

Data-Driven Predictive Maintenance and Failure Mode Recognition for Industrial

Zihan Shen^{1*}, Xiaodong Zhu¹, Jinpeng Sun¹, Xuanyu Shi¹, Haoyan Chen¹, Jue Wang²

¹ Business School, University of Shanghai for Science and Technology, Shanghai, China

² University of Southern California, Los Angeles, California, America

*Correspondence: Zihan Shen

Received: 20 May 2026; **Accepted:** 18 June 2026; **Published:** 3 August 2026

Abstract:

Predictive maintenance plays a pivotal role in minimizing industrial downtime and optimizing resource allocation. This study presents a comprehensive data-driven framework for failure prediction and mode recognition based on 10,000 operational records of industrial equipment, implementing a multi-stage analytical pipeline by integrating unsupervised and supervised learning approaches: K-Means clustering was employed to segment equipment operational conditions, the Apriori algorithm was utilized to mine association rules between operating conditions and failures, and multiple machine learning architectures including Gradient Boosting and XGBoost were constructed for predictive modeling. The results indicate that the overall baseline failure rate of the equipment is 3.39%, among which low-quality (Type L) products exhibit significantly higher failure vulnerability (3.92%) compared to other quality tiers. Unsupervised clustering identified four distinct operational regimes, revealing that the “High-Temperature, High-Wear” cluster carries the highest failure risk at 5.70%. Association rule mining further quantified that the synergistic effect of elevated temperatures and low-to-medium rotational speeds acts as a potent catalyst for Heat Dissipation Failures (lift > 7). In the modeling phase, the Gradient Boosting model achieved an optimal AUC of 0.9696 for binary failure early warning, while the XGBoost model demonstrated superior performance in multi-class failure diagnosis with an AUC of 0.9575. This study demonstrates that dynamic operational states exert a far more decisive influence on equipment health prognosis than static product grades. Accordingly, a stratified maintenance strategy combining real-time failure probability alerts with mode-specific interventions is proposed, which provides a robust and interpretable methodological support for industrial enterprises to transition from reactive maintenance to proactive, data-informed equipment management.

Keywords:

Predictive Maintenance, Failure Mode Recognition, Gradient Boosting, XGBoost, Cluster Analysis, Association Rules

1. Introduction

Compared to traditional “run-to-failure” [1,2] maintenance, predictive maintenance relies more heavily on equipment operational data to characterize health status, with the core objective of identifying risks and triggering maintenance before a failure

occurs. For industrial enterprises, critical equipment failures often lead to production line downtime, excessive energy consumption, delivery delays, and the waste of spare parts and human resources. With the advancement of the Industrial Internet[3], edge sensing, and manufacturing execution systems (MES), enterprises can continuously collect multidimensional data[4] such as temperature, rotational speed, torque, and tool wear, laying a practical foundation for the transition from experience-based maintenance to data-driven maintenance.

The data analysis of industrial scenarios presents obvious challenges: failures are low-frequency events, resulting in highly imbalanced data categories[5]; equipment performance varies significantly under different quality grades and operating conditions; and the same failure may be induced by multiple interacting factors. Merely predicting whether a machine will fail is insufficient to support on-site decision-making. Enterprises urgently need to clarify failure modes, trigger conditions, high-risk clusters, and corresponding maintenance strategies.

The introduction of this article is organized to address four core problems: identifying the overall failure level and differences among equipment of different quality types to build a failure profile; clustering equipment groups based on operating conditions to identify high-risk operational modes; mining association rules between operating condition features and failure modes to analyze the co-occurrence of failures; and constructing predictive models to achieve equipment failure early warning and auxiliary diagnosis of failure types. The overall analysis logic follows the sequence of “descriptive statistics → population clustering → rule mining → predictive modeling → maintenance recommendations”.

2. Materials and Methods

2.1. Dataset and Variables

The dataset utilized in this study comprises 10,000 operational records collected from industrial equipment, encompassing 14 original fields including equipment identifiers, product quality grades, air temperature, process temperature, rotational speed, torque, tool wear, a binary general machine failure label, and five discrete failure mode labels.

For the purpose of this analysis, variables are systematically categorized into input features and predictive target variables. The input features include: (1) product quality grade (denoted as L, M, and H), employed to quantify disparities in failure susceptibility across distinct quality tiers; (2) air temperature and process temperature, which characterize the ambient thermal environment and internal operating temperature differentials of the equipment; (3) rotational speed (unit: rpm), reflecting equipment load conditions and operational rhythm; (4) torque (unit: Nm), representing the output mechanical load level; and (5) tool wear (unit: min), serving as a quantitative proxy for long-term equipment usage and component aging degree.

Two distinct sets of target variables are defined for predictive modeling tasks: a binary “machine failure” label for early warning of impending equipment malfunction, and five multi-class labels corresponding to specific failure modes, namely Tool Wear Failure (TWF), Heat Dissipation Failure (HDF), Power Failure (PWF), Overstrain Failure (OSF), and Random Failures (RNF).

2.2. Data Preprocessing

A rigorous multi-step preprocessing pipeline was implemented to ensure data integrity, consistency, and compatibility with downstream machine learning algorithms. First, a data completeness audit was conducted using the `info()` and `isnull()` functions in the Python programming language, confirming the absence of missing values across all fields. Second, temperature unit conversion was performed to improve interpretability in descriptive statistical analysis: two new fields, `Air_temp_C` and `Process_temp_C`, were generated by converting the original Kelvin-scale temperature values to Celsius, while the raw Kelvin fields were retained to preserve numerical precision for specific model training tasks. Third, categorical encoding was applied to the nominal “Type” variable, which was one-hot encoded into three binary dummy variables (`Type_H`, `Type_L`, `Type_M`) to eliminate spurious ordinal relationships that could introduce bias into model performance. Finally, the preprocessed dataset was partitioned into training and testing subsets using an 80%/20% stratified random sampling strategy, which maintained a consistent binary failure ratio (3.39% in the training set and 3.40% in the testing set) to effectively mitigate the confounding impact of inherent class imbalance on model training and evaluation.

2.3. Overall Failure Rate and Quality Type Differences

Statistical results indicate that out of 10,000 records, there are 339 failure samples, yielding an overall equipment failure rate of 3.39%. This demonstrates a clear low-probability failure characteristic. This finding implies that in practical deployment, relying solely on overall accuracy could cause models to be dominated by “normal samples”; hence, subsequent modeling must prioritize AUC, recall[6], and minority class recognition capabilities (Table 1).

Table 1. Failure rate comparison across different quality types.

Type	Sample Size	Failure Count	Failure Rate (%)
H	1003	21	2.09
L	6000	235	3.92
M	2997	83	2.77
Overall	10000	339	3.39

2.4. Failure Mode Distribution and Sensor Differences

Among specific failure modes, HDF, OSF, and PWF were the most prevalent, occurring 115, 98, and 95 times respectively. TWF was recorded 46 times, and RNF occurred a mere 19 times, further underscoring the extreme class imbalance within failure modes (Table 2). Comparative analyses demonstrate statistically significant differences between normal and failed samples across all key metrics—air temperature, process temperature, rotational speed, torque, and wear time—with torque and tool wear displaying the most profound variances. The mean torque for failed samples was 50.17 Nm, compared to 39.63 Nm for normal samples. Similarly, average tool wear for failed samples was 143.78 min, significantly exceeding the normal cohort's 106.69 min (Table 3).

Table 2. Occurrence statistics of failure Modes.

Failure Mode	Occurrences	Share (%)
TWF	46	0.46
HDF	115	1.15
PWF	95	0.95
OSF	98	0.98
RNF	19	0.19

Table 3. Feature mean comparisons between normal and failure samples.

Feature	Normal Sample Mean	Failure Sample Mean	p-value
Air Temp (°C)	26.82	27.74	1.95E-14
Process Temp (°C)	36.85	37.14	1.15E-04
Rotational Speed (rpm)	1540.26	1496.49	3.76E-02
Torque (Nm)	39.63	50.17	3.64E-27
Tool Wear (min)	106.69	143.78	1.91E-18

It can be further observed from the distribution plot that the fault samples are significantly more concentrated in the high temperature, high torque and high wear regions, while accounting for a relatively small proportion in the higher rotational speed range. This indicates that the fault is not driven by a single feature, but is caused by the combined action of thermal environment, mechanical load and wear condition (Figure 1).

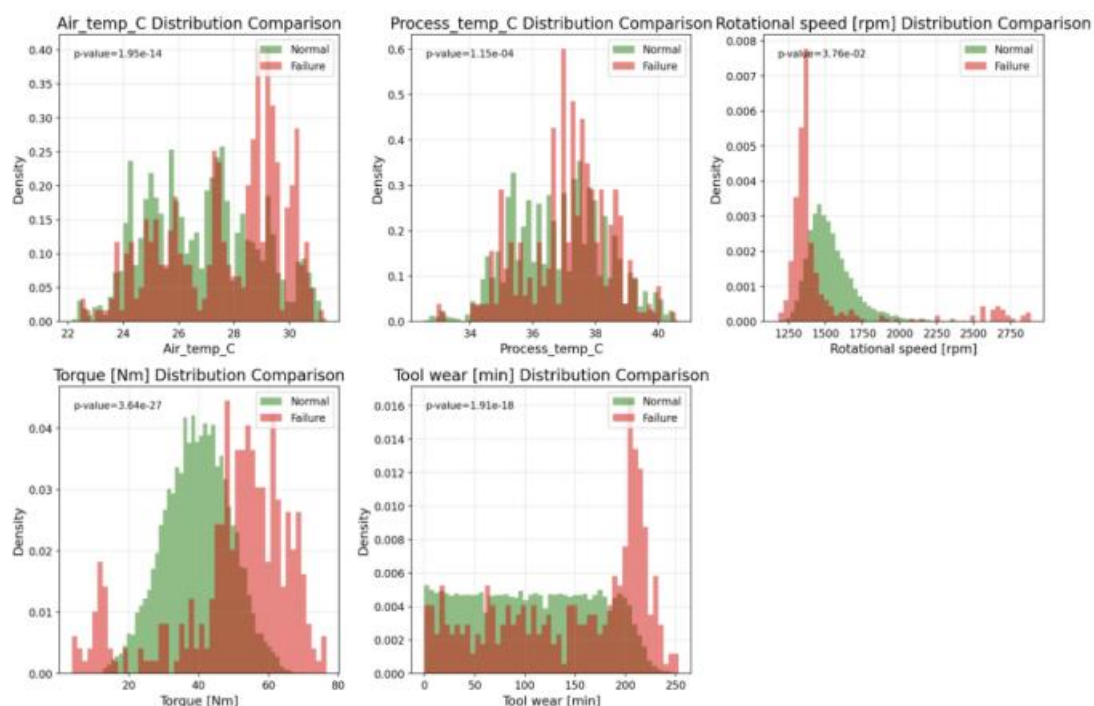


Figure 1. Distribution comparison of key sensor indicators between normal and failure samples.

2.5. Tool Wear Trends and Anomaly Interval Identification

As tool wear progresses, the failure rate universally manifests a “prolonged plateau followed by an exponential surge” trajectory. Figure 2 quantitatively characterizes the dynamic evolution of mean air temperature, rotational speed, torque, and failure rate across sequential 2.5-minute incremental intervals of cumulative tool wear duration (0–17.5 minutes). Air temperature exhibits remarkable thermal stability within the range of 26.8–27.0°C during the initial 15 minutes of operation, but undergoes a dramatic and statistically significant abrupt surge to 28.0°C exclusively in the terminal interval. Rotational speed displays pronounced stochastic variability with no discernible long-term monotonic trend, while torque manifests marked oscillatory fluctuations, consistently surpassing 40.0 Nm during periods of elevated mechanical loading.

The most salient finding is the failure rate’s characteristic “prolonged plateau followed by exponential surge” trajectory: it maintains a consistently low level below 5% (mean = 2.3%) throughout the 0–15 minute interval, then exhibits an exponential

escalation following a sharp inflection point at 15 minutes, peaking at 36.5% in the final interval—representing a 15.9-fold elevation in failure risk relative to the plateau phase. This confirms that equipment failure arises from the synergistic deterioration of multiple operational parameters in the late wear stage, which provides robust empirical validation for condition-based maintenance strategies that prioritize targeted interventions during this terminal high-risk window.

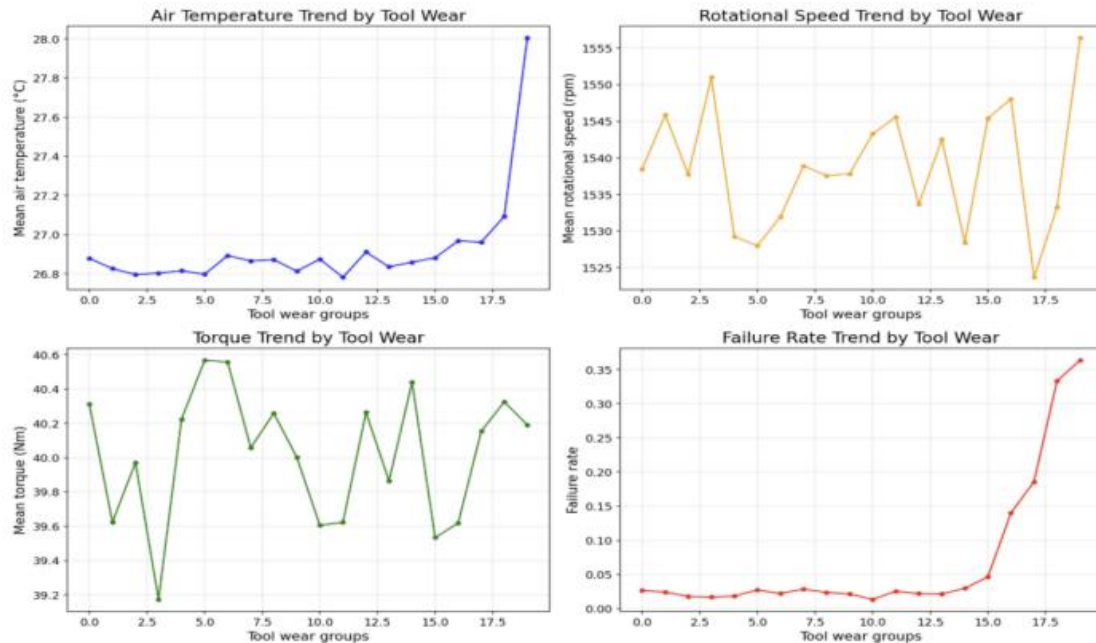


Figure 2. Trends of key indicators and failure rates with changing tool wear levels.

3. Results and Discussion

3.1. Equipment Population Analysis

To isolate equipment clusters demonstrating analogous operational paradigms, we applied standardization to five core features—air temperature, process temperature, rotational speed, torque, and tool wear—followed by K-Means clustering[7]. Utilizing the elbow method to evaluate partition counts between 2 and 10, the inertia curve began to stabilize noticeably at K=4; hence, a four-cluster taxonomy was adopted as the optimal configuration. (Figure 3)

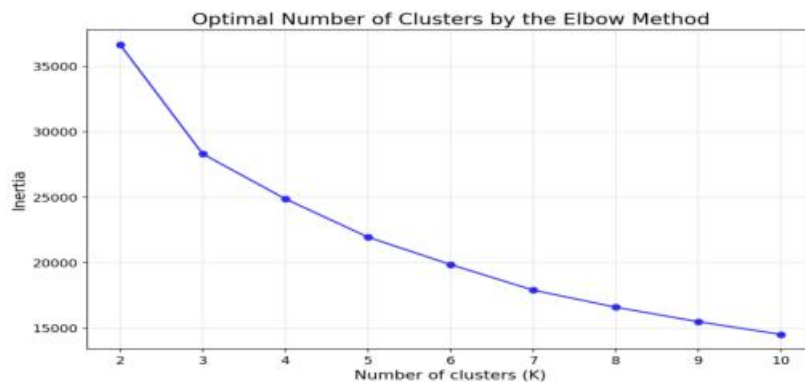


Figure 3. Optimal cluster number K determined by the elbow method.

The clustering results demonstrate that Cluster 1 has the highest failure rate at 5.70%, which is characterized by high temperature coupled with severe wear and can

be classified as the “high-risk aging group”. Cluster 2 presents a failure rate of 3.80%, featuring elevated temperature yet mild wear, and is defined as the “high-temperature mild-wear group”. Cluster 0 and Cluster 3 exhibit relatively low failure rates, at 2.04% and 2.39%, respectively. Among them, Cluster 3 is distinguished by high rotational speed, low torque, and moderate wear, which can be regarded as the “high-speed light-load group” (Table 4). Therefore, the classification of operating condition groups not only reveals the discrepancies in operating performance among individual equipment units, but also provides a solid foundation for the development of differentiated maintenance scheduling.

Table 4. Distribution of Sample Size, Failure Occurrences and Failure Rate for Each Cluster

Cluster	Sample Size	Failures	Failure Rate (%)
Cluster 0	3535	72	2.04
Cluster 1	2384	136	5.70
Cluster 2	2368	90	3.80
Cluster 3	1713	41	2.39

As can be seen from the PCA dimensionality reduction plot and cluster heatmap, the clustering boundaries exhibit favorable discriminability, indicating that temperature, torque, wear, and rotational speed play an effective role in cluster delineation (Figure 4). Meanwhile, the distribution proportions of different quality types across the four clusters are relatively consistent, demonstrating that the inter-cluster variations are mainly driven by operating condition features rather than the proportion of product types.

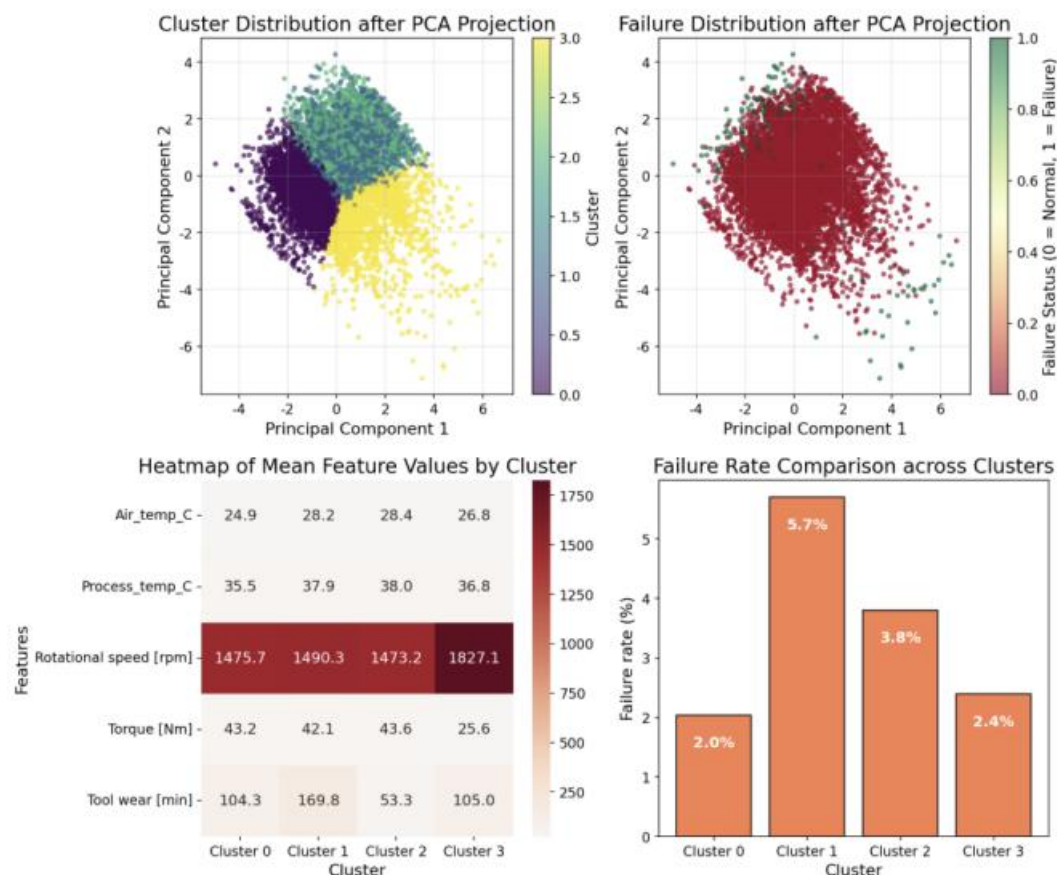


Figure 4. Cluster distribution in PCA space, group heatmap, and group failure rate comparison.

3.2. Association Rule Analysis

To deeper comprehend the preconditions for mechanical breakdown, continuous variables were discretized into categorical tiers for temperature, speed, torque, and wear, after which the Apriori algorithm[8] was deployed to extract frequent itemsets and association rules. Establishing a minimum support threshold of 1%, the algorithm successfully isolated 1,385 frequent itemsets and generated 10,589 association rules, demonstrating profound structural relationships between condition combinations and failure events.

Table 5. Detailed Metrics of Association Rules for Operating Conditions and HDF Outcomes.

Rule ID	Antecedent	Consequent	Support	Confidence	Enhancement
1	Low/Med Speed + High Temp	HDF	0.0114	0.0807	7.0206
2	High Temp	HDF	0.0114	0.0385	3.3445
3	High Temp	Low/Med Speed + HDF	0.0114	0.0385	3.3445
4	Low/Med Speed	HDF	0.0114	0.0234	2.0321
5	Low/Med Speed	High Temp + HDF	0.0114	0.0232	2.0321

The extracted rules unequivocally demonstrate that the confluence of high temperature and low-to-medium rotational speed acts as a potent catalyst for Heat Dissipation Failures (HDF). Notably, Rule 1 presents a lift exceeding 7, signifying that when this specific antecedent occurs, the probability of an ensuing HDF is vastly higher than its baseline probability. This aligns perfectly with the underlying mechanical logic of HDF generation, proving that rule mining can reliably reverse-engineer failure mechanisms (Table 5).

Regarding failure co-occurrences, PWF and OSF exhibited the strongest parallel relationship, co-occurring 12 times. HDF and OSF co-occurred 6 times, while HDF and PWF co-occurred 3 times (Figure 5). These linkages reveal a mechanical dependency between power anomalies and overstrain dynamics; consequently, real-world maintenance protocols dictating PWF interventions must synchronously assess overstrain and wear risks.

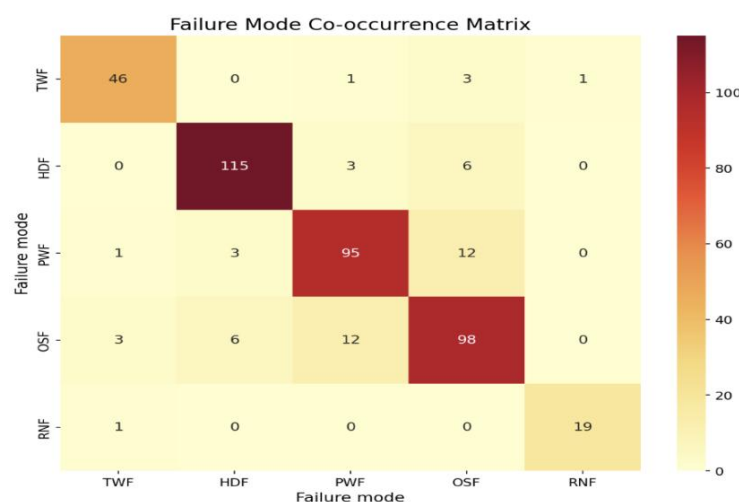


Figure 5. Failure mode co-occurrence matrix.

3.3. Binary Failure Early Warning Model

From the evaluation results, the Gradient Boosting model[9] achieves an AUC of

0.9696, delivering the best performance among the four models; its confusion matrix shows that, out of the 2,000 test samples, 1926 normal samples are correctly identified with 6 false alarms, while 45 fault samples are correctly identified with 23 missed detections. While maintaining an extremely low false alarm rate, this model can correctly recognize approximately 66.18% of fault samples, and its overall performance renders it the most suitable candidate for a real-time fault early warning model (Table 6).

Table 6. Performance Comparison of Different Machine Learning Models for Fault Prediction.

Model	AUC	Accuracy	Failure Precision	Failure Recall	Failure F1
Logistic Regression	0.9069	0.8285	0.14	0.82	0.24
Random Forest	0.9635	0.9775	0.96	0.40	0.56
Gradient Boosting	0.9696	0.9855	0.88	0.66	0.76
XGBoost	0.9669	0.9825	0.71	0.76	0.74

It is critical to note that while Logistic Regression yielded high recall, its severe lack of precision (0.14) ensures an unsustainable volume of false alarms, rendering it unfit for production. Conversely, Random Forest delivered supreme precision but suffered from depressed recall, risking catastrophic missed detections. XGBoost marginally outperformed Gradient Boosting in recall; thus, in industrial environments where the cost of a missed failure vastly eclipses the cost of a false alarm, XGBoost serves as an excellent alternative (Figure 6, Figure 7).

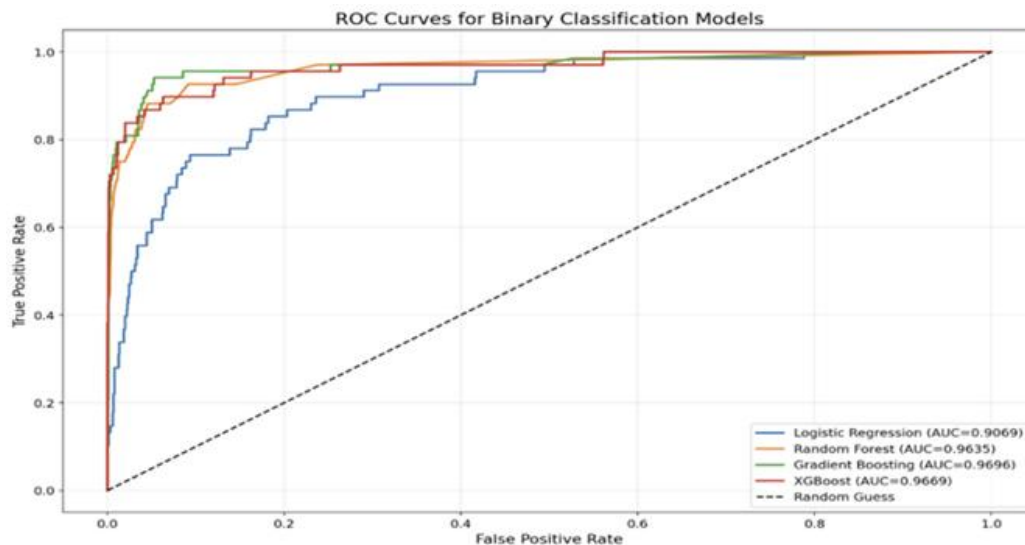


Figure 6. ROC curve comparison of binary models.

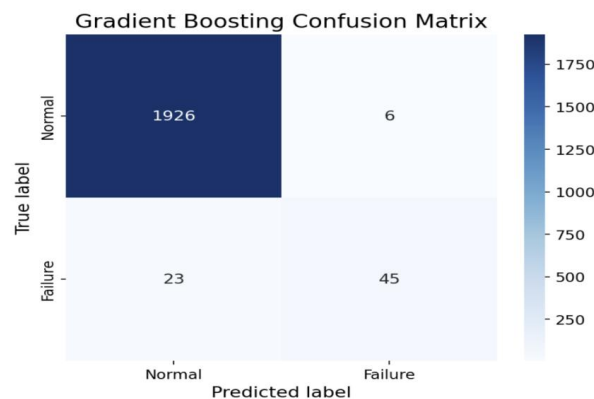


Figure 7. Confusion matrix of the optimal binary model, Gradient Boosting.

3.4. Multi-Class Failure Mode Diagnostic Model

In the multi-class diagnostic framework, XGBoost emerged superior[10], securing a One-vs-Rest (OvR) AUC of 0.9575 (Table 7). It effectively recognized HDF, OSF, and PWF; however, its recall for TWF was absolute zero, a direct artifact of extreme sample scarcity. This signifies that while current multi-class diagnosis reliably covers dominant failure modes, identifying ultra-rare failures mandates future incorporation of resampling techniques, cost-sensitive learning algorithms, or rule-based expert system enhancements.

Table 7. Performance Comparison of Machine Learning Models for Multi-class Fault Pattern Diagnosis.

Model	OvR AUC	Overall Accuracy
Random Forest	0.9235	0.98
Gradient Boosting	0.9349	0.99
XGBoost	0.9575	0.99

Feature importance mapping delineates rotational speed (0.3157), torque (0.2872), and tool wear (0.2006) as the paramount discriminative variables, vastly overshadowing the static categorical Type variable (Figure 8). This strongly validates the thesis that live dynamic conditions hold vastly greater prognostic power than inherent product quality, validating a maintenance philosophy that prioritizes real-time operational monitoring supplemented by static tiering.

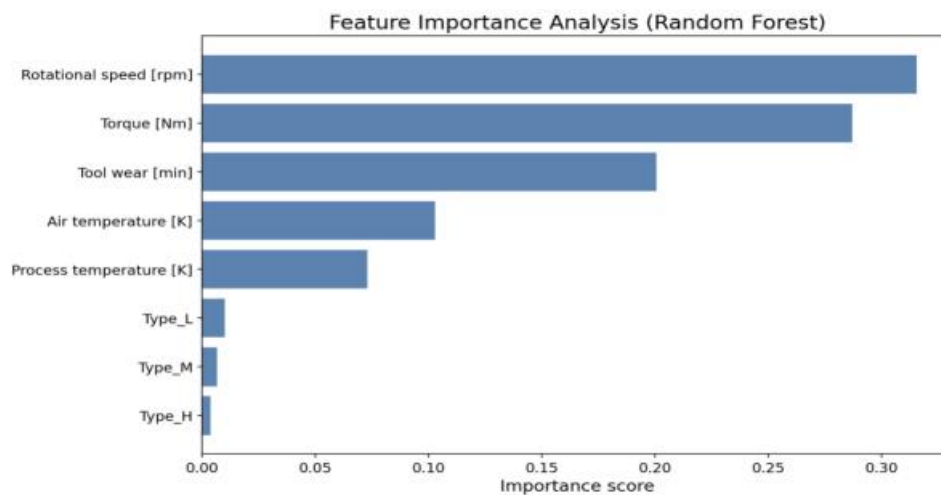


Figure 8. Key feature importance ranking via Random Forest.

4. Conclusions

The industrial equipment failure dataset presents clear low-probability and imbalanced characteristics. Type L products are at the highest risk, indicating a significant correlation between quality grades and failure susceptibility under coupled wear and load conditions. High temperature, high torque, and advanced wear are primary failure indicators. Unsupervised clustering successfully isolated a “high-temperature, high-wear” cluster that exhibits a failure rate of 5.70%, marking it as a priority target for maintenance resources. Association rule mining confirmed that the coupling of high temperature and medium-low speed acts as a strong trigger for Heat Dissipation Failures.

From a deployment perspective, we recommend a hierarchical early warning system utilizing Gradient Boosting for primary risk screening, and XGBoost for secondary

multi-class diagnosis. Differentiated maintenance scheduling should prioritize the high-risk clusters, while rule-driven troubleshooting protocols should be established .

Model Limitations and Future Directions: The current analysis relies on static cross-sectional records and lacks time-series evolutionary data. Future research should incorporate temporal sequences for Remaining Useful Life prediction. Furthermore, multi-label learning frameworks and cost-sensitive learning techniques should be explored to address the rare-event nature of specific failure modes like TWF and RNF.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Funding

This study is supported by the 2025 Graduate Teaching Project of University of Shanghai for Science and Technology (Undergraduate-Graduate Integrated Curriculum on “Object Engineering”). This work is also supported by the Teaching Research and Reform Project of Higher Continuing Education of the University of Shanghai for Science and Technology, under the project entitled “Research on the Innovation of Collaborative Education Models in Higher Continuing Education: A Case Study of the Artificial Intelligence Top-up Bachelor’s Program”.

Acknowledgments

We would like to express our gratitude to Business School, University of Shanghai for Science and Technology for providing the 6th Shanghai University Open Big Data Analysis Challenge and structural support necessary to conduct this data-driven predictive maintenance research.

References

- [1] Jardine, A.K.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* 2006, 20, 1483-1510. DOI: <https://doi.org/10.1016/j.ymssp.2005.09.012>.
- [2] Zonta, T.; da Costa, C.A.; da Rosa Righi, R.; de Lima, M.J.; da Trindade, A.M.; Li, G.P. Predictive maintenance in the Industry 4.0: A systematic literature review. *Comput. Ind. Eng.* 2020, 150, 106889. DOI: <https://doi.org/10.1016/j.cie.2020.106889>.
- [3] Lee, J.; Kao, H.A.; Yang, S. Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP*, 2014, 16, 3-8. DOI: <https://doi.org/10.1016/j.procir.2014.02.001>.
- [4] Carvalho, T.P.; Soares, F.A.; Vita, R.; Francisco, R.D.; Basto, J.P.; Alcalá, S.G. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* 2019, 137, 106024. DOI: <https://doi.org/10.1016/j.cie.2019.106024>.
- [5] Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* 2019, 115, 213-237. DOI: <https://doi.org/10.1016/j.ymssp.2018.05.050>.

- [6] He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* 2009, 21, 1263-1284. DOI: <https://doi.org/10.1109/TKDE.2008.239>.
- [7] Liao, T.W. Clustering of time series data—a survey. *Pattern Recognit.* 2005, 38, 1857-1874. DOI: <https://doi.org/10.1016/j.patcog.2005.01.025>.
- [8] Harding, J.A.; Shahbaz, M.; Srinivas; Kusiak, A. Data Mining in Manufacturing: A Review. *J. Manuf. Sci. Eng.* 2006, 128, 969-976. DOI: <https://doi.org/10.1115/1.2194554>.
- [9] Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 2001, 29, 1189-1232, DOI: <https://doi.org/10.1214/aos/1013203451>.
- [10] Zhang, D.; Qian, L.; Mao, B.; Huang, C.; Huang, B.; Si, Y. A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access*, 2018, 6, 21020-21031. DOI: <https://doi.org/10.1109/ACCESS.2018.2818678>.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)