified

as negative [31].

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

Higher values of recall are desirable, as this indicates that the model has a lower rate of false negatives. However, achieving a high recall often comes at the cost of more false positives, which is why it is essential to consider it alongside precision [30].

4) F1-SCORE

It serves as a balanced measure between precision and recall [30]. It is particularly useful when we have an imbalanced dataset and when we need a single metric to consider both False Positives and False Negatives. Mathematically, the F1-score is defined as in Equation 4. It takes into account both the precision and the recall values to compute the score.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

A higher F1-score is usually desirable as it indicates a balanced model that takes both False Positives and False Negatives into account [31].

D. WORKFLOW

Initial stages of the workflow were dedicated to data preprocessing. Given our dataset's completeness, the primary step was to scale the features using the Min-Max Scaler. This ensured uniformity across all attributes, making them suitable for classification without any particular feature dominating due to its scale.

Post-preprocessing, the dataset was partitioned into training and testing sets using an 80-20 split. This allocation means 80% of the data was reserved for training the classifiers, while the remaining 20% was set aside for testing.

With the data aptly segregated, the training set was employed to educate our classifiers. It's essential to note that most classifiers were invoked in their standard configurations. This approach adheres to our goal of establishing foundational benchmarks without venturing into hyperparameter optimizations. However, for the feed-forward neural network, which lacks a universally accepted standard configuration for its layers, we chose a straightforward three-layered architecture.

Moreover, given the absence of a standard configuration for the hidden layer in feed-forward neural networks, we utilized an approach from the scientific literature, specifically the rule proposed by Piramuthu et al. [32]. This rule stipulates that the number of neurons in the hidden layer will be $0.5(N_i + N_o)$, where N_i denotes the number of neurons in the input layer, and N_o represents the number of neurons in the output layer. Our rationale for this approach is the absence of a standard, prompting us to explore foundational configurations from the scientific community.

Therefore, the network consists of an input layer containing 14 neurons (aligned with the number of input features) utilizing a ReLU activation function. The output layer consists of 4 neurons, representing our classification categories and employing the softmax activation function for multi-class categorization. In between, there is a hidden layer with 9 neurons ($0.5(14 + 4)$), also utilizing the ReLU activation function. Training of the network was executed using the Adam optimizer and a sparse categorical crossentropy loss function, with batches of 16, over a span of 10 epochs. The graphical representation of the neural network used can be seen in Fig. 9.

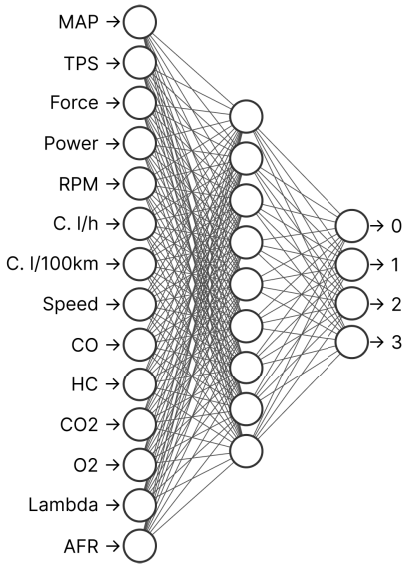


FIGURE 9. Neural network used for classification.

Once trained, each classifier was subjected to performance evaluation on the testing set. A suite of metrics, including accuracy, precision, recall, and F1-score, was utilized to provide a well-rounded perspective on each classifier's performance concerning EngineFaultDB.

Upon obtaining the evaluation metrics, a thorough analysis was undertaken to understand the behavior of each classifier in relation to the novel EngineFaultDB. This step was imperative to ascertain the strengths, weaknesses, and peculiarities of each classifier when confronted with the new dataset, providing insights into how the dataset may influence classification outcomes. The workflow followed for the classification can be seen in Fig. 10.

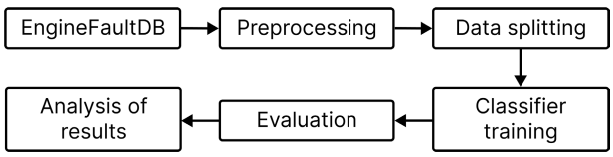


FIGURE 10. Workflow used for fault classification.

E. RESULTS

Table 6 presents the performance metrics of the classification models on the training set of our dataset. From this table,

several notable observations can be made. First, DT and RF have achieved perfect scores across all metrics, indicating a complete fit to the training data. While this might initially seem positive, there's a concern for potential overfitting, where the model could be too closely tailored to the training data, possibly reducing its generalizability on unseen data.

The KNN model also exhibits a strong performance, with accuracy, precision, recall, and F1-score all surpassing 0.84. This suggests that the feature space might have well-segregated clusters that the KNN algorithm can leverage. NB has the lowest performance among the models, particularly in precision and F1-score. This might be attributed to the assumptions that Naive Bayes makes about the data, such as the independence between features, which might not hold true in this dataset.

LR y SVC have moderately acceptable scores, with LR performing below SVC. Given that these are linear models, it could mean that the relationships between variables in the dataset might be non-linear or more complex than what these models can capture.

Lastly, the neural network exhibited a commendable performance across all metrics, specifically ranking as the second-best overall. Intriguingly, the neural network consistently scored 0.748 across all metrics, showcasing its stable and reliable performance in this particular evaluation.

TABLE 6. Performance of classification models on training.

Classifier	Accuracy	Precision	Recall	F1-score
LR	0.580	0.577	0.580	0.578
DT	**1.0**	**1.0**	**1.0**	**1.0**
RF	**1.0**	**1.0**	**1.0**	**1.0**
SVC	0.751	0.772	0.751	0.720
KNN	0.845	0.845	0.845	0.845
NB	0.396	0.374	0.396	0.356
Neural Net.	0.748	0.748	0.748	0.748

Moving forward, Table 7 showcases the performance the classification models evaluated on the testing set of our dataset. Several key insights can be drawn from this table.

LR has maintained a consistent performance from training to testing, suggesting a stable model without significant overfitting. However, its moderate scores indicate that there's room for potential improvement or that the data's structure might be too intricate for a linear classifier.

The performance of DT has shown a drop from a perfect score during training to 0.750 in testing. This decline confirms the earlier suspicion of overfitting during training. A similar pattern is observed for RF, albeit with a slightly lower performance drop, underlining the importance of ensemble methods in mitigating overfitting to some extent.

SVC and KNN both show a performance in the mid-0.74 to 0.75 range. KNN, based on the metrics, exhibits the best performance in testing. However, there is a noticeable drop in performance compared to what was seen in training, which may indicate overfitting. Moving on to the next, although SVC shows a slight drop in scores from training to testing, the minimal decline indicates consistent performance. NB continues to be the weakest performer, reinforcing the notion that its foundational assumptions might not be the best fit for this dataset's structure.

TABLE 7. Performance of classification models on testing.

Classifier	Accuracy	Precision	Recall	F1-score
LR	0.576	0.574	0.576	0.574
DT	0.750	0.750	0.750	0.750
RF	0.748	0.748	0.748	0.748
SVC	0.747	**0.768**	0.747	0.715
KNN	**0.751**	0.751	**0.751**	**0.751**
NB	0.394	0.370	0.394	0.353
Neural Net.	0.749	0.748	0.749	0.748

The neural network model demonstrates the best balance of performance between training and testing sets. Its scores on both the training and testing sets consistently hover around 0.748, showcasing remarkable consistency in its performance. The consistent performance of the neural network model can be attributed to its ability to effectively model the complexities and nuances present in the dataset.

The results from various classifiers, as showcased in Tables 6 and 7, provide comprehensive insights into the proposed dataset's characteristics. There are several significant observations and inferences that can be made.

The disparity in performance across different classifiers suggests that the dataset is multifaceted and captures a rich variety of information. For instance, the more tempered performance of LR could imply that the relationships among features in the dataset are not straightforwardly linear. This non-linearity can be perceived as a strength since real-world data, especially in complex domains such as engine diagnostics, often exhibit non-linear patterns.

The challenges faced by NB further shed light on the dataset. Given that NB assumes feature independence, its less-than-stellar results might indicate that the features in the dataset are interconnected, capturing complex relationships and dependencies. This interdependence is a sign of a dataset that encapsulates the multifarious dynamics of engine performance and faults.

However, the impeccable training results from DT and RF might hint at a different aspect of the dataset. Their capacity to perfectly fit the training data suggests that the dataset might have subsets of data that are uniquely distinct. On the flip side, the drop in their testing performance could indicate that some of these unique data points may not generalize well across all scenarios, highlighting areas where the dataset might benefit from further diversification.

The performance of the feed-forward neural network on our dataset underscores its complexity and depth. Neural networks excel at capturing intricate, non-linear relationships, and the commendable results indicate that our dataset embodies such patterns. Moreover, the modest difference between training and testing scores for the neural network suggests our dataset's robustness against overfitting.

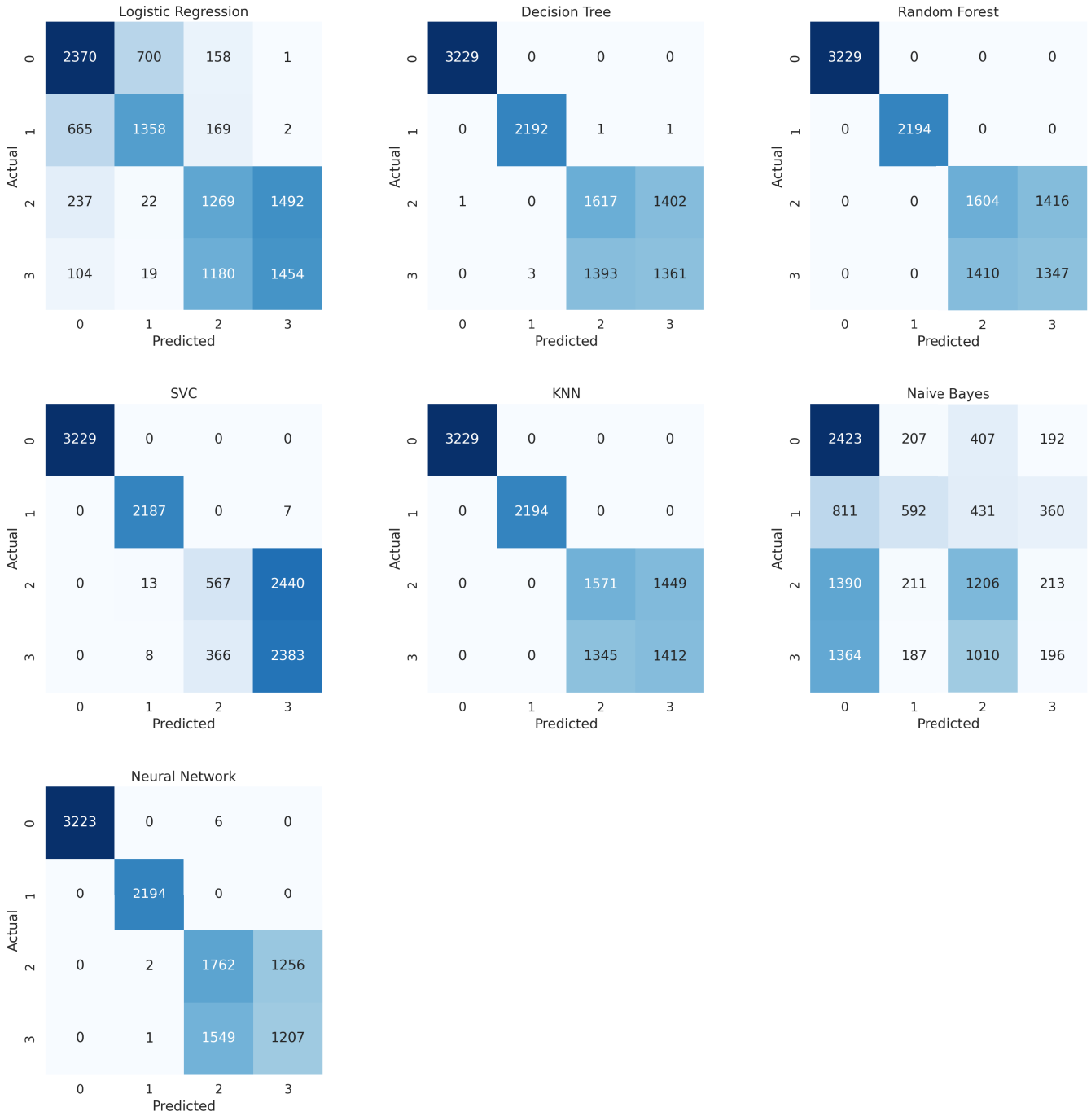


FIGURE 11. Confusion matrices for each classifier employed.

Now, turning our attention to Fig. 11, we can observe the confusion matrices of the models during the testing phase. Regarding LR, this model shows some misclassifications across all fault types. Notably, there's a considerable confusion between faults 2 and 3. However, fault 0 predictions appear relatively strong, with only a minor portion being misclassified as fault 1.

DT displays an impeccable classification for faults 0 and 1 but struggles slightly with differentiating between faults 2 and 3, as evidenced by the off-diagonal values in the corresponding rows and columns. Similarly, the RF perfectly classifies fault 0 and fault 1, with misclassifications for faults 2 and 3. The confusion between these two latter faults is slightly bigger compared to the DT, showing that

ensemble methods might not always guarantee improved class differentiation.

Regarding SVC, like RF, showcases flawless predictions for faults 0 and 1. Faults 2 and 3, however, have some overlap, although the misclassifications are relatively balanced between the two. KNN echoes the patterns observed in SVC and RF. It perfectly identifies faults 0 and 1 and shows limited misclassifications between faults 2 and 3.

NB displays the most diverse confusion across all fault types. While it maintains a reasonable classification rate for fault 0, it struggles considerably with the other faults, especially fault 2, which is spread across all predictions. The neural network demonstrates commendable classification capabilities. Faults 0 and 1 are classified with high accuracy. There is some confusion between faults 2 and 3, but the model seems to handle these better than most of the other classifiers, specially with fault type 2.

Overall, these findings accentuate the dataset's potential for facilitating advanced diagnostic models, reflecting its rich content and reliable structure.

Furthermore, in the majority of classifiers, the balanced performance metrics across precision, recall, and F1-score emphasize the dataset's well-distributed nature. This balanced distribution is promising, as it means that the dataset does not heavily favor one class over another, making it a valuable resource for developing unbiased diagnostic tools.

In sum, the proposed dataset's depth and complexity are evident from the classifier results. It captures a wide array of intricate relationships and patterns, making it a promising tool for research and applications in engine diagnostics. It's worth noting that the challenges posed by faults 2 and 3 are apparent, presenting a clear hurdle for both models and researchers due to their complexity. While the dataset has showcased substantial potential, like any robust dataset, it also highlights areas for continuous improvement and refinement.

V. CONCLUSION AND FUTURE WORKS

In this research, we have meticulously presented and scrutinized a pioneering dataset pertaining to automotive engine performance parameters. The foundation of our dataset lies in the thorough utilization of a C14NE spark ignition engine, incorporated within a representative vehicle. The choice of this engine, backed by its key specifications, assured the dataset's relevance to real-world automotive engines.

The data acquisition tools played an instrumental role in the quality and depth of our dataset. Instruments such as the NGA 6000 gas analyzer and the USB 6008 data acquisition card ensured that data captured was of high precision and encompassed a broad range of engine parameters. The comprehensive methodology founded on principles of experimental design and the response surface technique was pivotal in the data collection process. This structured approach enabled us to simulate and capture various engine conditions, encompassing normal operation and specific induced faults.

Our dataset, christened EngineFaultDB, consists of 55,999 entries, spanning 14 distinct attributes. Each variable within this dataset offers unique insights into engine behavior, performance, and potential malfunction indicators. The array of variables illuminates the intricate workings of an automotive engine and provides invaluable information for diagnostics and fault detection.

The classification results obtained from this dataset provide both validation for its applicability and insights into areas for potential model refinement. While the decision tree and random forest models achieved exemplary training performance, it is evident that the real-world variability and nuances of engine operations could be better captured. This is further substantiated by the neural network model's performance, highlighting that our dataset possesses underlying complexities and patterns that require more sophisticated modeling techniques.

In essence, our EngineFaultDB stands as a testament to the intricate processes of automotive engine operations, bridging the gap between theory and real-world applications. This research not only offers a detailed dataset for academic and industrial pursuits but also sets the stage for future studies in automotive diagnostics, fault detection, and engine performance optimization. Additionally, in our commitment to the academic and industrial communities, the EngineFaultDB dataset has been made available for unrestricted access at a GitHub repository.[1]

We plan to expand our research to curate similar datasets for different contexts, such as diesel engines. We also aim to enrich these datasets with a wider array of fault types and additional variables such as vibration metrics. Such enhancement of the datasets will significantly contribute to the academic and practical understanding of engine diagnostics and maintenance.

The insights garnered from our present research open several avenues for future exploration. One significant area lies in the realm of machine learning and artificial intelligence. Given the intricate patterns and underlying complexities of our dataset, as evidenced by the neural network model's performance, future endeavors could delve deeper into advanced deep learning techniques. Neural architectures like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) might be explored to harness the full potential of EngineFaultDB, possibly leading to more accurate fault detection and predictive maintenance models.

Furthermore, as connected and autonomous vehicles become more prevalent, integrating the findings from our dataset into real-time diagnostic systems could be a groundbreaking advancement. This would entail the development of lightweight, efficient algorithms capable of running on embedded systems within vehicles, offering offering real-time fault diagnosis and potentially preemptive solutions to drivers or autonomous vehicle systems.

[1] https://github.com/Leo-Thomas/EngineFaultDB

Lastly, collaborations with industry partners could lead to the practical application of our findings. By working hand in hand with automotive manufacturers and service providers, the insights from EngineFaultDB could be translated into tangible improvements in vehicle reliability, efficiency, and safety.

REFERENCES

[1] U. Nag, S. K. Sharma, and K. Govindan, "Investigating drivers of circular supply chain with product-service system in automotive firms of an emerging economy," *J. Cleaner Prod.*, vol. 319, Oct. 2021, Art. no. 128629.

[2] B. Vahedi-Nouri, H. Arbabi, F. Jolai, R. Tavakkoli-Moghaddam, and A. Bozorgi-Amiri, "Bi-objective collaborative electric vehicle routing problem: Mathematical modeling and matheuristic approach," *J. Ambient Intell. Hum. Comput.*, vol. 14, no. 8, pp. 10277–10297, Aug. 2023.

[3] M. Asghari and S. M. J. M. Al-e-hashem, "Green vehicle routing problem: A state-of-the-art review," *Int. J. Prod. Econ.*, vol. 231, Jan. 2021, Art. no. 107899.

[4] A. Kozina, G. Radica, and S. Nižetić, "Analysis of methods towards reduction of harmful pollutants from diesel engines," *J. Cleaner Prod.*, vol. 262, Jul. 2020, Art. no. 121105.

[5] S. Dey and N. S. Mehta, "Automobile pollution control using catalysis," *Resour., Environ. Sustainability*, vol. 2, Dec. 2020, Art. no. 100006.

[6] D. H. C. de Sá Só Martins, D. P. Viana, A. A. de Lima, M. F. Pinto, L. Tarrataca, F. L. E. Silva, R. H. R. Gutiérrez, T. de Moura Prego, U. A. B. V. Monteiro, and D. B. Haddad, "Diagnostic and severity analysis of combined failures composed by imbalance and misalignment in rotating machines," *Int. J. Adv. Manuf. Technol.*, vol. 114, nos. 9–10, pp. 3077–3092, Jun. 2021.

[7] Z. Long, X. Zhang, L. Zhang, G. Qin, S. Huang, D. Song, H. Shao, and G. Wu, "Motor fault diagnosis using attention mechanism and improved AdaBoost driven by multi-sensor information," *Measurement*, vol. 170, Jan. 2021, Art. no. 108718.

[8] P. Han, A. L. Ellefsen, G. Li, V. Æsøy, and H. Zhang, "Fault prognostics using LSTM networks: Application to marine diesel engine," *IEEE Sensors J.*, vol. 21, no. 22, pp. 25986–25994, Nov. 2021.

[9] C. Li, S. Zhang, Y. Qin, and E. Estupinan, "A systematic review of deep transfer learning for machinery fault diagnosis," *Neurocomputing*, vol. 407, pp. 121–135, Sep. 2020.

[10] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[11] T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, and E. Xu, "Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions," *ISA Trans.*, vol. 119, pp. 152–171, Jan. 2022.

[12] M. Xia, H. Shao, D. Williams, S. Lu, L. Shu, and C. W. de Silva, "Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning," *Rel. Eng. Syst. Saf.*, vol. 215, Nov. 2021, Art. no. 107938.

[13] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Hum. Res.*, vol. 5, no. 1, Dec. 2020, Art. no. 12.

[14] V. G. Costa and C. E. Pedreira, "Recent advances in decision trees: An updated survey," *Artif. Intell. Rev.*, vol. 56, no. 5, pp. 4765–4800, May 2023.

[15] R. A. Disha and S. Waheed, "Performance analysis of machine learning models for intrusion detection system using Gini impurity-based weighted random forest (GIWRF) feature selection technique," *Cybersecurity*, vol. 5, no. 1, Dec. 2022, Art. no. 1.

[16] A. V. Joshi, "Decision trees," in *Machine Learning and Artificial Intelligence*. Cham, Switzerland: Springer, 2023, pp. 73–87.

[17] M. Li, Y. Zhang, J. Wallace, and E. Campbell, "Estimating annual runoff in response to forest change: A statistical method based on random forest," *J. Hydrol.*, vol. 589, Oct. 2020, Art. no. 125168.

[18] K. Qi and H. Yang, "Elastic net nonparallel hyperplane support vector machine and its geometrical rationality," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7199–7209, Dec. 2022.

[19] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020.

[20] Z. Wang, L. Yao, and Y. Cai, "Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine," *Measurement*, vol. 156, May 2020, Art. no. 107574.

[21] M. Jena, B. Kabi, and S. Dehuri, "A faster lazy learner for data science," *Int. J. Inf. Technol.*, vol. 14, no. 4, pp. 2119–2128, Jun. 2022.

[22] M. O. Arowolo, M. O. Adebiyi, A. A. Adebiyi, and O. Olugbara, "Optimized hybrid investigative based dimensionality reduction methods for malaria vector using KNN classifier," *J. Big Data*, vol. 8, no. 1, pp. 1–14, Dec. 2021.

[23] W. A. van Eeden, C. Luo, A. M. van Hemert, I. V. E. Carlier, B. W. Penninx, K. J. Wardenaar, H. Hoos, and E. J. Giltay, "Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, Naïve Bayes classifier, and traditional logistic regression," *Psychiatry Res.*, vol. 299, May 2021, Art. no. 113823.

[24] R. R. Dalvi, S. B. Chavan, and A. Halbe, "Detecting a Twitter cyberbullying using machine learning," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Madurai, India, May 2020, pp. 297–301.

[25] S. M. Mallikarjunaiah, "A deep learning feed-forward neural network framework for the solutions to singularly perturbed delay differential equations," *Appl. Soft Comput.*, vol. 148, Nov. 2023, Art. no. 110863.

[26] H. Li and L. Zhang, "A bilevel learning model and algorithm for self-organizing feed-forward neural networks for pattern classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4901–4915, Nov. 2021.

[27] C. Li, Z. Wang, and H. Qi, "An efficient pipeline for pruning convolutional neural networks," in *Proc. 19th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Miami, FL, USA, Dec. 2020, pp. 907–912.

[28] E. Noroozi-Ghaleini and M. J. Shaibani, "Investigating the effect of vaccinated population on the COVID-19 prediction using FA and ABC-based feed-forward neural networks," *Heliyon*, vol. 9, no. 2, Feb. 2023, Art. no. e13672.

[29] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[30] E. Casas, L. Ramos, E. Bendek, and F. Rivas-Echeverría, "Assessing the effectiveness of YOLO architectures for smoke and wildfire detection," *IEEE Access*, vol. 11, pp. 96554–96583, 2023.

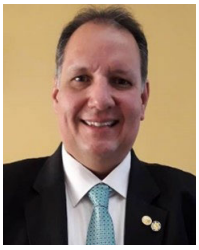
[31] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Niteroi, Brazil, Jul. 2020, pp. 237–242.

[32] S. Piramuthu, M. J. Shaw, and J. A. Gentry, "A classification approach using multi-layered neural networks," *Decis. Support Syst.*, vol. 11, no. 5, pp. 509–525, Jun. 1994.

MARY VERGARA received the bachelor's degree in mechanical engineering from Universidad de Los Andes, Venezuela, in 1991, the Master of Science degree in maintenance engineering, in 1997, and the Ph.D. degree in industrial engineering from Universitat Politècnica de València, Spain, in 2002. She is currently a Distinguished Mechanical Engineer with an extensive academic and research background. Building on her passion for engineering, she pursued further studies, obtaining the Master of Science degree. For over two decades, she was a Dedicated Professor and a Productive Researcher with the School of Mechanical Engineering, University of Los Andes, Venezuela, from 1993 to 2018. During this time, she shared her extensive knowledge with numerous students and contributed actively to the academic community through her research. In 2019, she took on a new role as a Professor and a Researcher with the Automotive Engineering Program, Universidad Politécnica Salesiana, Ecuador, where she continued her work, until 2022. Currently, she holds the position of a Professor and a Researcher with Valencian International University, Spain. Her research interests include failure analysis, design, and simulation in mechanical engineering, resulting in valuable design protections and utility models.

LEO RAMOS (Student Member, IEEE) received the degree (magna cum laude) in information technology engineering with a focus on artificial intelligence from Yachay Tech University, Ecuador, in 2023. He is currently pursuing the Ph.D. degree in computer science with the Computer Vision Center, Universitat Autónoma de Barcelona, Spain. He was a Research Assistant under the supervision of distinguished researchers on multiple occasions. He is also a Senior Research Engineer with Zeus Intelligent Solutions, USA, where he is engaged in the research and development of machine learning and deep learning models for various applications, with a special focus on computer vision. He is also a Researcher with Sleep Care Clinics, U.K., conducting research in artificial intelligence for medical data analysis. In addition, he is also an Applied Research Engineer with Kauel Inc., USA, contributing to the development of cutting-edge intelligent systems for the energy sector. He collaborates with renowned researchers from Venezuela, Spain, France, U.K., and USA, including collaborations with NASA scientists. He has broadened his skill set and network through internships in both industry and academia, nationally and internationally. His research interests include deep learning, computer vision, natural language processing, and data science.

NÉSTOR DIEGO RIVERA-CAMPOVERDE (Member, IEEE) received the master's degree in automotive systems from the National Polytechnic School and the Ph.D. degree in mechanical engineering from the Polytechnic University of Madrid. He is currently an Industrial Technician in automotive mechanics and an Automotive Mechanical Engineer with Salesian Polytechnic University. He serves as the Coordinator of the Research Line on Energy Efficiency and Environmental Pollution with the Transportation Engineering Research Group, Salesian Polytechnic University. He is also a member of the Machine Engineering Division, Higher Technical School of Industrial Engineers. His work involves the application of machine learning for estimating pollutants from mobile sources and their impact on air quality and assessing the influence of various factors on pollutant emissions from different modes of transportation. Currently, he is a Mentor of the Student Chapter of the IEEE Society of Vehicular Technology.



FRANCKLIN RIVAS-ECHEVERRÍA (Senior Member, IEEE) received the B.S. degree in systems engineering, in 1993, the M.S. degree in control engineering, in 1996, the Ph.D. degree in applied science, in 2000, the B.S. degree in law from Universidad de Los Andes, Venezuela, in 2017, and the master's degree in artificial intelligence and machine learning from the University of Purdue, USA, in 2023. He was a full-time Professor and a Researcher with Universidad de Los Andes, Venezuela (1994–2018), Universidad Técnica Federico Santa María, Chile (2018–2022), and Yachay Tech University, Ecuador (2022–2023). Currently, he is a Professor and a Researcher with Pontificia Universidad Católica del Ecuador Sede Ibarra, Ecuador, and MIU City University Miami, USA. He is also a Chief Research Officer with Kauel Inc., USA. His research interests include artificial intelligence and data science applications. He has received several international and national awards including recognition awarded by ''Revista Gerente'' as one of the 100 most successful managers in Venezuela, Halliburton awarded him a recognition for ''Contributions and dedication to the development of petroleum technology,'' Outstanding Professor Award by Universidad Técnica Federico Santa María (2021), Outstanding Leadership Award, Internet 2.0 Conference (2023), and Bicentennial Distinction by Universidad de Los Andes (2023).