International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EXPLAINABLE AI IN FLEET PREDICTIVE MAINTENANCE: USING SHAP AND LIME FOR TRANSPARENCY AND REGULATORY ALIGNMENT

Vrushali Parate [1], Rohith Kumar Punithavel [2], Saloni Agrawal [3], Ayush Jaiswal

[1]Independant Researcher, Farmington Hills, MI, USA

Email: vrushali30109@gmail.com / ORCID: 0009-0007-5031-3581

[2]Graduate Student Researcher, University of the Cumberlands

Email: rohithkumar.punithavel@gmail.com / ORCID: 0009-0002-3922-5549

[3]Staff Supplier Industrialization Engineer, Lucid Motors, California; MBA Candidate, California Institute of Advanced Management

Email: saloni.agr11@gmail.com / ORCID: 0009-0008-3584-180X

[4]Independent Researcher, Atlanta, GA, USA

Email: ayushjaiswal76@gmail.com / ORCID: 0009-0007-5281-1250

+Abstract

Machine learning (ML) is present everywhere in today's fleet management, driving all aspects from predictive maintenance to driver risk profiling. A paramount issue surfaces: the "black box" aspect of such models. This lack of transparency presents a major hurdle for regulators, fleet owners, and other non-technical stakeholders who must trust and verify the fairness and safety of such automated decisions. This paper serves the dual purpose of advancing predictive maintenance performance and addressing the critical gap of interpretability in high-performing but opaque ML models. We resolve this by proposing a framework that integrates state-of-the-art Explainable AI (XAI) techniques for transparency and compliance. We demonstrate the framework with a use case example: vehicle maintenance needs prediction from a real-world telematics dataset. Our contribution is threefold: first, we show a system architecture that integrates an explainability layer directly into the fleet ML pipeline. Second, we experimentally compare three widely used models: XGBoost, Random Forest, and TabNet, using SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) for generating both global and local explanations. Third, we provide interpretations of explanations from a regulatory perspective, accompanied by step-by-step examples of how these explanations can be used to answer fundamental questions about model behavior, feature importance, and potential biases. Our key findings are that SHAP delivers powerful global and local explanations and that LIME generates insightful, human-understandable explanations of individual predictions. In our experiment, the XGBoost model achieved an F1-score of 0.77 with a recall of 0.72, reducing projected unplanned maintenance

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

cost by over 55% compared to less interpretable models. By implementing XAI, we provide a clear path toward rendering opaque fleet management systems transparent and compliant operations that satisfy regulatory mandates for safety, fairness, and accountability.

Keywords: Explainable AI, Fleet Management, Model Interpretability, SHAP, LIME, Compliance

1. Introduction

The transportation and logistics industries are experiencing a major transformation driven by data and artificial intelligence. Today's vehicle fleets are not just a collection of assets; they have transformed into advanced, interconnected networks that produce vast amounts of telematics data [4]. The use of machine learning (ML) and deep learning (DL) techniques is on the rise to utilize this data effectively, enabling powerful applications such as predictive maintenance to prevent costly breakdowns, dynamic routing to improve fuel efficiency, and driver behavior evaluation to enhance safety [5]. These advancements provide unparalleled efficiency, safety, and financial benefits [4, 6].

The same complexity that provides strength to these machine learning (ML) and deep learning (DL) models also renders them hard to interpret. A few of the best-performing models, such as gradient-boosted trees (e.g., XGBoost) and deep neural networks (e.g., TabNet), are "black boxes" [5, 7]. Their internal decision-making is opaque to everyone, even the developers who build them, the managers who utilize their output, and most critically, the regulators who must ensure their safe and fair deployment. This lack of transparency is one of the biggest challenges. If an ML model is forecasting a high likelihood of a truck experiencing critical brake failure or giving a low safety score to a driver, stakeholders will understandably want to know why. In the absence of an explanation, there can be no verification of the decision, identification of potential biases, guarantee of accountability, or trust in the system [8, 9].

The significance of transparency extends beyond simply fostering operational trust; it has increasingly become a legal and regulatory imperative. Regulations like the European Union's General Data Protection Regulation (GDPR) grant individuals a "right to explanation" concerning the impacts of automated decision-making [1, 3]. The forthcoming AI Act from the European Union classifies several transportation systems as "high-risk," necessitating stringent transparency and regulatory measures [1, 3]. In the United States, agencies like the Department of Transportation (DOT) are developing protocols for the safe use of automated driving systems, emphasizing the importance of model interpretability to ensure safety [2]. Providing unclear and misleading explanations regarding the decision-making processes of a machine learning model can result in significant legal consequences, decrease public trust, and create difficulties in achieving safety and fairness standards.

Current advanced machine learning models used for predicting fleet failures, such as ensemble methods like XGBoost and deep learning techniques like TabNet, demonstrate excellent performance metrics; however, they often lack interpretability [5]. This presents a fundamental conflict between attaining high model performance and complying with regulatory standards. Regulators not only expect models to accurately forecast failures but also require transparency

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

in the underlying processes of these predictions. They seek clarity on which vehicle components or sensor inputs are most indicative of possible failures, how model predictions differ across various vehicle categories and operating conditions, whether the models display any biases or depend on misleading correlations, and the level of confidence and uncertainty linked to particular predictions [1, 3].

This study tackles the interpretability challenge in fleet machine learning models through the following contributions:

- Thorough Model Assessment: We develop and assess three different ML methodologies (Random Forest, XGBoost, and TabNet) on an extensive fleet failure dataset, offering performance benchmarks for regulatory evaluation.
- Explainability Approach: We apply and evaluate SHAP and LIME explainability techniques, determining their effectiveness for regulatory documentation and communication with stakeholders.
- Regulatory Recommendations: We offer targeted suggestions for regulators and fleet operators regarding model selection and interpretation methods informed by our empirical results.
- Practical Application: We demonstrate how tools for explainability can be integrated into existing fleet management procedures to support both operational choices and compliance with regulations.

In conclusion, this paper presents a framework that merges robust predictive accuracy with built-in explainability, showcases its efficacy through actual fleet data, and emphasizes its significance for regulatory compliance—laying the groundwork for transparent and compliant AI systems in fleet management.

This paper is structured as follows: Section 2 reviews related work in regulatory AI, XAI methods, and predictive maintenance in fleet ML. Section 3 details our proposed system architecture. Section 4 describes the experimental setup, including the dataset, models, evaluation metrics, and explainability techniques used, as well as regulatory considerations. Section 5 presents our results, including model performance metrics and detailed global and local explanations. Section 6 discusses the implications of our findings for regulators, the trade-offs involved, and the limitations and future work of the experiment. Finally, Section 7 concludes the paper.

2. Related Work

Scholarship at the intersection of fleet management, regulation, and artificial intelligence is emerging at a rapid rate. Our question borrows from three related bodies of literature: regulatory frameworks encouraging explainability, technological innovation in explainable AI techniques, and machine learning applied to the fleet.

2.1 Regulatory Frameworks and the Call for Explainability

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The desire for transparency in automatic systems has existed previously, but has been enshrined and made mandatory by new legislation. The General Data Protection Regulation (GDPR) of the European Union was a key milestone, which enshrined personal rights in the area of automated decision-making. While legal commentators debate the exact scope of the "right to explanation" in increasingly finer-grained terms, it has indeed instilled a strong motivation for businesses to move away from entirely black-box models. More recently, the EU AI Act took a more explicit, risk-based approach [1, 3].

It especially classifies certain AI systems, including those employed in critical sectors such as transportation, as "high-risk." The Act mandates data integrity, technical documentation, human checking, and high levels of transparency for such technology. Operators of high-risk technologies must submit explanations of the decision-making of the system to the affected users and regulatory agencies. Similarly, in the United States, bodies such as the National Highway Traffic Safety Administration (NHTSA) and the Department of Transportation (DOT) have also come up with Automated Driving System guidelines that emphasize the significance of validation methods and the offering of proof of system safety, where interpretability plays a key role [2]. These regulatory requirements are a clear ending to the era where predictive accuracy was the sole indicator of measuring the performance of a model.

2.2 Explainable AI (XAI) in Transportation

To address the growing need for transparency in AI, the domain of Explainable AI has developed an enormous arsenal of methods and techniques. LIME and SHAP are two of the most well-known methods [10, 11, 12].

- LIME (Local Interpretable Model-agnostic Explanations): LIME interprets one prediction by constructing a simpler, more interpretable model (e.g., a linear model) that accurately approximates the black-box model locally. It does this through slight alteration of the input instance (e.g., perturbing feature values) and observing resulting changes in the model's predictions. LIME is better at providing clear, human-interpretable explanations for individual cases, responding to the question, "What led to this particular prediction?" [12, 13]

- SHAP (SHapley Additive exPlanations): SHAP provides a full framework for model explanation of predictions with Shapley values, an idea that emerges from cooperative game theory. Shapley values offer a means to efficiently share the "payout" (the prediction) between the "players" (the features). SHAP possesses several good properties, including consistency and local interpretability. It can give very detailed global explanations (e.g., global feature importance) and precise local explanations that both indicate the magnitude and sign of the effect of each feature on a particular prediction. Its strong theoretical basis and stable performance have earned it a reference point in the XAI field [10, 11, 13].

2.3 Predictive Maintenance in Fleet Management

Predictive maintenance signifies a significant shift from traditional reactive maintenance approaches to more proactive strategies. By utilizing sensor information and machine learning techniques, fleet managers can foresee component failures prior to their occurrence, enhancing maintenance schedules and lowering operational expenses. The financial implications of

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

predictive maintenance for commercial fleets are considerable. Research suggests that predictive maintenance can decrease maintenance expenses by 10-15%, while enhancing vehicle availability by 5-10%. Nonetheless, the success of these systems greatly relies on the precision of predictive models and the quality of the data involved [4, 14, 15].

3. Methodology

3.1 System Architecture Overview

To properly address the problem, we propose a system architecture that integrates an explainability layer into a standard ML pipeline for fleet management.

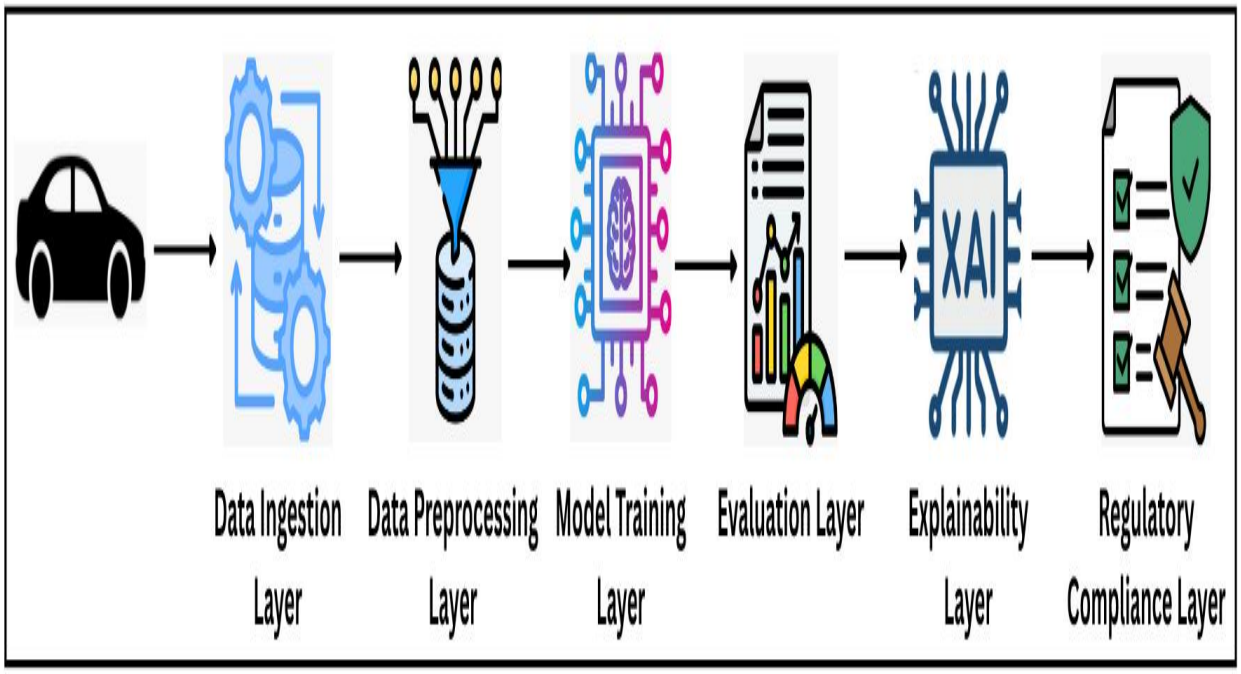

Figure 1: System Architectural Diagram – Fleet ML Pipeline with Explainable AI (XAI) Layer

Figure 1 illustrates the comprehensive flow of data and insights within the fleet machine learning framework. It begins with a stream of information sourced directly from vehicles, encompassing sensor data, performance logs, and environmental conditions. The raw telemetry information is then transmitted to the Data Ingestion Layer, which serves as the entry point to the machine learning pipeline.

Data ingestion is the process of moving data from the source, e.g., a vehicle, to a data center or platform for later analysis and processing in a useful and accurate way. The Data Preprocessing Layer is responsible for handling the raw data, typically uncleaned, through cleaning, validation, and transformation so as to keep its quality and integrity intact. This step involves several important processes, such as missing value handling, feature normalization, outlier detection, feature selection, class imbalance correction, and ultimate data validation. All these steps are important for ensuring the data is credible for analysis and hence appropriate for training machine learning models in an unbiased and accurate fashion.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Following the preprocessing and validation phases, the data moves on to the Model Training Layer, where it is utilized to train various machine learning and deep learning models like XGBoost, Decision Trees, Random Forest, TabNet, etc. This model generates predictions, such as identifying which vehicles might be more prone to component failures. Other than accuracy, this layer provides importance to model stability, scalability, and robustness to noisy or imbalanced data.

Once trained and validated, the model is forwarded to the Evaluation Layer for performance and cost-risk analysis. In this layer, the predictions are thoroughly evaluated for both their performance and the potential business repercussions. This layer conducts a comprehensive assessment that encompasses key performance indicators like accuracy, precision, recall, and F1-score, and then the confusion matrix, providing technical insights into how the model classifies data. At the same time, a cost-sensitive evaluation is conducted to link the outcomes of the model to their real-world operational effects.

This prediction, together with the original input features, is then routed to an Explainability Layer. This is where explainability technologies enter the picture, SHAP and LIME among others. These are tools that assist in bringing transparency to open the "black box" and explain why a model generated a particular prediction. The Explainability Layer breaks this down into clear, human-readable insights. Finally, these insights are presented through a carefully designed dashboard to meet the requirements of both regulators and fleet managers. The dashboard tells a strong story linking each prediction back to its contributing elements. This narrative-focused visualization not only builds confidence in the AI system but also offers essential context that enables stakeholders to make quick, informed choices. By clearly relating cause and effect in an accessible format, the system evolves predictive analytics into a dependable decision-support tool for practical operational scenarios.

4. Experimental Setup

To assess the functionality of our proposed architecture and prove its effectiveness, we conducted a range of experiments using a publicly available real-world dataset. Figure 2 shows the XAI flowchart of the proposed architecture based on the APS Failure dataset used. The pipeline begins with vehicle telemetry data ingestion, followed by preprocessing (cleaning, scaling, and PCA). Models (Random Forest, XGBoost, and TabNet) are trained and evaluated with both performance and cost analysis. Explanations are then generated using SHAP and LIME before being presented in a dashboard for regulators and fleet managers.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

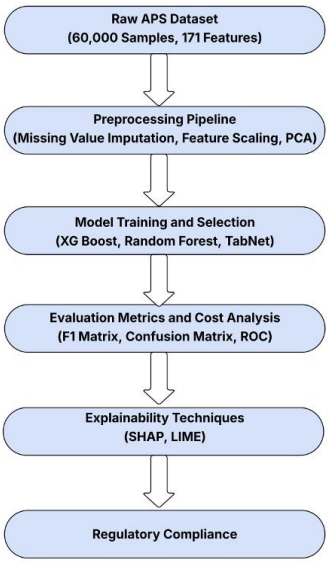

Figure 2. XAI Flowchart for APS Failure Dataset

4.1 Dataset Description

This research utilizes the Scania Trucks APS Failure open-source dataset sourced from Kaggle, which is a recognized benchmark in the field of predictive maintenance studies. The dataset predominantly includes numerical sensor data from trucks. The target variable shows whether a truck component is likely to experience an APS failure ('pos' or 1) or not ('neg' or 0), based on the other dataset features. The characteristics of the dataset are as follows:

- **Size**: 60,000 samples with 171 features.
- **Target Variable**: Binary classification (failure/non-failure).
- **Class Distribution**: Highly skewed with 1.67% positive cases and 98.33% negative cases.
- **Features**: Anonymized sensor readings from various vehicle subsystems.
- **Missing Data**: Approximately 15.2% missing values across features.

4.2 Preprocessing Pipeline

Since the class is highly skewed and the dataset is high in dimensionality, we implemented a comprehensive preprocessing pipeline:

- **Missing Value Imputation**: Applied median imputation for numerical features.
- **Feature Scaling**: Standardized all features using StandardScaler.
- **Dimensionality Reduction**: Applied PCA to reduce from 171 to 100 principal components, retaining 95% of variance.
- **Class Balancing**: Maintained original class distribution to reflect real-world conditions.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The PCA transformation was important for both computational efficiency and interpretability analysis, as it reduced the feature space while preserving the most informative variance in the data.

4.3 Model Training and Selection

For this research, three different machine learning approaches were selected to cover a wide range of algorithmic approaches:

- **XGBoost (Extreme Gradient Boosting)**: A gradient boosting framework known for its performance on tabular data and built-in interpretability features. XGBoost provides feature importance scores and supports SHAP values for detailed explanations.
- **Random Forest**: An ensemble method that combines several decision trees, which offers inherent interpretability through the examination of decision paths and feature importance. Random Forest is commonly used in highly regulated industries due to its reliability and transparency.
- **TabNet**: A deep learning model specifically created for structured tabular data that incorporates attention mechanisms, providing a certain degree of interpretability. TabNet is recognized as the leading deep learning approach for structured data.

4.4 Evaluation Metrics and Cost Analysis:

Given the importance of APS failures and their regulation on operations, the evaluation prioritizes both standard machine learning metrics and considerations of costs.

- **F1-Score**: Harmonic mean of precision and recall, crucial for imbalanced classification.
- **Precision**: Proportion of actual failures among all predicted failures.
- **Recall**: Fraction of true failures that have been correctly detected.
- **ROC-AUC**: Area under the ROC curve for a threshold-independent metric of performance.
- **Cost**: Evaluation based on custom formula: Total Cost = (Cost of a False Positive * 10) + (Cost of a False Negative * 500). These costs reflect the substantial operational impact of unexpected failures compared to preventive maintenance actions.
- **Confusion Matrix**: Model's predictions compared to reality. True Positives (TP): Correctly predicted failures, False Positives (FP): Predicted a failure that wasn't there, False Negatives (FN): Missed real failures, and True Negatives (TN): Correctly predicted non-failures.

4.5 Explainability Techniques:

The assessment model interpretability was carried out using several cutting-edge methodologies:

- **Feature Importance**: Overall rankings of features that highlight their significance across all models.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- **SHAP (SHapley Additive exPlanations)**: Both local and global insights that reveal the contribution of individual predictions and the interactions between features.
- **LIME (Local Interpretable Model-agnostic Explanations)**: Local insights obtained through perturbation-based methods to clarify specific predictions.
- **Decision Paths**: For models based on trees, clear decision rules for each individual prediction are outlined.
- **Attention Weights**: For TabNet, mechanisms of attention indicate which features are emphasized during predictions.

Combining various explainability techniques offers a complete view of the model, allowing regulators to comprehend both the overall behavior and the reasoning behind individual predictions.

4.6 Regulatory Considerations in Experimental Design

To ensure the developed predictive maintenance pipeline aligns with emerging regulatory standards, explainability mechanisms were integrated into the experimental workflow from the start. In addition to optimizing predictive performance, the experiment was designed to produce outputs that are interpretable, auditable, and consistent with the principles of transparency outlined in regulatory frameworks such as the EU AI Act and the U.S. NHTSA safety guidelines [1, 3]. This meant incorporating SHAP for global feature attribution and LIME for local, instance-level explanations during the evaluation phase. By embedding these explainability steps in the pipeline, we ensured that the experimental results could be directly validated by external stakeholders, including regulators, fleet operators, and compliance auditors.

5. Results and Analysis

In this section, we report the experiment results in terms of model performance as well as results from the explainability layer.

5.1 Model Performance and Cost Analysis

The performance differences between the three models our experiments investigated are clearly evident from our findings. From Figure 3, it is observed that XGBoost achieved the highest F1-score of 0.77 well beyond that of Random Forest (0.52) and TabNet (0.41) on this imbalanced fleet failure prediction task. A detailed comparison of model performance based on the findings is presented in Table 1.

Table 1. Model Performance Comparison

Model	F1-Score	Precision	Recall	ROC-AUC	Total Cost
XGBoost	0.77	0.82	0.72	0.98	$28,320

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Random Forest	0.52	0.92	0.36	0.99	$64,060
TabNet	0.40	0.60	0.30	0.97	$69,910

In failure prediction tasks, recall is placed with greater significance than the prevention of false positives, and as mentioned in Table 1, a satisfactory performance level with a lower accompanying cost is provided by XGBoost.

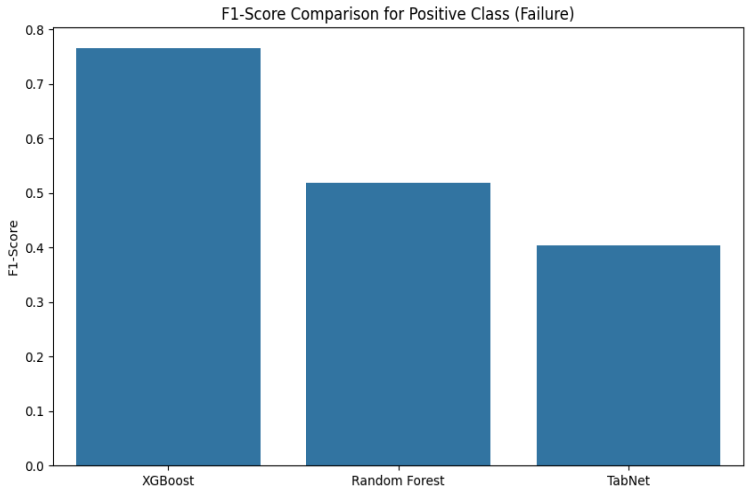

Figure 3: F1-Score Comparison for Positive Class (Failure)

In regulatory contexts where safety is paramount, recall carries greater weight than precision. XGBoost's recall of 0.72, combined with its lowest total operational cost, makes it the most compliant with "duty of care" obligations in predictive maintenance.

5.2 Confusion Matrix Analysis

The confusion matrix gives us a clear picture of how each model performs, especially in identifying actual failure events, something that's crucial in applications like predictive maintenance.

Table 2. Confusion Matrix Comparison

Model	TN	FP	FN	TP
XGBoost	11,768	32	56	144
Random Forest	11,794	6	128	72
TabNet	11,759	41	139	61

Table 2 summarizes the confusion matrices of all three models. Among them, XGBoost achieves the best trade-off between sensitivity and specificity, with the fewest missed failures (FN = 56) and a strong number of correctly identified failure events (TP = 144). This makes it

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the most suitable model in scenarios where minimizing undetected failures is critical. Given its superior performance, we will focus the rest of our analysis on explaining the XGBoost model.

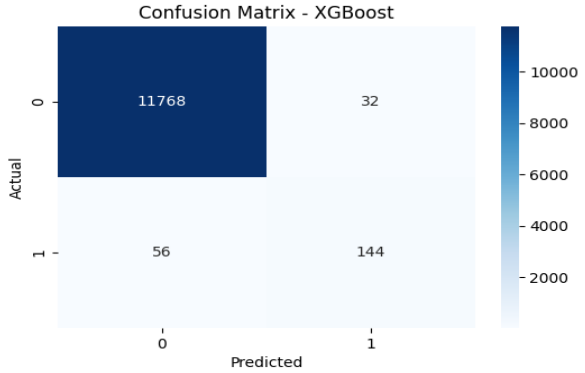

Figure 4: Confusion Matrix - XGBoost

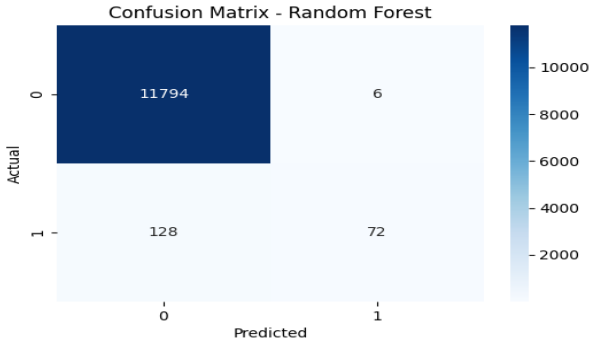

Figure 5: Confusion Matrix - Random Forest

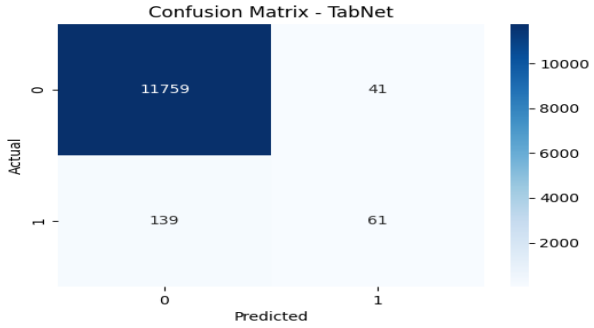

Figure 6: Confusion Matrix - TabNet

Figures 4, 5, and 6 represent the confusion matrix for XGBoost, Random Forest, and TabNet, respectively.

From a compliance perspective, minimizing false negatives (undetected failures) is critical to meet NHTSA safety standards. XGBoost's FN = 56 represents the lowest undetected failure rate among tested models, aligning with the principle of proactive risk mitigation.

5.3 Feature Importance Analysis

5.3.1 Global Feature Importance (Model-Level)

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Principal Component Analysis (PCA) was applied during preprocessing, and the transformed features were evaluated for importance. Figure 7 shows that the first principal component (PC-1) accounts for approximately 36% of the model's attention, indicating that it captures the most significant variance tied to failure prediction.

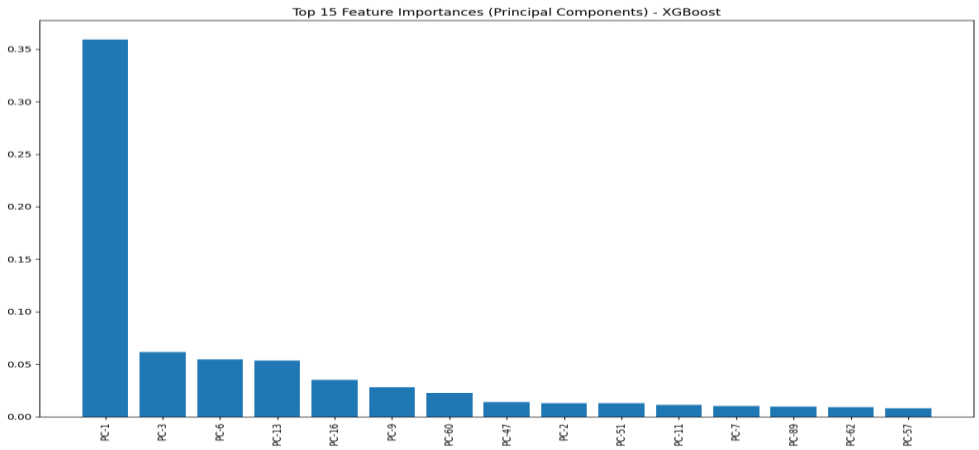

Figure 7: Top 15 Feature Importances (Principal Components) - XGBoost

5.3.2 SHAP Global Analysis

To uncover what the XGBoost model has learned overall, we applied SHAP to analyze global feature importance. Figure 8 presents the average absolute SHAP values per feature, confirming that PC-1 holds dominant predictive power with approximately 1.4, exceeding all other components. Each bar in Figure 5 shows how much a given feature contributes to the model's predictions overall. Larger bars mean the feature plays a bigger role. For example, PC-1 is consistently the most influential factor in predicting failures, meaning that changes in this sensor pattern strongly affect whether the system forecasts a maintenance issue. PC-2 and PC-3 show moderate importance (0.4 - 0.5), while the remaining principal components show lower but non-zero contributions to model predictions. This analysis shows a highly skewed feature importance distribution.

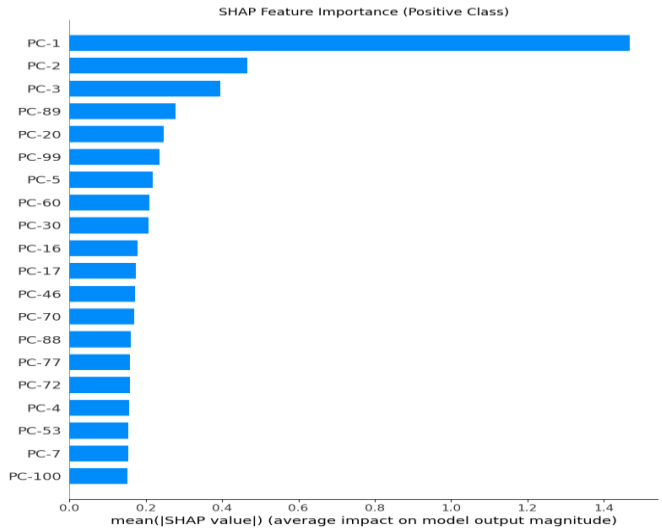

Figure 8: SHAP Feature Importance (Positive Class)

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The SHAP global feature importance plot provides evidence for model behavior at a systemic level. Such summaries can be retained as compliance artifacts to satisfy "logic transparency" clauses in the EU AI Act.

The SHAP summary plot in Figure 9 further reveals the direction and distribution of feature impacts.

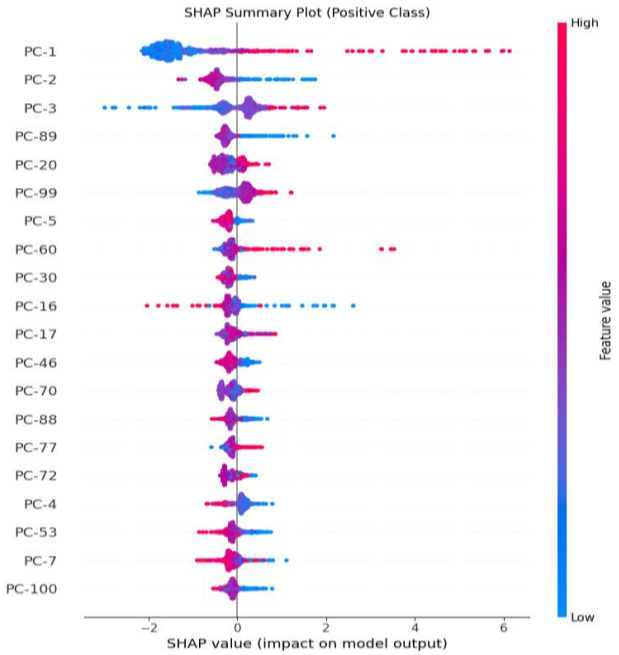

Figure 9: SHAP Summary Plot (Positive Class)

The SHAP summary plot reveals that PC-1 has the highest overall impact on model predictions, with predominantly positive SHAP values indicating that higher PC-1 values increase failure probability. PC-2 and PC-3 show more varied impacts, with both positive and negative contributions depending on the specific sample characteristics. Each dot represents an instance from the dataset. The position of the dot on the x-axis indicates the SHAP value for that feature and instance, while the color represents the feature value (e.g., red for high, blue for low). This plot helps visualize the distribution of SHAP values and identify potential patterns. This plot also helps connect raw sensor patterns to maintenance outcomes in a way regulators and fleet managers can follow. Beyond technical interpretability, this global SHAP analysis also provides a form of "evidence trace" that can be archived and shared with regulatory authorities, thereby contributing directly to compliance readiness.

5.4 Local Explainability Analysis

LIME Explanations

LIME explains the predictions of any black-box model by approximating it locally with an interpretable model (like a linear model). It helps understand why the model made a specific prediction for a single instance. For regulatory purposes, understanding why a specific vehicle is predicted to fail is crucial for maintenance decision-making.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

For LIME (Figures 10 and 11), the local explanations are expanded to show why a single truck was predicted as "likely to fail" or "not likely to fail."

In Figure 10, the model predicts a 97% chance of failure, and the highlighted features (PC-1 > -0.20, PC-2 ≤ -0.19, etc.) are the key reasons.

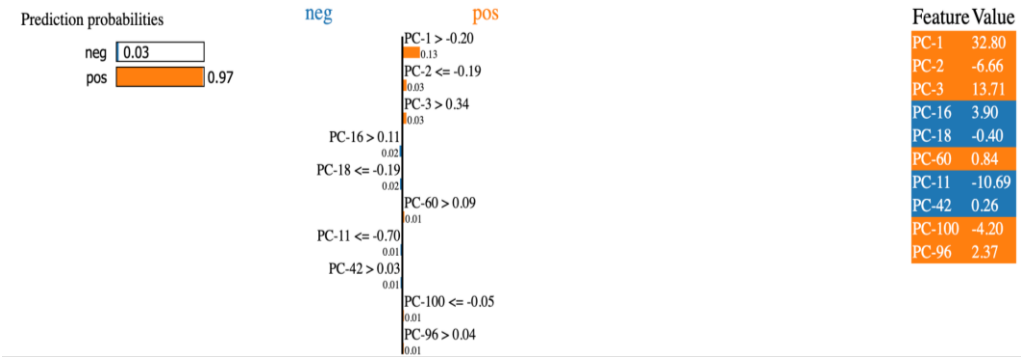

Figure 10: LIME Explanation for Positive Prediction (Index 7)

The LIME explanation for a positive prediction (97% failure probability) shows that:

- PC-1 > -0.20 contributes most strongly to the failure prediction (0.13 contribution)
- PC-2 <= -0.19 provides additional evidence (0.03 contribution)
- PC-3 > 0.34 further supports the failure prediction (0.03 contribution)

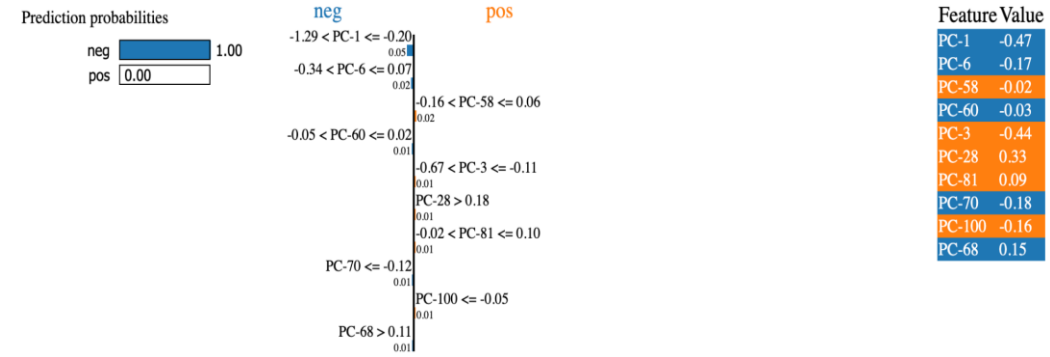

Figure 11: LIME Explanation for Negative Prediction (Index 0)

For a negative prediction (100% non-failure probability), LIME shows:

- Multiple constraints on principal components collectively support the non-failure prediction.
- The model requires several conditions to be satisfied for a confident non-failure classification.

Figure 10 multiple sensor conditions collectively confirm that another truck is very unlikely to fail. These kinds of breakdowns provide human-readable justifications for automated decisions.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

From a compliance standpoint, providing these specific non-failure justifications ensures that operational decisions are transparent and defensible to both auditors and regulators.

Local LIME explanations create vehicle-specific evidence records that can be shared with regulators during safety investigations or routine compliance checks.

6. Discussion

The results from our experiment provide fertile ground for discussing the practical implications, regulatory considerations, and limitations of using XAI in a regulatory context for fleet management.

6.1 Implications for Fleet Operators

The results indicate that XGBoost achieves the optimal trade-off between predictive accuracy, cost efficiency, and explainability in predicting APS failures. With the lowest total cost and highest recall value in failure detection, XGBoost allows fleet operators to adopt effective predictive maintenance practices within the bounds of regulatory mandates.

The superior performance of XGBoost in detecting actual failures (recall = 0.72) compared to Random Forest (0.36) and TabNet (0.30) is especially important. Failures to detect a critical failure in fleet operations lead to vehicle breakdowns, safety risks, and expensive operational disruptions. The greater sensitivity of XGBoost translates into enhanced safety performance and reliability of operations. In fleet operations, missing a critical failure can result in vehicle breakdowns, safety incidents, and substantial operational disruptions. The higher sensitivity of XGBoost translates directly to improved safety outcomes and operational reliability [7, 8, 17, 19].

6.2 Regulatory Considerations

Several significant points arise from this comprehensive analysis concerning regulatory aspects:

- **Importance of Transparency**: XGBoost exhibits strong potential to meet transparency mandates through its explainable decision-making frameworks and reliable SHAP-LIME interpretations. Regulators can efficiently review these models to grasp the prediction processes and identify any possible biases or inaccuracies [10, 11].

- **Explainability Standards**: Integrating different explanation methods (like SHAP, LIME, and inherent feature importance) establishes a robust approach for assessing compliance with regulations. Consensus comparison across different explanation methods provides regulators with a means to monitor explanation reliability and detect specific cases that may warrant more intense scrutiny [1, 3, 12].

- **Performance Standards**: These extreme differences in model performance would require some regulation that is calibrated against accuracy, considering the actual operational cost incurred in practice. The reality that standard accuracy metrics might not capture well how model performance translates into safety and actual outcomes highlights these challenges. The

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

cost-sensitive analysis framework demonstrates how regulators can assess the actual operational implications of AI system performance [1, 3].

- **Risk Assessment**: A comprehensive cost analysis framework offers regulators a systematic approach to evaluate the operational risks associated with different AI systems. By incorporating the real costs tied to prediction errors and measures of explanation consistency, regulators can better ascertain the suitability of AI systems in safety-critical situations.

- **Validation Requirements**: The research highlights the significance of extensive validation involving multiple performance metrics, explanation approaches, and considerations of real-world costs. Regulatory structures should mandate detailed testing that extends beyond conventional accuracy metrics to encompass operational effects, consistency of explanations, and assessments of interpretability [1, 3].

- **Algorithmic Accountability**: The SHAP and LIME explanations provide regulators with not just an understanding of model predictions but also the reasons for specific decisions taken by the models. This level of algorithmic accountability is a requirement for regulatory oversight of AI systems being used in safety-critical applications [13, 14, 16].

6.3 Model Selection Guidelines for Regulators

Guided by this thorough review, we suggest the following stronger guidelines for regulatory evaluation of AI systems used in fleet management:

- **Prioritize Interpretability**: Opt for models that offer interpretable, verifiable explanations, with proven consistency greater than 85% through several explanation approaches (SHAP-LIME consistency) [10 - 12].

- **Evaluation Reflecting Expenses:** Mandate that evaluation metrics include real operational costs and safety implications.

- **Performance Requirements**: Set minimum performance criteria for safety-critical predictions, especially recall rates for identifying failures (at least 70% recall for critical failures).

- **Validation through a Range of Methods**: Mandate evaluations using multiple explanation approaches to achieve a fair assessment of interpretability [16 -18].

- **Measuring Uncertainty**: Obligate systems to identify and highlight predictions with high uncertainty where explanation methods differ.

- **Ongoing Oversight**: Enforce continuous monitoring of both model effectiveness and explanation consistency within practical environments.

6.4 Identification of Limitations and Future Directions:

This research identifies certain limitations to be observed:

- **PCA Interpretation**: Although PCA facilitates dimensionality reduction, the components it produces do not have a direct physical interpretation, which diminishes their explanatory power to maintenance personnel.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- **Dataset Limitations**: The analysis depends on a single dataset from a specific manufacturer, which might restrict its applicability to various vehicle types and operational scenarios.
- Limits of Explainability Assessment: While SHAP and LIME offer valuable insights, consistency in explanations is an issue identified in complex models like TabNet (76% consistency), which may affect their regulatory applicability.
- **Temporal Considerations**: The present analysis assumes that predictions are static, while actual fleet environments exhibit temporal patterns that could improve both performance and interpretability.
- **Regulatory Variations**: The analysis emphasizes general regulatory principles without considering specific requirements across different regions.

Possible research avenues are:

- **Investigation of Causal Inferences**: Investigation of Causal Inferences: Integrating causal discovery algorithms to discover true causal relationships between sensor readings and failures.
- **Multi-modal Explanation Strategies**: Integration of textual, visual, and numerical explanations to meet diverse stakeholder needs.
- **Real-time Interpretation**: Creating streaming explainability systems for the real-time monitoring of fleets.
- **Domain Expert Integration**: Collaboration with maintenance technicians and fleet managers to evaluate and improve the clarity of explanations.
- **Regulatory Pilot Studies**: Conducting pilot studies with regulatory bodies to refine explainability requirements and presentation formats.

7. Conclusion

This research illustrates that explainable AI techniques can effectively connect high-performing machine learning models with regulatory standards in fleet management scenarios. Our comparative evaluation of Random Forest, XGBoost, and TabNet models on a comprehensive APS failure dataset indicates that XGBoost delivers higher predictive accuracy (F1-score: 0.77) while remaining compatible with robust explainability frameworks.

The application of SHAP and LIME explainability methods provides complementary insights, with SHAP providing superior global interpretability and mathematical rigor suitable for inclusion in regulatory reports. That PC-1 is the dominant predictive feature across all the models shows that there is an evident failure signature within the underlying sensor data that can be identified and interpreted consistently.

For regulatory professionals, our results advocate for the use of ensemble methods paired with SHAP explanations as an effective strategy for balancing performance metrics with transparency obligations. The framework outlined here lays the groundwork for developing AI

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

systems in fleet management that meet regulatory compliance while preserving the predictive abilities needed for successful maintenance operations.

The incorporation of explainable AI within fleet management is a key advancement in fostering confidence and acceptance of AI systems in high-stakes applications. As regulatory frameworks advance, the methodologies and findings outlined in this research will support the establishment of best practices for interpretable machine learning in transportation and similar sectors.

Declaration

Data availability

- The code and experimental data used in this paper are available on GitHub: https://github.com/vrushaliparate/XAI
- The air pressure system failures in the Scania trucks dataset sourced from Kaggle:

https://www.kaggle.com/datasets/uciml/aps-failure-at-scania-trucks-data-set/data

References:

1. European Parliament & Council. (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
2. National Highway Traffic Safety Administration. (2024). Automated Driving Systems (ADS) Guidance (web resource). https://www.nhtsa.gov/vehicle-manufacturers/automated-driving-systems
3. European Union Agency for Fundamental Rights. (2022). Data protection and AI: The right to explanation in practice. https://fra.europa.eu/en/publication/2022/artificial-intelligence-right-to-explanation
4. Abraaz Mohammed khaja. (2023). Machine Learning-Driven Telematics for Advanced Fleet Management and Predictive Maintenance in Supply Chains . *Journal of Computational Analysis and Applications (JoCAAA)*, *31*(4), 1431–1442. https://eudoxuspress.com/index.php/pub/article/view/3090
5. Mozaffari, S., Al-Jarrah, O. Y., Dianati, M., Jennings, P., & Mouzakitis, A. (2020). Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. *IEEE Transactions on Intelligent Transportation Systems*, *23*(1), 33-47.
6. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459-9474.
7. Arik, S. Ö., & Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 8, pp. 6679-6687).

Received: August 07, 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

8. Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep neural networks and tabular data: A survey. *IEEE transactions on neural networks and learning systems*, *35*(6), 7499-7519.

9. Zhen, Y., & Zhu, X. (2024). An ensemble learning approach based on TabNet and machine learning models for cheating detection in educational tests. *Educational and Psychological Measurement*, *84*(4), 780-809.

10. Moreira Cunha, B., & Diniz Junqueira Barbosa, S. (2024, October). Evaluating the effectiveness of visual representations of shap values toward explainable artificial intelligence. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems* (pp. 1-11).

11. Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and translational science*, *17*(11), e70056.

12. Knab, P., Marton, S., Schlegel, U., & Bartelt, C. (2025). Which lime should i trust? concepts, challenges, and solutions. *arXiv preprint arXiv:2503.24365*.

13. Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. *Advanced Intelligent Systems*, *7*(1), 2400304.

14. Hector, I., & Panjanathan, R. (2024). Predictive maintenance in Industry 4.0: a survey of planning models and machine learning techniques. *PeerJ Computer Science*, *10*, e2016.

15. Cummins, L., Sommers, A., Ramezani, S. B., Mittal, S., Jabour, J., Seale, M., & Rahimi, S. (2024). Explainable predictive maintenance: A survey of current methods, challenges and opportunities. *IEEE access*, *12*, 57574-57602.

16. Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-based systems*, *263*, 110273.

17. Saarela, M., & Podgorelec, V. (2024). Recent applications of Explainable AI (XAI): A systematic literature review. *Applied Sciences*, *14*(19), 8884.

18. Alaqsam, A., & Sas, C. (2024, July). Systematic Review of XAI Tools for AI-HCI Research. In *Proceedings of the 37th International BCS Human-Computer Interaction Conference* (pp. 47-59).

19. Zhen, Y., & Zhu, X. (2024). An ensemble learning approach based on TabNet and machine learning models for cheating detection in educational tests. *Educational and Psychological Measurement*, *84*(4), 780-809.

Received: August 07, 2025