IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 June 2024, accepted 6 August 2024, date of publication 16 August 2024, date of current version 28 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3444907

RESEARCH ARTICLE

Secure Explainable-AI Approach for Brake Faults Prediction in Heavy Transport

MUHAMMAD AHMAD KHAN[1], MAQBOOL KHAN[1,2], (Senior Member, IEEE), HUSSAIN DAWOOD[3], (Senior Member, IEEE), HASSAN DAWOOD[4], AND ALI DAUD[5]

[1]Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur, Khyber Pakhtunkhwa, Pakistan
[2]Software Competence Center Hagenberg, 4232 Hagenberg, Austria
[3]School of Computing, Skyline University College, Sharjah, United Arab Emirates
[4]Software Engineering Department (SED), University of Engineering and Technology Taxila 47080, Pakistan
[5]Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates

Corresponding author: Ali Daud (alimsdb@gmail.com)

ABSTRACT Ensuring the safety of vehicles requires the critical responsibility of diagnosing and correcting brake faults. Implementing this proactive measure to address brake faults not only ensures the protection of lives but also enhances the efficiency and cost-effectiveness of repair processes conducted on-site. Machine learning technology has recently contributed to a significant rise in the popularity of predictive maintenance. The objective of this study is to provide a method for identifying issues with the air pressure system (APS) of air brake systems in heavy-duty vehicles. The data obtained by sensors has been used to analyse the APS failure in this Scania Truck. After examining numerous classification methods, Random Forest was determined to have the greatest performance, with a classification accuracy of 99.4%. Moreover, the implementation of eXplainable Artificial Intelligence has included the use of SHapley Additive exPlanation (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) to provide explanations for the contributions of features in model predictions. We picked 20 features from the wheel speed sensor data received from several Internet of Things (IoTs) sensors, which significantly influenced our final selection. By repeatedly applying random forest to these 20 features, we achieved the same degree of accuracy as previously. Consequently, our suggested approach used a reduced amount of computer resources and was less intricate to execute in terms of calculation.

INDEX TERMS Predictive maintenance, machine learning, eXplainable AI (XAI), SHAP, smart transportation system.

I. INTRODUCTION

Explainable Artificial Intelligence, often known as XAI, is a forward-thinking and essential initiative within field of artificial intelligence (AI) [1]. Its primary objective is to tackle a significant obstacle that AI systems must surmount, namely capability of providing humans with explanations that are easy to comprehend for their judgments and forecasts. Demand for people to believe and have faith in results that are derived from AI models is driving force behind relevance of XAI. Increasing interpretability and transparency of results produced by machine learning (ML) algorithms is major focus of XAI [2]. It is acknowledged that "black-box" character of certain AI models might generate suspicion and limit their acceptance in important fields like healthcare industry [3], financial industry, academics [4] and autonomous cars [5], [6], [7]. Goal of XAI is to enable consumers to understand why a given choice was taken, rather than taking it as an unintelligible result, by revealing underlying mechanisms of AI algorithms and making their decision-making process more visible. This is accomplished through discovering internal processes of AI algorithms [8], [9]. XAI has potential to democratize AI technology and allow its responsible and ethical use across various fields by providing consumers with a greater knowledge of AI systems. Continued development of XAI has a possibility of bringing about a new age of AI-human cooperation, one in which AI systems will become more understandable, trustworthy, and

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

seamlessly integrated into our everyday lives [10], [11], [12], [13]. Additionally, prioritizing the alignment of XAI with privacy and security factors is imperative [14], [15], [16].

Heavy Commercial Road Vehicles (HCRVs) rely primarily on air braking systems because of their dependability and safety advantages they provide due to usage of compressed air to actuate braking mechanism. However, just like any other complicated system, air brake systems may run into problems that need to be fixed to keep passengers of vehicle as well as the vehicle itself safe [17]. Problem of an increased stroke length in pushrod is one of major challenges that are related to air brake system [18]. Natural wear and tear of brake lining as well as thermal expansion of braking drum are primary causes of this longer stroke. Cumulative effect of these elements might, over time, result in an excessive gap between brake lining and brake drum. This reduces the effectiveness of brakes as a whole and can cause vehicle to become unstable. This is a major cause for worry because it may result in longer stopping distances, which makes it more difficult for drivers to maintain control of vehicles while they are using brakes [19].

A compressor that is linked to engine is responsible for producing necessary amount of air pressure for braking system. This compressor works to increase air pressure before storing it in vehicle's air storage tanks until braking system needs it again [20], [21]. Multiple air pressure sensors are included in air brake system to facilitate management of various facets of technology. This facilitates operation that is both dependable and smooth. Application of air pressure to braking mechanism is controlled in large part by these circuits, which also play an important role in releasing pressure. A simple schematic diagram of an air brake system with APS is given in Figure 1. This was presented in [22] as a base for illustration of air brake systems in HCRVs. Pressure sensors are mounted on air reservoirs to monitor pressure level and leakage if happens. Air brake chamber, shown in Figure 2, is an important part of air brake system. This space serves two purposes simultaneously. It is what slows or stops a vehicle by turning air pressure created by brake pedal into mechanical force applied to braking components like brake shoes or brake pads [23].

Mechanical failures and delays not only result in increased expenses but also put precious lives in danger. Traditional brake maintenance and repairs may not always be feasible, especially in situations when there is limited time and resources. Predictive maintenance (PdM) is an essential component in ensuring maximum up-time and minimizing risk of unanticipated breakdowns as a result of this challenge [24]. To improve upkeep of HCRVs, PdM makes use of several different technologies, including ML, big data analysis, constraint programming, and route optimization. It is possible to get useful insights about individual maintenance needs of each product by collecting data from embedded sensors in vehicles and evaluating that data. This data is subsequently used in training of ML algorithms, such as predictive random forest (R.F) model, which can accurately forecast requirements for vehicle maintenance based on operational data collected from the vehicle [25]. When operational data from vehicles is entered into predictive model, performance of vehicle can be analyzed, and the most effective maintenance strategy can be devised for each specific vehicle. This procedure is performed on each vehicle, which results in a maintenance strategy that is more accurate and effective since it is suited to specific requirements of each vehicle. Use of methodologies that can be explained in validation of model guarantees that findings are clear and visible to maintenance workers, which further increases the level of faith that can be placed in AI suggestions of system [26].

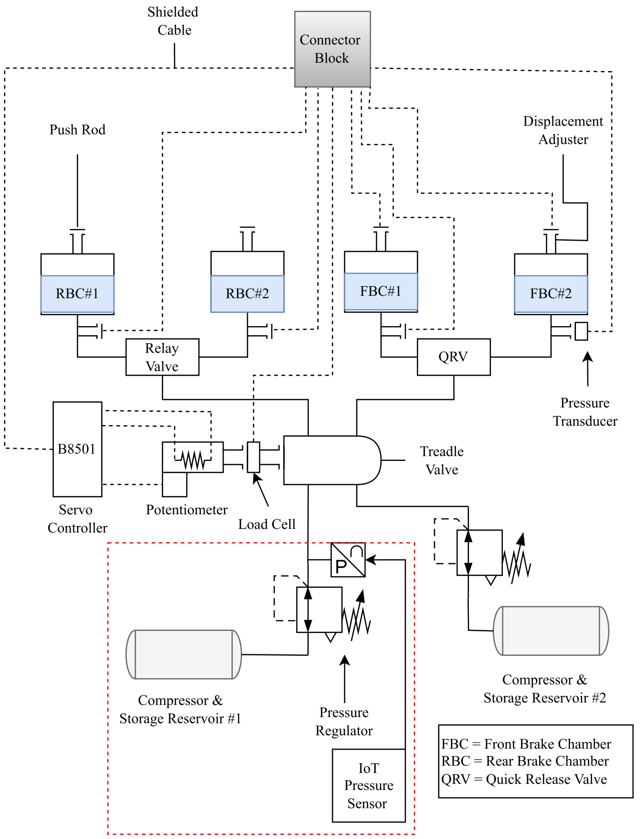


FIGURE 1. IoT-enabled intelligent air brake system: Schematic diagram for sensor integration and data collection.

A viable strategy to improve HCRVs maintenance is to combine PdM strategies with AI models that are explicable. Maintenance schedules may be tailored to individual requirements of each vehicle with the use of ML and big data analysis [27], hence minimizing the number of breakdowns and unplanned breaks. Use of AI that is capable of being explained guarantees that decision-making process that lies behind suggestions for maintenance is open to scrutiny and can be comprehended, therefore striking a balance between risks of failure and need for maintenance. This strategy will, in long run, result in increased productivity decreased expenses, and improved security [28] in the movement of commodities. This work offers vital insights into the explainability of AI in the context of analyzing brake faults in HCRVs. However, it is crucial to acknowledge that the scope of XAI extends beyond the specific area of focus in this study. Explainability plays a vital role in establishing confidence and understanding in the usage of AI technology. To better understand how XAI contributes to these elements, it is important to delve into its practical implementation and user interaction

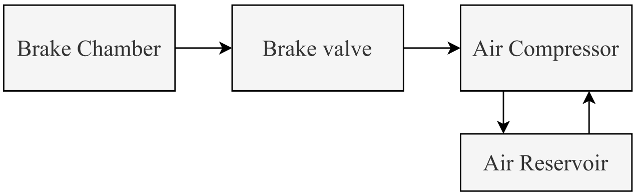


FIGURE 2. Operational framework air pressure system.

Following are the key contributions of this study:

- Data preprocessing with imputation and imbalance class handling with Synthetic Minority Over-sampling Technique and Tomek Links (SMOTE)
- Classification of fault using ML algorithms
- Interpretability of the model using SHapley Additive exPlanation
- Sequential modeling for validation of SHAP results
- Implementation of LIME for sequential model interpretation

Remaining sections of this paper are organized as follows: In Section II, we discuss related work on brake fault prediction of HCRVs and XAI. Section III details the methodology for our proposed model. Section III presents our experimental results and Section V concludes the paper.

II. RELATED WORK

Researchers have made great strides in the field of vehicle brake fault prediction, both in terms of their knowledge of air braking systems and their ability to perform diagnostic testing on such systems. To precisely establish length of pushrod's stroke and pressure inside the brake chamber, Kandt et al. [29] included a numerical model in their research. This model is an invaluable tool for assessing operation of air brake system, as it offers crucial insights into the performance of system as well as its qualities. Through development of model-based diagnostic approaches, Subramanian et al. [22] were able to further enhance diagnostic capabilities of air brake systems. Their study centered on developing methods to predict pressure fluctuations that occur within braking chamber of an air brake system. It is possible to identify potential problems or abnormalities in braking system early on by forecasting pressure changes. This enables timely maintenance and reduces danger of unanticipated failures. When it comes to improving safety and dependability of HCRVs when they are being used for smart transportation missions, this diagnostic method is an essential component. Dhar et al. [30] proposed a novel diagnostic method for anticipating out-of-adjustment problems in pushrod's stroke by building a brake chamber pressure and energy model. These numerical models and model-based diagnostic approaches offer crucial tools for precisely assessing operation of air brake system as well as its overall health and provide predictions on possibility of difficulties linked to pushrods. These approaches help to enhancement of vehicle safety, operating efficiency, and overall transportation dependability. Maintenance staff may spot differences in pushrod's stroke adjustment by monitoring pressure and energy dynamics in brake chamber. This allows for quick maintenance steps to be taken, which helps avoid potentially dangerous situations on road. Ramarathnam et al. [31] developed another model-based approach to fault diagnosis in air brake systems that include leakage. Robust machinery is required to put these concepts into action; yet, these machines are not favorable to environment or economy. These designs only include usage of brakes, and they have not been tested for a diverse array of vehicle weights and driving circumstances. This investigation into a model for use of airborne pollutants using an algorithm that is based on data began with these participants, who served as motivation for project. This approach applies to a broad variety of applications and makes use of cycle velocity data to test and identify pushrod's stroke.

ML has become a prominent technology in the field of PdM, where fault detection and diagnosis (FDD) is a key application. To develop an ML model for prediction, a reasonable amount of data is required. This dataset is typically divided into a training set and a testing set for model training and evaluation, respectively [32], [33], [34]. Raveendran et al. [35] focused on fault identification, and various ML techniques including decision tree (D.T) and R.F, were applied to a wheel speed sensor dataset. Results showed that R.F model outperformed other methods. Building on this work, researchers extended their analysis and tested additional algorithms such as Naïve Bayes, Support Vector Machines (SVM), SVM (Linear), SVM (Gaussian), and K-Nearest Neighbor (K-NN) on the same dataset with variations in the amount of training and testing data. Once again, R.F model demonstrated superior performance. In the context of stability control, study [36] introduced a model called Sliding Mode Observer (SMO) for brake applications. SMO model is designed to enhance vehicle stability control, further improving safety and performance. Another study applied Gaussian Kernel SVM (G-SVM) to predict faults in brake system of front right vehicle [37]. This demonstrates versatility of ML techniques in various aspects of PdM. Preference for R.F models in these studies can be attributed to their resistance to model overfitting issues [38]. Overfitting occurs when a model is too complex and learns noise in data, leading to poor generalization. R.F is less prone to overfitting due to its ensemble nature, which aggregates multiple decision trees. Development of AI has been heavily influenced by our understanding of natural human intelligence. However, path toward Artificial General Intelligence (AGI) requires integration of common sense, cognitive models, and computing approaches that emulate human behaviour [39]. AGI aims to create intelligent systems capable of generalizing and understanding world similar to human intelligence, rather than being limited to specific tasks like current narrow AI systems.

Explaining these models is becoming more and more important due to the extensive usage of ML for critical predictions. A novel method known as SHAP has been used to maintain confidence in their forecasts. Finding out how much of an impact each attribute has on the model's predictions is the goal of this approach [40], [41], [42], [43]. The prevailing opinion is that the Shapley value frequently arises as the foremost approach capable of satisfying specific criteria in XAI. Significance of this option stems from its distinctive attributes and numerous benefits, rendering it an irresistible alternative that is difficult to disregard in sector. Although the output of the Shapley value computation is unique, specific value obtained can vary greatly based on the properties of model, data used for training, and the context in which the explanation is sought. Kwon et al. [44] found which features were important for which tasks without having to retrain the system. They did this by using the Shapley value to find the feature's variance explained. To solve the attribution issue, Å trumbelj et al. [45] established the significance of traits for certain forecasts by using the Shapley value. The first research finds the Shapley value by retraining the model with every conceivable subset of features. Without retraining the model, the second research applies the Shapley value to its conditional expectation. No matter how independent characteristics may be, both methods treat them as if they were randomly distributed.

Shapley value for the conditional expectancies of a model's function was calculated by Datta et al. [46] using a constructed distribution. This distribution was generated by multiplying the marginals of the distribution of underlying features. A study by Lundberg and Lee [47] looked at how to use conditional expectancies to find the Shapley value, how to get different approximations based on function or distribution assumptions, and how to combine these to use them in deep network modules. The significance of features was assessed in their study using the Shapley value. In particular, Matzka S. used an easy-to-understand XAI method for such PdM [48], [49]. But these methods do not do a good job of dealing with major problems like class imbalance and missing data, which our suggested ways fix. Additionally, none of these approaches offer a clear-cut, simple solution.

Our research is the first of its kind to focus on the unique problem of predicting when brakes may fail while also placing a premium on how well models can be explained. We use a dataset of medium size that has imbalanced classes and missing values. Our goal is to create a plan that addresses the problems of imbalanced classes and missing data in this field by improving our approaches. By providing solutions to these unique problems within the framework of predicting when brakes will fail, our study adds to the existing literature.

III. RESEARCH METHODOLOGY

Standard procedures for using SHAP to examine black box models and ensure model interpretability usually include the following:

- **Preparing Data:** First step is to prepare data used to train black box model. We need to make sure the dataset is preprocessed, cleaned, and ready for analysis [50], [51].
- **Defining Baseline Model:** Choose a black box model (e.g., ensemble algorithm) that we want to interpret using SHAP. Train model on prepared dataset to create baseline model.
- **Compute Shapley Values:** Calculate Shapley values for each feature in dataset. Shapley values contribute to each feature to prediction of model for a specific instance.
- **Interpret Shapley Additive Feature Attribution Values:** Analyze Shapley values to understand how each feature affects output of model. Positive Shapley values indicate a feature contributes positively to prediction, while negative values suggest a negative contribution.
- **Visualize Explanations:** To gain better insights and communicate explanations effectively, we create visualizations of SHAP values. Plotting individual feature attributions or summary plots can help us understand feature importance.
- **Validate and Refine Explanations:** It is essential to validate explanations provided by SHAP to ensure they are accurate and reliable. This may involve checking results against known ground truths or comparing explanations across different model instances. If necessary, we can refine explanations by adjusting parameters or using alternative techniques.

  By implementing these steps, SHAP can shed light on inner workings of black box models and provide valuable insights into feature importance and model behaviour. This understanding is critical for building trust in AI systems, ensuring fairness, and identifying potential biases in decision-making process.

Process flow of the suggested method is shown in Figure 3.

A. DATASET DESCRIPTION

Scania Trucks APS failure dataset has been used for fault classification and model evaluation. Scania is a diversified firm that produces HCRVs, coaches, and engines for use in industrial and marine applications. The data has been collected using different IoT sensors like Accelerometers, Gyroscopes, Motion sensors, Pressure sensors, Proximity sensors, Temperature sensors and different Internet of Vehicles (IoVs) sensors. Training set and testing set that makeup Scania Trucks APS failure dataset are both included in the set. There are 1,000 failure examples out of 60,000 in the training set that are marked as positive. This dataset is made up of 171 columns, one of which is specified to be used as the class label. Seventy different features are broken up and placed in seven different histograms; each histogram has ten different bins to place characteristics.

Notably, characteristics were encrypted to protect security since the information was initially meant for industrial reasons. Furthermore, a large number of attributes had missing values; in fact, some attributes were missing as much as 82% of their data. Additional complications arose from the study due to the existence of several outliers. We used KNN imputation, a non-parametric method that finds values for missing attributes by comparing them to comparable occurrences, to solve the problem of missing data.

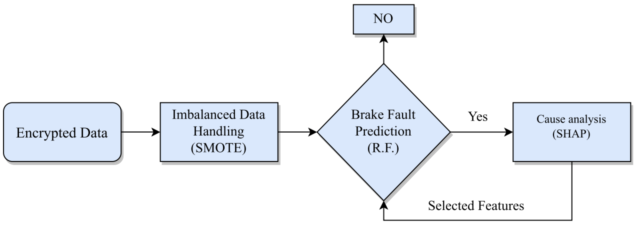

NO
Encrypted Data
Imbalanced Data Handling (SMOTE)
Brake Fault Prediction (R.F.)
Yes
Cause analysis (SHAP)
Selected Features

FIGURE 3. Workflow for the suggested methodology.

B. DATA PREPROCESSING

The dataset is highly imbalanced and it is difficult to use it without some preprocessing techniques for better results of ML algorithms. Therefore for handling class imbalances and missing data respectively, implementation of SMOTE and KNN imputation proved to be advantageous, which contributed to model prediction for possible APS failures. Furthermore, dataset is an invaluable resource for tackling maintenance difficulties in industrial applications, and it highlights significance of processing data and imbalances that might occur in real-world situations. A short description of applied techniques is given below:

1) SMOTE

To rectify dataset class imbalances, SMOTE is used. Using feature-space similarity as a metric, SMOTE creates synthetic instances for underrepresented groups. We utilised Scikit-learn to apply SMOTE and normalise the input data. The dataset was balanced when SMOTE was applied, with 59,000 occurrences for positive and negative classes combined.

2) KNN IMPUTATION

To solve the problem of missing data we utilised KNN imputation, which stands for k-nearest neighbours imputation. Using an average of the attribute values of the nearest instances in the feature space, this method recovers missing values. Similar cases or pieces of data will have comparable values; that is the basic premise. We used KNN imputation to successfully deal with missing values as our dataset includes instances with a large quantity of missing data.

C. MODEL DESCRIPTION

R.F classifier has been used as a base model for fault classification. Then for interpretation of classification results, we applied SHAP [52]. For comparative analysis of feature attribution, we used Local Interpretable Model-Agnostic Explanations (LIME) [53]. Below is the outline of our proposed framework as depicted in Algorithm 1.

Algorithm 1 Random Forest With SHAP and Sequential Random Forest, Along With LIME

```
Let $D_{\text{train}} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ be the training dataset.
Let $D_{\text{test}} = \{x_{\text{test}_1}, x_{\text{test}_2}, \ldots, x_{\text{test}_m}\}$ be the test dataset.
Initialize $\hat{y}_{\text{test}}$ as the predicted labels for $D_{\text{test}}$ using R.F.
Apply SHAP to interpret $R.F$ predictions on $D_{\text{test}}$.
Enrich $D_{\text{augmented}}$ by appending tuples $(x_{\text{test}_i}, \text{SV}_i)$.
Generate $\hat{y}_{\text{augmented}}$ with $RF_{\text{augmented}}$ on $D_{\text{augmented}}$.
Train $Sequential\ RF_{\text{augmented}}$ $on$ $\hat{y}_{\text{augmented}}$.
Report performance improvement of Sequential Random Forest with SHAP.
Apply LIME to interpret $RF_{\text{augmented}}$ predictions on $D_{\text{test}}$.
Analyze LIME explanations for feature importance insights.
```

Note:SV_i: Shapley Values

D. MACHINE LEARNING MODELS

A brief explanation of each ML model implemented is provided below:

1) DECISION TREE

D.T are widely used for classification and regression tasks due to their interpretability and clarity. They partition dataset based on input feature values, making sequential judgments from the root node to leaf nodes to make predictions. Root node represents entire dataset and is split into subgroups based on feature values. At each node, algorithm selects the best feature and threshold to divide data, aiming to maximize information gain and reduce impurity. Child nodes are created after each split, and process continues recursively until a stopping criterion is met. Leaf nodes mark end of D.T, providing final predictions. Decision rules derived from features explain how model arrives at its predictions.

2) RANDOM FOREST

R.F bagging technique excels in noisy or poorly classified data. R.F is reliable in many scenarios since its parameter order is generally unaffected. Bootstrap is used to randomly sample datasets. Each sample generates a D.T without pruning. This process creates a D.T forest. Each D.T votes for class with the most support in R.F model. Permutation-based feature selection strategy using R.F. works well on datasets with large dimensions and strongly correlated variables, which has led to its widespread adoption across disciplines. The Gini coefficient's decline in purity determines a variable's relevance. Permutation approach breaks connection between a variable a_i and outcome variable Y by randomly replacing alternative values for all occurrences of a_i. R.F uses

Gini coefficients to identify variables that reduce predictor purity deterioration. R.F is less sensitive to parameters compared to other predictive models. It depends on two main parameters: *ntrees* and *mtrees*. The *ntrees* parameter determines the number of trees in the R.F model, while the *mtrees* parameter controls the number of variables used to split a D.T node. Determining the optimal value for *mtrees* often involves testing various options. The generalization error of R.F tends to converge with *ntrees*, unlike many other classifiers [54]. More trees improve R.F model. Optimal *ntrees* balance classification accuracy and processing speed.

3) GRADIENT BOOSTING

Gradient Boosting is a powerful ensemble learning technique that has gained widespread popularity for both regression and classification tasks. It operates by constructing multiple weak learners, often D.T, in a sequential manner to iteratively correct errors made by the previous models. Key concept behind Gradient Boosting lies in its focus on residuals or gradients of target variable from previous model. Each new weak learner is designed to capture and learn from these residuals, making subsequent model more adept at addressing remaining errors. This iterative process continues until final model is created, which combines predictions of all weak learners using optimized weights. Strength of Gradient Boosting lies in its ability to handle complex relationships within data.

4) LOGISTIC REGRESSION

Logistic regression is a popular statistical approach for binary classification in ML. It classifies, not regresses. It works well for cases where result variable is yes or no, spam or not spam etc. Logistic Regression models how input characteristics affect binary outcome probability. Logistic function (sigmoid function) transfers every real-valued input to a range between 0 and 1. This range represents instance's positive class probability.

5) KNN CLASSIFIER

K-Nearest Neighbors (KNN) is an easy-to-use machine-learning classifier for regression and classification. It uses idea that related feature space instances have similar labels. KNN uses a user-defined parameter to classify closest neighbors. KNN uses Euclidean distance to create predictions. It then chooses K closest neighbors and applies input point's expected label to majority class label. KNN is a slow, non-parametric approach that defers training until prediction time. It is good for intricate interactions and shifting patterns. KNN's performance depends on K and may be computationally expensive on big datasets. It also assigns equal priority to all characteristics, which might restrict high-dimensional data.

E. FAULT DETECTION USING SHAP VALUES

In mathematical terms, the following describes the local feature attribution problem: The objective is to find an importance vector $\phi \in \mathbb{R}^d$ given a decision function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ obtained from a machine learning model, where d is the number of input features. Concerning each input x, the model's prediction $f(x)$ is affected by the input phi_i in this vector.

Explanatory capacity of methods that attempt to address this problem is often severely limited since they generate relevance ratings based on raw input attributes x. To illustrate the point, inferring precise patterns or finding clear explanations for how the model detects problems becomes hard when trained on raw input [55].

Ribeiro et al. [53] suggested a way to solve this issue by figuring out how important a feature is about an understandable representation $y \in \mathcal{Y}$ of the instance $x \in \mathcal{X}$ that is being explained. To achieve this, it is essential to establish a predefined mathematical context that enables accurate conversion between interpretable representations and original ones. In this regard, we propose following formulation:

Definition 1: The feature space is represented by $\mathcal{X}$ and the mapping from $\mathcal{X}$ to $\mathcal{Y} \times R$ is easily written as $\phi_{x\rightarrow y}$, where R is the residual component and $\mathcal{Y}$ is the interpretable domain.

$$\phi_{x\rightarrow y} : \mathcal{X} \rightarrow (\mathcal{Y} \times R) \tag{1}$$

Similarly, $\phi_{y\rightarrow x}$ represents the mapping from $(\mathcal{Y} \times R)$ to $\mathcal{X}$, defined as:

$$\phi_{y\rightarrow x} : (\mathcal{Y} \times R) \rightarrow \mathcal{X} \tag{2}$$

For every value of x:

$$\phi_{y\rightarrow x}(\phi_{x\rightarrow y}(x)) = x \tag{3}$$

By including the residual component R, we may limit the evaluation of feature significance to certain parts of the representation while still keeping enough data to restore the original input. With the initial model f and some post-hoc tweaks to its input domain, deriving attributions from the representation y and the residual r becomes easy.

An enhanced model, denoted as $\widetilde{f}$, is the result of this procedure:

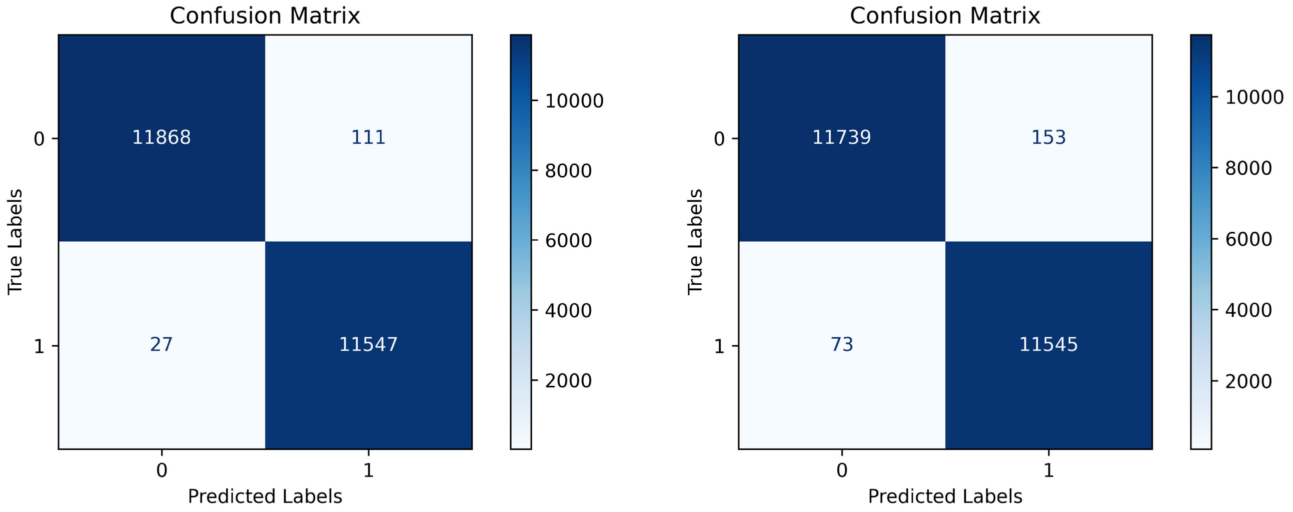
$$\widetilde{f}(y; r) = f(\phi_{y\rightarrow x}(y, r)) \quad \text{where} \quad (y, r) = \phi_{x\rightarrow y}(x) \tag{4}$$

A linear function of binary variables is used in the explanatory model using additive feature attribution methods:

$$g(y') = \phi_0 + \sum_{i=1}^{M}(\phi_i y'_i) \tag{5}$$

The simplified explanation is represented by $g(y')$, where $y' \in \{0, 1\}^M$ is the number of simplified input characteristics and $\phi_i \in \mathbb{R}$. Each feature is given an effect ϕ_i by an explanatory model in approaches that follow Definition 1. After that, $f(x)$ is almost equal to the output of the original model, as it is the result of combining the effects of all feature attributions.

By incorporating this approach, the original model is extended to function within a distinct input domain while



(a). Confusion matrix with complete dataset

(b). Confusion matrix with features withheld

FIGURE 4. Confusion matrix obtained from random forest results.

preserving all essential characteristics of f without necessitating any retraining. This facilitates the computation of feature attributions based on the modified input domain while maintaining the effectiveness and performance of the original model.

According to Scott et al. [55], the only additive technique that meets the requirements of local precision, missingness, and consistency assigns ϕ_i an effect (Shapley value) to each variable x'_i.

$$\phi_i(f) = \sum_{x' \subseteq N \setminus \{i\}} \frac{|x'|! \cdot (|N| - |x'| - 1)!}{|N|!} \cdot (f(x' \cup \{i\}) - f(x')) \tag{6}$$

where the Shapley value of feature i is represented by $\phi_i(f)$. All features in the dataset are represented by N, which is equal to $N = \{1, 2, \ldots, n$. The representation of the number of features in coalition x' is given by $|N|$, which stands for the entire number of features in the dataset, i.e., $|N| = n$, and feature i is not included in this coalition (i.e., $x' \subseteq N \setminus \{i\}$. The characteristic function that represents the overall value of the coalition is denoted as $f(x')$, and the total value of the coalition is $f(x' \cup \{i\})$ when feature i joins.

The formula essentially sums up the marginal contributions of feature i to all possible coalitions x', weighted by the number of ways each coalition can be formed. It considers all possible orders in which a feature can join a coalition to ensure fairness.

Going beyond only estimating brake failure is vital for achieving a holistic view of brake fault prediction. Gaining a more comprehensive understanding of the dynamics at play is essential for equipping domain specialists with the information needed for PdM. In Section IV, we define "cause" as the primary factors that are expected to lead to failures. Just so there is no confusion, we are not saying these are the actual causes; the algorithm just came up with these justifications.

Estimating the relevance of each parameter for fault prediction can be approached in various ways, involving specific steps for exploration and deeper analysis. One widely used explainable method is SHAP, which quantifies parameter importance independently of any particular model. SHAP values are employed in SHAP method to describe the relevance of individual parameters in prediction model.

To successfully utilize SHAP approach, output of prediction model, for example, R.F must align with total SHAP values for a given input. Determining the most crucial parameters for fault prediction entails computing SHAP values of all parameters under consideration. This investigation allows us to delve into underlying reasons for any air brake's air pressure problems.

Given the complexities involved in precisely computing SHAP values, methods such as Tree SHAP, Kernel SHAP, and Deep SHAP are commonly utilized. Despite the intricacies of SHAP value calculations, these methods provide valuable insights. In this study, we implemented a prediction model based on ML and Tree SHAP. A comprehensive mathematical formulation of SHAP can be found in [47], [56], and [57].

F. LIME

Black box ML models use the Locally Interpretable Model Agnostic Explanations (LIME) technique to post-hoc explain their predictions. Its main objective is to approximate the behaviour of any complex model locally by constructing an interpretable model that explains individual predictions [53]. LIME is intentionally designed to be model-agnostic, enabling its application to any classifier, regardless of the specific algorithm employed for predictions.

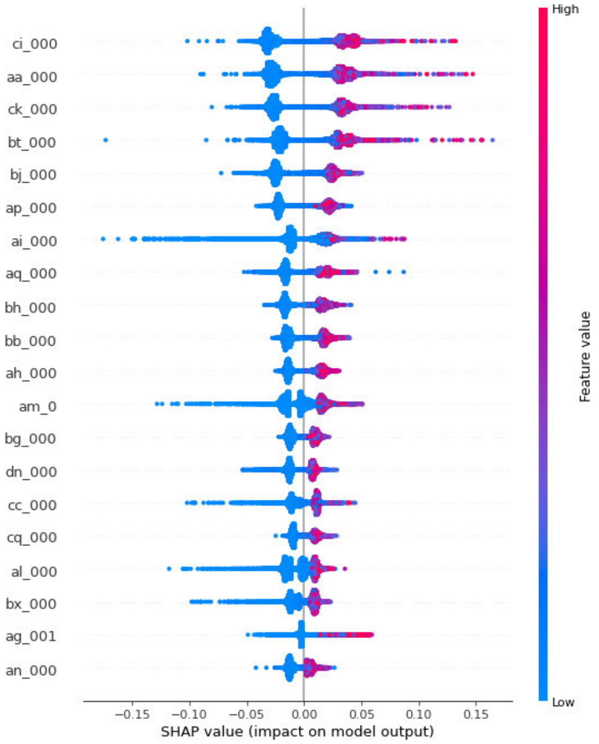


FIGURE 5. Overview plot of SHAP values for the predicted fault categories.

Similar to the SHAP approach, LIME primarily highlights the local context. For each observation, LIME tries to build a localised model so it can explain it. Using a set of data points that are statistically close to the one being described, this method is carried out. D.T and linear models are two examples of the many shapes these interpretable models may take, which helps to make them more user-friendly.

To explain a particular observation x, LIME constructs a local model by utilizing a subset of data points that are similar to x. This local model is then analyzed to interpret the prediction for x. This method enables LIME to shed light on the decision-making process of black-box models, making it a crucial tool for model interpretation and enhancing transparency in complex ML systems. The objective function minimized by LIME to determine ϕ_i is expressed as follows:

$$\psi = \arg \min_{g \in \mathcal{G}(f,g,\pi_x)} (L(f, g, \pi_x) + \Omega(g)) \tag{7}$$

In this formulation:

- An explanation model is represented as $g \in \mathcal{G}$, where $\mathcal{G}$ is a set of theoretically interpretable models, which includes linear models and decision trees.
- The function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ maps input instances in $\mathbb{R}^d$ to an output with a real value.
- The proximity metric $\pi_x(y)$ takes into account how near an instance y is to instance x.
- To determine how complicated the explanation is, the function $\Omega(g)$ is used.
- To show that LIME is model-agnostic, we aim to minimise the locality-aware loss L without supposing any particular shape for f. How closely the function f is approximated within the locality described by $\pi(x)$ is quantified by the loss L.

Ultimately, LIME seeks to identify a simple model g from the set $\mathcal{G}$ that closely approximates the behaviour of the complex black-box model f in the vicinity of instance x, as indicated by the proximity measure $\pi_x(y)$. This methodology allows LIME to deliver model-agnostic explanations for individual predictions, independent of any prior knowledge about the underlying black-box model.

IV. RESULTS AND DISCUSSIONS

As a practical demonstration, we have selected Scania Trucks APS failure dataset, which serves as an ideal example for model explainability and feature contribution evaluation. Input features in this dataset have been anonymized to protect proprietary information, making it suitable for studying and assessing feature importance without revealing sensitive details. This dataset allows us to apply the earlier approach and gain insights into how the model operates within modified input domain while preserving its explainability and performance.

A. COMPARISON OF MACHINE LEARNING CLASSIFIERS

We can determine efficiency of our ML model in terms of its accuracy, precision, recall and F1-Score by using various assessment parameters. R.F had the highest overall testing accuracy 99.4% compared to other methods we tested i.e. Decision Tree, Gradient Boosting, Logistic Regression, K-Neighbors Classifier, XGBClassifier, CatBoosting Classifier, and AdaBoost Classifier. Table 1 below provides results of a comparison study of all Classifiers with minority sampling strategy used by SMOTE.

TABLE 1. Classifiers performance after SMOTE.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Random Forest	0.99	0.99	0.99	0.99
Decision Tree	0.98	0.98	0.99	0.98
Gradient Boosting	0.98	0.97	0.98	0.98
Logistic Regression	0.59	0.55	0.95	0.70
K-NN Classifier	0.97	0.96	0.98	0.97
XGB Classifier	0.98	0.98	0.99	0.97
CatBoosting Classifier	0.98	0.99	0.98	0.98
AdaBoost Classifier	0.97	0.97	0.97	0.98

Confusion matrix is constructed using predicted labels of trained machine-learning model and actual labels of each record in test dataset. It is now possible to get ML model assessment parameters. Confusion matrix, shown in Figure 4.(a) is a result of R.F model applied to experimental dataset.

B. CAUSE ANALYSIS WITH SHAP

After assessing multiple classification techniques, we opted for R.F as a base model to analyze relevant feature attribution. Subsequently, we applied SHAP and observed the following results.

For illustration of features contribution in decision process of fault classification, a summary plot is given below in Figure 5. While class 0 indicates no fault and class 1 indicates prediction of fault in classification.

A SHAP decision plot in XAI provides insights into relative impact of each input variable on model predictions. It reveals factors that influence model's outcomes and sheds light on decision-making process. In Figure 6, SHAP decision plot ranks attributes based on their importance in fault identification. In plot, the most significant features in model's prediction are represented towards rightmost side and represent positive class prediction. It is essential to note that contribution of each variable in positive class (class 1) is emphasized in plot. As dataset's details are not transparent, we can only speculate that the most helpful feature likely belongs to family of data that includes air pressure, engine load, gear selections, and air consumption from Scania trucks during actual operation. These features appear to play a crucial role in fault identification based on SHAP decision plot.

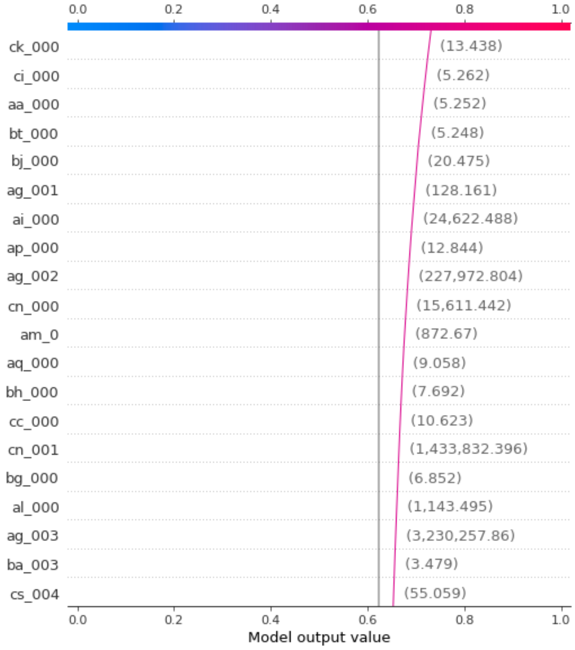


FIGURE 6. Decision plot of SHAP values for predicted class faults.

C. SEQUENTIAL RANDOM FOREST

Upon evaluating SHAP and identifying the most influential feature for model predictions, we proceeded to utilize only 20 out of initial 171 selected features. Remarkably, after replicating process, we found that these chosen features yielded identical results compared to using entire dataset's features.

This highlights effectiveness of SHAP in selecting essential features. By doing so, we not only reduced model complexity and conserved computational resources but also gained valuable insights into inner workings of ML models, which are often perceived as black boxes.

Results for the sequential R.F are shown below in Table 2.

Confusion matrix shown in Figure 4.(b) is a result of sequential R.F model applied to experimental dataset.

TABLE 2. Classification report from sequential random forest.

	Accuracy	Precision	Recall	F1 Score
Random Forest	0.99	0.99	0.98	0.99
Class Negative	0.99	0.99	0.98	0.98
Class Positive	0.99	0.99	0.97	0.99

D. LIME

Both LIME and SHAP get parameters for feature contribution at observation level (local explanation); however, techniques that achieve this task are different. LIME obtains parameters locally, whereas SHAP obtains them globally. Figure 7 shows a visual representation of the results from LIME, comparing two techniques in terms of their capacity to assess the effect of variables at the regional level. These findings were obtained using LIME.

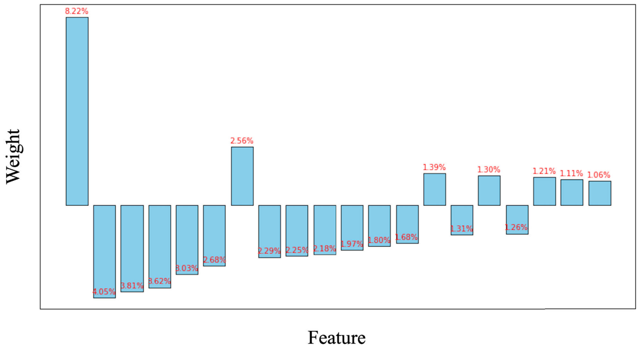


FIGURE 7. Representation of features attribute from LIME.

E. ALGORITHM EFFICIENCY REVOLUTION

The computational efficiency of our training and prediction processes was investigated in our research about optimisation measures. We achieved significant reductions in space and temporal complexity by exhaustive analysis and repeated refining. As an example, the training time complexity was lowered from O(10 * 171 * log(171)) to O(10 * 20 * log(20)), which is an 88.24% reduction. The prediction phase also saw a drop, going from $O(10 \times \log(171))$ to $O(10 \times \log(20))$, which is a reduction of 54.1%. Furthermore, improvements in space complexity were seen during training, when the need decreased from $O(10 \times 171)$ to $O(10 \times 20)$, indicating a decrease of 86.35%. The findings show that our optimisation methods were successful in making our computing processes more efficient, which allowed us to train and forecast models faster without compromising accuracy.

V. CONCLUSION

Primary focus of our study is predicting APS issues within air brake systems, employing various classification methods. We utilized multiple classification algorithms for fault detection, with the R.F classifier emerging as the most effective. To understand the feature contribution within the R.F model, we leveraged SHAP to compute overall feature attribution, revealing that only 20 out of 171 features significantly influenced model prediction. Subsequently, we utilized R.F to implement these SHAP-identified features, finding that the accuracy of these features was consistent with previous results. Our proposed solution not only streamlines the process but also reduces the demand for computational resources. The most notable achievement of our study is our concerted effort to delve into the black box of ML models, shedding light on the inner workings and enhancing transparency. To enhance interpretability in future research, we propose to employ DeepLIFT, DeepSHAP (DeepLIFT+Shapley Values).

REFERENCES

[1] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai-explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, 2019, Art. no. eaay7120.

[2] A. Barredo Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[3] I. Masood, Y. Wang, A. Daud, N. R. Aljohani, and H. Dawood, "Towards smart healthcare: Patient data privacy and security in sensor-cloud infrastructure," *Wireless Commun. Mobile Comput.*, vol. 2018, no. 1, Jan. 2018, Art. no. 2143897.

[4] A. Badshah, A. Ghani, A. Daud, A. Jalal, M. Bilal, and J. Crowcroft, "Towards smart education through Internet of Things: A survey," *ACM Comput. Surveys*, vol. 56, no. 2, pp. 1–33, Feb. 2024.

[5] V. Balasubramaniam, "Artificial intelligence algorithm with SVM classification using dermascopic images for melanoma diagnosis," *J. Artif. Intell. Capsule Netw.*, vol. 3, no. 1, pp. 34–42, Mar. 2021.

[6] C. Singh and W. Lin, "Can artificial intelligence, RegTech and CharityTech provide effective solutions for anti-money laundering and counter-terror financing initiatives in charitable fundraising," *J. Money Laundering Control*, vol. 24, no. 3, pp. 464–482, Jul. 2021.

[7] N. A. Hamad, K. M. A. Alheeti, and S. S. Al-Rawi, "Intrusion detection system using artificial intelligence for internal messages of robotic cars," in *AIP Conf. Proc.*, 2022, pp. 1–20.

[8] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[9] X. Wang and M. Yin, "Are explanations helpful? A comparative study of the effects of explanations in AI-assisted decision-making," in *Proc. 26th Int. Conf. Intell. User Inter.*, Apr. 2021, pp. 318–328.

[10] R. Hamon, H. Junklewitz, I. Sanchez, G. Malgieri, and P. De Hert, "Bridging the gap between AI and explainability in the GDPR: Towards trustworthiness-by-design in automated decision-making," *IEEE Comput. Intell. Mag.*, vol. 17, no. 1, pp. 72–85, Feb. 2022.

[11] F. Emmert-Streib, O. Yli-Harja, and M. Dehmer, "Explainable artificial intelligence and machine learning: A reality rooted perspective," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 6, p. e1368, Nov. 2020.

[12] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, and W. Samek, "Explainable ai methods-a brief overview," in *Proc. Int. Workshop Extending Explainable AI Beyond Deep Models Classifiers*, 2020, pp. 13–38.

[13] S. Nazat, L. Li, and M. Abdallah, "XAI-ADS: An explainable artificial intelligence framework for enhancing anomaly detection in autonomous driving systems," *IEEE Access*, vol. 12, pp. 48583–48607, 2024.

[14] C. N. Spartalis, T. Semertzidis, and P. Daras, "Balancing XAI with privacy and security considerations," in *Proc. Eur. Symp. Res. Comput. Secur.*, 2023, pp. 111–124.

[15] H. Montenegro, W. Silva, A. Gaudio, M. Fredrikson, A. Smailagic, and J. S. Cardoso, "Privacy-preserving case-based explanations: Enabling visual interpretability by protecting privacy," *IEEE Access*, vol. 10, pp. 28333–28347, 2022.

[16] Z. Miftioglu, M. A. Kızrak, and T. Yıldırım, "Privacy-preserving mechanisms with explainability in assistive AI technologies," in *Learning and Analytics in Intelligent Systems*. Cham, Switzerland: Springer, 2022, pp. 287–309.

[17] I. C. Guleryuz and Ö. Baser, "Modelling the longitudinal braking dynamics for heavy-duty vehicles," *Proc. Inst. Mech. Engineers, Part D, J. Automobile Eng.*, vol. 235, nos. 10–11, pp. 2802–2817, Sep. 2021.

[18] I. A. K. Soudagar and P. D. Tota, "Modelling and simulation of electro-pneumatic parking brake system for real time estimation of pressure inside parking brake chamber," Tech. Rep., 2022.

[19] J. S. Avliyokulov, M. S. Pulatovich, and M. I. Rakhmatov, "Main failures of the vehicle brake system, maintenance and repair," *Central ASIAN J. Math. Theory Comput. Sci.*, vol. 4, no. 3, pp. 63–69, 2023.

[20] X. Hua, J. Zeng, H. Li, J. Huang, M. Luo, X. Feng, H. Xiong, and W. Wu, "A review of automobile brake-by-wire control technology," *Processes*, vol. 11, no. 4, p. 994, Mar. 2023.

[21] T. M. Alamelu Manghai, R. Jegadeeshwaran, and G. Sakthivel, "Real time condition monitoring of hydraulic brake system using naive Bayes and Bayes net algorithms," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 624, no. 1, Oct. 2019, Art. no. 012028.

[22] S. C. Subramanian, S. Darbha, and K. R. Rajagopal, "A diagnostic system for air brakes in commercial vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 360–376, Sep. 2006.

[23] D. Leontiev, V. Klimenko, M. Mykhalevych, Y. Don, and A. Frolov, "Simulation of working process of the electronic brake system of the heavy vehicle," in *Proc. Int. scientific-Practical Conf.*, 2019, pp. 50–61.

[24] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Comput. Ind. Eng.*, vol. 150, Dec. 2020, Art. no. 106889.

[25] M. H. Abidi, M. K. Mohammed, and H. Alkhalefah, "Predictive maintenance planning for industry 4.0 using machine learning for sustainable manufacturing," *Sustainability*, vol. 14, no. 6, p. 3387, Mar. 2022.

[26] N. Omrani, G. Rivieccio, U. Fiore, F. Schiavone, and S. G. Agreda, "To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts," *Technological Forecasting Social Change*, vol. 181, Aug. 2022, Art. no. 121763.

[27] M. Khan, A. Ahmad, F. Sobieczky, M. Pichler, B. A. Moser, and I. Bukovský, "A systematic mapping study of predictive maintenance in SMEs," *IEEE Access*, vol. 10, pp. 88738–88749, 2022.

[28] R. Alharbey, A. Shafiq, A. Daud, H. Dawood, A. Bukhari, and B. Alshemaimri, "Digital twin technology for enhanced smart grid performance: Integrating sustainability, security, and efficiency," *Frontiers Energy Res.*, vol. 12, Jun. 2024, Art. no. 1397748.

[29] L. D. Kandt, P. G. Reinhall, and R. R. Scheibe, "Determination of air brake adjustment from air pressure data," *Proc. Inst. Mech. Engineers, Part D, J. Automobile Eng.*, vol. 215, no. 1, pp. 21–29, Jan. 2001.

[30] S. Dhar, *Development of Diagnostic Algorithms for Air Brakes in Trucks*. Austin, TX, USA: Univ. of Texas Press, 2010.

[31] S. Ramarathnam, S. Dhar, S. Darbha, and K. R. Rajagopal, "Development of a model for an air brake system with leaks and a scheme for the estimation of the steady-state pushrod stroke," *Vehicle Syst. Dyn.*, vol. 49, no. 8, pp. 1267–1282, Aug. 2011.

[32] G. Bode, S. Thul, M. Baranski, and D. Muller, "Real-world application of machine-learning-based fault detection trained with experimental data," *Energy*, vol. 198, May 2020, Art. no. 117323.

[33] F. Sobieczky, I. Bukovsky, O. BudíK, and M. Khan, "A predictive maintenance cost-model," Rep., 2022.

[34] M. A. Khan, F. Ahmad, K. Khan, and M. Khan, "Pashto language handwritten numeral classification using convolutional neural networks," in *Proc. Int. Conf. Netw. Sustainability AIoT Era*, 2024, pp. 287–297.

[35] R. Raveendran, K. B. Devika, and S. C. Subramanian, "Intelligent fault diagnosis of air brake system in heavy commercial road vehicles," in *Proc. Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2020, pp. 93–98.

[36] R. Raveendran, K. B. Devika, and S. C. Subramanian, "Brake fault identification and fault-tolerant directional stability control of heavy road vehicles," *IEEE Access*, vol. 8, pp. 169229–169246, 2020.

[37] R. Raveendran, K. Devika, and S. C. Subramanian, "Learning-based fault diagnosis of air brake system using wheel speed data," *Proc. Inst. Mech. Engineers, Part D, J. Automobile Eng.*, vol. 236, no. 12, pp. 2598–2609, Oct. 2022.

[38] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst. Appl.*, vol. 134, pp. 93–101, Nov. 2019.

[39] M. Mitchell, "Why AI is harder than we think," 2021, *arXiv:2104.12871*.

[40] S. M. Srinivasan, T. Truong-Huu, and M. Gurusamy, "Machine learning-based link fault identification and localization in complex networks," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6556–6566, Aug. 2019.

[41] J. Zareei, M. Haseeb, K. Ghadamkheir, S. A. Farkhondeh, A. Yazdani, and K. Ershov, "The effect of hydrogen addition to compressed natural gas on performance and emissions of a DI diesel engine by a numerical study," *Int. J. Hydrogen Energy*, vol. 45, no. 58, pp. 34241–34253, Nov. 2020.

[42] J. Liu, C. Pan, F. Lei, D. Hu, and H. Zuo, "Fault prediction of bearings based on LSTM and statistical process analysis," *Rel. Eng. Syst. Saf.*, vol. 214, Oct. 2021, Art. no. 107646.

[43] J. Yao, B. Lu, and J. Zhang, "Tool remaining useful life prediction using deep transfer reinforcement learning based on long short-term memory networks," *Int. J. Adv. Manuf. Technol.*, vol. 118, nos. 3–4, pp. 1077–1086, Jan. 2022.

[44] Y. Kwon, M. A. Rivas, and J. Zou, "Efficient computation and analysis of distributional Shapley values," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2021, pp. 793–801.

[45] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowl. Inf. Syst.*, vol. 41, no. 3, pp. 647–665, Dec. 2014.

[46] A. Datta, S. Sen, and Y. Zick, "Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2016, pp. 598–617.

[47] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[48] S. Matzka, "Explainable artificial intelligence for predictive maintenance applications," in *Proc. 3rd Int. Conf. Artif. Intell. Industries (AI4I)*, Sep. 2020, pp. 69–74.

[49] M. Kisten, A. E.-S. Ezugwu, and M. O. Olusanya, "Explainable artificial intelligence model for predictive maintenance in smart agricultural facilities," *IEEE Access*, vol. 12, pp. 24348–24367, 2024.

[50] M. Qasim and M. Khan, "A comparative analysis of anomaly detection methods for predictive maintenance in SME," in *Proc. Database Expert Syst. Appl.*, vol. 1633, 2022, pp. 22–31.

[51] M. Khan, X. Wu, X. Xu, and W. Dou, "Big data challenges and opportunities in the hype of industry 4.0," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[52] A. B. Owen, "Sobol' indices and Shapley value," *SIAM/ASA J. Uncertainty Quantification*, vol. 2, no. 1, pp. 245–251, Jan. 2014.

[53] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?'Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144.

[54] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata J., Promoting Commun. Statist. Stata*, vol. 20, no. 1, pp. 3–29, Mar. 2020.

[55] Y. Zhou, S. Booth, M. T. Ribeiro, and J. Shah, "Do feature attribution methods correctly attribute features?" in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 9, pp. 9623–9633.

[56] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, and A. Mohammadian, "Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis," *Accident Anal. Prevention*, vol. 136, Mar. 2020, Art. no. 105405.

[57] G. Van den Broeck, A. Lykov, M. Schleich, and D. Suciu, "On the tractability of SHAP explanations," *J. Artif. Intell. Res.*, vol. 74, pp. 851–886, Jun. 2022.

MUHAMMAD AHMAD KHAN received the bachelor's degree in electronic engineering from The Islamia University of Bahawalpur, Pakistan, in 2021. He is currently pursuing the master's degree in artificial intelligence with the Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology (PAF-IAST), Pakistan. He was a Research Assistant with the Department of IT and Computer Science, PAF-IAST. He was also an Intern with the Software Competence Center, Hagenberg, Austria. His research interests include machine learning, neural networks, the IoT, and eXplainable Artificial Intelligence (XAI).

MAQBOOL KHAN (Senior Member, IEEE) received the M.S. degree from HUST, Wuhan China, and the Ph.D. degree from Nanjing University, China, on fully funded CSC Scholarship, in 2011 and 2013, respectively, and the Ph.D. degree from the Software Competence Center Hagenberg (SCCH), Austria. He is currently an Assistant Professor with the Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology (PAF-IAST), Haripur, Pakistan, and as an Adjunct Researcher with SCCH. He has multi-disciplinary expertise and working experience on diverse topics of big data analytics, cloud computing, predictive maintenance, explainable AI, knowledge graphs, data science, and machine learning. He worked in multinational companies like Siemens and Atos. He also won a project during Pakistan Scientific Foundation (PSF) CRP4 call as a Principal Investigator.

HUSSAIN DAWOOD (Senior Member, IEEE) received the master's and Ph.D. degrees in computer application technology from Beijing Normal University, Beijing, China, in 2012 and 2015, respectively. He is currently an Associate Professor with the School of Computing, Skyline University College, Sharjah, United Arab Emirates. His research interests include artificial intelligence, image restoration, and image classification.

HASSAN DAWOOD received the M.S. and Ph.D. degrees in computer application technology from Beijing Normal University, Beijing, China, in 2012 and 2015, respectively. He is currently an Associate Professor with the Software Engineering Department (SED), University of Engineering and Technology (UET), Taxila, Pakistan. His research interests include image restoration, feature extraction, and image classification.

ALI DAUD received the Ph.D. degree in computer science from Tsinghua University, Beijing, China, in July 2010. Currently, he is a Full Professor with the Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates. He has 13 years' post-Ph.D. experience of teaching, supervision, and research at B.S., M.S., and Ph.D. level. He has published more than 100 research papers in reputed international impact factor journals and conferences. He has taken part in many research projects as well and have written and acquired many research funding's. He has proven and extensive experience in data mining, artificial intelligence (machine learning/deep learning) applications to social networks, data science, natural language processing, and the Internet of Things.