IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 April 2024, accepted 20 April 2024, date of publication 2 May 2024, date of current version 10 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3395927

RESEARCH ARTICLE

An Ensemble Deep Learning Model for Vehicular Engine Health Prediction

ISINKA JOSEPH CHUKWUDI[1], NAFEES ZAMAN[2], MD ABDUR RAHIM[3], MD ARAFATUR RAHMAN[1], (Senior Member, IEEE), MOHAMMED J. F. ALENAZI[4], (Senior Member, IEEE), AND PRASHANT PILLAI[1], (Senior Member, IEEE)

[1]School of Mathematics and Computer Science, University of Wolverhampton, WV1 1LY Wolverhampton, U.K.
[2]Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia
[3]Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Pahang 26600, Malaysia
[4]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia

Corresponding author: Md Arafatur Rahman (arafatur.rahman@wlv.ac.uk)

This work was partially supported by the University of Wolverhampton, UK, UK-Saudi Challenge Fund 2023-24 and Researcher Supporting Project number (RSPD2024R582), King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Predictive maintenance has gained importance across various industries, including the automotive sector. It is very challenging to detect vehicle failures in advance due to the intricate composition of various components and sensors. The vehicle's reliability is of utmost importance for ensuring the absence of fatalities or malfunctions to foster economic development. This study introduces an innovative method for developing a predictive framework for vehicle engines with faster and higher decision accuracy. The framework is specifically designed to recognize patterns and abnormalities that may suggest prospective engine problems in real-time and allow proactive maintenance. We assessed the performance of the developed vehicular engine health monitoring systems using a deep learning model based on essential measures like root mean square error, root mean square deviation, mean absolute error, accuracy, confusion matrix, and area under the curve. In this case, the deep learning models are developed by following ensemble techniques using the most prominently used machine learning techniques. Significantly, *Stacked Model 1* outperformed other stacked models (Models 2 and 3) and achieved an impressive AUC value of 0.9702 with a low root mean square error (RMSE) of 0.3355, a high accuracy rate of 0.9470, and a precision of 0.9486. It happens due to the effective incorporation of different approaches into *Stacked Model 1*, which signifies a significant advancement in predicting vehicular engine failures. The model can be used in real-time monitoring systems to continuously monitor the health of vehicular engines and provide early warnings of potential failures, thereby reducing maintenance costs and improving safety.

INDEX TERMS Vehicular engine health monitoring system, machine learning, deep learning, ensemble stacking, vulnerability assessment, decision strategy, micro services.

I. INTRODUCTION

The use of artificial intelligence (AI) and other data-driven methods to realize Industry 4.0 is increasingly being used in the automotive industry around vehicle fault diagnosis systems [1]. The reliability and performance of vehicle engines are critical factors for a safe and effective transportation system. Early detection and diagnosis of engine faults are essential to prevent vehicle breakdowns and reduce maintenance costs. Traditional methods of monitoring vehicle health involve scheduled inspections or reactive maintenance after a failure has occurred, which can be expensive and time-consuming. The emergence of AI and the Internet of Things (IoT) has paved the way for the real-time collection and analysis of substantial sensor data from vehicles. This is called an AI-enabled vehicle health monitoring system (VHMS) [2]. This capability presents prospects for predictive maintenance and fault diagnosis,

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

offering a proactive approach to identify potential problems and ensure the efficient functioning of vehicles [3]. However, analyzing such large datasets and detecting complex patterns and relationships in the data require sophisticated machine learning, deep learning techniques, and an end-to-end data architecture framework. One of the main challenges in vehicular health prediction is the need for high-quality and diverse training data. Acquiring and categorizing extensive data that precisely represents the diverse conditions of a vehicle might pose a challenge due to its inherent difficulty and time-intensive nature [2]. In this case, a novel VHMS was presented in [4]. Also, deep learning has demonstrated significant promise in scrutinizing time-series data and forecasting future values. It is important to note that no singular deep learning model can encompass all the intricacies inherent in engine sensor data. Therefore, there is a need to explore ensemble deep-learning models that can combine multiple models to improve prediction accuracy and robustness [5]. However, for the complete vehicular system, developing a complete VHMS using ensemble deep-learning models is challenging due to its complexities and massive data management, which should be moved forward by considering several subsections like the engine system, transmission system, or vehicular chassis. Then, the fusion of individual outcomes from different subsystems could be used to make better and more robust VHMS decisions [4]. As part of VHMS, Rahim et al. proposed an AI-based vehicular engine health monitoring system (VEHMS) using deep learning (MLP) techniques [2].

However, the decision accuracy of this proposed method for predicting vehicular engine health was within 81%. Additionally, the performance of the proposed models was analyzed based on decision accuracy only instead of an extensive discussion by considering root mean square error, root mean square deviation, mean absolute error, confusion matrix, and area under the curve with computation time. These are significant factors in implementing the VEHMS in real-time to meet Industry 4.0 requirements. Thus, to overcome the addressed problems, we developed a stacked ensemble combining Random Forest, support vector machine,Gradient Boosting, Decision Tree and K-Nearest Neighbors as the final estimator for vehicular engine health prediction. We found that this model can effectively analyze large amounts of engine sensor data and provide accurate and reliable predictions of engine health, enabling proactive maintenance and avoiding costly breakdowns in real-time. Therefore, the contributions outlined in this paper include the following:

- Developed a real-time model that can accurately predict potential failures or issues with a vehicle's engine and classify them into Good, Minimal, Moderate, and Critical.
- Introduced a data pre-processing technique and scaled it by using sklearn, label encoder, and standard scaler to get better outcomes.
- Analyzed and validated the performance of the proposed models based on root mean square error, root mean square deviation, mean absolute error, and area under the curve with computation time.

Furthermore, achieving these goals can be notably enhanced through the utilization of a Stacked Ensemble Model. This model proves effective by leveraging the strengths of various machine learning models to generate predictions that are both more accurate and dependable. The ensemble model integrates predictions from multiple individual models, thereby mitigating errors and biases and ultimately enhancing the overall precision of predictions. The subsequent sections of this study are organized as follows: Section II delves into past related works; Section III elaborates on the methodology; Section IV presents the results and findings; and Section V draws out the conclusion and future research direction.

II. RELATED WORK

Vehicles have been gaining tremendous popularity due to their excellent transport capacity, fast, efficient, flexible, pleasant journey, minimal physical effort, and substantial economic effect [4]. According to Industry 4.0, it is essential to develop a system that monitors and informs the structural condition of a vehicle intelligently so that maintenance expenses can be minimized and longevity can be increased significantly [6]. As a part of these, [7] presented a data-driven predictive maintenance approach for a vehicle powertrain using machine learning algorithms. The approach involves collecting sensor data and using machine learning algorithms to process and analyze the data to predict when maintenance is needed. Similarly, [8] implemented a real-time data processing framework for VHMS that involves collecting sensor data, pre-processing the data, and using machine learning algorithms to analyze and interpret the data. The framework is designed to be scalable and can handle large volumes of data in real-time. Also, [9] suggested a big data analytics approach for VHMS that involves collecting and processing data from various sources, such as sensors and maintenance records. The approach involves using machine learning algorithms as AI to analyze and interpret the data to predict when maintenance is needed. Additionally, [10] presented a data-driven framework for prognostics and health management of power electronics in electric vehicles. The framework involves collecting sensor data, pre-processing the data, and using machine learning algorithms to analyze and interpret the data to predict when maintenance is needed.

In the afore-discussed VHMS, various deep learning techniques were employed in addition to machine learning techniques. Reference [11] work presents a predictive maintenance framework for VHMS that uses deep learning algorithms to predict potential vehicle failures based on sensor data. The framework was designed to improve failure prediction accuracy and reduce maintenance costs. More so, [11] proposed a predictive maintenance approach for electric vehicles that involves using machine learning algorithms to predict potential failures based on sensor data. The approach was designed to improve vehicle reliability and reduce maintenance costs. Reference [12] suggested a predictive maintenance approach for heavy-duty vehicles that involves using machine learning algorithms to predict potential failures based on sensor data. The approach is designed to reduce maintenance costs and improve vehicle uptime. However, [11] implemented a predictive maintenance approach for autonomous vehicles that uses machine learning algorithms to predict potential failures based on sensor data. The approach is designed to improve vehicle safety and reduce maintenance costs.

Due to the complexities of vehicular data and computational time constraints, several studies have recently explored the effectiveness of stacking ensemble models across diverse domains, encompassing finance, healthcare, automotive, computer vision, and more. Nevertheless, the performance of these models can hinge on several factors, including the selection of base models, the quality of training data, and the complexity of the specific problem at hand. For example, [13] compared the performance of different ensemble models, including stacking, bagging, and boosting, for predicting stock prices. The results showed that the stacking ensemble model outperformed the other methods in terms of both accuracy and robustness. In another study, [14] proposed a stacked autoencoder ensemble model for predicting the risk of heart disease. The result showed that the model achieved higher accuracy and AUC.

In the field of computer vision, [15]proposed a novel approach for object detection using a deep learning-based stacking ensemble model. The findings indicated that the model proposed attained higher accuracy in comparison to other cutting-edge models. In a review article by [16], the authors highlighted the advantages of using stacking ensemble models in healthcare applications. They argued that this approach can improve the accuracy and reliability of disease diagnosis, drug discovery, and personalized medicine. Lastly, to predict vehicular engine health in real-time as a part of VHMS, [2] proposed using the conventional stacked ensemble technique, but the decision accuracy bounded within 80.3 % compared with our approach which achieved higher accuracy and AUC.

Also, recently employed conventional stacked ensemble techniques are presented in Table 2, and Table 1 shows a comparative analysis of the existing approaches with our proposed scheme.

However, to meet the requirements of Industry 4.0, it is necessary to enhance the performance of the vehicular engine health monitoring systems (VEHMS) Decision strategy by raising decision accuracy and computational time with other performance criteria. In this case, a stacked ensemble VEHMS combining Random Forest, Support vector machine, Gradient Boosting, Decision Tree, and K-Nearest Neighbors addressed the limitations of the previous study and showed its novelty applicability in the real world.

III. METHODOLOGY

A. VEHMS DECISION STRATEGY

The individual components/points Severity Value of a vehicular engine from [1] is

$$SV_i(t) = S_i(t)W_i\{1-\lambda\}^k \left\{\sum_{j=1}^{m} I_j \times IRI_j\right\} \tag{1}$$

where S_i is data from sensors of individual components or point, λ is the individual components or points degradation performance for the kth 10,000 Km vehicle movement. W_i is the weightage of these components or points, obtained as

$$Wi = \frac{RS_i}{\sum RS_i} \tag{2}$$

using relative significance RSi based on Value Focused Thinking (VFT) from 0 to 10 [2], and $\sum RSi = 1$. I_j is the individual value of various intensifying factors that may affect individual component points, obtained using the Vulnerability Assessment Questionnaire Technique based on an intuitive questionnaire and *CSVS* Qualitative Severity Evaluation Scale [3], [4], with *CSVS* Qualitative Severity Evaluation Scale [5] for rating and indexing as impact severity on engine performances. Here, the rating and indexing scale is 0.0 for No impact, 0 < Low impact $\geq$ 3.99 for, 3.99 < Medium impact $\geq$ 6.99, 6.99 < High impact $\geq$ 8.99, and 8.99 < Critical impact $\geq$ 10. Finally, the intensity of relative importance (IRI_j) of intensifying factors of individual components is given by

$$IRI_j = \frac{RI_j}{\sum RI_j}. \tag{3}$$

RI_j is calculated based on *VFT* [2] from 0 to 10, and $\sum Wi = 1$. is the ratio between the relative importance of an intensifying factor to the total relative importance of all intensifying factors of individual VC/VP.

In addition, $\sum_{j=1}^{m} I_j \times X \times IRI_j$ is vulnerability affecting value, and $Wi \times \left(\sum_{j=1}^{m} I_j \times IRI_j\right)$ is the VC/VP's vulnerability value. However, equation (1) helps to define the health of an individual VC/VP by following the stated condition with the threshold as follows for the t time interval.

$$\begin{cases} \text{Critical}, & \text{if } SV_i(t) \geq TH_C \\ \text{Moderate}, & \text{if } TH_C < SV_i(t) \geq TH_M \\ \text{Minor}, & \text{if } TH_M < SV_i(t) \geq TH_{MN} \\ \text{Good}, & \text{if } SV_i(t) < TH_{MN} \end{cases} \tag{4}$$

The ***Threshold*** (TH) as TH_C for ***Critical problem***, TH_M for ***Moderate problem***, TH_{MN} for ***Minor problem*** and TH_G for ***Good condition***. Since the above-stated boundary limits define the individual VC/VP health conditions in the engine subsystem and for t time, the following matrix from equation (1) provides the decision about whole engines by aggregating column-wise of stated elements of Table 3, i.e.

$$\sum VEHMS_{\mathrm{D}}(t) = \sum SV_{iC} + \sum SV_{iM} + \sum SV_{iMN} + \sum SV_{iG} \tag{5}$$

TABLE 1. Comparative analysis of the existing approaches with the proposed scheme.

Ref	Aim/Objective	Pros	Cons
[17]	An Optimal Stacking Ensemble for Remaining Useful Life Estimation of Systems Under Multi-Operating Conditions	Reduce the influence of multi-noise and time-varying environments and improve accuracy.	High computing cost and limited to lithium-ion battery dataset.
[18]	Proposed a Blockchain and AI-based Engine Fault Detection Scheme for Autonomous Vehicles	Improved accuracy and secure data transfer.	Other performance measurements are not considered for comparison.
[19]	Ensemble learning for Fault Condition Prediction and Health Status Monitoring in Military Ground Vehicles.	Improve prediction accuracy.	Fewer variable features were used in the training dataset and not targeting the engine critical component.
[20]	Predicting Vehicle Behavior Using Multi-task stacked Ensemble Learning	Improve prediction accuracy.	No specific area of focus in vehicular behaviors prediction.
[11]	Survey paper on ML-based PdM for automotive systems	Combining data from multiple sources can improve accuracies and enable new applications.	Literature survey did not allow for a quantitative comparison of the results due to entirely different problem settings and datasets.
[13]	Proposed a Tree-based ensemble models and deep learning algorithms	An improved Stacking framework for predicting the stock price index direction.	Limited to three major U.S. stock indices.
[14]	Proposed a RBPs prediction tool based on deep learning and ensemble learning	Shows promising prediction using stacked ensemble classifier composed of bidirectional long short-term memory (BiLSTM), gated recurrent unit (GRU), and support vector machine (SVM).	Limited to biological processes for RNA synthesis, protein folding, and alternative splicing.
[21]	An ensemble stacked generalization (ESG) approach for better prediction of electric vehicles (EVs) energy consumption	Improve prediction accuracy.	Limited to energy consumption of electric vehicles.
[22]	Machine learning models applied to predictive maintenance in automotive engine components	The applicability of machine learning methods, trained with simulation data, can be used in predictive maintenance to recognize failures in automotive engine components.	The computational cost is also a limiting factor for real-life applications, which may lead to unfeasibility depending on the embedded technology used on such applications.
[2]	An Intelligent Risk Management Framework for Monitoring Vehicular Engine Health	An efficient Decision model for monitoring and diagnosing vehicular engine health and condition in real-time using vulnerable components.	80.3% decision accuracy for the 80% training data.
Proposed	Proposed an Ensemble Deep Learning Model for Vehicular Engine Health Prediction	Improved accuracy and Vehicle engine prediction architecture.	-

TABLE 2. Recently employed various conventional stacked ensemble techniques.

S/No	Ref	Item	Technique	Model
1	Huang et al	Stacked Ensemble	Stacking, Bagging and Boosting	Diagnosis Prediction
2	Yu, Y. et al	Data Collection and Process	ML Algorithms	Engine Health Management Prognostics
3	Park, J. et al	Data Storage	Hadoop, Distribute File System, ML Blockchain, Cassandra	Heterogenous Data Storage
4	Park, J. et al	Predictive Analysis	ML Algorithms	Predictive Maintenance
5	Al Nasseri, A. et al	Application	ML Algorithms	Fault Diagnosis, Real Time Monitoring

TABLE 3. Complete engine decision matrix.

Critical($SVic$)	Moderate($SViM$)	Minimal($SViMN$)	Good($SViG$)
$SV1c$	$SV1Mc$	$SV1MN$	$SV1G$
$SV2c$	$SV2Mc$	$SV2MN$	$SV2G$
$SV3c$	$SV3Mc$	$SV3MN$	$SV3G$
$SV4c$	$SV4Mc$	$SV4MN$	$SV4G$
$SV5c$	$SV5Mc$	$SV5MN$	$SV5G$
$SV6c$	$SV6Mc$	$SV6MN$	$SV6G$
$SVnC$	$SV4nM$	$SVnMN$	$SVnG$

where, engine condition is

$$\begin{cases} \text{Critical}, & \text{if } \sum VEHMS_D(t) \geq TH_{DC} \\ \text{Moderate}, & \text{if } TH_{DC} < \sum VEHMS_D(t) \geq TH_{DM} \\ \text{Minor}, & \text{if } TH_{DM} < \sum VEHMS_D(t) \geq TH_{DMN} \\ \text{Good}, & \text{if } \sum VEHMS_D(t) < TH_{DMN} \end{cases}$$

Here, the ***Threshold*** (TH) as TH_{DC} for ***Critical***, TH_{DM} for ***Moderate***, TH_{DMN} for ***Minor*** and TH_{DG} for ***Good*** condition of engine. However, this Decision strategy can decide whether the engine health as good condition or has a minor, moderate, or critical problem by utilising the stated thresholds. Another notable matter is that i denotes the individual number of VC/VP, and j denotes the number of intensifying factors for i-th VC/VP.

B. VEHMS DATA PROCESSING AND ANALYTIC MODEL ARCHITECTURE

The architecture of the VEHMS data life-cycle for predicting real-time vehicle engine health using a newly developed stacked ensemble combining KNN, SVM, RF, Ada, and XGB is presented in Fig.1. However, the proposed architecture is presented as follows:

C. VHMS DATA FLOW

Overheating, piston, misfire, starter, and lubricant information and baseline conditions need to be recorded from vehicle engine sensors and communicated to IoT Hub via a cellular-enabled device, which was discussed in detail in [2]. However, in this case, we introduced Stream Analytics which could receive the message from IoT Hub in real-time, process it according to the business logic, and transmit the data to the serving layer for storage. Depending on the data, several databases are used. The messages are stored in Azure Cosmos

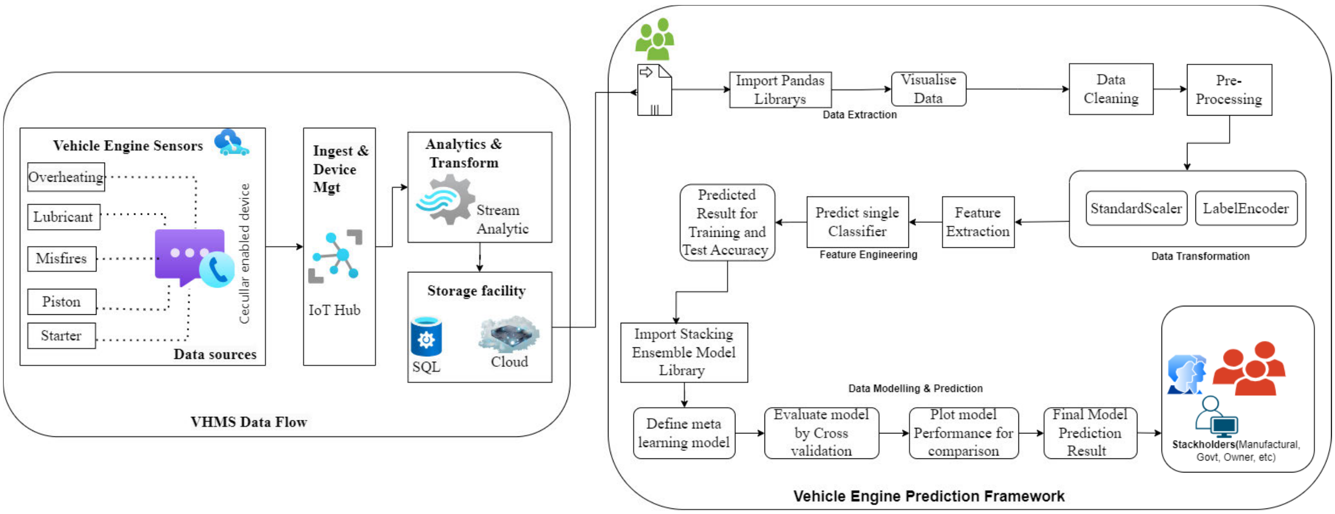


FIGURE 1. Vehicle engine prediction architecture.

TABLE 4. Dataset head before LabelEncoder.

Crankshaft	Overheating	Lubricant	Misfires	Piston	Starter	Decision
8.296	12.333	10.242	4.778	8.140	377.835	Minor
817.231	15.070	1249.033	4.843	7.806	3.778	Critical
7.354	11.929	12.492	3.232	717.916	3.382	Moderate
9.387	11.499	12.459	4.164	8.301	437.046	Moderate
8.052	14.746	11.962	5.337	7.055	483.639	Moderate

TABLE 5. Dataset tail before LabelEncoder.

Crankshaft	Overheating	Lubricant	Misfires	Piston	Starter	Decision
8.051	15.713	9.394	5.122	8.573	520.98	Moderate
10.97	11.024	12.57	4.661	10.373	2.914	Good
8.877	14.445	12.249	5.441	8.14	501.867	Moderate
9.52	11.596	12.749	4.379	8.682	378.091	Minor
8.683	13.312	14.609	5.122	8.299	4.367	Minor

DB, whereas SQL DB contains relational data and serves as a data source for the display and action layers.

D. DATA EXTRACTION

Vehicle engine performance analysis and predictive maintenance necessitate access to accurate and relevant data from engine systems [23]. Quickly extracting data from car engine systems is critical for generating useful insights for performance analysis and predictive maintenance [11]. There are various methods and techniques for data extraction from vehicle engine systems, such as onboard diagnostics (OBD) ports, Controller Area Network (CAN) bus, and sensor-based methods; however, this paper uses a sensor-based method considering data quality, data availability, and data security [24].

In implementing an Ensemble Deep Learning Model for Vehicular Engine Health Prediction, the quality and quantity of the extracted data play a pivotal role. The DataFrame dimensions (3003, 7) indicate that the dataset comprises 3003 rows with 7 columns (features). However, several potential limitations in the data collection process could impact the model's effectiveness. In this case, Sample Size of 3003 of 7 instances, the dataset size may be relatively small for training ensemble deep learning models effectively. Adequate sample size is essential to prevent overfitting and ensure model generalization to unseen data. Techniques like data augmentation or transfer learning may help address this limitation.

E. DATA TRANSFORMATION

The process of data transformation plays an important role in both data analysis and machine learning processes. This is because the quality and suitability of the data for analysis significantly influence the accuracy and reliability of the results obtained. [25]. Proper data transformation techniques can improve data quality, reduce noise, deal with missing data, and allow data to be successfully processed for decision-making [26]. In this case, we used StandardScaler as a data transformation and data pre-processing technique to scale and normalized the dataset features to a standard scale to ensure that each feature has a mean of 0 and a standard deviation of 1. It is a popular pre-processing procedure

TABLE 6. Dataset head after converting to dummy variables.

Crankshaft	Overheating	Lubricant	Misfires	Piston	Starter	Decision
-0.156146	-0.205369	-0.358201	-0.396882	-0.257603	0.860096	2
6.031942	-0.194697	2.864693	-0.396495	-0.259228	-0.920210	0
-0.163352	-0.206944	-0.352347	-0.406088	3.195291	-0.922095	3
-0.147800	-0.208621	-0.352433	-0.400538	-0.256820	1.141908	3
-0.158013	-0.195960	-0.353726	-0.393553	-0.262881	1.363665	3

TABLE 7. Dataset tail after converting to dummy variables.

Crankshaft	Overheating	Lubricant	Misfires	Piston	Starter	Decision
-0.158020	-0.192189	-0.360406	-0.394833	-0.255496	1.541387	3
-0.135691	-0.210473	-0.352144	-0.397578	-0.246739	-0.924322	1
-0.0151701	-0.197134	-0.352979	-0.392933	-0.25760	1.450421	3
-0.146783	-0.208242	-0.0351678	-0.399257	-0.254966	0.861314	2
-0.153185	-0.201551	-0.346839	-0.394833	-0.256829	-0.917406	2

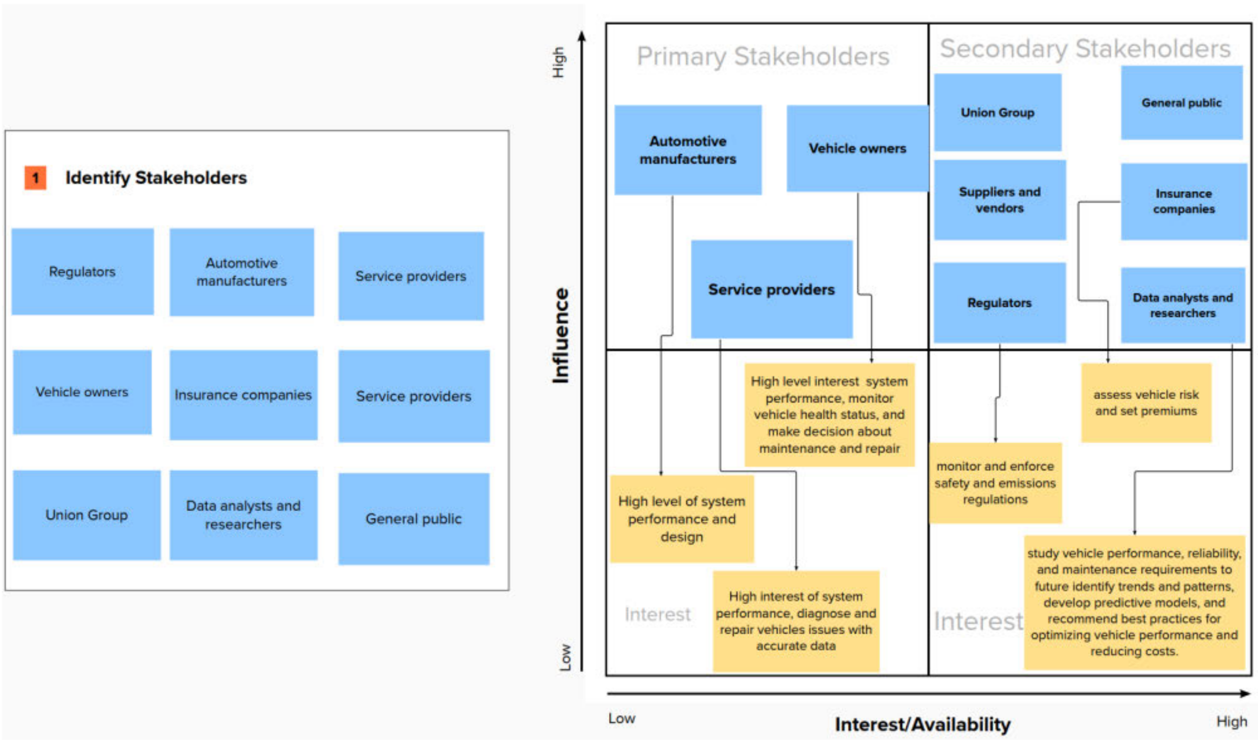


FIGURE 2. VHMS stakeholder management framework.

used before applying ML/DL algorithms to numerical data analysis. Also, we Encoded the Targeted Variable (Decision) with the LabelEncoder. This is appropriate for classification tasks where the target variable is categorical as shown in Table 4 and Table 5. It is a pre-processing technique that assigns a unique numerical label to each distinct category in a categorical feature, making it easier for ML/DL algorithms to interpret such data [27]. furthermore after encoding the target variable, the categorical variables in the feature matrix are being converted into dummy variables using *pd.get_dummies*(), see Table 6 and 7 This ensures that each categorical variable is represented as a set of binary columns, where each column indicates the presence or absence of a particular category.

F. FEATURE ENGINEERING

This stage holds substantial significance within the machine learning pipeline, as the excellence and pertinence of features can markedly influence the effectiveness and interpretability of ML/DL models. Properly built features can lead to more accurate predictions, enhanced model interpretability, and faster model training times. However, this study considered

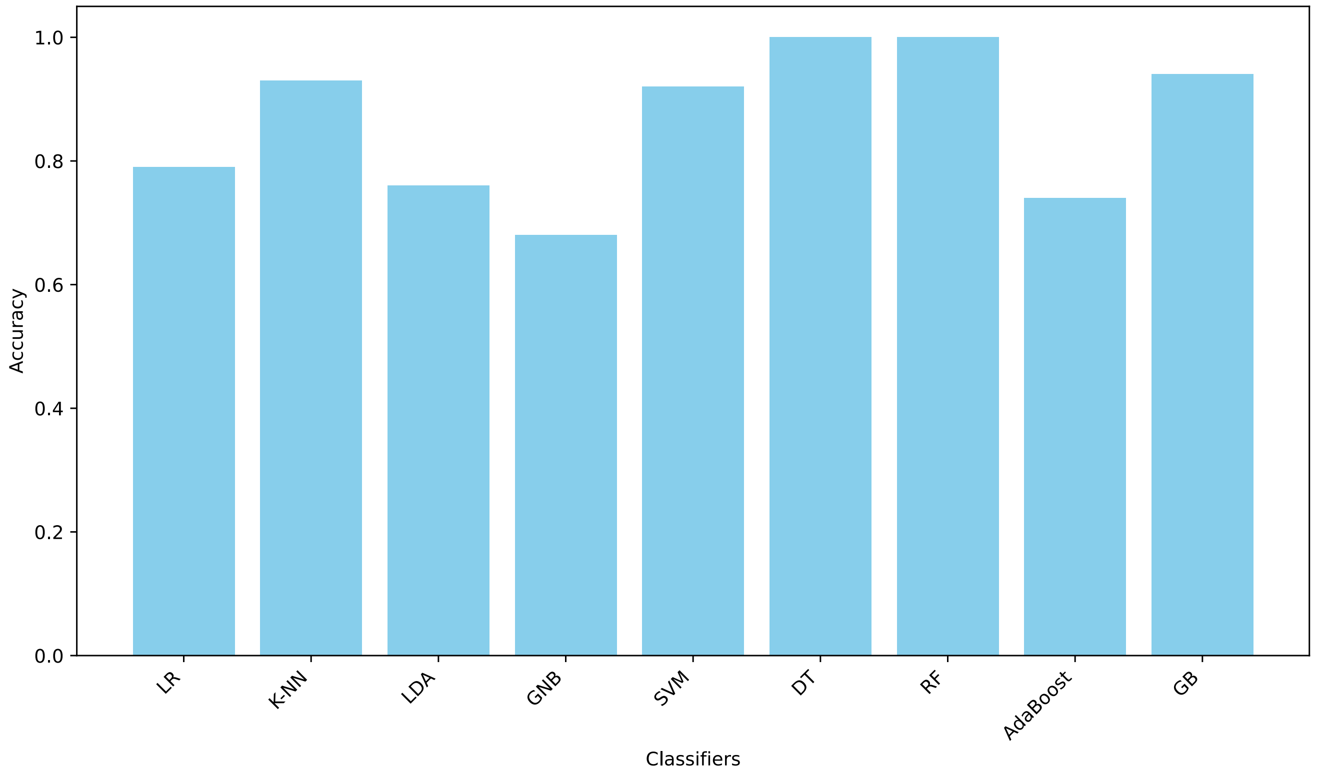


FIGURE 3. Plotting training set accuracies.

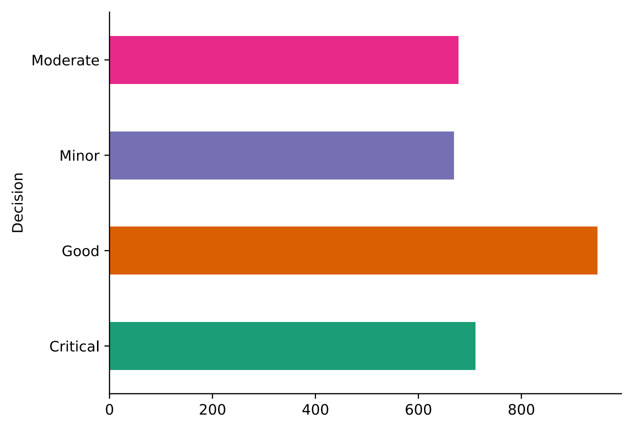


FIGURE 4. Decision strategy (A).

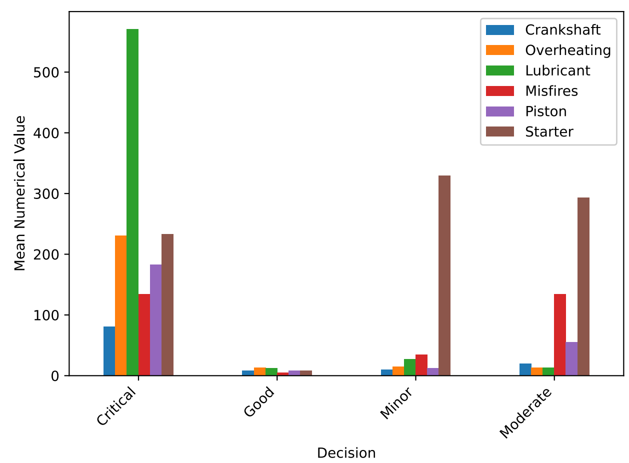


FIGURE 5. Decision strategy (B).

only 7 features in the DataFrame see Fig. 10 shows the feature prediction's scatter matrix for each input variable for the feature prediction [28]. Given the constraint of a small number of features, it's crucial to carefully analyze the existing features and potentially engineer new ones that might provide additional predictive power.

G. DATA MODELLING AND PREDICTION

In stack ensemble techniques, data modeling and prediction often entail integrating the predictions of numerous base models (also known as weak models or base learners) to get a more accurate and robust prediction [29]. The outputs of the base models are utilized as input features for a higher-level model, known as the meta-model, which subsequently produces the ultimate prediction [30].

Assume we have N base models represented by $h1, h2 \ldots,$ hN and a meta-model represented by H. X represents the input data used to train the basic models, and y represents the associated output (i.e., target variable). The base models'

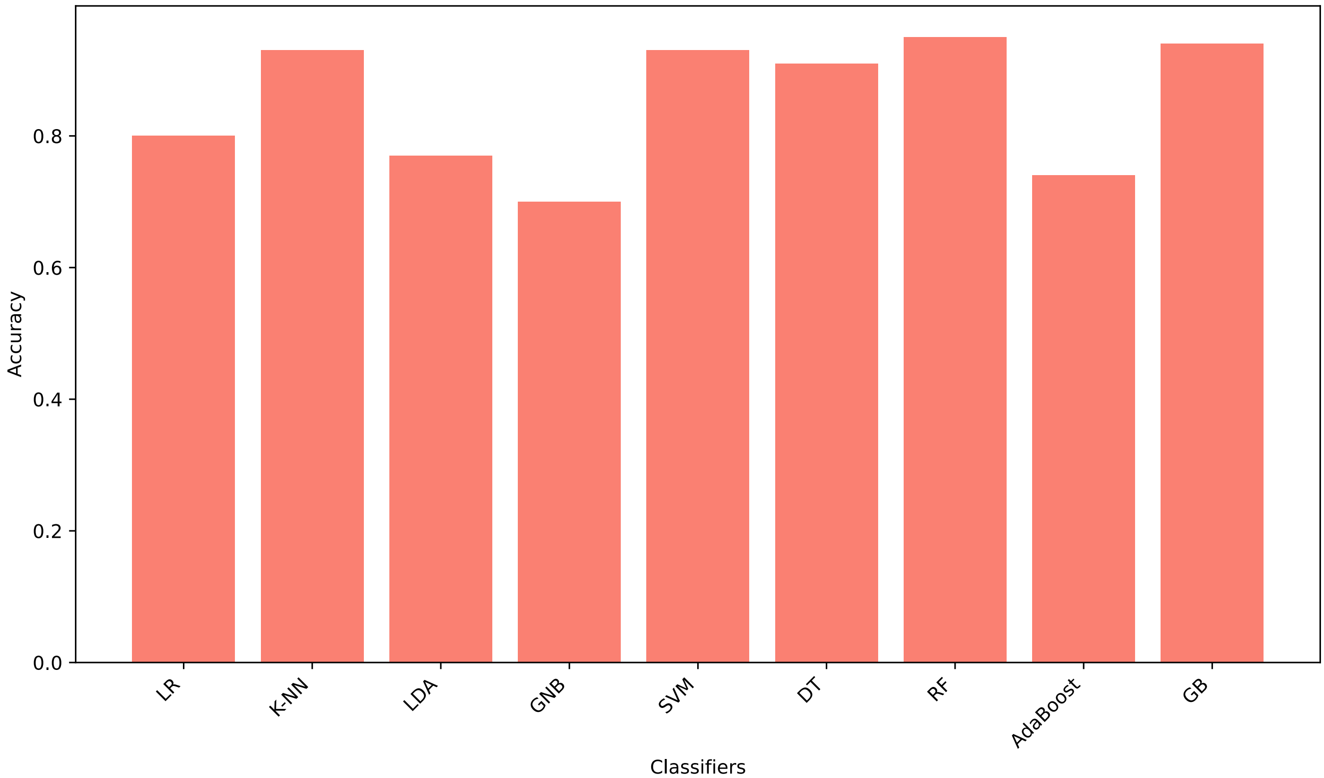


FIGURE 6. Plotting test set accuracies.

predictions for a given input X denoted by $y1, y2 \ldots, yN$ respectively. These predictions are then combined to generate a new feature matrix, designated by *Xmeta*, which is fed into the meta-model H. The final prediction is made by the meta-model H, which is denoted by *Ymeta*.

The equations are presented below:

$$\begin{cases} y_1 = h_1(X), \quad y_2 = h_2(X), \quad \ldots, \quad y_N = h_N(X) \\ Xmeta = [y_1, y_2, \ldots, y_N] \\ Ymeta = H(X_{\text{meta}}, y) \end{cases} \tag{6}$$

H. STAKEHOLDER

A vehicle health monitoring system could benefit from their support and input if stakeholders are managed properly. Stakeholders can be classified based on their function and interest [31]. Also, effective stakeholder management is a vital success factor in the automotive business. It entails identifying stakeholders, prioritising their problems, designing a communication strategy, and measuring stakeholder satisfaction. [32]. Regular stakeholder meetings, collaborative decision-making, and effective communication channels can assist in overcoming stakeholder management difficulties and accelerate industrial innovation, which is presented in Fig 2.

I. PERFORMANCE EVALUATION

We evaluated individual ensemble approaches using a range of performance metrics, including RMSE, RMSD, MAE, Root square error, Accuracy, Confusion Matrix, and AUC respectively, see Table11.

J. DEFINITION OF EVALUATION METRICS USED

Root Mean Square Error(RMSE): is a metric that quantifies the average magnitude of the discrepancies between projected and actual values. The algorithm computes the square root of the mean of the squared discrepancies between the expected and actual values. Smaller RMSE values suggest superior model performance, as they correspond to less prediction errors.

Root Mean Square variation(RMSD): is a metric that quantifies the average variation between projected values and actual values, similar to RMSE (Root Mean Square Error). The statement describes how the measure calculates the extent to which data points deviate from the regression line. Smaller RMSD values suggest superior model performance, as they reflect reduced variability in prediction accuracy.

Mean Absolute Error (MAE): The Mean Absolute Error is a metric that quantifies the average absolute deviation between projected and actual data. It offers a more understandable and precise measure of error in comparison to RMSE. Smaller Mean Absolute Error (MAE) values imply superior model performance, as they correspond to reduced average prediction errors.

Accuracy: Accuracy is a metric that quantifies the ratio of accurately predicted instances to the total number of

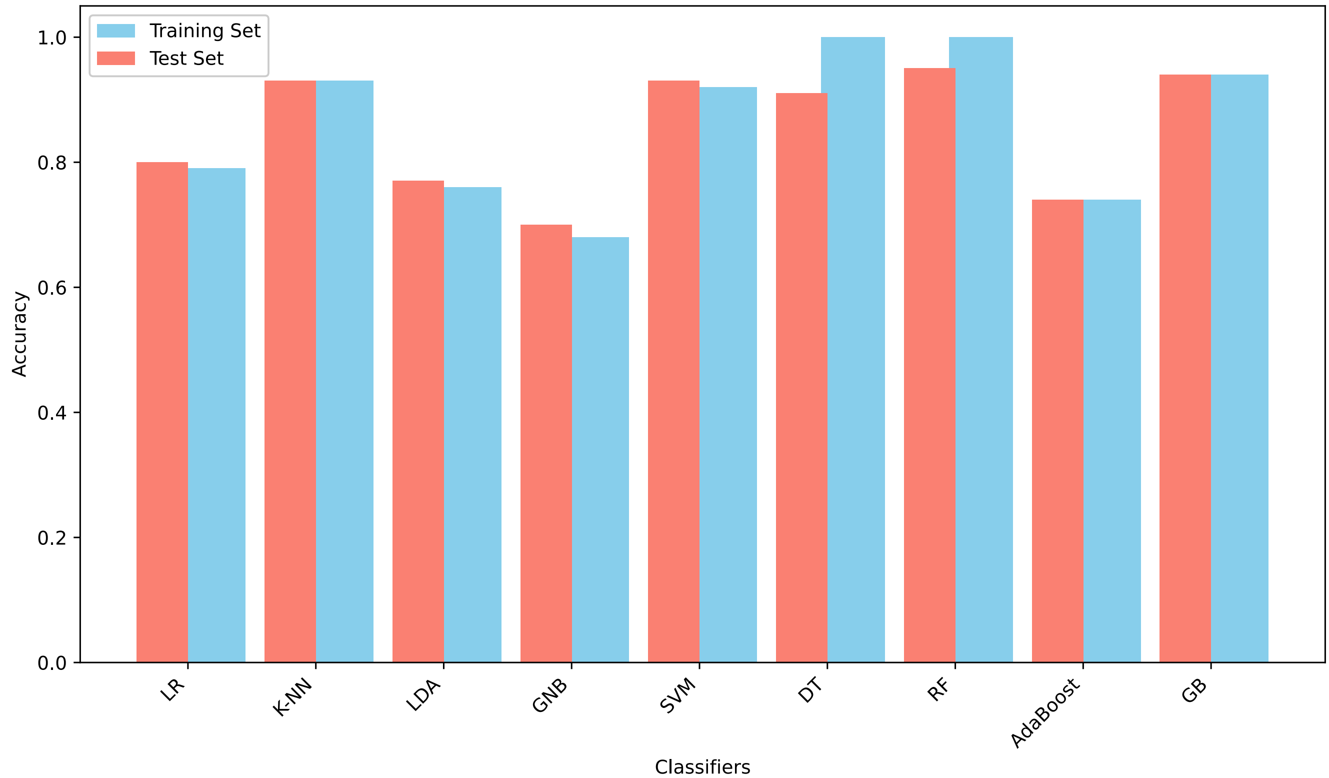


FIGURE 7. **Comparative accuracies of different machine learning.**

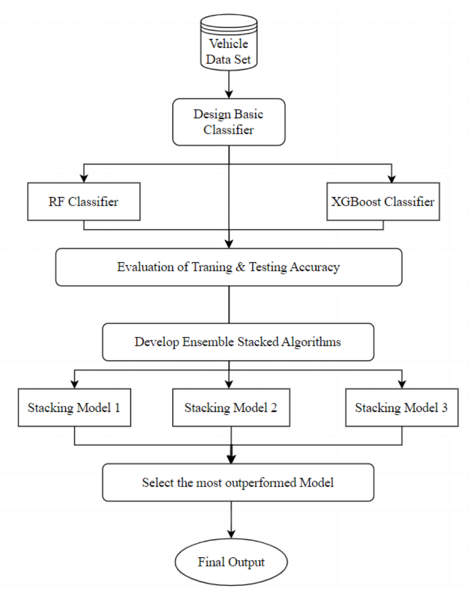


FIGURE 8. **The architecture of the proposed stacked ensemble technique.**

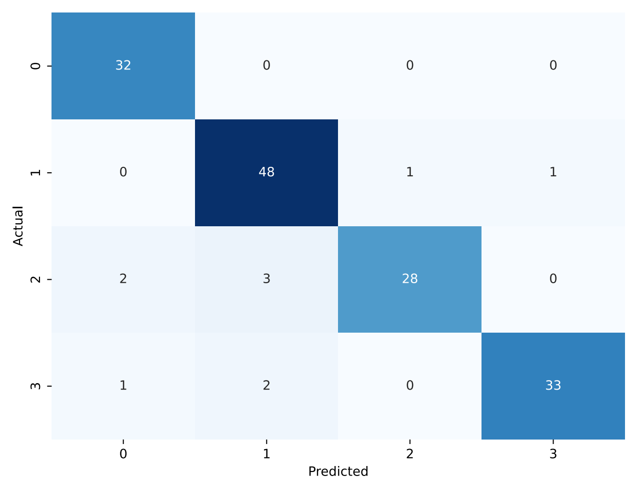


FIGURE 9. **Classification report for ML: Confusion matrix.**

instances. It is frequently employed in classification tasks and offers a comprehensive evaluation of model performance. Greater accuracy values indicate superior model performance in accurately classifying engine health statuses.

confusion matrix: A confusion matrix is a comprehensive representation of a model's predictions compared to the actual results for several classes. The framework is comprised of four quadrants, namely true positive, true negative, false positive, and false negative. It aids in assessing the model's effectiveness in terms of its ability to accurately classify and detect any instances of misclassification or biases.

Area Under the Curve (AUC): The AUC is a metric that quantifies the performance of a classification model across all possible categorization levels. The plot illustrates the relationship between the true positive rate and the

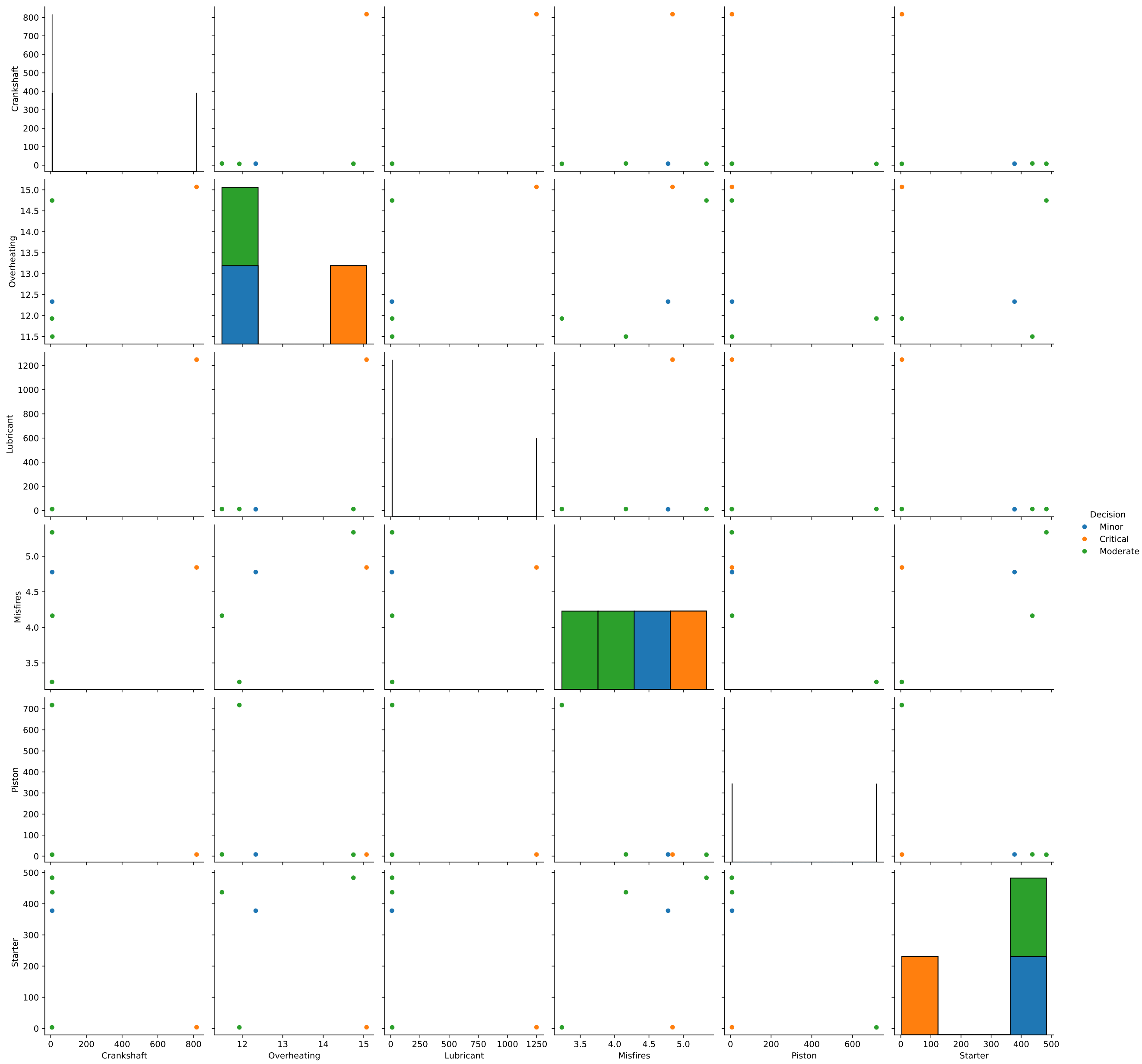


FIGURE 10. Scatter-matrix for each input variable.

false positive rate, yielding a single numerical value that quantifies the model's capacity to differentiate across classes. AUC values that are higher suggest superior discrimination and overall performance of the model.The logic for the selection of these particular metrics was based on their ability to provide a thorough evaluation of the model's effectiveness in predicting the health of vehicle engines. These metrics offer extensive insights into many aspects of model accuracy, precision, and resilience. RMSE, RMSD, and MAE quantify the size of prediction errors, accuracy gauges the overall performance of classification, confusion matrix offers specific classification outcomes, and AUC quantifies the discriminatory power of the model, [33], [34], [35]. By considering a range of metrics, this study ensure a thorough evaluation of the ensemble deep learning model's effectiveness in predicting engine health states accurately and reliably.

1) EXPERIMENTAL SETUP

Figure 5 and 5 described the variables utilised in this study and, further categorised them into Critical-0, Good-1, Minor-2 and Moderate-3 which are encoded into dummy variable see Table 6 and 7. We started by testing each machine learning algorithm to understand their predictive capacity etc, see Fig. 3, Fig. 6, Fig. 7 and Fig. 9. The result indicated that SVM, K-NN, and Gradient Boosting demonstrate strong

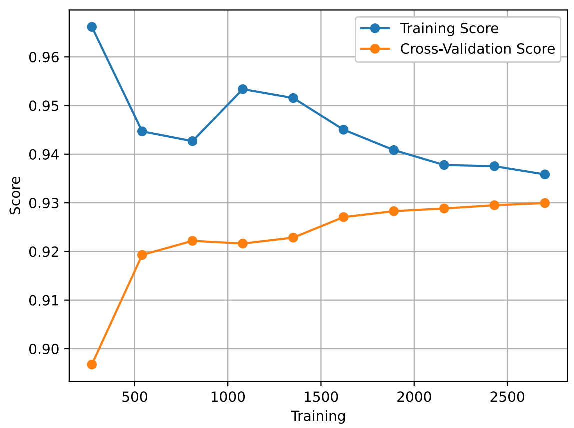


FIGURE 11. Stacked model 1 learning curve.

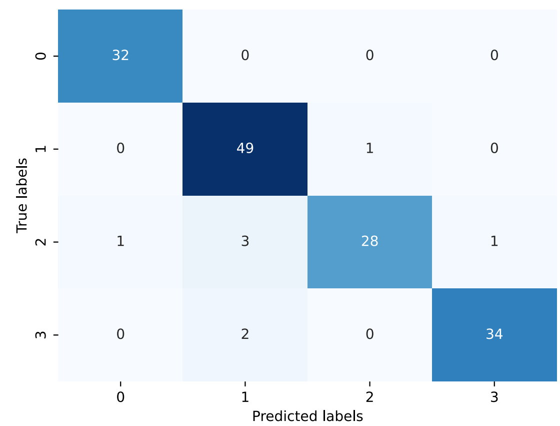


FIGURE 12. Stacked model 1 confusion matrix.

performance and generalization, while Decision Tree and Gaussian Naive Bayes exhibited lower performances in capturing the underlying patterns. To improve the overall predictive performance, this further necessitates exploring an ensemble stacked method. The integration of each ML to form (stacked models 1, 2, and 3 respectively) an ensemble method was randomly sampled and based on expert experience. Stacked model 1: is a combination of Random Forest, support vector machine, Gradient Boosting, Decision Tree, and K-Nearest Neighbors algorithm. Stacked model 2: is a combination of Logistic Regression, support vector machine, Linear Discriminant Analysis, Gradient Boosting, and AdaBoost algorithm. Stacked model 3: is a combination of Logistic Regression, K-Nearest Neighbors, support vector machine, Linear Discriminant Analysis, Gradient Boosting, AdaBoost, Decision Tree, Random Forest, and Gaussian Naive Bayes algorithm. However, the ensemble model utilizes a weighted combination of predictions from individual machine-learning models. By combining the predictions from multiple models, the ensemble aims to improve overall predictive performance and robustness. The evaluation metrics of the ensemble, such as accuracy, precision, recall, and AUC, are similar to or better than those of individual models, indicating the effectiveness of the ensemble approach. It is worth mentioning that the integration of different machine learning techniques into the ensemble model contributes to improved predictive performance and reliability compared to using a single model alone.

The Ensemble Stacking technique is promising and utilized in this study to increase prediction accuracy for the vehicular engine health monitoring Decision strategy. It acquires the ability to integrate predictions from two or more fundamental machine-learning algorithms through a meta-learning approach, in this case, the three primary classes of ensemble learning methods are stacking, bagging, and boosting as *Stacked Model 1*, *Stacked Model 2*, and *Stacked Model 3*, respectively. These three ensemble stacked models were designed in addition to the base classifiers such as logistic regression, AdaBoost, RF, and XGBoost for prediction accuracy based on sample vehicle engine features to identify good, minimal, moderate, and critical components, see Fig 5.

The mathematical formula for ensemble stacking in VHMS thus can be represented as follows:

Let's say we have N base models denoted as $M1, M2, \ldots, Mn$ and one meta-model denoted as S. Let's denote the training dataset as D which consists of input features denoted as X and corresponding target labels denoted as y.

2) TRAINING PHASE

For each base model $M1(1 <= i <= N$: Train the base model $M1$ using the training dataset D to generate predictions and estimate the heteroscedastic uncertainties. Denote the predictions as $Pi = Mi(X)$ and the estimated uncertainties as$Ui = Ui(X)$, where $Ui$is a vector of uncertainties for each prediction. Combine the predictions and uncertainties from all the base models into a new feature matrix denoted as $P = [P1, P2, \ldots PN] =$ and $U = [U1, U2, \ldots UN]$,where each column of$P$represents the predictions from one base model, and each column of U represents the estimated uncertainties from one base model.Train the meta-model S using the combined predictions P, the target labels y, and the estimated uncertainties U to create a meta-model denoted as $S = S(P, y, U)$.

3) PREDICTION PHASE

For making predictions on a new dataset D, Where each base model $Mi(1 <= i <= N)$: is considered. Generate predictions denoted as $Pi = Mi(X)$, where X represents the input features of the new dataset D. Combine the predictions from all the base models into a new feature matrix denoted as $P = [P1, P2, \ldots PN]$, where each column of P represents the predictions from one base model. Use the trained meta-model S to generate the final prediction denoted as $y = S(P)$. The design of the proposed ensemble technique is illustrated in Fig. 8.

The selection of the meta-model and the hyper-parameters of the base models and the meta-model are critical considerations in VHMS ensemble stacking. Furthermore, precisely measuring the heteroscedastic uncertainty for each base model's predictions is critical in VHMS ensemble stacking. However, based on the existing literature [2], in this study, three layered models are developed for the prediction of vehicle engine health monitoring systems. A stacking ensemble architecture, named *Stacked Model 1*, is constructed by combining Random Forest, AdaBoost, and Gradient Boosting, and *Stacked Model 2* is created by merging the Logistic Regression Classifier, Random Forest Classifier, and Gradient Boosting Classifier, while *Stacked Model 3* is the combination of Logistic Regression, KNeighborsClassifier, SVC, Random Forest Classifier, AdaBoost Classifier and Gradient Boosting Classifier. Moreover, Logistic Regression is used as the final estimator. The base estimators comprise Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Support Vector Machine (SVM), and Naive Bayes. The evaluation includes various techniques such as Regression Metrics (RMSE, RMSD, MAE, R^2), Classification Metrics (Accuracy, Confusion Matrix, Precision, Recall, AUC), and Cross-validation is conducted using RepeatedStratifiedKFold with 10 splits and 3 repeats to ensure reliable model evaluation. Additionally, learning curves are generated to visualize the model's performance.

K. EXPERIMENT AND PERFORMANCE ANALYSIS

1) STACKED MODEL 1

The *Stacked Model 1* obtained a mean accuracy of 92.93 %, with a low standard deviation of 1.46 %, suggesting a consistent and stable predicting ability.

The model demonstrates a high level of accuracy, see Fig. 11, as evidenced by its low regression metrics. The model's performance metrics are as follows: RMSE = 0.3355, MAE = 0.0728, R^2 = 0.9021. The metrics show that the model effectively reduces prediction errors, leading to accurate decision predictions. The confusion matrix shows the model's accurate predictions for all four classes, with minimal misclassifications in Fig. 12.

The performance of VEHMS is important since it requires accurate identification of various engine health states to make informed decisions. The model also obtained a precision score of 94.86 % and a recall exceeding 94.70 % in Fig. 11.

These measures prioritize the model's precision in accurately detecting positive occurrences, which is essential for reliable decision-making in vehicular situations. This model's ability to distinguish between different classes is measured by the Area Under the Curve (AUC) score, which achieves a value of 97.02 %. The Model has a higher capacity to accurately differentiate between classes in VEHMS, which is important for identifying small variations in engine health states. However, despite such achievement of Stacked model 1, Table 8 further suggests its limitations, and area of improvement.

TABLE 8. Evaluation of stacked model 1.

Strengths	Weaknesses	Areas for Improvement
High accuracy and AUC indicate good overall performance.	The model may have to overfit the training data, as indicated by the high-performance metrics on the training set. It's essential to evaluate the model on unseen data to assess its generalization ability accurately.	Regularization techniques could be applied to prevent overfitting and improve generalization.
Low RMSE, RMSD, and MAE suggest the model's predictions are close to the actual values.	The confusion matrix indicates some misclassifications in predicting certain classes.	Further feature engineering or selection might enhance the model's performance and interpretability.
High precision and recall show the model's ability to correctly identify positive cases while minimizing false positives and false negatives.		

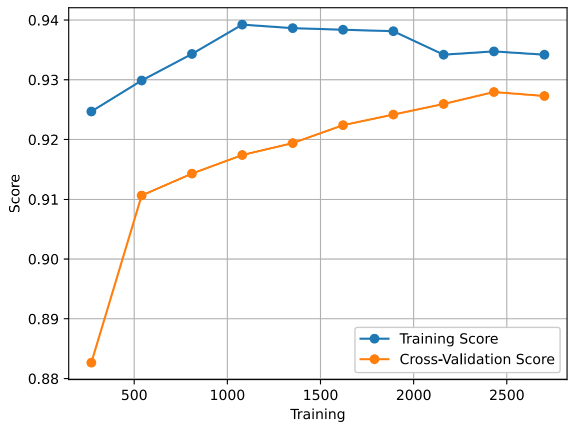


FIGURE 13. Stacked model 2 learning curve.

2) STACKED MODEL 2

As depicted in Fig. 13, the accuracy of *Stacked Model 2* in recognizing occurrences across different engine health statuses is 94.70 %, making it important for precise decisions in VEHMS.

The model's efficacy in reducing prediction errors is evident from the low values of significant metrics–Mean Absolute Error 0.0795, Root Mean Squared Deviation 0.1325, and Root Mean Squared Error 0.3639, respectively. These metrics illustrate the model's capacity to reduce disparities between feature and target values, hence ensuring accurate decision predictions. The model's robustness is shown in its high R-squared value of 0.8849 which indicates its ability to explain a significant proportion of the variation in engine health data patterns. This is crucial for facilitating decision-making in the VEHMS system. The confusion matrix, in contrast, offers comprehensive data regarding the classification accuracy of the model, see Fig. 14.However, Table 9, underscores the examination of

TABLE 9. Evaluation of stacked model 2.

Strengths	Weaknesses	Areas for Improvement
The model maintains a high level of accuracy and AUC, comparable to Stacked Models 1.	Slightly higher RMSE, RMSD, and MAE compared to Stacked Models 1 suggesting that this model's predictions might have more significant deviations from the actual values.	Investigate potential sources of error that lead to higher RMSE and other error metrics.
Similar precision and recall values indicate consistent performance in correctly classifying positive cases.		Consider feature selection or engineering techniques to improve model performance and reduce prediction errors.

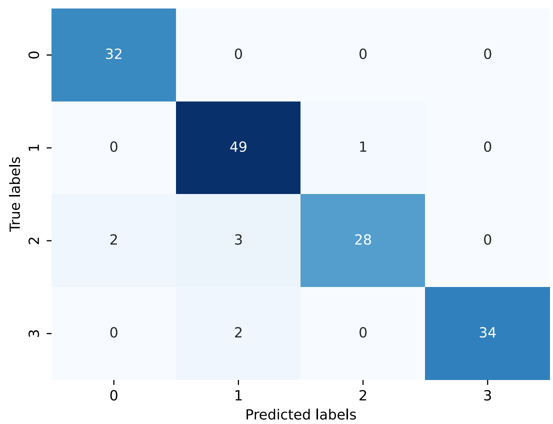


FIGURE 14. Stacked model 2 confusion matrix.

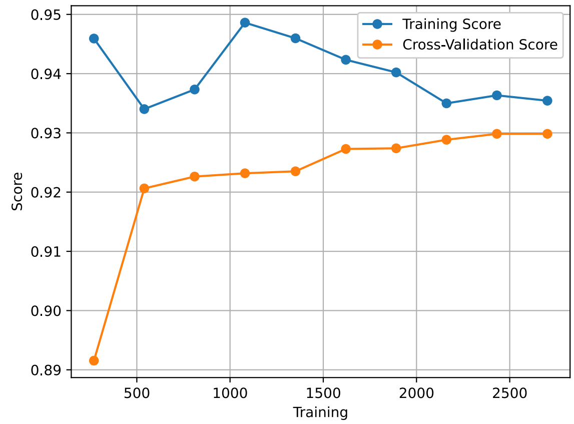


FIGURE 15. Stacked model 3 learning curve.

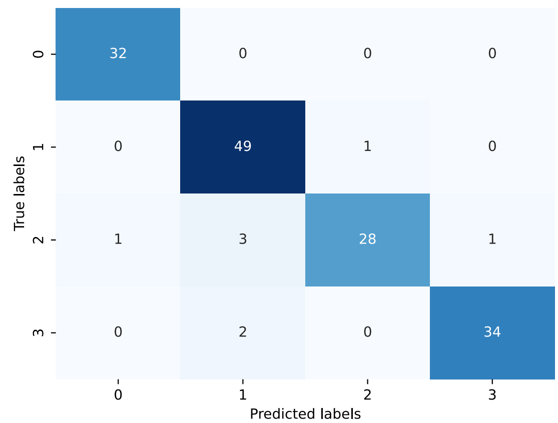


FIGURE 16. Stacked model 3 confusion matrix.

the model highlighting it strength, Weaknesses and area of improvement. Despite this indications, the model had a precision of 94.93 % and a recall of 94.70 % in successfully predicting class labels. Accurate detection of positive cases in vehicular scenarios requires an unbiased examination of the actual positive rates. However, the Area Under the Curve (AUC) score of 0.9665 emphasized the model's capacity to accurately identify subtle engine health issues in VEHMS, which is important for accurate identification.

3) STACKED MODEL 3

The *Stacked Model 3* in this experiment achieved a class label prediction accuracy of 93.01 % as shown in Fig. 15.

The outcome additionally demonstrates a decrease in predicted mistakes through the utilization of low-key indicators such as RMSE (0.3355), RMSD (0.1126), and MAE (0.0728), which indicates a high level of precision. The accuracy of VEHMS decisions relies on the model's ability to minimize differences between projected and actual outcomes. The model's strong R-squared score of 0.9021 demonstrates its effectiveness in explaining a significant amount of variance and its ability to accurately capture and show data patterns. This is particularly essential for making informed decisions regarding engine health monitoring.

The confusion matrix Fig. 16 provides a comprehensive evaluation of the model's classification performance across several classes. The recall and precision rates are 94.86 % and 94.70 % respectively. The result further demonstrates the model's proficiency in precisely recognizing positive instances—a crucial aspect in intricate traffic scenarios. The AUC value of 0.9610 provides suggestive evidence of the model's ability to accurately categorize data. A high Area Under the Curve (AUC) of 0.9653, indicates that the model is capable of effectively differentiating labels across different classes and identifying even minor vehicle health problems.

The findings underscore the model's proficiency in precisely recognizing positive instances—a crucial aspect in intricate traffic scenarios.

IV. DISCUSSION

The comprehensive evaluation of three Stacked Ensemble Models, namely *Stacked Model 1*, *Stacked Model 2* and *Stacked Model 3* within the context of Vehicle Engine Health Monitoring Systems (VEHMS) provided valuable insights into their effectiveness for decision accuracy prediction as showed in Table 11.

TABLE 10. Evaluation of stacked model 3.

Strengths	Weaknesses	Areas for Improvement
Similar performance to Stacked models 1 in terms of accuracy, precision, recall, and AUC.	Convergence Warning suggests that the logistic regression model failed to converge with the default number of iterations. This may indicate issues with the data or model configuration. Despite the warning, performance metrics are consistent with Stacked models 1 and 2.	Investigate the cause of the convergence issue and address it through adjustments such as increasing the maximum number of iterations or scaling the data. Perform further analysis to ensure the model's reliability and stability.

TABLE 11. Comparison of stacked models result.

Model	Performance Metrics
Stacked Model 1	RMSE: 0.3355, RMSD: 0.1126, MAE: 0.0728, R^2: 0.9021, Accuracy: 0.9470, Precision: 0.9486, Recall: 0.9470, AUC: 0.9702
Stacked Model 2	RMSE: 0.3639, RMSD: 0.1325, MAE: 0.0795, R^2: 0.8849, Accuracy: 0.9470, Precision: 0.9493, Recall: 0.9470, AUC: 0.9665
Stacked Model 3	RMSE: 0.3355, RMSD: 0.1126, MAE: 0.0728, R^2: 0.9021, Accuracy: 0.9470, Precision: 0.9486, Recall: 0.9470, AUC: 0.9653

Stacked Model 1 demonstrates a consistent mean accuracy of 92.98%, with a low standard deviation of 1.47%, indicating stable predicting ability. Its low regression metrics RMSE, MAE, and R^2 underscore its efficiency in reducing prediction errors, ensuring highly accurate decision predictions. The confusion matrix Fig. 12 reveals high accuracy in predicting all four classes, emphasizing its suitability for VEHMS. Remarkable precision of 94.86% and recall of 94.70% further highlight its capacity for accurate identification of positive occurrences. The AUC score of 97.02% indicates exceptional discriminative ability, crucial for nuanced engine health monitoring. *Stacked Model 1* emerges as a precise prediction instrument, especially for applications prioritizing precision. In contrast, *Stacked Model 2* showcases noteworthy performance metrics, including an accuracy of 94.70%, demonstrating its ability to identify occurrences across multiple classes. Its low values of MAE, RMSD, and RMSE signify precise decision prediction, essential for reliable vehicular health monitoring. The R-squared value of 0.8849 emphasizes its robustness in explaining large variations. The confusion matrix affirms its accuracy in predicting class labels with a precision of 94.93% and recall of 94.70%. The exceptional AUC score of 0.9665 highlights its superior discriminative ability, making it a preferred choice for applications prioritizing discrimination. *Stacked Model 2* proves to be a dependable forecasting tool for enhancing decision prediction accuracy in VEHMS. However, *Stacked Model 3* contributes significantly to decision prediction accuracy in VEHMS with a commendable accuracy of 94.70%. Its precision in minimizing prediction errors is evident through low values of RMSE, RMSD, and MAE. The R-squared score of 0.9021 reinforces its effectiveness in capturing data patterns crucial for decision correctness. High recall 94.70% and precision 94.86% values in the confusion matrix 16 highlight its reliable recognition of positive events. The AUC score of 0.9653 further supports its ability to handle classification tasks, emphasizing its capacity to capture subtle vehicle health issues. *Stacked Model 3* is validated as a precise prediction tool, enhancing decision-making precision and dependability in VEHMS. Finally, the decision to adopt a particular Stacked Ensemble Model into VEHMS should align with specific application goals. For precision-centric applications, *Stacked Model 1* outperformed the rest of the model in terms of decision accuracy prediction. *Stacked Model 2*, with its superior discriminative ability, is the preferred choice for applications prioritizing discrimination. *Stacked Model 3*, with consistent performance and high accuracy, serves as a reliable option for general-purpose applications. The integration of the selected model into VEHMS ensures robust decision accuracy prediction, contributing to more effective and reliable engine health monitoring and risk management systems.

In summary, each stacked model demonstrates high accuracy, precision, and robustness in predicting engine health states. The comprehensive evaluation through performance metrics, confusion matrices, and AUC scores enhances transparency and credibility by offering a comprehensive insight into the prediction process of the models and their dependability in real-world automotive situations. Additionally, acknowledging the models' limitations and areas for improvement further strengthens transparency and trustworthiness by promoting critical examination and continuous refinement of the predictive models.

A. COMPARISON OF STACKED MODEL 1 WITH EXISTING APPROACH

Again, result from the experiment appears that the Stacked Model 1 outperformed the current methods found in [2] for predicting the health of vehicle engines in multiple important aspects as shown below.

High Accuracy and Stability: The Stacked Model 1 demonstrates exceptional precision and consistency, with a mean accuracy of 92.93%, surpassing the conventional approach's accuracy of 80.3%. Furthermore, the fact that it has a low standard deviation of 1.46% demonstrates a reliable predictive capability, which is essential for real-time monitoring when consistent performance is required.

Minimized Forecasting Inaccuracies: The regression metrics, with RMSE = 0.3355, MAE = 0.0728, and $R^2 = 0.9021$, indicate that the model successfully minimizes prediction errors. This demonstrates a significant degree of accuracy in its forecasts, which is crucial for making precise decisions.

Precise Detection of Engine Health Conditions: The confusion matrix and precision-recall scores indicate that the Stacked Model 1 achieves high prediction accuracy for all engine health states, with minimum misclassifications. Precise accuracy is essential for promptly detecting different engine health conditions, guaranteeing dependable decision-making.The model demonstrates a high discriminative capacity, as evidenced by its AUC score of 97.02%. This suggests its exceptional ability to reliably distinguish between various classes of engine health conditions. The ability to distinguish between different engine health states is crucial for promptly maintaining and preventing potential failures, particularly when it comes to recognizing minor deviations.

The unique advantages of the ensemble approach are: The improved performance of Stacked Model 1 is presumably due to its ensemble character. Through the integration of various models, it is possible to capture a wide range of patterns and interconnections within the data, resulting in improved forecast accuracy and the ability to apply knowledge to new situations. Ensemble methods are recognized for their capacity to address overfitting and decrease variation, which may account for the model's stability and consistent performance.The Stacked Model 1 provides a substantial enhancement compared to current methods for predicting the health of vehicle engines. This is mainly because of its high accuracy, stability, exact identification of engine health conditions, and strong ability to differentiate between different states of engine health. The advantages mentioned above demonstrate the efficacy of ensemble deep learning methods in improving predicted accuracy for intricate systems such as automotive engines.

B. IMPLEMENTATION CHALLENGES

Data Quality Assurance: One of the primary challenges encountered during the implementation of the experiment was ensuring the quality and reliability of the input data. Vehicular engine health data are subject to various sources of noise, outliers, and inconsistencies, which can affect the performance of predictive models. Ensuring the accuracy, completeness, and consistency of the dataset required extensive data preprocessing and cleaning efforts. Additionally, integrating heterogeneous data sources and formats posed challenges in standardizing and harmonizing the data for analysis.

Model Complexity: Implementing and evaluating three Stacked Ensemble Models within the context of Vehicle Engine Health Monitoring Systems (VEHMS) involved dealing with complex and high-dimensional data. The complexity of the models, including the integration of multiple machine learning algorithms and ensemble techniques, added computational overhead and resource requirements. Balancing model complexity with interpretability was crucial to ensure practical utility and deployment in real-world settings.

Computational Resources: The computational resources required for training, evaluating, and deploying the Stacked Ensemble Models posed significant challenges. Deep learning models and ensemble techniques often require substantial computational resources, including high-performance computing (HPC) infrastructure and specialized hardware accelerators. Limited access to such resources may hinder the scalability and accessibility of the predictive framework. Stacked model 3 in particular highlighted a warning message which suggests that the lbfgs solver used in logistic regression failed to converge, reaching the maximum number of iterations, see Table 10. This warning is particularly relevant to computational resources. Increasing the number of iterations, as suggested, would demand more computational power and time during the training phase.

1) ADDRESSING IMPLEMENTATION CHALLENGES

Data Quality Assurance: Implementing robust data quality assurance procedures, including data validation, cleansing, and preprocessing, was essential to ensure the reliability and integrity of the input data. Leveraging automated data validation tools, domain knowledge, and expert insights helped identify and address data quality issues effectively. Additionally, conducting sensitivity analyses and sensitivity tests helped assess the robustness of the predictive models to variations in data quality and input parameters.

Model Complexity Management: Managing model complexity and interpretability was critical to ensure practical utility and deployment of the predictive framework in real-world settings. Employing model regularization techniques, feature selection methods, and model simplification strategies helped mitigate overfitting and improve model interpretability. Moreover, conducting model sensitivity analyses and sensitivity tests helped assess the impact of model complexity on predictive performance and generalization.

Optimizing Computational Resources: Optimizing the utilization of computational resources, including parallel computing, distributed computing, and cloud computing, helped mitigate the computational challenges associated with training and evaluating the Stacked Ensemble Models 3. Leveraging scalable and efficient algorithms, model parallelism, and hardware accelerators (e.g., GPUs) helped accelerate model training and inference tasks, improving scalability and efficiency. Researchers and practitioners should consider optimizing the computational resources allocated for model training to ensure convergence and improve overall efficiency. Additionally, scaling the data as recommended in the warning message can also affect computational requirements, as preprocessing steps may introduce additional computational overhead. Therefore, the availability and optimization of computational resources play a crucial role in effectively implementing and training machine learning models like Stacked Model 3. By addressing these implementation challenges and limitations related to data quality, model complexity, and computational resources, researchers and practitioners can develop more robust, scalable, and effective predictive frameworks for vehicular engine health monitoring, contributing to improved reliability and safety in automotive systems.

C. REAL WORLD APPLICATION AND SCALABILITY

The developed model holds promise for various real-world applications and scalability. In automotive manufacturing plants, the model can be integrated into quality control processes to identify faulty engines before installation. In fleet management systems, it can enable predictive maintenance scheduling, optimizing vehicle uptime and reducing maintenance downtime. Moreover, the model can adapt to different types of vehicles or engine configurations by retraining on data specific to those vehicles or configurations. Its scalability allows for deployment across diverse vehicle fleets, ranging from passenger cars to heavy-duty trucks, enhancing safety and reliability across the transportation sector. Although this paradigm shows promise, its adoption in the automobile sector has significant practical obstacles that must be solved to ensure feasibility and efficacy.

1) DATA ACQUISITION

Data Quality, Availability and Security The utilization of sensor-based methodology guarantees the accuracy and accessibility of data. However, it is crucial to ensure broad coverage of pertinent sensor data from various vehicle kinds and configurations. Standardization efforts may be necessary to provide compatibility and integration with different data sources, such as onboard diagnostics (OBD) ports and Controller Area Network (CAN) bus. Data security is of utmost importance in safeguarding critical vehicle data from unauthorized access or cyber attacks. To ensure data privacy and security, it is imperative to have strong data encryption, access restrictions, and compliance with industry requirements such as GDPR, CCPA and to establish explicit standards and guidelines to oversee the collection, storage, and utilization of vehicle data, with a strong focus on ensuring transparency and obtaining user consent.

2) BIG DATA PROCESSING MECHANISMS

Scalability The ability to process huge quantities of sensor data from various vehicle fleets necessitates the use of scalable big data processing mechanisms. Utilizing distributed computing frameworks such as Apache Spark or harnessing cloud-based technologies can effectively manage data processing needs. Real-time Processing: The immediate examination of continuously flowing sensor data allows for prompt detection of engine health problems. Utilizing stream processing frameworks, such as Apache Kafka, and implementing real-time analytics models can expedite decision-making in automobile manufacturing and fleet management.

3) MODEL DEPLOYMENT

Operationalization: The process of deploying predictive maintenance models into production environments necessitates the smooth integration with pre-existing systems and procedures. By implementing resilient model deployment pipelines, version control systems using GitHub, and automated testing frameworks, the seamless deployment and maintenance of the predictive maintenance solution are ensured. Edge computing is another aspect, it involves implementing models at the edge, such as onboard car systems, in situations where there is limited network connectivity or a need for low latency. This approach improves responsiveness and minimizes dependence on centralized data processing infrastructure.

4) INTEGRATION WITH EXISTING MAINTENANCE PROCESSES

Organizational Alignment: Integrating predictive maintenance solutions with existing maintenance processes requires organizational buy-in and alignment across departments. Collaborating with maintenance teams, providing training on new technologies, and demonstrating the value proposition of predictive maintenance enhance acceptance and adoption.

5) ORGANIZATIONAL BUY-IN

Stakeholder Engagement: Organizational buy-in refers to the level of support and commitment that an organization demonstrates towards a certain idea, initiative, or decision. Engaging stakeholders is crucial for the success of the predictive maintenance effort. This involves obtaining support and agreement from important stakeholders, such as management, operations, and IT departments. Conveying the advantages, return on investment (ROI), and how the solution aligns strategically with company goals encourages support and dedication. Promoting a culture that values data-driven decision-making and innovation is essential for fostering the adoption and ongoing enhancement of practices. Facilitating cultural transformation can be achieved by implementing programs that offer incentives, recognition, and training opportunities to increase data literacy and encourage experimenting with new technologies.

To tackle these difficulties, a multidisciplinary strategy is necessary, which involves the participation of data scientists, engineers, domain specialists, and stakeholders. The suggested predictive maintenance system has the potential to improve safety, reliability, and efficiency in the automobile sector by utilizing sensor-based methodologies, ensuring data quality, availability, and security, and addressing the practical obstacles of implementation.

6) ETHICAL AND SOCIETAL IMPLICATIONS

Impact on Employment: Predictive maintenance technologies have the potential to enhance vehicle performance and minimize the occurrence of unforeseen failures. However, they may also cause significant changes to the conventional employment arrangements within the automotive repair and maintenance industry. With the increasing capabilities of automobiles to perform self-diagnosis and preventative maintenance, there may be a reduced need for human mechanics and technicians. This change has the potential to result in job displacement for workers in these sectors, emphasizing the necessity for retraining initiatives and assistance for those impacted to successfully move into different positions or industries.

Environmental Sustainability: Predictive maintenance can enhance the longevity of vehicles and decrease emissions by preventing unnecessary repairs and replacements. However, it also prompts concerns regarding the environmental consequences associated with the production and disposal of advanced automotive components. With the increasing complexity of cars due to improved sensors and electrical systems, it is necessary to evaluate the environmental impact of manufacturing and disposing of these components. Furthermore, it is important to take into account the energy usage linked to data processing and analysis for predictive maintenance within the broader framework of environmental sustainability initiatives.

Biases in Data and Models: Predictive maintenance models are trained using historical data, which may contain biases that were present during the data gathering process or biases that are ingrained in the underlying systems due to societal factors. For instance, if the training data predominantly comprises automobiles from specific demographics or geographic regions, the prediction models may not effectively apply to different populations. Additionally, if diagnostic algorithms are not thoroughly verified across varied datasets, biases in these algorithms could have a disproportionate impact on specific demographic groups. To tackle these biases, it is crucial to meticulously focus on the quality, diversity, and fairness of data at every stage of developing and implementing predictive maintenance technology. Ultimately, predictive maintenance technologies provide substantial advantages in terms of enhancing vehicle performance and minimizing maintenance expenses. However, they also give rise to crucial ethical and societal concerns. To ensure ethical and equitable deployment of these technologies, it is crucial to address employment, privacy, environmental sustainability, and biases in data and models.

V. CONCLUSION

In conclusion, a vehicle engine prediction architecture is presented in the paper. This study utilized a computer-generated dataset to execute an improved machine and deep learning model that alerts the user promptly and with the utmost priority of fault in a vehicular engine in real time. The effectiveness of machine learning algorithms and ensemble approaches was evaluated using measures including RMSE, RMSD, MAE, R^2, Accuracy, Confusion Matrix, and AUC. *Stacked Model 1* outperformed the individual algorithms and the other stacked models 2 and 3. The AUC of 0.9702 was impressive. The ensemble model combined Random Forest, support vector machine,Gradient Boosting, Decision Tree and K-Nearest Neighbors for prediction. A low root mean square error (RMSE) of 0.3355, a high accuracy rate of 0.9470, and a precision of 0.9486 shows that the model performs well in decision prediction. The confusion matrix proves engine health issue categorization works. The drawback of this study is that the results may differ if conducted in the real world. The study's findings, however, could serve as a guide for the automotive industry and a standard procedure for enhancing prediction accuracy.

Future research should focus on increasing the efficiency of the novel technique by exploring alternative ensemble methods. Techniques such as gradient boosting, random forests, and ensemble stacking offer opportunities to enhance prediction accuracy, scalability, and computational efficiency. Comparative studies evaluating the performance of different ensemble methods can provide valuable insights into their effectiveness for vehicular engine health monitoring. Another avenue for future research is to test the prediction accuracy of the predictive framework using real-world vehicle datasets from heterogeneous sources. Incorporating diverse datasets representing a wide range of vehicle types, engine configurations, and operating conditions can help evaluate the robustness and generalization capabilities of the model. Additionally, investigating the impact of additional sensor data, such as temperature, pressure, or vibration, on prediction accuracy can further enhance the model's effectiveness in real-world scenarios. Optimizing model hyperparameters, including ensemble composition and tuning parameters, to improve prediction performance and generalization should be other area for future research Furthermore, exploring the integration of domain knowledge or expert systems into the predictive framework can enhance interpretability and reliability. Incorporating domain-specific insights and constraints can help tailor the model to specific application scenarios, improving its practical utility and adoption in real-world settings. By addressing these future research directions, researchers can advance the field of predictive maintenance for vehicular engines, addressing emerging challenges in automotive reliability and safety, and ultimately improving the performance and longevity of automotive systems.

REFERENCES

[1] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, ''Industrial artificial intelligence in Industry 4.0–systematic review, challenges and outlook,'' *IEEE Access*, vol. 8, pp. 220121–220139, 2020.

[2] Md. A. Rahim, Md. A. Rahman, Md. M. Rahman, N. Zaman, N. Moustafa, and I. Razzak, ''An intelligent risk management framework for monitoring vehicular engine health,'' *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 3, pp. 1298–1306, Sep. 2022.

[3] L. Decker, D. Leite, L. Giommi, and D. Bonacorsi, ''Real-time anomaly detection in data centers for log-based predictive maintenance using an evolving fuzzy-rule-based approach,'' in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2020, pp. 1–8.

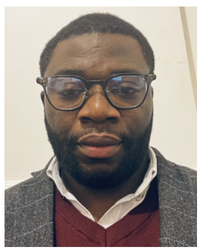
[4] Md. A. Rahman, Md. A. Rahim, Md. M. Rahman, N. Moustafa, I. Razzak, T. Ahmad, and M. N. Patwary, ''A secure and intelligent framework for vehicle health monitoring exploiting big-data analytics,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19727–19742, Oct. 2022.

[5] M. R. Sarkar, S. G. Anavatti, T. Dam, M. M. Ferdaus, M. Tahtali, S. Ramasamy, and M. Pratama, ''GATE: A guided approach for time series ensemble forecasting,'' *Expert Syst. Appl.*, vol. 235, Jan. 2024, Art. no. 121177.

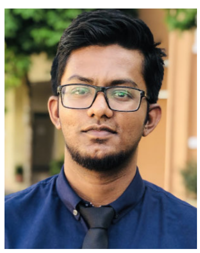
[6] H. Nordal and I. El-Thalji, ''Modeling a predictive maintenance management architecture to meet Industry 4.0 requirements: A case study,'' *Syst. Eng.*, vol. 24, no. 1, pp. 34–50, Jan. 2021.

[7] F. Arena, M. Collotta, L. Luca, M. Ruggieri, and F. G. Termine, ''Predictive maintenance in the automotive sector: A literature review,'' *Math. Comput. Appl.*, vol. 27, no. 1, p. 2, Dec. 2021.

[8] K. Kanagaraj and S. Geetha, "Data analytics framework based on cloud environment, integration of cloud computing with Internet of Things: Foundations," *Analytics Applications*, pp. 251–275, 2021.
[9] K. L. Tsui, Y. Zhao, and D. Wang, "Big data opportunities: System health monitoring and management," *IEEE Access*, vol. 7, pp. 68853–68867, 2019.
[10] W. Chen, L. Zhang, K. Pattipati, A. M. Bazzi, S. Joshi, and E. M. Dede, "Data-driven approach for fault prognosis of SiC MOSFETs," *IEEE Trans. Power Electron.*, vol. 35, no. 4, pp. 4048–4062, Apr. 2020.
[11] A. Theissler, J. Pérez-Velázquez, M. Kettelgerdes, and G. Elger, "Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry," *Rel. Eng. Syst. Saf.*, vol. 215, Nov. 2021, Art. no. 107864.
[12] S. Katreddi, S. Kasani, and A. Thiruvengadam, "A review of applications of artificial intelligence in heavy duty trucks," *Energies*, vol. 15, no. 20, p. 7457, Oct. 2022.
[13] M. Jiang, J. Liu, L. Zhang, and C. Liu, "An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms," *Phys. A, Stat. Mech. Appl.*, vol. 541, Mar. 2020, Art. no. 122272.
[14] Q. Wei, Q. Zhang, H. Gao, T. Song, A. Salhi, and B. Yu, "DEEPStack-RBP: Accurate identification of RNA-binding proteins based on autoencoder feature selection and deep stacking ensemble classifier," *Knowl.-Based Syst.*, vol. 256, Nov. 2022, Art. no. 109875.
[15] D. Cao, H. Xing, M. S. Wong, M.-P. Kwan, H. Xing, and Y. Meng, "A stacking ensemble deep learning model for building extraction from remote sensing images," *Remote Sens.*, vol. 13, no. 19, p. 3898, Sep. 2021.
[16] C. Patino-Alonso, M. Gómez-Sánchez, L. Gómez-Sánchez, E. Rodríguez-Sánchez, C. Agudo-Conde, L. García-Ortiz, and M. A. Gómez-Marcos, "Diagnosing vascular aging based on macro and micronutrients using ensemble machine learning," *Mathematics*, vol. 11, no. 7, p. 1645, Mar. 2023.
[17] F. Li, L. Zhang, B. Chen, D. Gao, Y. Cheng, X. Zhang, Y. Yang, K. Gao, and Z. Huang, "An optimal stacking ensemble for remaining useful life estimation of systems under multi-operating conditions," *IEEE Access*, vol. 8, pp. 31854–31868, 2020.
[18] Y. Patel, S. Patel, R. Gupta, M. S. Obaidat, R. Kakkar, N. K. Jadav, and S. Tanwar, "Blockchain and AI-based engine fault detection scheme for autonomous vehicles," in *Proc. Int. Conf. Commun., Comput., Cybersecurity, Informat. (CCCI)*, Oct. 2023, pp. 1–8.
[19] S. Mun, J. Hwang, L. Bian, T. Falls, and W. G. Bond, "Ensemble learning for fault condition prediction and health status monitoring in military ground vehicles," in *Proc. IIE Annual Conf.*, 2023, pp. 1–6.
[20] R. Khoshkangini, P. Mashhadi, D. Tegnered, J. Lundström, and T. Rögnvaldsson, "Predicting vehicle behavior using multi-task ensemble learning," *Expert Syst. Appl.*, vol. 212, Feb. 2023, Art. no. 118716.
[21] I. Ullah, K. Liu, T. Yamamoto, M. Zahid, and A. Jamal, "Electric vehicle energy consumption prediction using stacked generalization: An ensemble learning approach," *Int. J. Green Energy*, vol. 18, no. 9, pp. 896–909, Jul. 2021.
[22] I. Tessaro, V. C. Mariani, and L. D. S. Coelho, "Machine learning models applied to predictive maintenance in automotive engine components," in *Proc. 1st Int. Electron. Conf. Actuator Technol., Mater., Devices Appl.*, Nov. 2020, p. 26.
[23] D. Giordano, F. Giobergia, E. Pastor, A. La Macchia, T. Cerquitelli, E. Baralis, M. Mellia, and D. Tricarico, "Data-driven strategies for predictive maintenance: Lesson learned from an automotive use case," *Comput. Ind.*, vol. 134, Jan. 2022, Art. no. 103554.
[24] R. Young, S. Fallon, P. Jacob, and D. O'Dwyer, "Vehicle telematics and its role as a key enabler in the development of smart cities," *IEEE Sensors J.*, vol. 20, no. 19, pp. 11713–11724, Oct. 2020.
[25] V. Kolisetty and D. Rajput, "A review on the significance of machine learning for data analysis in big data," *Jordanian J. Comput. Inf. Technol.*, vol. 6, no. 1, pp. 1–12, 2019.
[26] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," *J. Big Data*, vol. 8, no. 1, pp. 1–37, Oct. 2021.
[27] R. Thomas, "A novel ensemble method for detecting outliers in categorical data," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 4947–4953, Aug. 2020.
[28] J.-W. Chen, "Analysis of subsampled image size for detection and identification of brake pad contours by using deep learning," in *Proc. IEEE Eurasia Conf. IoT, Commun. Eng. (ECICE)*, Oct. 2020, pp. 174–176.
[29] N. El-Rashidy, S. El-Sappagh, T. Abuhmed, S. Abdelrazek, and H. M. El-Bakry, "Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model," *IEEE Access*, vol. 8, pp. 133541–133564, 2020.
[30] M. Garouani, A. Ahmad, M. Bouneffa, and M. Hamlich, "Autoencoder-kNN meta-model based data characterization approach for an automated selection of AI algorithms," *J. Big Data*, vol. 10, no. 1, pp. 1–40, Feb. 2023.
[31] J. Linåker, B. Regnell, and D. Damian, "A method for analyzing stakeholders' influence on an open source software ecosystem's requirements engineering process," *Requirements Eng.*, vol. 25, no. 1, pp. 115–130, Mar. 2020.
[32] S. Singh, S. C. Misra, and F. T. S. Chan, "Establishment of critical success factors for implementation of product lifecycle management systems," *Int. J. Prod. Res.*, vol. 58, no. 4, pp. 997–1016, Feb. 2020.
[33] U. Mohseni, C. B. Pande, S. Chandra Pal, and F. Alshehri, "Prediction of weighted arithmetic water quality index for urban water quality using ensemble machine learning model," *Chemosphere*, vol. 352, Mar. 2024, Art. no. 141393.
[34] Z. Rahmatinejad, T. Dehghani, B. Hoseini, F. Rahmatinejad, A. Lotfata, H. Reihani, and S. Eslami, "A comparative study of explainable ensemble learning and logistic regression for predicting in-hospital mortality in the emergency department," *Sci. Rep.*, vol. 14, no. 1, p. 3406, Feb. 2024.
[35] V. Villa, G. Bruno, K. Aliev, P. Piantanida, A. Corneli, and D. Antonelli, "Machine learning framework for the sustainable maintenance of building facilities," *Sustainability*, vol. 14, no. 2, p. 681, Jan. 2022.



ISINKA JOSEPH CHUKWUDI received the Bachelor of Science degree in computer science from the University of Greenwich, U.K., in 2014, and the master's degree from the University of Coventry, U.K. He is currently pursuing the Ph.D. degree in computer science with the University of Wolverhampton, U.K. His research interests include big data analytics, machine, and deep learning, predictive modeling, electric and autonomous vehicles, the Internet of Things (IoT), cloud computing, distributed computing, and computer programming. He is a member of the IET and has a professional experience in IT consulting.



NAFEES ZAMAN received the B.Sc. degree in computer science from International Islamic University Malaysia, in 2020. He is currently pursuing the joint master's degree in intelligent field robotic systems with the University of Girona, Spain, and the University of Zagreb, Croatia. He is also a Visiting Researcher with the KTH Royal Institute of Technology, Stockholm, Sweden. Previously, he was a Machine Learning Research Intern with IBM CoE Malaysia and VICOROB, Spain. His research interests include intelligent systems (robotics), computer vision, 3D deep learning, and computer graphics. He was a recipient of the Erasmus Mundus Scholarship (2021–2022) and other awards, including the Best Innovation Award in Malaysia Technology Expo Special Edition, in 2020, Gold Medal in National Innovation and Invention Competition Malaysia, in 2019, and Second Runner-Up in Nasa Space App Challenge Malaysia, in 2019.

MD ABDUR RAHIM received the B.Sc. degree in mechanical engineering from KUET, Bangladesh, and the M.Sc. degree in automotive engineering from UMP, Malaysia. His research interests include AI, traditional, connected, electric, autonomous vehicles, ITS, AI-based SHM, the Internet of Vehicles (IoV), intelligent robotic systems, big-data analytics, machine learning, deep learning, renewable energy, and ICE.

MD ARAFATUR RAHMAN (Senior Member, IEEE) received the Ph.D. degree in electronic and telecommunications engineering from the University of Naples Federico II, Naples, Italy, in 2013. He has around 15 years of research and teaching experience in the domain of computer and communications engineering. Currently, he is a Reader in cyber security with the School of Engineering, Computing & Mathematical Sciences, University of Wolverhampton, U.K. He was an Associate Professor with the Faculty of Computing, Universiti Malaysia Pahang. He was a Postdoctoral Research Fellow with the University of Naples Federico II, in 2014, and a Visiting Researcher with the Sapienza University of Rome, in 2016. He has coauthored around 150 prestigious IEEE and Elsevier journals [e.g., IEEE Transactions on Industrial Informatics, IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Green Communications and Networking, IEEE Transactions on Services Computing, *IEEE Communications Magazine*, JNCA (Elsevier), and FGCS (Elsevier)]; and conference publications (e.g., IEEE Globecom and IEEE DASC). His research interests include the Internet of Things (IoT), wireless communication networks, cognitive radio networks, 5G, vehicular communication, cyber-physical systems, big data, cloud-fog-edge computing, machine learning, and security. He has served as a Specialty Chief Editor for IoT Theory and Fundamental Research (specialty section of Frontiers in the Internet of Things); an Advisory Board Member; an Editor *Computers* (MDPI); a Lead Guest Editor (IEEE Access and *Computers*); an Associate Editor of IEEE Access and *Patron*; the General Chair; an Organizing Committee Member; the Publicity Chair; the Session Chair; a Programme Committee Member; and a member of Technical Programme Committee (TPC) in numerous leading conferences worldwide (e.g., IEEE Globecom, IEEE DASC, IEEE iSCI, and IEEE ETCCE) and journals. His name was enlisted inside the World's Top 2% Scientists list released by Stanford University under the category of "Citation Impact in Single Calendar Year, in 2019 and 2021." He was awarded a Higher Education Academy (HEA) Fellowship from the U.K. He was endorsed by the Royal Academy of Engineering, U.K., as Global Talent under the category of "Exceptional Talent," in 2022. He is the Founder of RITECHS, a U.K.-based research organization.

MOHAMMED J. F. ALENAZI (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer engineering from the University of Kansas, in 2010, 2012, and 2015, respectively. He is currently an Associate Professor in computer engineering with King Saud University and a reviewer of several international journals. His research interests include cybersecurity, focusing on network security, encryption, vulnerability analysis, and machine learning, where he applies AI to enhance network security and performance. He also works on the design and analysis of resilient networks, network routing, and mobile ad-hoc network (MANET) protocols. A member of ACM, his work contributes to the intersection of cybersecurity and machine learning for developing adaptive and threat-resistant systems.

PRASHANT PILLAI (Senior Member, IEEE) is currently an Indian-Born British Computer Scientist. He is also the Pro Vice-Chancellor of Research and Knowledge Exchange with the University of Wolverhampton, U.K., where he was the Associate Dean of Research and Knowledge Exchange and a Professor in cyber security. He is also the Director of Cyber Quarter -Midlands Centre for Cyber Security. He specializes in security and privacy for communication networks. He has led several national/international research projects in the areas of networking protocols and cyber security. He has been invited for talks at various international conferences and has published over 100 papers within peer-reviewed journals and conferences. He is a fellow of British Computer Society (FBCS) and the Higher Education Academy (FHEA) and a member of the Institution of Engineering and Technology (MIET).