SMOTE-Enhanced Ensemble Learning Framework for Industrial Asset Reliability: Human-Centric and Socio-Technical Perspectives

Mohd Fauzi Ab Rahman[1]*, Abdul Halim Dahalan[1], Shairazi Akmal Anuar[2]

[1]Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer (FTKEK), Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

[2]Labuan Shipyard and Engineering Sdn. Bhd., 87000 Labuan, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101000255

Received: 12 October 2026; Accepted: 17 October 2026; Published: 01 November 2026

ABSTRACT

Predictive maintenance (PdM) plays an important role in modern smart manufacturing by reducing unplanned machine downtime, optimizing operational costs, and fostering human-centric industrial environments. However, real-world industrial monitoring records commonly encounter severe class imbalance, as operational failure events occur far less frequently than normal operating states. Standard predictive models often suffer from high false-alarm rates, which can potentially induce alarm fatigue and lower user trust among maintenance technicians in industrial settings. To resolve this issue, this paper proposes an integrated machine learning framework combining physics-informed feature engineering with the Synthetic Minority Over-sampling Technique (SMOTE). Utilizing 10,000 multi-sensor operational records from the UCI AI4I 2020 Predictive Maintenance Dataset, domain-specific physical features, namely Temperature Gradient (ΔT) and Mechanical Power Output (P), were derived to enrich data representation. Two ensemble models, Random Forest and XGBoost, were trained on SMOTE-balanced data and validated on an independent 20% holdout test set ($N = 2{,}000$) using 5-fold cross-validation. The empirical findings demonstrate that SMOTE significantly enhances minority failure detection for both classifiers. Proposed XGBoost achieved superior overall performance with 98.2% Accuracy ($\pm 0.3\%$), 71.1% Precision ($\pm 1.2\%$), 79.4% Recall ($\pm 1.5\%$), 75.1% F1-Score ($\pm 1.1\%$), and an ROC-AUC of 0.891 (± 0.008). Crucially, XGBoost reduced false positive cases by 42.1% compared to Random Forest, demonstrating potential operational benefits for minimizing false alarm fatigue and unnecessary diagnostic interventions for workforce management. Feature importance evaluation confirmed that mechanical power output and thermal gradient served as key physical indicators of incipient degradation. This socio-technical decision-support paradigm directly aligns with UN SDG 9 (Industry, Innovation, and Infrastructure) and SDG 12 (Responsible Consumption and Production).

Keywords: Predictive maintenance, machine learning, socio-technical systems, workforce management, SMOTE, XGBoost.

INTRODUCTION

In the modern era of Industry 4.0 and human-centric smart manufacturing, continuous operational monitoring of critical machinery is essential for organizational sustainability, shop-floor safety, and workflow reliability [1]. Traditional maintenance management strategies generally rely on reactive repairs after a physical breakdown occurs or preventive maintenance executed at rigid, time-based calendar intervals. However, both paradigms carry significant economic and operational inefficiencies [2]. Reactive approaches cause unexpected equipment downtime, catastrophic financial losses, and heightened safety risks for frontline operators, whereas preventive schedules frequently lead to premature component replacements, inflated operational expenditure, and excessive labour demands [3]. To overcome these operational bottlenecks, Predictive Maintenance (PdM) offers a proactive decision-support strategy that leverages real-time digital monitoring logs and data-driven algorithms to forecast incipient equipment degradation before functional breakdown occurs [4].

Despite presenttantial operational advantages of predictive analytics, modelling real-world industrial operational records presents major technical and socio-technical hurdles, highlighting extreme class imbalance [5]. In automated manufacturing settings, modern equipment operates continuously under nominal conditions, causing failure instances to represent less than or about 5% of total digital monitoring logs [6]. Machine learning algorithms trained on such severely imbalanced class distributions exhibit a heavy bias toward the majority class of normal operation, yielding deceptively high overall accuracy while failing to detect critical, low-frequency failure events, a phenomenon known as the "accuracy paradox" [7]. In industrial settings, this imbalance can potentially lead to socio-technical friction. A predictive model that routinely generates false alarms risks triggering alarm fatigue among maintenance technicians, potentially decreasing user trust in automated systems and causing workers to ignore critical diagnostic warnings [8]. Conversely, missed failures severely disrupt production schedules and demoralize maintenance personnel.

To resolve class imbalance while maintaining workforce trust, recent studies have explored synthetic resampling and feature enrichment techniques [9]. The Synthetic Minority Over-sampling Technique (SMOTE) has proven effective in balancing minority failure distributions across diverse industrial domains [9, 10]. When combined with ensemble architectures such as Random Forest (RF) [11, 12] and Extreme Gradient Boosting (XGBoost) [13, 14], SMOTE significantly enhances diagnostic sensitivity and minority failure recognition. Furthermore, empirical research highlights that gradient boosting algorithms effectively capture complex, non-linear failure patterns while balancing precision and recall, thereby significantly reducing false alarm rates [15].

However, there is still a gap in combining SMOTE-balanced ensemble models with physical machinery metrics for effective socio-technical decision-making. Existing data-driven approaches often treat feature extraction purely as a statistical exercise, overlooking essential physical degradation indicators such as thermal gradient and mechanical power output [13]. Incorporating domain-informed physical principles into predictive models provides interpretable boundaries that bridge the gap between algorithmic predictions and human decision-making on the factory floor [1].

This paper fills this gap by introducing an integrated machine learning framework evaluated on multi-sensor operational records from the UCI AI4I 2020 Predictive Maintenance Dataset [6]. The primary objectives of this study are: (i) to construct domain-engineered physical parameters, specifically Temperature Gradient (ΔT) and Mechanical Power Output (P), to enrich feature representation; (ii) to systematically rebalance training distributions via SMOTE resampling without data leakage; and (iii) to perform a comparative performance evaluation between SMOTE-enhanced Random Forest, XGBoost, and baseline resampling/weighting schemes on an independent holdout test dataset ($N = 2{,}000$) with emphasis on false alarm reduction and human-centric operational efficiency.

METHODOLOGY

The structured workflow of this study is illustrated in Figure 1. The operational pipeline moves from data acquisition through preprocessing, physics-informed feature engineering, SMOTE resampling, model training, and performance evaluation.

Data and Preprocessing

The study utilizes the UCI AI4I 2020 Predictive Maintenance Dataset comprising 10,000 operational monitoring records designed to reflect real industrial machinery conditions [6]. Non-informative identifier columns (Unique Device Identifier [UDI] and Product ID) as well as failure-mode indicators (Tool Wear Failure [TWF], Heat Dissipation Failure [HDF], Power Failure [PWF], Overstrain Failure [OSF], and Random Failures [RNF]) were explicitly removed from the input feature matrix to prevent target-derived variable leakage. The binary operational state (*Target*) served as the sole classification target (0 representing nominal operation, 1 representing failure state).

To prevent data leakage, all data preprocessing transformations, including Z-score standardization parameters and one-hot encoders for categorical attributes (*Product Type*), were fitted strictly on the training partition (80%, $N = 8{,}000$) and applied to transform the holdout test set (20%, $N = 2{,}000$). A fixed random seed

(random_state=42) was applied across all data partitioning, resampling, and algorithm initialization steps to ensure full computational reproducibility.

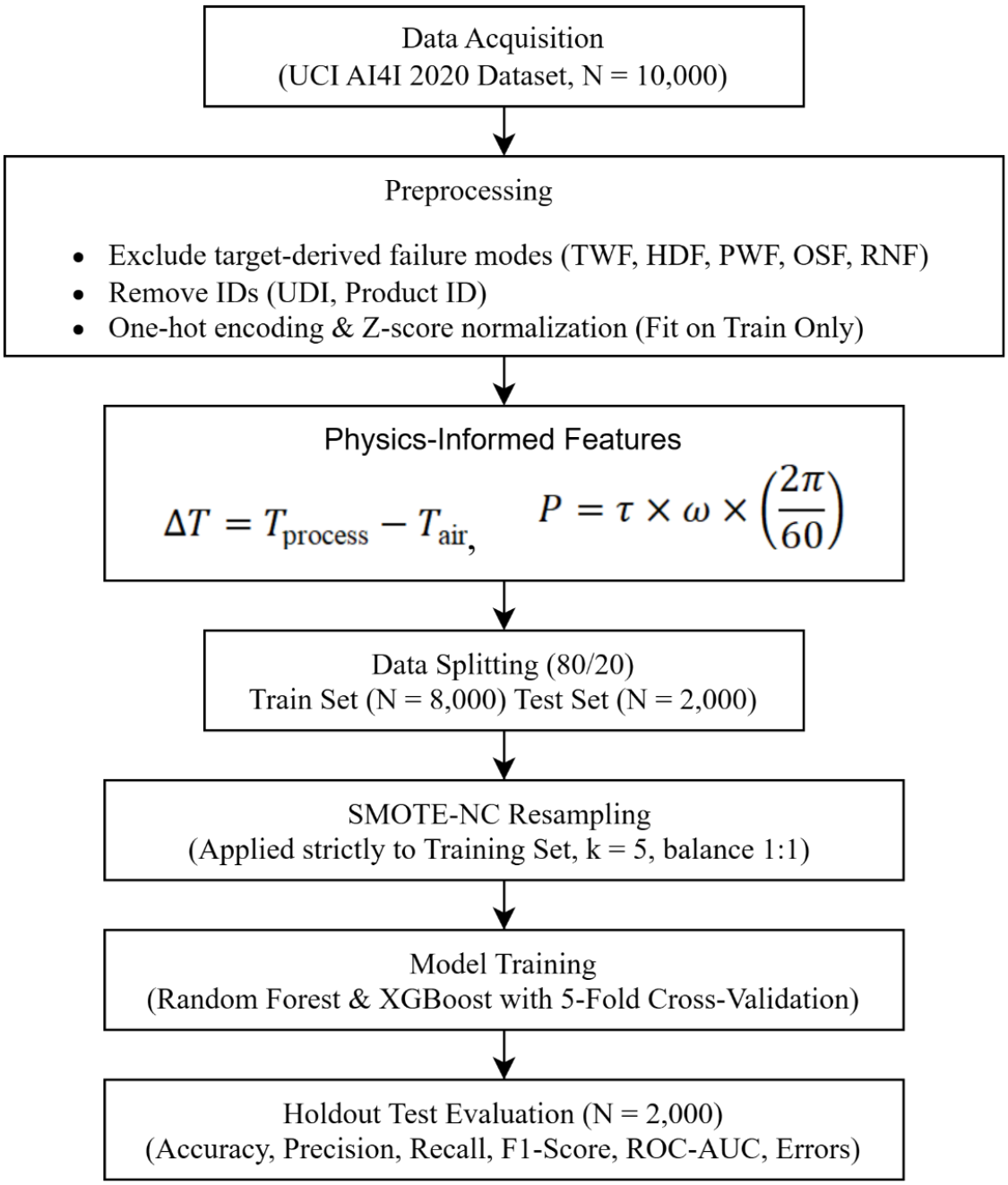

Figure 1. Technical and socio-technical workflow for predictive maintenance framework

Physics-Informed Feature Engineering

To provide clear physical context for decision makers, two domain-specific physical parameters were mathematically derived. The first parameter is the Temperature Gradient (ΔT), calculated as:

$$\Delta T = T_{process} - T_{air} \quad (1)$$

where $T_{process}$ represents internal process temperature in Kelvin and T_{air} represents ambient room temperature in Kelvin. A narrowing temperature difference (ΔT) indicates thermodynamic degradation and heat accumulation due to friction or cooling system decline, serving as a direct physical precursor to Heat Dissipation Failure (HDF).

The second parameter is Mechanical Power Output (P), expressed as:

$$P = \tau \times \omega \times \left(\frac{2\pi}{60}\right) \qquad (2)$$

where τ denotes rotational torque in Newton-metres (N · m) and ω is rotational speed in revolutions per minute (RPM). Mechanical power (P) reflects instantaneous motor loading ($\tau \cdot \omega$), directly capturing power spikes and overstrain conditions associated with Power Failure (PWF) and Overstrain Failure (OSF).

SMOTE Resampling and Model Training

During model training, 5-fold cross-validation was conducted on the training set to optimize hyperparameters. For Random Forest, key hyperparameters were set to n_estimators=200, max_depth=12, min_samples_split=5, and random_state=42. For XGBoost, hyperparameters were set to n_estimators=200, max_depth=6, learning_rate=0.05, subsample=0.8, colsample_bytree=0.8, reg_alpha=0.1, reg_lambda=1.0, and random_state=42. Evaluation relied on Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics on the unseen test set ($N = 2{,}000$), with specific focus on false positive reduction and standard deviation metrics across validation folds.

RESULTS

The quantitative empirical performance comparison across different resampling strategies, baseline classifiers, and the proposed framework on $N = 2{,}000$ unseen operational test logs are presented in Table 1.

Table 1. Quantitative performance benchmark across resampling and weighting strategies on unseen test dataset ($N = 2,000$)

Model	Strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
XGBoost	Unbalanced Baseline	97.1 ± 0.2	78.6 ± 2.1	32.4 ± 2.5	45.8 ± 2.1	0.812 ± 0.011
XGBoost	Class-Weighted	96.8 ± 0.3	52.3 ± 1.8	76.5 ± 2.0	62.1 ± 1.6	0.865 ± 0.009
XGBoost + Random Under-Sampling (RUS)	RUS	93.4 ± 0.5	31.5 ± 1.5	82.4 ± 2.2	45.6 ± 1.8	0.871 ± 0.010
XGBoost + Random Over-Sampling (ROS)	ROS	97.6 ± 0.3	62.2 ± 1.6	75.0 ± 1.9	68.0 ± 1.4	0.878 ± 0.008
Random Forest + SMOTE	SMOTE	97.4 ± 0.4	58.7 ± 1.5	79.4 ± 1.8	67.5 ± 1.3	0.887 ± 0.007
XGBoost + SMOTE (Proposed)	SMOTE (Proposed)	98.2 ± 0.3	71.1 ± 1.2	79.4 ± 1.5	75.1 ± 1.1	0.891 ± 0.008

As detailed in Table 1, models trained on unbalanced data or simple under sampling (RUS) suffer from either low recall (32.4%) or poor precision (31.5%). While both SMOTE-enhanced Random Forest and XGBoost achieved high Recall (79.4%), proposed XGBoost + SMOTE achieved a significantly higher Precision of 71.1% compared to 58.7% for Random Forest. This precision advantage improved the F1-Score of XGBoost to 75.1% (versus 67.5% for Random Forest) and yielded an ROC-AUC of 0.891. Figure 2 illustrates the relative performance comparison between the two models. To evaluate classification errors from an operational management standpoint, confusion matrices for both classifiers are detailed in Figure 3.

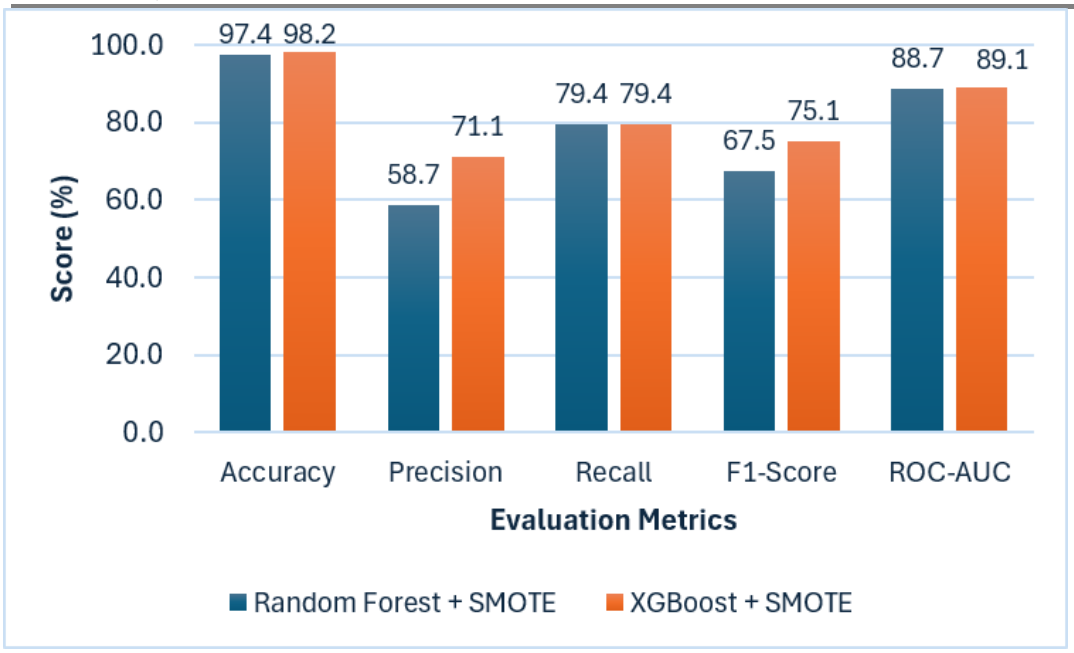

Figure 2. Performance metrics comparison between Random Forest + SMOTE and XGBoost + SMOTE models.

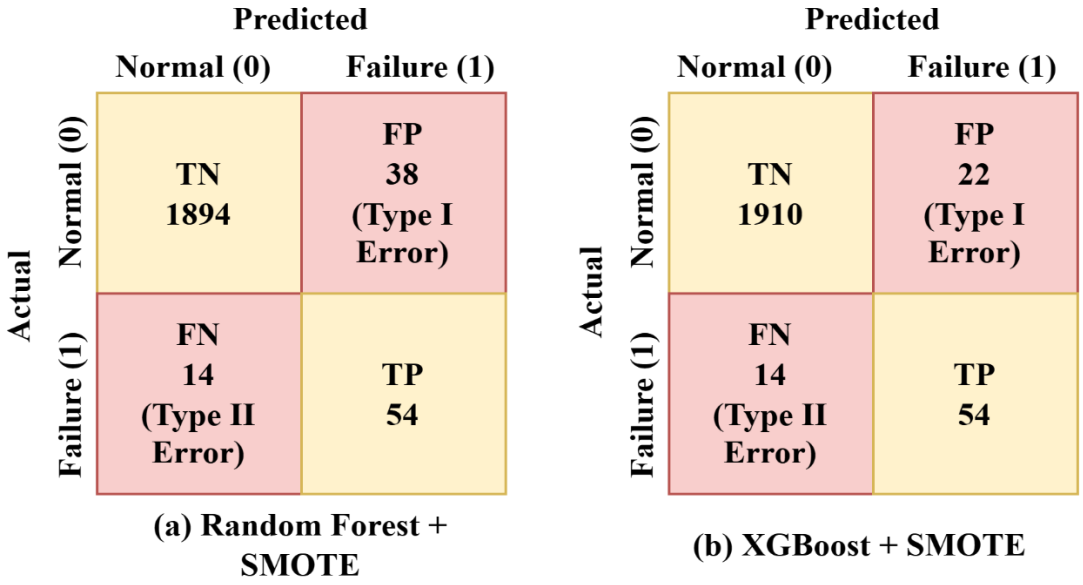

Figure 3. Confusion matrix comparison demonstrating classification outcomes and error-type distribution for (a) Random Forest + SMOTE and (b) XGBoost + SMOTE (N = 2,000)

DISCUSSION

The empirical results yield vital socio-technical insights for smart manufacturing operations and human-technology interaction. Relying solely on overall accuracy in imbalanced industrial data is dangerous, as high accuracy frequently masks total model inability to detect rare failure events. By applying SMOTE resampling without data leakage, both ensemble classifiers successfully learned minority failure features, capturing 79.4% of actual machinery faults before functional breakdown occurred.

The primary operational value of this framework lies in the substantial reduction of false alarms achieved by XGBoost. In industrial plant environments, false alarms can impose severe potential human and economic costs. Frequent false alarms are hypothesized to induce alarm fatigue, potentially causing technicians to distrust AI recommendations, spend labour hours on unneeded inspections, and disrupt scheduled production workflows.

XGBoost reduced false alarms from 38 to 22 (a 42.1% reduction) due to its built-in L1 (L_1) and L2 (L_2) regularisation parameters (reg_alpha, reg_lambda), which penalize model complexity and smooth decision boundaries against synthetic sample noise.

Furthermore, feature importance analysis confirmed that Mechanical Power Output (P) and Temperature Gradient (ΔT) served as primary physical predictors of failure. Incorporating explicit thermodynamic and mechanical variables into data-driven decision tools provides clear physical rationale for predictive alerts. This transparency fosters worker trust, enhances decision confidence, and supports human-AI collaboration on the shop floor. Mitigating unnecessary diagnostic interventions while ensuring asset protection directly supports sustainable production standards aligned with UN Sustainable Development Goals (SDG 9: Industry, Innovation, and Infrastructure; SDG 12: Responsible Consumption and Production).

CONCLUSION

This study introduced an integrated, socio-technical machine learning framework combining physics-informed feature engineering with SMOTE resampling to address extreme class imbalance in industrial predictive maintenance. Evaluated on an independent holdout test dataset ($N = 2{,}000$), both Random Forest and XGBoost achieved 79.4% Recall, capturing approximately four out of five incipient machinery failures. XGBoost demonstrated superior decision capability, attaining 98.2% Accuracy, 71.1% Precision, 75.1% F1-Score, and an ROC-AUC of 0.891, while reducing false alarms by 42.1% compared to Random Forest. Derived physical attributes, specifically Mechanical Power Output and Thermal Gradient, proved critical in providing transparent, physical failure context that enhances human-machine trust. Future research will examine shop-floor user acceptance, human-in-the-loop validation, and real-time edge deployment for sustainable asset management.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to Universiti Teknikal Malaysia Melaka (UTeM) for research facilities and support, as well as the Centre for Research and Innovation Management (CRIM), UTeM, for publication funding support.

REFERENCES

1. Hu, Y., Jia, Q., Yao, Y., Lee, Y., Lee, M., Wang, C., Zhou, X., Xie, R. and Yu, F. R. 2024: Industrial Internet of Things Intelligence Empowering Smart Manufacturing: A Literature Review. IEEE Internet of Things Journal 11(11). 19143–19167. DOI : http://dx.doi.org/10.1109/JIOT.2024.3367692
2. Firdaus, N., Ab-Samat, H. and Prasetyo, B. T. 2023: Maintenance Strategies and Energy Efficiency: A Review. Journal of Quality in Maintenance Engineering 29(3). 640–665. DOI : http://dx.doi.org/10.1108/JQME-06-2021-0046
3. Fordal, J. M., Schjølberg, P., Helgetun, H., Skjermo, T. O. and Wang, C. 2023: Application of Sensor Data Based Predictive Maintenance and Artificial Neural Networks to Enable Industry 4.0. Advances in Manufacturing 11. 248–263. DOI : http://dx.doi.org/10.1007/s40436-022-00433-x
4. Khan, M., Khan, M. A., Moser, B., Rafique, W., Xu, X. and Dai, W. 2026: AI-Driven Predictive Maintenance in Industrial IoTs: A Comprehensive Survey. IEEE Internet of Things Journal 13. 1–20. DOI : http://dx.doi.org/10.1109/JIOT.2026.3660780
5. Ayyat, M., Osman, M. and Nadeem, T. 2025: Opportunities and Challenges of Foundation Models in Industrial Manufacturing. IEEE Access 13. 138745–138775. DOI : http://dx.doi.org/10.1109/ACCESS.2025.3589777
6. Ileri, U., Altun, Y. and Narin, A. 2026: From Data to Fault Diagnosis: Using Predictive Maintenance Image Transformation and Nature-Inspired Optimization Techniques for Fault Classification. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 240(3). DOI : https://doi.org/10.1177/1748006X261430835
7. Manga, E., Marinagi, C., Skourlas, C., Karanikolas, N. and Markopoulos, D. 2026: Handling Small and Imbalanced Datasets in Government Social Media: A Comparative Study of Machine Learning and Deep Learning Models. SSRN Electronic Journal. DOI : http://dx.doi.org/10.2139/ssrn.6097171
8. Dangut, M. D., Skaf, Z. and Jennions, I. 2022: Handling Imbalanced Data for Aircraft Predictive

Maintenance Using the BACHE Algorithm. Applied Soft Computing 123. 108924. DOI : http://dx.doi.org/10.1016/j.asoc.2022.108924

9. Elreedy, D., Atiya, A. F. and Kamalov, F. 2024: A Theoretical Distribution Analysis of Synthetic Minority Oversampling Technique (SMOTE) for Imbalanced Learning. Machine Learning 113(7). 4903–4923. DOI : https://doi.org/10.1007/s10994-022-06296-4
10. Wang, J. and Awang, N. 2025: A Novel Synthetic Minority Oversampling Technique for Multiclass Imbalance Problems. IEEE Access 13. 6054–6066. DOI : http://dx.doi.org/10.1109/ACCESS.2025.3526673
11. Prasojo, R. A., Putra, M. A. A., Apriyani, M. E., Rahmanto, A. N., Ghoneim, S. S., Mahmoud, K. and Darwish, M. M. 2023: Precise Transformer Fault Diagnosis via Random Forest Model Enhanced by Synthetic Minority Over-Sampling Technique. Electric Power Systems Research 220. 109361. DOI : http://dx.doi.org/10.1016/j.epsr.2023.109361
12. Ghaleb, F. A., Saeed, F., Al-Sarem, M., Qasem, S. N. and Al-Hadhrami, T. 2023: Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection. IEEE Access 11. 89694–89710. DOI : http://dx.doi.org/10.1109/ACCESS.2023.3306621
13. Sampath, P., Elangovan, G., Ravichandran, K., Shanmuganathan, V., Pasupathi, S., Chakrabarti, T. and Margala, M. 2024: Robust Diabetic Prediction Using Ensemble Machine Learning Models with Synthetic Minority Over-Sampling Technique. Scientific Reports 14(1). 28984. DOI : http://dx.doi.org/10.1038/s41598-024-78519-8
14. Hassan, M., Nasiri, H., Esmaeili, M. and Dorrigiv, M. 2025: Stroke Prediction Using Synthetic Minority Over-Sampling Technique and Extreme Gradient Boosting. Computer Methods in Biomechanics and Biomedical Engineering 28(2). 1–12. DOI : https://doi.org/10.1080/10255842.2025.2570510
15. Kim, M. and Hwang, K.-B. 2022: An Empirical Evaluation of Sampling Methods for the Classification of Imbalanced Data. PLOS ONE 17(7). e0271260. DOI : http://dx.doi.org/10.1371/journal.pone.0271260