SETSCI Conference
Proceedings
22, 20-26, 2025

International Symposium on AI-Driven
Engineering Systems

June 19-20, 2025, Tokat, Türkiye

https://doi.org/10.36287/setsci.22.21.001
2687-5527/ © 2015 The Authors. Published by SETSCI

A Comparative Evaluation of Interpolation and Generative Oversampling Techniques for Predictive Maintenance

Abdazeez Atere, Hasan Kivrak

Northumbria University, Newcastle, UK {abdazeez.atere, hasan.kivrak}@northumbria.ac.uk

Abstract: Predictive maintenance (PdM) enhances industrial operational efficiency by facilitating timely detection of equipment failures using machine learning models developed from historical maintenance data. Real-world industrial datasets frequently exhibit significant class imbalance, as failures are infrequent occurrences. This imbalance substantially diminishes predictive accuracy for the minority class (failures). This study systematically evaluates three data augmentation techniques—Synthetic Minority Oversampling Technique (SMOTE), SMOTETomek, and Conditional Tabular Generative Adversarial Networks (CTGAN)—to address this challenge, utilising the AI4I 2020 Predictive Maintenance dataset. A Random Forest classifier was trained on augmented data, with a comparison of augmentation methods conducted through various performance metrics, including precision, recall, F1-score, ROC-AUC, and PR-AUC. The findings indicate that both SMOTE and SMOTETomek significantly enhance failure detection performance, with F1-scores and recall rates surpassing 0.99. In contrast, CTGAN demonstrates marginally lower classification performance (F1-score ≈ 0.88) while effectively generating realistic synthetic samples that maintain the original data distributions and inter-variable relationships. These results underscore the trade-offs between oversampling methods and generative models: SMOTE-based approaches optimise raw predictive accuracy for rare failures, whereas CTGAN demonstrates significant potential for improving model generalisation in complex industrial applications.

Keywords: —Predictive Maintenance, Class Imbalance, SMOTE, SMOTETomek, Generative Adversarial Networks, Random Forest, Data Augmentation.

I. INTRODUCTION

The handling and maintenance of technical systems in present-day businesses have been entirely revolutionised by the advancement of machine learning (ML) and data-driven methodologies. In today's data-driven industrial era, predictive maintenance has become a cornerstone of operational efficiency, enabling the proactive identification of system faults and optimising equipment longevity and reliability. Predictive maintenance is a proactive strategy that utilises real-time data to detect probable issues before to their occurrence, representing one of the most notable developments. Leveraging data from sensors, logs, and diagnostics, PdM systems aim to forecast potential failures before they escalate into costly downtimes or safety hazards. It can enhance the scheduling of maintenance activities, save expenses, augment equipment availability, and elevate operational safety [1].

Traditional maintenance strategies, such as reactive (run-to-failure) and preventive (planned) maintenance, often result in unnecessary costs or fail to prevent significant downtime effectively. In contrast, PdM utilises sensor-based monitoring and sophisticated machine learning algorithms to forecast and avert equipment failures by categorising fault types, evaluating the remaining useful life (RUL) of components, and identifying early indicators of anomalies, thus substantially lowering maintenance expenses and improving operational reliability [2].

However, a critical barrier to succesfull PdM implementation is limiting the accuracy and applicability of such systems is data scarcity and class imbalance, especially when it comes to critical failure events that occur infrequently [3]. Most real-world PdM datasets are dominated by healthy operational data, while failure cases particularly early and rare failure modes are significantly underrepresented [4]. This imbalance skews the training of machine learning (ML) models, which often develops a bias toward the majority class, leading to high false-negative rates. This mismatch limits the effectiveness of supervised machine learning models, as they are often overfit the majority class and exhibit poor performance in recognising rare yet critical failure patterns. Traditional classifiers thus tend to misclassify minority events, undermining the reliability of PdM systems and potentially resulting in equipment degradation, safety incidents, and financial losses [1] - [8].

To mitigate this issue, data augmentation techniques have emerged as vital tools for addressing class imbalance. These methods enhance the representation of minority class instances, thereby improving ML model sensitivity and predictive accuracy. Among these, Synthetic Minority Oversampling Technique (SMOTE) and its variant SMOTETomek are established approaches known for interpolating between existing minority class samples. More recently, deep learning-based Conditional Tabular Generative Adversarial Networks (CTGAN) have gained attention for their capability to synthesize realistic tabular data, offering promising generalization capabilities [9].

This study aims to evaluate these three data augmentation techniques, SMOTE, SMOTETomek, and CTGAN, using the AI4I 2020 PdM dataset. Specifically, we examine the effectiveness of these methods through comprehensive metrics including precision, recall, F1-score, ROC-AUC, and PR-AUC, providing critical insights for industry practitioners in selecting appropriate data augmentation strategies to enhance predictive maintenance outcomes.

II. LITERATURE REVIEW

Recent improvements in predictive maintenance (PdM) utilise machine learning techniques, including Random Forest, XGBoost, and deep learning models such as LSTM and autoencoders, to identify equipment breakdowns using sensor and operational data. Ensemble techniques such as Random Forest and Gradient Boosting are proficient in structured data contexts, but deep learning models effectively identify intricate temporal patterns in multivariate time series. Research indicates that the incorporation of data imbalance mitigation strategies, such as SMOTE and localised loss, significantly enhances failure prediction efficacy in practical industrial contexts [10],[6],[11],[12],[13],

Initial approaches to class imbalance involved algorithmic changes and data-level adjustments. Algorithmic techniques endeavor to modify the machine learning process to address the imbalance. Cost-sensitive learning, for instance, imposes penalties on classification errors of minority class samples during model training [14]. Alternative methodologies encompass ensemble techniques, which entail training many models on diverse balanced or resampled data subsets and subsequently amalgamating their predictions [15]. Algorithmic solutions are efficient but require considerable hyperparameter tuning and are contingent upon data distribution.

On the data level, resampling techniques have been widely adopted. Traditional approaches like random oversampling and undersampling often lead to issues such as overfitting or loss of critical data [16]. Reference [17] introduced SMOTE, an interpolation-based oversampling method that synthetically generates minority instances. While effective, SMOTE sometimes produces synthetic samples near class boundaries, potentially reducing classifier accuracy due to ambiguous decision boundaries. The hybrid approach SMOTETomek addresses this limitation by combining SMOTE with Tomek links, effectively cleaning noisy samples near decision boundaries [18].

Recent research has explored Generative Adversarial Networks (GANs), particularly Conditional Tabular GAN (CTGAN), for synthetic data generation [19]. CTGAN uses conditional generation to ensure realistic interrelationships between features, thus outperforming traditional oversampling techniques like SMOTE in terms of preserving feature distributions [19]. GAN-based methods are recognized for their potential to preserve complex feature relationships and generate highly realistic synthetic data, thus offering superior generalization capabilities for imbalanced classification problems [6]. Despite their theoretical advantages, GAN-based methods require careful hyperparameter tuning and computational resources, limiting widespread industrial adoption [9]. Nonetheless, its practical implementation is constrained by issues including training instability, mode collapse, hyperparameter sensitivity, and elevated computing expenses [9],[19]-[21]. Moreover, research demonstrates that high-quality synthetic data does not consistently enhance model performance, particularly in industrial contexts characterised by intricate and varied failure modes [20], [21].

While numerous studies have individually evaluated data augmentation techniques, comprehensive comparative evaluations, particularly within the PdM domain, remain limited on benchmark datasets that may not fully reflect the temporal and operational complexities of real industrial environments. More empirical studies using domain-specific datasets are required. Moreover, the comparative analysis of SMOTE, SMOTETomek, and GANs in a unified framework has not been explored yet. Few studies systematically compare these methods across multiple performance dimensions—accuracy, recall, robustness, and computational efficiency. This study aims to fill that gap by evaluating all three augmentation techniques on real-world PdM data, evaluating their effectiveness, and drawing practical insights for industry adoption.

III. METHODOLOGY

The methodology involves several key steps: data collection, preprocessing, augmentation, model training and evaluation.

A. Dataset Used

This study utilised the AI4I 2020 Predictive Maintenance Dataset, a synthetic dataset designed to replicate realistic industrial predictive maintenance conditions. It includes 10,000 records, each with ten attributes such as rotational speed, torque, temperature measurements, tool wear and other operational metrics. Five types of failures are labeled as tool wear failure, heat dissipation Failure, power failure, overstrain failure and random failures [22]. The dataset suffers from significant class imbalance, with failure events comprising a small fraction of total records.

Key features include:

- **Operational metrics** such as *rotational speed (rpm)*, *torque (Nm)*, *air temperature (K)*, *process temperature (K)*, and *tool wear (minutes)*.
- **Categorical identifiers** like *Product ID*, *Type*, and *Failure Type*.
- A **binary target variable** (*Target*) that indicates the presence (1) or absence (0) of a failure event.

Descriptive statistics reveal important characteristics of the dataset. The Target variable shows a mean of 0.0399, indicating that only around 3.99% of the records represent failure events. This severe class imbalance highlights the need for data augmentation techniques such as SMOTE, SMOTETomek, and GANs to improve minority class detection and model performance.

Furthermore, significant variability is observed in rotational speed, torque, and tool wear:

- Rotational speed ranges from 1158 to 2886 rpm with a standard deviation of 179.28, suggesting substantial fluctuations in operational dynamics.
- Torque values vary from 3.0 to 76.6 Nm, accompanied by a standard deviation of 9.97, indicating diverse load conditions across machines.

- Tool wear spans from 0 to 253 minutes, with a high standard deviation of 63.65 minutes, highlighting varied levels of tool degradation.

These wide ranges and high standard deviations indicate that operational behaviours are highly diverse across entries, making these features critical predictors for machine failure modelling. In contrast, air temperature (mean: 300.00 K, std: 2.00) and process temperature (mean: 310.09 K, std: 1.48) show minimal variation, reflecting stable environmental conditions.

Given the structure, variability, and imbalance of the dataset, it is well-suited for the objectives of this study, particularly the evaluation of augmentation methods aimed at improving predictive maintenance outcomes. The descriptive statistics confirm that while the dataset is clean and well-structured, the inherent imbalance and feature variability still present significant challenges.

TABLE 1. DATASET DESCRIPTION

Variables	**Measure**	**Data type**	**Count**	**Mean**	**Standard deviation (SD)**
UDI	Nil	Integer	1000	Nil	Nil
Product ID	Nil	Object	1000	Nil	Nil
Type	Low (L) Medium (M) High (H)	Object	1000	Nil	Nil
Air temperature	Kelvin (k)	Float	1000	300	2
Process temperature	Kelvin (k)	Float	1000	310	1.48
Rotational speed	Revolutions per minute (RPM)	Integer	1000	1538.78	179.28
Torque	Newton-meters (Nm)	Float	1000	39.99	9.97
Tool wear	Minutes (min)	Integer	1000	107.95	63.65
Target	0=No Failure; 1=Failure	Integer	1000	Nil	Nil
Failure Type	Heat Dissipation Failure Power Failure Overstrain Failure Tool Wear Failure Random Failure	Object	1000	Nil	Nil

B. Data Preparation

The dataset was examined utilising pandas tools, including *isnull()*. Utilise *sum()* to detect missing values and employ *describe()* in conjunction with *boxplot()* visualisations to identify probable outliers. No discrepancies were identified, affirming the integrity and consistency of the data for analysis. To address class imbalance, we implemented three augmentation techniques:

1. SMOTE generates synthetic samples of the minority class by interpolating between existing minority instances. This method equilibrates class distribution by generating new synthetic minority examples using interpolation between existing samples and their nearest neighbours [17].
2. SMOTETmek improves class differentiation by eliminating overlapping samples at decision boundaries and boosting the minority class through the integration of SMOTE and Tomek Link under-sampling [18].
3. CTGAN generates synthetic data that closely resembles the minority class distribution through conditional generative adversarial networks is particularly effective with mixed-type and high-dimensional tabular data [19].

C. Random Forest Classifier

We selected the Random Forest classifier for its interpretability, robustness to overfitting, and capability to handle complex and non-linear interactions [23]. To assess model performance, we employed the following metrics:

Precision measures the proportion of true positives among predicted positives. This is a crucial metric here, as in cases of highly imbalanced datasets, a model with high precision will indicate fewer false positives.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall (Sensitivity) measures the proportion of true positives among actual positives. High recall implies that the model can identify most of the instances from the minority class, making it a crucial metric here.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-Score computes the harmonic mean of precision and recall, providing a balance between the two. A high F1-score indicates that there are few false positives and false negatives.

$$\text{F1\_Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

ROC-AUC indicates the area under the Receiver Operating Characteristic curve, indicating the model's ability to distinguish between classes. The AUC ranges from 0 to 1, where a higher value indicates a good balance in predicting both classes.

Finally, PR-AUC measures the area under the Precision-Recall curve, especially informative for imbalanced datasets.

These criteria ensure that the evaluation process adequately represents the minority class (failures). The model was trained and evaluated on both the original and augmented datasets to assess the impact of data augmentation techniques on predictive performance.

IV. FINDINGS AND DISCUSSIONS

This section presents the experimental results obtained from evaluating the impact of data augmentation techniques, SMOTE, SMOTETomek, and CTGAN, on the predictive performance of a Random Forest classifier. The goal is to compare how these augmentation methods influence the model's ability to detect rare failure events within an imbalanced dataset.

The findings are organized to demonstrate the model's baseline performance using the original imbalanced dataset and to compare this with results obtained from models trained on synthetically balanced datasets.

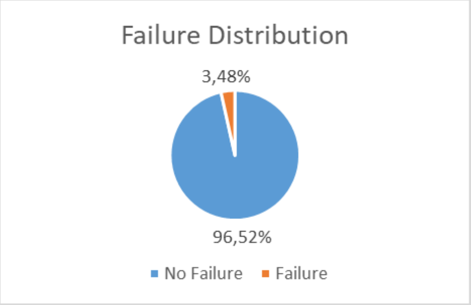
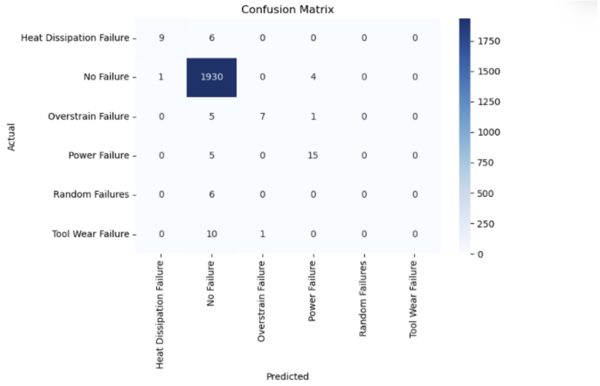


Fig. 1 Failure distribution. The pie chart shows that 96.52% of cases had no failure, while 3.48% experienced a failure, highlighting significant class imbalance in the dataset.

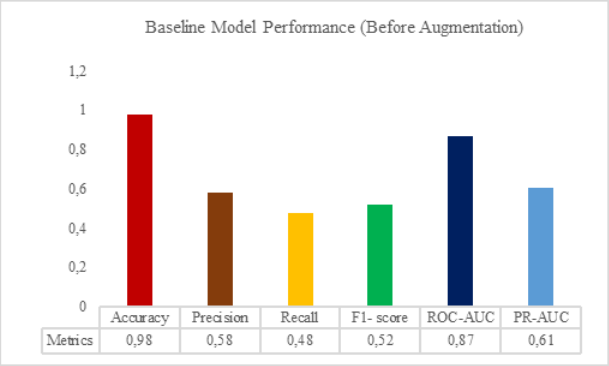
A. Baseline Model Performance (Before Augmentation)

Figure 2 illustrates the confusion matrix for the Random Forest model trained on the original imbalanced dataset. The Random Forest classifier trained on the imbalanced dataset demonstrated high overall accuracy of 0.98. However, the lower However, the precision and recall were notably lower (0.58 and 0.48, respectively), indicating the model's inability to correctly detect minority class instances. While common failure types were somewhat detectable, rarer categories such as Tool Wear and Random Failures had near-zero recall. The PR-AUC was 0.61, also indicating limited value in minority class predictions.



	Heat Dissipation Failure	No Failure	Overstrain Failure	Power Failure	Random Failures	Tool Wear Failure
Heat Dissipation Failure	9	6	0	0	0	0
No Failure	1	1930	0	4	0	0
Overstrain Failure	0	5	7	1	0	0
Power Failure	0	5	0	15	0	0
Random Failures	0	6	0	0	0	0
Tool Wear Failure	0	10	1	0	0	0

Fig. 2 Confusion Matrix for Imbalance Dataset (Before Augmentation)

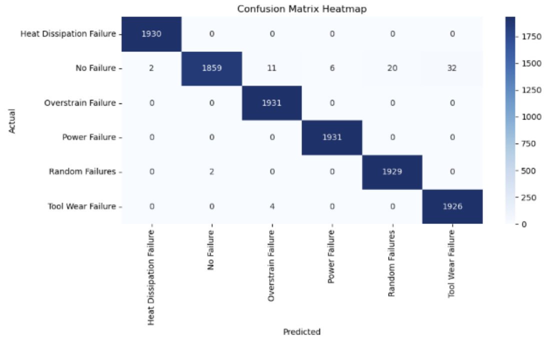


	Accuracy	Precision	Recall	F1- score	ROC-AUC	PR-AUC
Metrics	0,98	0,58	0,48	0,52	0,87	0,61

Fig. 3 Random forest result (Before Augmentation). Baseline model performance metrics before data augmentation, highlighting initial accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC.

B. Model Performance After SMOTE Augmentation

Figure 4 shows the confusion matrix after applying SMOTE augmentation to the dataset. Following SMOTE augmentation, the dataset was rebalanced to ensure that each failure category was equally represented with 9652 instances. Applying SMOTE led to a fully balanced dataset significantly enhancing recall and F1-score values for all failure types. Precision and recall scores exceeded 0.98, and both ROC-AUC and PR-AUC approached 0.99. Misclassification rates minority class instances such as Tool Wear Failure and Overstrain Failure were considerably reduced. Feature importance analysis shown in Figure 5 revelaled that torque, rotational speed, and tool wear remained the most influential features of failure as Environmental variables such as Air Temperature and Process Temperature demonstrated moderate importance. In conclusion, SMOTE proved highly effective for increasing sensitivity to rare events.



	Heat Dissipation Failure	No Failure	Overstrain Failure	Power Failure	Random Failures	Tool Wear Failure
Heat Dissipation Failure	1930	0	0	0	0	0
No Failure	2	1859	11	6	20	32
Overstrain Failure	0	0	1931	0	0	0
Power Failure	0	0	0	1931	0	0
Random Failures	0	2	0	0	1929	0
Tool Wear Failure	0	0	4	0	0	1926

Fig. 4 Confusion Matrix from Random Forest after SMOTE Augmentation. Confusion matrix from Random Forest after SMOTE augmentation, showing correctly classified failure types.

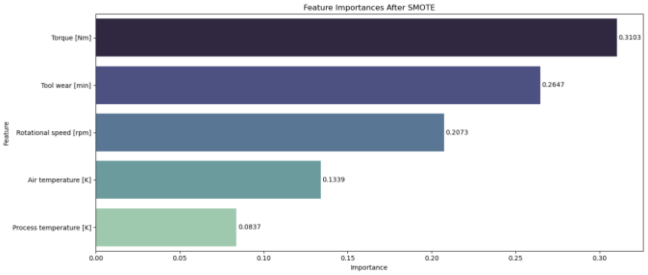


Fig. 5 Feature Importances. Feature importance ranking after SMOTE, highlighting Torque [Nm] as the most influential predictor in the dataset.

C. Model Performance After SMOTETomek Augmentation

Following SMOTETomek augmentation, the dataset was adjusted to achieve a near-balanced distribution across all failure classes. Specifically, each of the five failure types was represented with instance counts ranging between 9627 to 9652. With SMOTETomek, the model achieved similarly high classification metrics. Its hybrid nature of augmenting the minority class while removing noisy boundary instances resulted in slightly cleaner class separability and better generalization.

Precision, recall, and F1-score values exceeded 0.98, and ROC-AUC and PR-AUC remained close to 0.99. The corresponding confusion matrix in Figure 6 shows the improved separability across failure types. Feature importance anlysis remained aligned with SMOTE-augmented results.

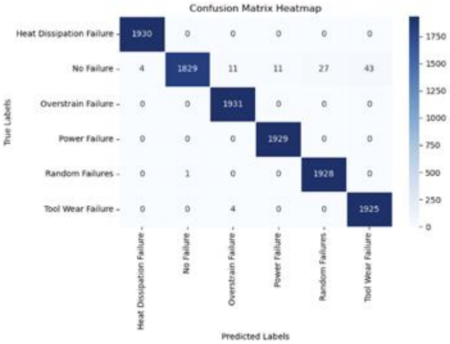


Fig. 6 Confusion Matrix from Random Forest after SMOTETomek Augmentation. Confusion matrix from Random Forest after SMOTETomek augmentation, showing correctly classified failure types.

D. Model Performance After CTGAN Augmentation

CTGAN synthetically generated high-fidelity samples for minority classes by modeling complex relationships between numerical and categorical features. This allowed for a more realistic and representative augmentation of the dataset. After CTGAN augmentation, each failure type achieved equal representation, with approximately 9,652 instances per class.

The use of CTGAN also improved model sensitivity to rare failure events. The classifier achieved an F1-score of 0.88 and PR-AUC of 0.91, slightly lower than SMOTE-based augmentation but provided the benefit of generating more realistic and diverse synthetic samples. Figure 7 presents the confusion matrix for the dataset augmented using CTGAN, highlighting diverse pattern capture across classes. Nevertheless, this diversity suggests better generalization capabilities, particularly in real-world scenarios where preserving complex inter-feature relationships is essential.

Fig. 7 Confusion Matrix from Random Forest after CTGAN Augmentation. Confusion matrix from Random Forest after CTGAN augmentation, showing correctly classified failure types.

E. Comparative analysis

A quantitative summary of the Random Forest model's performance metrics across the baseline and all augmentation methods is presented in Table 2. All three augmentation techniques improved model performance over the baseline. SMOTE and SMOTETomek achieved the highest classification metrics, with CTGAN showing slightly reduced numerical performance but offering stronger data diversity. While SMOTE is simple and effective, SMOTETomek's denoising offers additional generalization. CTGAN is particularly useful for its generative capabilities, especially when feature relationships are non-linear and multidimensional [19].

Overall, SMOTETomek proved to be the most effective technique based on the AI4I 2020 dataset, as shown in Figure 18, delivering the highest and most consistent performance across all evaluation criteria, while CTGAN offered promising generalisation capabilities suitable for complex real-world predictive maintenance applications.

These findings demonstrate the critical role of data augmentation in PdM model performance, particularly when detecting infrequent but consequential failure types. Selecting an augmentation method should depend on the specific operational context prioritizing raw performance, interpretability, or synthetic data fidelity.

TABLE 2. COMPARISON OF MODEL PERFORMANCE METRICS

Model	Accuracy	Precision	Recall	F1-Score
Baseline	0.98	0.58	0.48	0.58
SMOTE	0.99	0.99	0.99	0.99
SMOTETomek	0.99	0.99	0.99	0.99
CTGAN	0.88	0.88	0.88	0.88

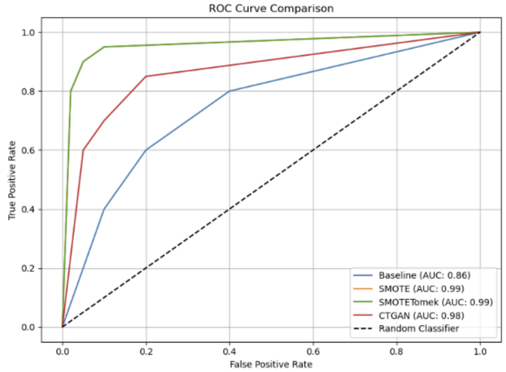


Fig. 8 ROC curve Comparison: Baseline vs SMOTE vs SMOTETomek VS CTGAN.

Additionally, Table 3 summarizes the ROC-AUC and PR-AUC values for each dataset condition, helping to evaluate the trade-offs in model performance and generalization capacity across different augmentation strategies. It highlights, for instance, that while SMOTE and SMOTETomek achieve near-perfect AUC scores, CTGAN provides strong generalization with slightly lower AUCs, and the imbalanced baseline performs weakest, especially in PR-AUC.

Despite the sophisticated nature of CTGAN, it did not surpass lean interpolation-based methods such as SMOTE and SMOTETomek in classification performance. This suggests that while CTGAN excels in generating complex and diverse synthetic data, traditional oversampling techniques remain highly effective for structured, tabular datasets used in predictive maintenance. In such domains, where feature distributions are often well-captured through standard measurements, the added complexity of GANs may not yield proportional gains in accuracy.

TABLE 3. PERFORMANCE COMPARISON INTERPRETATION IN TERMS OF THE ROC-AUC AND PR-AUC

Dataset Condition	ROC - AUC	PR - AUC	Strengths	Limitations
Imbalanced	0.86	0.60	High overall accuracy; useful baseline	Poor minority detection; biased towards majority class
SMOTE	0.99	0.99	Excellent recall and precision; effective minority coverage	May generate borderline synthetic samples near class overlap
SMOTETomek	0.99	0.99	Balanced learning with noise reduction; refined decision boundaries	Slightly lower PR-AUC than SMOTE; higher computational cost
CGAN-Augmented	0.98	0.91	Strong generalization; realistic data synthesis via deep learning	Slightly less effective than SMOTE/SMOTE Tomek; tuning required

V. CONCLUSION AND RECOMMENDATIONS

This study presented a comparative evaluation of three data augmentation techniques, specifically SMOTE, SMOTETomek, and CTGAN for addressing class imbalance in predictive maintenance applications using the AI4I 2020 dataset and Random Forest classifier. The findings revealed the importance of data balancing in improving model sensitivity to rare failure events. Among the evaluated methods, SMOTE and SMOTETomek delivered the highest classification performance across key metrics such as precision, recall, F1-score, ROC-AUC, and PR-AUC. CTGAN**,** while not outperforming interpolation-based methods in raw metrics, demonstrated potential in generating realistic and diverse synthetic samples that preserve feature relationships.

There are drawbacks despite these advantages. The dependence on Random Forest, while interpretable and effective, may not fully leverage temporal or nonlinear patterns present in high-frequency sensor data. Synthetic data techniques, such as GANs, while potent, may produce implausible patterns when trained on a restricted number of minority class instances. Nonetheless, the paper provides a reasonable contribution by empirically evaluating the trade-offs among various augmentation strategies and delivering practical recommendations for constructing more balanced and successful prediction models in maintenance applications.

Future research could investigate the application of next-generation GAN architectures, such as Wasserstein GANs (WGANs) or Attention-based GANs, to further improve synthetic data quality and training stability. Furthermore, hybrid augmentation approaches that combine SMOTETomek boundary cleaning with CTGAN generative capabilities may yield superior results by leveraging the strengths of both oversampling and generative approaches. Additional research could also focus on multivariate time-series data augmentation and online learning models, enabling real-time predictive maintenance in dynamically evolving industrial settings. A broader exploration across diverse industrial datasets would further validate the robustness and scalability of these augmentation strategies.

REFERENCES

[1] Altalhan, M., Algarni, A. and Alouane, M.T.H. (2025) 'Imbalanced data problem in machine learning: A review', *IEEE Access*, 11. doi:10.1109/ACCESS.2025.3531662.

[2] Batista, G.E.A.P.A., Prati, R.C. and Monard, M.C. (2004) 'A study of the behavior of several methods for balancing machine learning training data', *SIGKDD Explorations Newsletter*, 6(1), pp. 20–29.

[3] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research*, 16, pp. 321–357.

[4] Chen, C., Liaw, A. and Breiman, L., 2004. Using random forest to learn imbalanced data. University of California, Berkeley.

[5] Chen, Irene & Joshi, Shalmali & Ghassemi, Marzyeh. (2020). Treating health disparities with artificial intelligence. Nature Medicine. 26. 16-17. 10.1038/s41591-019-0649-2.

[6] Davis, J. and Goadrich, M., 2006. The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd international conference on Machine learning*, pp.233–240.

[7] Drummond, C. and Holte, R.C., 2003. C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. Workshop on Learning from Imbalanced Data Sets II.

[8] Elkan, C., 2001. The foundations of cost-sensitive learning. In Proceedings of the 17th international joint conference on Artificial intelligence (Vol. 1, pp. 973-978).

[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014) 'Generative adversarial nets', *Advances in Neural Information Processing Systems*, 27, pp. 2672–2680.

[10] Hakami, A., 2024. Strategies for overcoming data scarcity, imbalance, and feature selection challenges in machine learning models for predictive maintenance. *Scientific Reports*, 14(9645).

[11] ICO, 2022. *Explaining decisions made with AI*. [online] Available at: https://ico.org.uk [Accessed 14 Apr. 2025].

[12] Japkowicz, N. and Stephen, S., 2002. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5), pp.429–449.

[13] Jardine, A.K.S., Lin, D. and Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), pp.1483–1510.

[14] Kulkarni, A., Batarseh, F.A. and Chong, D. (2023) 'Foundations of data imbalance and solutions for a data democracy', in *Foundations of Data Science for Engineers*. Springer, pp. 135–157.

[15] Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. R News, 2(3), pp.18-22.

[16] Liu, R., Yang, B., Zio, E. and Chen, X. (2018) 'Artificial intelligence for fault diagnosis of rotating machinery: A review', *Mechanical Systems and Signal Processing*, 108, pp. 33–47.

[17] Mahale, Y., Kolhar, S. and More, A.S. (2025) 'Enhancing predictive maintenance in automotive industry: addressing class imbalance using advanced machine learning techniques', *Discover Applied Sciences*, 7:340. doi:10.1007/s42452-025-06827-3.

[18] Matzka, S. (2020). AI4I 2020 Predictive Maintenance Dataset. UCI Machine Learning Repository.

[19] Mobley, R.K., 2002. An Introduction to Predictive Maintenance. Elsevier.

[20] Saito, T. and Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3), p.e0118432.

[21] Sipos, R., Fradkin, D., Moerchen, F. and Wang, Z., 2014. Log-based predictive maintenance.

[22] Wang, Zhe & Wu, Chunhua & Zheng, Kangfeng & Niu, Xinxin & Wang, Xiujuan. (2019). SMOTETomek-based Resampling for Personality Recognition (July 2019). IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2940061.

[23] Xu, L., Skoularidou, M., Cuesta-Infante, A. and Veeramachaneni, K., 2019. Modeling Tabular Data Using Conditional GAN. *Advances in Neural Information Processing Systems*, 32

[24] Yoon, J., Jarrett, D. and van der Schaar, M., 2020. Anonymization through data synthesis using generative adversarial networks (ADS-GAN). *IEEE Journal of Biomedical and Health Informatics*, 24(8), pp.2378–2388.

[25] Zhang, W., Yang, D. and Wang, H. (2019) 'Data-driven methods for predictive maintenance of industrial equipment: A survey', *IEEE Systems Journal*, 13(3), pp. 2213–2227.