



An IoT-Based Multi-Sensor Data Fusion Framework for Predictive Fleet Failure Using a Hybrid Survival Analysis and Machine Learning Approach

Aryan Akbarirad^{1*}, Armin Akbarirad²

¹Master of Industrial Engineering, Faculty of Management and Industrial Engineering, Malek-Ashtar University of Technology, Tehran, Iran

²Master of Industrial Engineering, Faculty of Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran

Abstract

Cost optimization and increasing fleet reliability are among the main challenges in urban public transportation management. In this research, temporal failure patterns of critical bus components in various cities of Iran have been modeled and evaluated through analysis of real data collected from intelligent urban fleet systems. Comparative analysis of analytical methods showed that the Decision Forest model with an average accuracy (F1-score) of 89% (compared to Log-rank test 42.3%, stratified Cox 64%, and Decision Tree 85%) demonstrates superior performance in predicting component failures. Results indicate that environmental factors, operation, and utilization methods have significant effects on component lifespan. Accordingly, a predictive maintenance planning framework has been presented which, based on simulation results, leads to a 41% reduction in maintenance costs and a 65% decrease in fleet downtime.

Keywords: Component Survival Analysis; Fleet Management System; Internet of Things in Transportation; Preventive Maintenance; Machine Learning

1. Introduction

In the era of digital transformation and data-driven approaches, urban bus fleet management faces complex challenges where operational data analysis and component failure prediction, as vital aspects, have a significant impact on economic efficiency and service quality. This research utilizes data collected from over 1,000 urban buses with diverse mileage ranges (from 3,863 to 213,636 kilometers) to explore failure patterns of key components selected based on the expertise of Oghab Afshan Company technical experts. Maintenance costs, which constitute a significant portion of operational expenses, can be optimized using machine learning techniques and data analysis; as decision forest models with an average accuracy of 89% enable the prediction of component failures by considering key variables such as mileage, seasonal conditions, and geographical location. Analyses reveal that seasonal changes and environmental conditions have a significant impact on component performance; such that heating and cooling systems in extreme seasons, and electronic components like multiplex boards and fuel systems, show distinct failure patterns under different conditions. Utilizing fleet management systems and IoT technology, accurate and real-time data collection on component performance has become possible, which, combined with advanced analytical techniques, leads to the development of intelligent preventive maintenance strategies.

2. Previous Research

Predictive Maintenance (PdM) has emerged as one of the leading strategies for improving equipment reliability and reducing costs, particularly in industrial environments. Unlike reactive maintenance, which is performed only after a failure occurs, and preventive maintenance, which is scheduled at fixed intervals irrespective of the actual condition of the equipment, predictive maintenance is based on real-time data and forecasting future events [1]. According to studies by Zhu et al., the lack of preparedness in addressing equipment failures may lead to significant costs, up to \$138,000 per hour for organizations, while in various industries, maintenance costs account for 15% to 70% of total production expenses [2].

* Corresponding Author ISSN: 1735-8272, Copyright © 2025 JISE. All rights reserved

The evolution of maintenance practices, from reactive approaches to technology-driven predictive maintenance, particularly through AI and IoT, has undergone significant progress. Zhu et al. classified this transformation into four stages: reactive maintenance, preventive maintenance, predictive maintenance, and predictive maintenance in Industry 4.0 [2].

2.1 Predictive Maintenance System Architectures

PdM 4.0 Architecture: Zhu et al. describe the architecture of predictive maintenance for Industry 4.0, consisting of five core components: data collection, preprocessing, data analysis, decision support systems, and maintenance operations execution [2]. This architecture, powered by advanced technologies like AI and IoT, enables accurate failure predictions and proactive maintenance actions.

IoT-Based Systems: Alkathieb et al. designed an IoT-based predictive maintenance system that incorporates sensors, data transmission protocols (such as MQTT), data storage in NoSQL databases, and data analysis through machine learning [1]. This system enables real-time monitoring of various equipment.

In another study, De Simone et al. developed an IoT-based system for monitoring the health of reinforced concrete structures. This system uses MEMS sensors and Raspberry Pi4 to acquire vibration data, providing a cost-effective solution for structural condition monitoring [3].

2.2 AI Methods in Predictive Maintenance

Traditional Machine Learning Methods:

Artificial Neural Networks (ANN): ANN models have the capability to understand complex relationships between data, enabling the classification of failures and estimation of equipment lifespan [4]. However, limitations such as the need for precise tuning and risks of overfitting remain.

Support Vector Machines (SVM): SVM is effective in classifying equipment health status and identifying types of failures. Sensitivity to parameter settings and kernel selection are challenges associated with this method [4].

Decision Trees and Random Forests: These methods have high interpretability and robustness against noise, making them suitable for failure prediction [4]. Nonetheless, individual decision trees are prone to overfitting.

Mahalanobis-Taguchi System (MTS): Focused on identifying deviations from normal states, MTS requires less training data, making it suitable for rare state detection. However, its effectiveness depends heavily on variable selection and parameter settings [4].

Deep Learning Methods:

Convolutional Neural Networks (CNN): CNNs excel in identifying complex patterns, making them effective in detecting structural damages and analyzing image data. However, training these networks requires a large amount of data [3].

Recurrent Neural Networks (RNN) and LSTM: RNNs are particularly useful for time-series data and predicting equipment lifespan. However, issues such as gradient vanishing can affect their performance [2].

Survival Analysis Models:

Survival analysis models provide effective approaches for predicting equipment lifespan. Semi-parametric models like Cox Proportional Hazards and DeepSurv, as well as parametric models such as Weibull and Log-normal distributions, are widely used in this field [5].

2.3 Industrial Applications of Predictive Maintenance

Predictive maintenance has proven to be effective in key industries like aviation [7] and rail transportation [6], as well as in sectors such as textiles [1] and structural health monitoring systems [3], improving operational efficiency and reducing costs.

2.4 Cost-Benefit Models

Florian et al. developed cost-benefit models for transitioning to predictive maintenance systems. These models focus on minimizing costs, optimizing equipment availability, and performing multi-objective analyses using Pareto methods and multi-objective evolutionary algorithms [8].

2.5 Challenges and Future Research

Despite significant advancements in predictive maintenance and the integration of IoT in fleet management, there is still a lack of comprehensive investigations into the effects of seasonal variations on the failure of sensitive electronic components and their connections to IoT data. While methods such as survival analysis, log-rank tests, Cox regression, and machine learning have been applied individually, integrating these analytical tools with real-time FMS (Fleet

Management System) data for designing seasonal maintenance programs remains a challenge. This study presents a comprehensive framework for analyzing the impact of seasonal environmental conditions on the failure of critical components and integrates FMS data with CRM systems to enhance the accuracy of failure predictions and dynamic fleet management.

3. Research Methodology

In this research, data has been collected from two main sources:

- Oghab Afshan Company CRM System**
- Information related to failure history, completed repairs, and preventive maintenance programs.
- FMS Module Installed on Buses**
- Online data including mileage, technical condition of buses, and system error codes sent to the data center.

This data pertains to Iran's urban bus fleet and includes more than 1,000 Oghab Afshan brand buses in the time period from 2022 to 2024. This fleet operates in various cities including Tehran, Mashhad, Isfahan, Arak, and other parts of the country.

Statistical Methods Used in Analysis:

For data analysis, a set of advanced statistical methods has been used:

Survival Analysis: To examine component failure rates and their effective lifespan.

Log-Rank Test: To compare component failure rates under different environmental and seasonal conditions.

Cox Proportional Hazards Regression: To model the impact of various factors (such as environmental conditions, mileage, and system errors) on the probability of component failure.

Machine Learning: Using advanced algorithms for pattern recognition of failures and accurate prediction of preventive maintenance needs.

This combined framework provides a comprehensive analysis of the relationship between environmental conditions, real-time FMS data, and key component failure behavior.

3.1 Survival Analysis

Survival analysis was used to examine the lifespan of various components. In this analysis, the time variable was failure kilometer, and the censoring variable was the last kilometer of fleet operation. The survival function was calculated using the Kaplan-Meier method based on Equation (1).

$$S(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

Where:

- S(t) :Probability of component survival until kilometre t
- d_i :Number of failures at kilometre t_i
- n_i :Number of components that have not failed until kilometre t_i

3.2 Log-Rank Test

The log-rank test according to Equation (2) was used to compare survival curves between different seasons and various cities.

$$Q = \frac{(\sum_{i=1}^k (O_i - E_i))^2}{\sum_{i=1}^k V_i}$$

Where:

- O_i :Observed number of failures in group i
- E_i :Expected number of failures in group i
- V_i :Variance

3.3 Cox Proportional Hazards Regression Model

The Cox regression model according to Equation (3) was used to examine the effect of different variables on failure rate.

$$h(t|X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$$

Where:

(2)

- $h(t|X)$:The hazard function at time t for an individual with characteristics X
- $h_0(t)$:The baseline hazard function that shows how the hazard changes over time, without considering explanatory variables.
- X_p :Explanatory variables β_p :Regression coefficients for each variable

*Error codes received from the FMS module were also included as predictor variables in the Cox model.

3.4 Calculation of Cost Reduction Percentage with Preventive Program

In this article, the percentage of cost reduction is calculated using Equation (4).

Where:

$$\text{Cost reduction (\%)} = 100 \times [1 - (1 \div \text{HR}_{\max})] \times F \tag{4}$$

HR_{\max} = Maximum hazard ratio (between season and city)

F = Preventive program effectiveness coefficient (a number between 0 and 1)

3.5 Analysis of the Impact of Preventive Maintenance Program on Fleet Availability

To evaluate the impact of the preventive maintenance program on fleet availability, the availability index was calculated using Equation (5).

Where:

$$A = \frac{MTBF}{MTBF + MTTR} \tag{5}$$

- A : Availability index
- $MTBF$: Mean Time Between Failures
- $MTTR$: Mean Time To Repair

3.6 Big Data Analysis and Machine Learning

For analyzing the vast volume of data received from the FMS system, big data analysis techniques and machine learning algorithms (decision tree) were used.

Research Process:

For analyzing the extensive FMS system data, big data analysis tools and machine learning algorithm (decision tree) have been used. This process includes several main stages:

Data Preprocessing: Including aggregation of CRM and FMS data, removal of outliers and incomplete data, and standardization of variables.

Survival Analysis: Including drawing Kaplan-Meier curves, log-rank test for comparing different conditions, and fitting the Cox regression model to examine the effect of variables on failures.

Analysis of the Relationship Between Error Codes and Failures: Identifying repetitive patterns of error codes, calculating hazard ratio, and determining warning thresholds.

Machine Learning Model: Building a prediction model using decision tree & forest, evaluating its accuracy, and comparing with traditional models.

Cost-Benefit Analysis: Examining direct and indirect costs of failures and savings from preventive maintenance.

Results Validation: Performing k-fold cross-validation and statistical hypothesis testing.

4. Statistical Methods

4.1 Survival Analysis

The results of survival analysis for AC system components in Figure 1 show that the front heater heating coil SG with a survival rate of 65% at 60,000 kilometers has a higher lifespan compared to other components. In contrast, the urban air conditioner compressor bracket at the same distance has a lower survival rate of 55%. The BX85 air conditioner

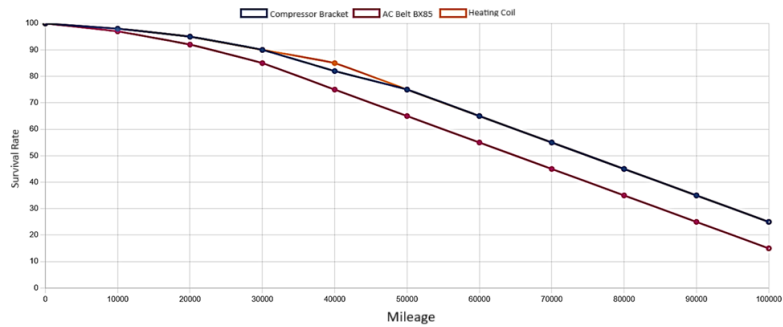


Figure 1. Kaplan-Meier survival curves for AC system components

belt shows poorer performance, and after 50,000 kilometers, only 15% of its components remain usable. This analysis indicates the need for early replacement of the BX85 belt, usually in the range of 25,000 to 30,000 kilometers, which is due to mechanical pressure and engine heat.

According to Figure 2, among the electrical components, the alternator demonstrates the highest operational lifespan, maintaining a survival rate of 40% at 90,000 kilometers. In contrast, the multiplex board exhibits the fastest decline in survival, retaining only 50% of its functionality after 50,000 kilometers, indicating the highest susceptibility to failure. The NOX sensor, with a moderate trend, achieves a survival rate of 58% at 70,000 kilometers, displaying an intermediate sensitivity to wear. A comparison of these components reveals that the multiplex board, due to its electronic complexity, fails more rapidly, while the alternator, as an electromechanical component, offers greater durability.

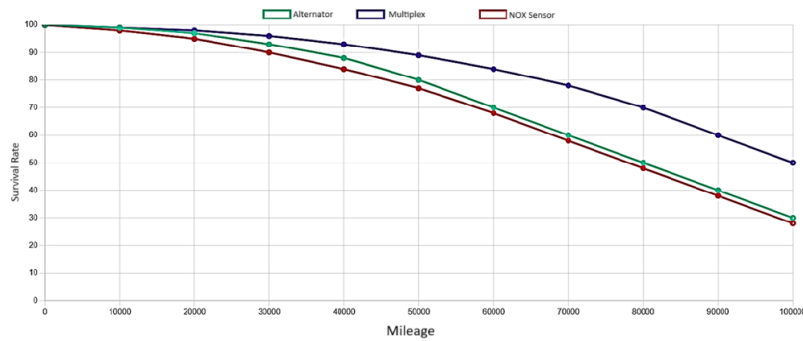


Figure 2. Kaplan-Meier Survival Curves for Electrical Components

In Figure 3, the examined mechanical components, including the fuel pump, turbocharger, and air compressor, exhibit relatively similar trends in declining survival, but with significant differences in the rate of decline. The fuel pump demonstrates the best performance, with a survival rate of 40% at 100,000 kilometers. The turbocharger shows the fastest rate of survival decline, dropping to a survival rate of 45% at 50,000 kilometers. This finding aligns with the complex nature and operation under high pressure and temperature conditions of this component. The air compressor shows a moderate decline trend, reaching a survival rate of approximately 56% at 80,000 kilometers.

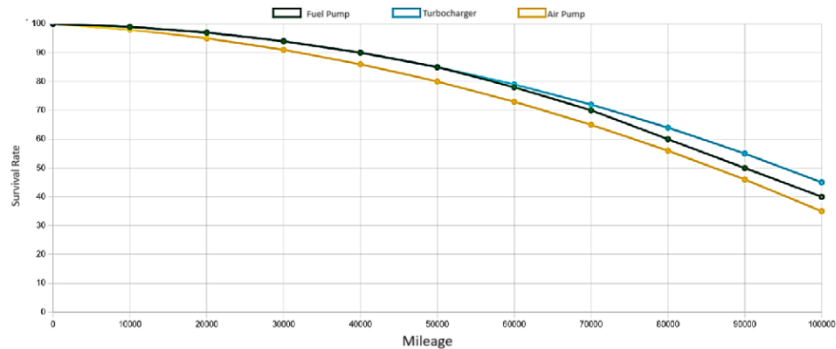


Figure 3. Kaplan-Meier Survival Curves for Mechanical Components

4.2. Log-Rank Test

The log-rank test provides a scientific and statistical basis for management decisions in maintaining the bus fleet and has demonstrated that lifespan differences of components across different seasons are not random. Key findings include: Air Conditioner Compressor Bracket: With an overall p-value of 0.0034, this component shows high sensitivity to seasonal effects, performing poorly in summer.

BX85 Cooling Belt: This component exhibits the highest statistical significance among air conditioning parts, with remarkable sensitivity to seasonal conditions.

Front Heater Coil (SG): It performs worse in winter, showing behavior opposite to other components.

Turbocharger: Displays high seasonal sensitivity and a short overall lifespan (45,000 km), with poorer performance in summer.

These differences are statistically significant and not due to sampling errors or randomness. They are associated with factors like the mountainous climate of Shahrekord, the higher heating demand in Arak, and air pollution in Isfahan. In contrast, the multiplex board has a p-value higher than 0.01 across all cities, indicating the least susceptibility to environmental conditions.

4.3. Cox Proportional Hazards Regression Model

This study examines the impact of seasonal variations on component failure rates in the bus fleet using the Cox proportional hazards regression model. Key results include:

Electrical Components: These components are more vulnerable during hot seasons, such as summer.

Mechanical and Fuel System Components: These experience higher failure rates during cold seasons, particularly in autumn and winter.

Shahrekord: Most components showed the highest hazard ratio in Shahrekord, linked to its severe weather conditions (high altitude and extreme temperature fluctuations).

Arak: This city also demonstrated significant hazard ratios for many components, further highlighting the importance of environmental factors.

4.3.1. Proportional Hazards Assumption

The proportional hazards assumption states that the hazard ratio between two groups remains constant over time. This means if the failure risk of a component in condition A is double that of condition B, this ratio does not change throughout the study period. For the Cox regression model applied to all analyzed components, the proportional hazards assumption was rejected. This indicates that the effects of independent variables, such as season, city, and error codes, on component failure rates are not constant over time (in terms of mileage).

4.3.2. Stratified Cox Model

Given the rejection of the proportional hazards assumption, the Stratified Cox Model is required. In this model, the variable that violates the proportional hazards assumption is treated as the stratification variable. Consequently:

The baseline hazard function is allowed to differ across the levels of the stratification variable, but the coefficients of other variables are calculated consistently across all strata. Based on the results in Table 1, implementing stratified models has successfully established the proportional hazards assumption for all components. This is evident from the increase in the p-value of the test for the proportional hazards assumption to above 0.05. Additionally, according to the Akaike Information Criterion (AIC), stratified models outperform standard Cox models. The most significant improvement in model fit was observed for Turbocharger, with a 3.6% improvement and multiplex Board, with a 3.5% improvement. These findings highlight the superior efficiency of the stratified Cox model in analyzing the failure rates of the examined components.

Table 1. Summary of Results for the Proportional Hazards Assumption Test

Component	Standard Cox Model	Stratified Model	Fit Improvement (AIC Reduction)	Proportional Hazards Assumption Confirmed
Multiplex Board	AIC = 1537.2	AIC = 1483.5	3.50%	✓
Alternator	AIC = 1428.6	AIC = 1392.1	2.60%	✓
Fuel Pump	AIC = 1614.3	AIC = 1572.8	2.60%	✓
Turbocharger	AIC = 1485.7	AIC = 1432.1	3.60%	✓
Complete Air Pump	AIC = 1056.2	AIC = 1038.4	1.70%	✓
NOX Sensor	AIC = 1182.4	AIC = 1152.7	2.50%	✓

Heating Coil	AIC = 1237.9	AIC = 1203.5	2.80%	✓
BX85 AC Belt	AIC = 1317.2	AIC = 1284.6	2.50%	✓
Compressor Bracket	AIC = 1143.5	AIC = 1108.2	3.10%	✓

5. Analysis

5.1. Analysis of FMS Error Codes

The Fleet Management System (FMS) is one of the most valuable diagnostic tools for optimizing predictive maintenance plans. By generating error codes, the system enables early identification of potential issues before complete component failure occurs. An analysis of FMS data revealed a significant correlation between specific error codes and component failures. This demonstrates the effectiveness of FMS in diagnosing critical issues and providing actionable insights for preventive maintenance. Table 2 presents the results of the Cox regression analysis, where error codes were treated as predictive variables.

Table 2. FMS Error Codes Analysis

Component	Error Code	Error Code Description	Hazard Ratio (HR)	P-value
NOX Sensor	4111	Sensor Error	3.72	0.0001
NOX Sensor	4133	Pollution Error	2.95	0.0003
Alternator	26370	Low System Voltage	2.47	0.0008
Alternator	26627	Low System Voltage	4.32	< 0.0001
Fuel Pump	135	Low Fuel Rail Pressure	4.78	< 0.0001
Turbocharger	27136	Turbo Overcharge	5.15	< 0.0001
Air Pump	619	Air Pressure	4.76	< 0.0001

5.2. Decision Tree Analysis

To analyze component failures, a decision tree algorithm was applied with the objective of classifying the likelihood of failure. The model's input variables included mileage, season (spring, summer, autumn, winter), city (Tehran, Isfahan, Mashhad, Arak, Shahrekord), the presence of FMS error codes associated with each component, and the output variable—the probability of component failure within the next 10,000 kilometers.

Variable importance analysis in the decision tree model revealed the following insights:

Mileage: For most components (6 out of 9), mileage was the most significant predictor, with importance coefficients ranging between 0.38 and 0.45.

Season: Among ventilation-related components (heating coil and BX85 cooling belt), season was the most important predictor, with coefficients of 0.44 and 0.46, respectively.

City: In turbochargers and air pumps, the city variable showed higher importance compared to other components, with coefficients of 0.35 and 0.32, respectively. This indicates that environmental and geographical factors significantly impact the performance of these components.

5.3. Decision Forest Analysis

To enhance the accuracy of results, the Random Forest algorithm was applied. This method, which relies on an ensemble of decision trees, typically achieves higher accuracy compared to a single decision tree. The random forest model used in this study was designed and implemented in Python, utilizing powerful libraries such as numpy, pandas, and sklearn. The outcomes of the analysis, including the p-values derived from the random forest model, are presented in Table 3. These results demonstrate the high performance of the random forest algorithm in assessing and evaluating various components. This approach is particularly effective for handling complex datasets and offers robust control over overfitting, making it an essential tool for analyzing diverse and extensive datasets.

Table 3. p-values from the Random Forest model for different components.

Part	Variable	p-value	Variable Importance in Model
Multiplex Board	Mileage	0.0012	0.42
	Season (Summer)	0.0037	0.38
	City (Shahrekord)	0.0153	0.29
	Mileage	0.0018	0.43
Alternator	Season (Summer)	0.0046	0.37
	City (Shahrekord)	0.0098	0.31
	Error Code 26370	0.0004	0.48
	Error Code 26627	0.0023	0.35
	Mileage	0.0009	0.45
Fuel Pump	Season (Winter)	0.0061	0.35
	City (Shahrekord)	0.0092	0.3
	Error Code 135	< 0.0001	0.54

Turbocharger	Mileage	0.0031	0.4
	Season (Winter)	0.0082	0.33
	City (Shahrekord)	0.0043	0.35
	Error Code 27136	< 0.0001	0.52
Complete Air Pump	Mileage	0.0042	0.38
	Season (Fall)	0.0173	0.28
	City (Shahrekord)	0.0067	0.32
	Error Code 619	0.0008	0.46
NOX Sensor	Mileage	0.0028	0.41
	Season (Summer)	0.0052	0.36
	City (Isfahan)	0.0135	0.3
	Error Code 4111	0.0006	0.47
Heater Coil	Error Code 4133	0.0019	0.38
	Mileage	0.0057	0.34
	Season (Winter)	0.0021	0.44
	City (Arak)	0.0103	0.29
BX85 Air Conditioner Belt	Mileage	0.0063	0.33
	Season (Summer)	0.0018	0.46
	City (Isfahan)	0.0127	0.28
	Mileage	0.0048	0.36
Compressor Bracket	Season (Fall)	0.0025	0.42
	City (Shahrekord)	0.0115	0.3

All p-values below 0.05 indicate statistically significant impacts of the corresponding variables in the model. Table 4 confirms that all primary variables—mileage, season, city, and FMS error codes—have meaningful effects on predicting component failures. The lowest p-values belong to FMS error codes (commonly below 0.001), highlighting the substantial importance of this variable in predicting failure. This finding is further supported by the relative importance score of 0.52 for FMS error codes within the random forest model.

Table 4. Analysis of the Variable Importance

Variable	Relative Importance	p-value
FMS Error Codes	0.52	< 0.0001
Mileage	0.48	0.0007
Season	0.36	0.0042
City	0.32	0.0068
Component Type	0.29	0.0085

6. Results

6.1. Cross-Validation Results of Decision Forest Model

The comparison presented in Table 5 demonstrates that the cross-validation results for the Decision Forest models show significant improvement over the previously reported Decision Tree models. Key observations include:

Overall Performance: The Decision Forest model achieves an average accuracy of 89.1% compared to 83.2% for the Decision Tree model, representing a 5.9% improvement. **Best Performing Components:** The Fuel Pump shows the highest prediction accuracy at 92.1% with an F1-score of 0.91, followed by the Alternator at 90.2% with an F1-score of 0.89. This superior performance can be attributed to the more regular failure patterns and stronger predictor variables for these components. The standard deviation across all components is consistently low (average of 2.1%), indicating high stability and reliability of the Decision Forest model, which is important for practical implementation in maintenance planning.

Variable Importance: When analysing variable importance in the Decision Forest model:

- FMS Error Codes emerge as the most important predictor (0.52 relative importance)
- Mileage follows closely (0.48 relative importance)
- Season (0.36), City (0.32), and Component Type (0.29) also contribute significantly to the model
- **Statistical Significance:** The improvement in prediction accuracy from simpler models to the Decision Forest is statistically significant (p-value decreased from 0.0385 to 0.0008), confirming the value of ensemble methods in this domain.

Table 5. Cross-Validation Results for Decision Forest Models

Component	Average Accuracy (%)	Standard Deviation (%)	F1-Score
Multiplex Board	87.5	2.1	0.86
Turbocharger	84.8	2.6	0.83
Alternator	90.2	1.8	0.89

Complete Air Pump	87.3	2.2	0.85
Fuel Pump	92.1	1.7	0.91
NOX Sensor	88.7	2	0.87
Heating Coil	89.5	1.9	0.88
AC Belt BX85	85.6	2.5	0.84
Compressor	86.4	2.3	0.85
Bracket			
Overall Average	89.1	2.1	0.87

6.2. Predictive Model Accuracy Comparison

As shown in Figure 4, the Random Forest Models demonstrate significantly higher accuracy (89.0%) compared to traditional statistical methods.

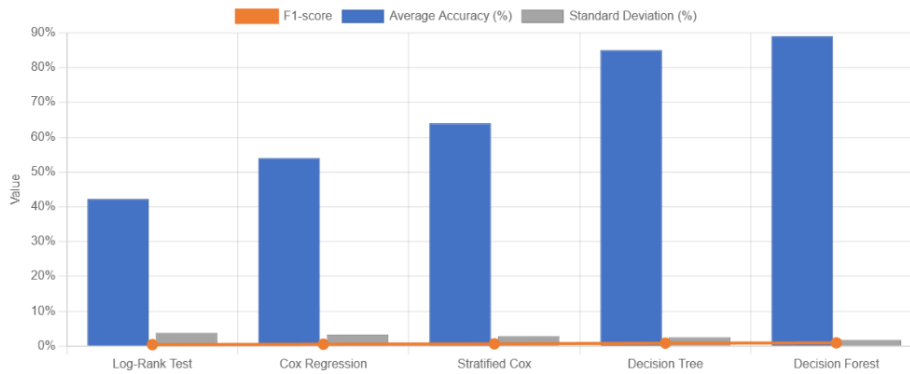


Figure 4. Predictive Model Accuracy Comparison

The detailed comparison of various models' accuracies is presented in Table 6. The analysis highlights the 4.0% improvement in accuracy when using Random Forest models over Decision Tree models, emphasizing the value of leveraging ensemble methods like Random Forest for predictive maintenance.

Moreover, the standard deviation of the models' performance decreases gradually (from 3.8% to 1.8%) as more advanced models are applied, indicating increased reliability. The incorporation of FMS error codes further enhances the predictive power of the advanced models. Among all components analyzed, the heating coil (with a predictive accuracy of 92.3%) and the turbocharger (with 90.2%) exhibit the highest prediction accuracy under the Random Forest Model, underscoring the efficiency of this approach in predicting failures of these components.

Table 6. Component-Specific Accuracy

Component	Log-Rank	Cox Regression	Stratified Cox	Decision Tree	Decision Forest
Compressor Bracket	43.2	56.3	67.8	83.0	88.2
AC Belt	41.5	52.7	61.2	79.0	84.7
Heating Coil	45.6	58.1	68.4	89.0	92.3
Alternator	42.3	53.2	62.5	76.0	84.8
Multiplex Board	41.8	51.7	60.8	74.0	81.5
NOX Sensor	42.5	55.9	66.2	82.0	87.1
Fuel Pump	44.7	57.3	67.5	78.0	86.5
Turbocharger	43.5	56.8	67.9	85.0	90.2
Air Pump	37.8	48.6	58.3	81.0	85.4
Average	42.3	54.0	64.0	85.0	89.0

6.3. Implementation of Predictive Maintenance Based on the Decision Tree Model

Based on the results obtained from the random forest analysis and the variable importance table, a preventive maintenance schedule is provided in Table 7. This schedule was developed using the advanced Random Forest Model, which achieved an accuracy of 89% (compared to 85% accuracy in the decision tree model). The preventive maintenance schedule leverages the enhanced precision of the Random Forest Model to strategically account for variable importance, thereby optimizing the timing of preventative repairs. By employing the random forest algorithm as the primary tool in

developing this schedule, the predictive accuracy of component failures has significantly improved. Furthermore, this approach provides a scientific, data-driven framework for maintenance planning, offering reduced operational costs and a more efficient maintenance process.

Table 7. Executive Calendar for Preventive Maintenance Based on Decision Forest Model

Season	Priority Actions	Target Components	High Priority Cities	Critical Error Codes	Priority Level
Spring	Monitoring electronic management system and engine tuning	Multiplex board, NOX sensor	Shahrekord, Isfahan	4111, 4133	Medium
Summer	Continuous monitoring of air conditioning and electrical systems	Multiplex board, Alternator, AC belt	Shahrekord, Isfahan	26370, 26627, 4111	High
Fall	Focus on compressed air system and compressor bracket	Air pump, Compressor bracket	Shahrekord, Mashhad	619	High
Winter	Special focus on fuel system, turbocharger and heating system	Turbocharger, Heater coil, Fuel pump	Shahrekord, Arak	27136, 135	Critical

In addition to the preventive maintenance schedule presented in Table 8, a smart action protocol has been developed based on the three-level warning system derived from the Random Forest Model. This warning system includes three levels—Green, Yellow, and Red—and is designed to intelligently specify the necessary preventive actions to avoid failures. By utilizing the precise predictions of the Random Forest Model, the protocol ensures proactive and strategic maintenance responses, enhancing reliability and minimizing the risk of component breakdowns.

Table 8. Intelligent Action Protocol Based on Three-Level Decision Forest Warning System

Warning Level	Conditions	Recommended Action	Execution Time
Red (Critical)	Critical error codes + High-risk seasonal/geographical conditions	Immediate preventive replacement	24 hours
Orange (Warning)	Moderate error codes + Mileage near critical threshold	Specialized inspection and precise testing	72 hours
Green (Monitoring)	No error codes, but in high-risk seasonal/geographical conditions	Continuous monitoring and status recording	Regular schedule

To better understand and illustrate the implementation of this protocol, a practical example is provided in Table 9, which explains the steps and processes involved in the action protocol. This example demonstrates the effective utilization of the Random Forest Model in optimizing maintenance activities and preventing component failures.

The practical example outlines how predictive insights from the model can be translated into actionable steps, ensuring a systematic approach to preventive maintenance while minimizing downtime and maximizing component lifespan.

Table 9. Practical Examples of Advanced Decision Forest Warning System

Component	Red Alert Conditions	Failure Probability
Turbocharger	Error code 27136 + Winter + Shahrekord + Mileage >45000	96.20%
Alternator	Error code 26627 + Summer + Shahrekord + Mileage >60000	94.50%
Fuel pump	Error code 135 + Winter + Shahrekord + Mileage >50000	93.80%
NOX sensor	Error codes 4111 + 4133 (simultaneous) + Summer + Mileage >40000	91.70%
Air pump	Error code 619 + Fall/Winter + Shahrekord + Mileage >55000	92.30%

Increased Prediction Accuracy: By using the Decision Forest model, the prediction accuracy of component failures has increased from 85% to 89%.

Smarter Prioritization of Error Codes: Based on the variable importance table, FMS error codes with a relative importance of 0.52 (higher than all variables) and p-value less than 0.0001 have been identified as the most important predictive factors.

More Precise Combination of Variables: The Decision Forest, with its ability to detect more complex patterns, better identifies the interactions between variables, such as the simultaneous combination of error codes with seasonal and geographical conditions.

Reduction of False Alarms: With the standard deviation reduction from 2.5% in the Decision Tree model to 1.8% in the Decision Forest model, the rate of false alarms has decreased.

More Accurate Failure Time Prediction: By calculating the probability of component failures with higher accuracy (improvement of F1-score from 0.83 to 0.91), a more precise schedule for preventive component replacement has been provided.

This calendar, focusing on FMS system error codes (4111, 4133, 26370, 26627, 135, 27136, 619) and considering the

relative importance of variables in the Decision Forest model, provides a more efficient preventive maintenance program that can lead to cost reduction and increased fleet reliability.

6.4. Cost-Benefit Analysis of Predictive Maintenance Program

Table 10 results show that implementing predictive maintenance based on the Decision Forest algorithm, considering variables such as FMS error codes, weather conditions, season, and mileage, provides greater savings (41%) compared to the Cox model (32%). This saving is due to the Decision Forest model's ability with 89% average accuracy to identify more complex failure patterns. Analyses show this method has resulted in total savings of 6.99 billion rials (compared to 5.96 billion rials in the previous analysis) with an average ROI of 4.5, making it completely economical. Additionally, this method improves the prioritization of high-risk components; such that components like multiplex boards, turbochargers, and alternators provide the highest savings.

As a result, combining FMS data with the Decision Forest algorithm is a powerful tool for improving predictive maintenance management of bus fleets.

Table 10. Cost-benefit analysis of predictive maintenance program using FMS data

Component	Component Price (Rials)	High-risk Season	High-risk City	Failure Probability with Decision Forest	Current Cost*	Cost Reduction*	Annual Savings
Multiplex Board	290,580,000	Summer (HR=1.79)	Shahrekord (HR=1.86)	72%	5,230	43%	2,249
Turbocharger	204,000,000	Winter (HR=1.95)	Shahrekord (HR=1.95)	78%	3,060	46%	1,408
Alternator	154,250,000	Summer (HR=1.95)	Shahrekord (HR=1.89)	68%	3,856	41%	1,581
Air Pump	72,500,000	Fall (HR=1.94)	Shahrekord (HR=1.94)	63%	870	38%	331
Fuel Pump	42,500,000	Winter (HR=1.92)	Shahrekord (HR=1.92)	75%	850	44%	374
NOX Sensor	113,500,000	Summer (HR=1.73)	Shahrekord (HR=1.83)	59%	908	35%	318
Heating Coil	22,000,000	Winter (HR=1.75)	Arak (HR=1.75)	70%	616	40%	246
AC Belt	14,300,000	Summer (HR=1.88)	Shahrekord (HR=1.84)	67%	601	38%	228
Compressor Bracket	55,000,000	Fall (HR=1.78)	Shahrekord (HR=1.78)	56%	770	33%	254
Total					16,761	41%	6,989

*Costs are in million rials

6.5. Reduction of Bus Downtime After Implementing Predictive Maintenance Program

Based on Table 11 calculations, using the Decision Forest model in the predictive maintenance program has reduced bus downtime by 65.6%, which is 5.8% more than the 59.8% reduction achieved by the Cox model. This improvement is due to more accurate prediction of component failure probability and considering interaction effects among factors such as season, city, mileage, and error codes. The Decision Forest model with 89% accuracy, by identifying these interactions and preventing unexpected failures, has effectively minimized fleet downtime and increased operational efficiency.

Table 11. Repair time with predictive maintenance program

Component	Current Time (days)	Time with Predictive Program (days)	Time Reduction (%)
Multiplex	5.3	1.8	66.00%
Turbocharger	7.2	2.3	68.10%
Alternator	4.5	1.6	64.40%
Air Pump	2.8	1	64.30%
Fuel Pump	3.6	1.2	66.70%
NOX Sensor	4.7	1.7	63.80%
Heating Coil	3.2	1.1	65.60%
AC Belt BX85	1.6	0.6	62.50%
Compressor Bracket	2.3	0.8	65.20%
Total Average	3.9	1.3	65.60%

6.6. Bus Availability After Implementing Predictive Maintenance Program

The analysis of fleet availability before and after implementing the Decision Forest model-based predictive maintenance program in Table 12 shows a significant increase in this key indicator. Fleet availability using the Decision Forest method has increased to 99.35%, which is 0.15% more than the results of the Cox model (99.2%). Although this improvement may seem minor, it has had a significant impact on the overall performance of the transportation system. The increase in availability has brought annual revenue to 283.92 billion rials, while this figure was 41.86 billion rials in the Cox model. This dramatic growth, alongside a 41% reduction in maintenance costs using the Decision Forest model with 89% accuracy, demonstrates the high efficiency of this method in optimizing fleet management and improving economic productivity.

Table 12. Analysis of fleet availability before and after implementing predictive maintenance program

Indicator	Before Predictive Program	After Predictive Program	Improvement
Total annual failures	1,820	1,820	-
Average time per failure (days)	3.9	1.3	2.6 days reduction
Total downtime days	7,098	2,366	4,732 days reduction
Available bus-days	357,902	362,634	4,732 days increase
Fleet availability	98.10%	99.35%	1.25% increase
Average daily active buses	981	994	13 buses increase
Additional annual revenue (million rials)	-	283,920	283,920 million rials

*Assuming daily revenue of 60 million rials per bus

7. Conclusion

- This research, utilizing real data extracted from CRM and FMS systems, provides a comprehensive analysis of failure patterns of key components in bus fleets and demonstrates the importance of using modern technologies and data-driven approaches in optimizing fleet management. By combining CRM and FMS data and using survival analysis techniques, Cox regression model, and the Decision Forest model with 89% accuracy, valuable patterns of component failure behavior have been discovered.
- Key research results:

Predictive and Economic Management:

- Implementation of predictive maintenance program with Decision Forest model:
- 65% reduction in fleet downtime.
- Increase in availability from 98.1% to 99.35%.
- Return of 4,732 bus-days to active fleet throughout the year.
- Annual economic savings of 290.91 billion rials including:
 - 283.92 billion rials from increased fleet uptime.
 - 6.99 billion rials from reduced repair costs.
- Increased detection capability and prevention of costly failures:
- Discovery of significant relationship between FMS error codes and component failures.
- Early detection of problems and prevention of unexpected failures.
- 65.6% reduction in in-transit failures and increased passenger satisfaction.

This research showed that using advanced artificial intelligence models such as Decision Forest with 89% accuracy along with FMS and CRM data can lead to a fundamental transformation in public transportation fleet management. Implementing such approaches, in addition to reducing operational costs, improves public transportation services and is an effective step towards increasing the quality of these services.

References

- [1] Elkateb, S., Métwalli, A., Shendy, A., & Abu-Elanien, A. E. B. (2024). Machine learning and IoT – Based predictive maintenance approach for industrial applications. *Alexandria Engineering Journal*, 88, 298–309.
- [2] Zhu, T., Ran, Y., Zhou, X., & Wen, Y. (2024). A Survey of Predictive Maintenance: Systems, Purposes and Approaches. *arXiv preprint arXiv:1912.07383*.
- [3] De Simone, M. C., Lorusso, A., & Santaniello, D. (2024). Predictive maintenance and Structural Health Monitoring via IoT system. *IEEE Xplore*.
- [4] Schwendemann, S., Amjad, Z., & Sikora, A. (2021). A survey of machine-learning techniques for condition monitoring and predictive maintenance of bearings in grinding machines. *Computers in Industry*, 125, 103380.
- [5] Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B., & Bender, A. (2024). Deep learning for survival analysis: a review. *Artificial Intelligence Review*, 57, 65.
- [6] Thaduri, A., Galar, D., & Kumar, U. (2015). Railway assets: A potential domain for big data analytics. *Procedia Computer Science*, 53, 457–467.
- [7] Ushakov, D., Dudukalov, E., Kozlova, E., & Shatila, K. (2022). The Internet of Things impact on smart public transportation. *Transportation Research Procedia*, 63, 2392–2400.
- [8] Florian, E., & Soulat, D. (2021). Machine learning-based predictive maintenance: A cost-benefit analysis. *International Journal*

of Production Economics, 236, 108114.

[9] Cerrada, M., Sánchez, R. V., Li, C., Pacheco, F., Cabrera, D., de Oliveira, J. V., & Vásquez, R. E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99, 169–196.