© January 2024| IJIRT | Volume 10 Issue 8 | ISSN: 2349-6002

Anomaly Resilient Balancer: A Pioneering Integration of Anomaly Detection and Class Imbalance Mitigation for Enhanced Predictive Maintenance Modeling

[1]Prof. Veena R.Pawar, [2]Dr.Dev Ras Pandey
[1,2]Kalinga University, Raipur

Abstract-Predictive maintenance plays a pivotal role in industrial settings, relying extensively on machine learning for the timely detection of equipment failures. However, persistent challenges arise from imbalances and anomalies present in the datasets. Current solutions tend to concentrate on singular aspects, anomaly removal or class imbalance correction without addressing the intricate interplay between the two. This paper introduces the Anomaly Resilient Balancer algorithm, a novel and specialized approach designed to create a balanced dataset free from anomalies. The algorithm leverages robust deviation-aware metrics, employing median absolute deviation and median values to distinguish anomalies from normal instances effectively. This paper not only surpasses existing methods in the context of predictive maintenance datasets but also introduces a groundbreaking, integrated solution for anomaly-resilient balancing. The Anomaly Resilient Balancer sets a new standard for robust machine learning models in real-world applications, showcasing its effectiveness in addressing the complexities of industrial predictive maintenance through a comprehensive and integrated approach. This advancement marks a significant step in developing resilient and reliable machine-learning models for critical applications.

Keywords: Predictive Maintenance, Machine Learning, Anomaly Resilient Balancer Algorithm, Class Imbalance, Deviation-Aware Metrics.

1 INTRODUCTION

In industrial settings, adopting predictive maintenance has become imperative for ensuring equipment's continuous and efficient operation [1]. This paradigm relies extensively on machine learning techniques to detect potential equipment failures before they occur, minimizing downtime and optimizing maintenance efforts [2]. However, the effectiveness of predictive maintenance models is often hampered by challenges inherent in the datasets, particularly the presence of anomalies and imbalances [3], [4].

The datasets used in predictive maintenance scenarios frequently exhibit anomalies, representing irregularities or unexpected events in the equipment's behaviour [5].

The drawbacks of current approaches further underscore the need for a more comprehensive solution. One of the significant limitations lies in the limited integration of strategies, where existing methods predominantly focus on anomaly removal or class imbalance correction. This narrow scope fails to provide a holistic solution that considers the intricate interdependence of these challenges. Consequently, the resulting models may lack the resilience required for robust predictive maintenance.

The Anomaly Resilient Balancer algorithm makes noteworthy contributions through its holistic approach, concurrently addressing anomaly detection and class imbalance challenges.

- Holistic Approach: The Anomaly Resilient Balancer algorithm presents a comprehensive approach by simultaneously addressing anomaly detection and class imbalance challenges, ensuring the creation of a balanced dataset representative of both classes.
- Robust Deviation-Aware Metrics: The algorithm employs novel deviation-aware metrics, such as median absolute deviation and median values, to discern anomalies effectively, enhancing the robustness of predictive maintenance models.
- Superior Performance: Comparative experiments demonstrate the excellent performance of the Anomaly Resilient Balancer algorithm, highlighting its innovative strategy in tackling the complexities of anomaly detection and class imbalance.

- Integrated Solution: This paper introduces a groundbreaking, integrated solution for anomaly-resilient balancing, setting a new standard for robust machine learning models in real-world predictive maintenance applications.

This research aims to advance the development of resilient and reliable machine-learning models for predictive maintenance in industrial settings. The Anomaly Resilient Balancer algorithm is designed for broad applicability across diverse industrial domains, ensuring its utility in addressing the complexities associated with anomaly detection and class imbalance.

2 RELATED WORKS

In recent years, anomaly detection and class imbalance mitigation have witnessed a surge in research endeavours. Various authors have proposed innovative techniques to address these challenges in different domains. The following paragraphs provide detailed insights into the methodologies proposed by each author.

Xu et al. [6] introduced a data-driven intrusion and anomaly detection approach in the Internet of Things (IoT). Their methodology employs automated machine learning techniques to discern irregularities in IoT data. The study focuses on leveraging sophisticated algorithms to enhance anomaly detection accuracy, particularly in the context of IoT devices.

Maleki et al. [7] proposed an unsupervised anomaly detection method utilizing Long Short-Term Memory (LSTM) autoencoders with statistical data filtering. By employing deep learning architectures, the authors aimed to capture intricate patterns in data and identify anomalies without requiring labelled training data.

Ripan et al. [8] presented a data-driven heart disease prediction model incorporating K-means clustering-based anomaly detection. The authors utilized clustering techniques to identify abnormal patterns in heart-related data, contributing to the early prediction of potential cardiovascular issues.

Liu et al. [9] proposed a data mining-based framework for identifying daily electricity usage patterns and anomaly detection in building electricity consumption data. Their approach aimed to enhance the understanding of daily energy consumption patterns and accurately detect anomalies in the context of building management.

Perales Gómez et al. [10] introduced Madics, a methodology for anomaly detection in industrial control systems. Their work focused on developing a systematic approach tailored to the unique challenges presented by industrial settings, emphasizing the need for specialized techniques in anomaly detection.

While these works make valuable contributions to anomaly detection and class imbalance mitigation, a standard limitation is the lack of integration between anomaly removal and class imbalance correction. Existing methods tend to focus on either aspect in isolation, leading to suboptimal solutions that may either miss anomalies or fail to address class imbalances adequately.

To overcome these limitations, our proposed Anomaly Resilient Balancer algorithm takes a holistic approach by concurrently addressing anomaly detection and class imbalance challenges.

3 ANOMALY RESILIENT BALANCER

The Anomaly Resilient Balancer algorithm is a groundbreaking solution for predictive maintenance datasets, addressing challenges posed by anomalies and class imbalances. In the first phase of its workflow, deviation-aware metrics are computed, including the median and median absolute deviation, offering a robust foundation for subsequent anomaly detection.

In the final phase, the algorithm combines and evaluates the balanced datasets using a Support Vector Machine (SVM) classifier. The dataset with superior accuracy is selected for further utilization, emphasizing the algorithm's practical applicability and performance-driven approach. Overall, the Anomaly Resilient Balancer algorithm introduces a new standard for anomaly-resilient balancing, offering a comprehensive and integrated solution that outperforms existing methods in the complex landscape of predictive maintenance datasets. Algorithm 1 shows the Anomaly Resilient Balancer algorithm.

3.1 Deviation-Aware Metric Calculation

In the initial phase, the Anomaly Resilient Balancer algorithm calculates robust deviation-aware metrics to establish a solid foundation for subsequent anomaly detection. For each feature in the predictive maintenance dataset, the algorithm computes the median, a central tendency measure derived by sorting the elements in the feature in ascending order and selecting either the middle element or the average of the two middle elements, accommodating both odd and even numbers of elements. Simultaneously, the algorithm calculates each feature's median absolute deviation (MAD), offering a robust measure of the spread of data points around the median. The MAD of a feature (i) is determined by finding the median of the absolute deviations of individual data points (Xi) from their respective medians, which is shown in Eq. (1).

$$MAD[i] = Median(|Xi - Median[i]|) \quad (1)$$

Furthermore, the algorithm introduces additional measures for a more nuanced understanding of data distribution.

3.2 Robust Normalized Deviation (RND) Threshold Calculation

The algorithm calculates each feature's Robust Normalized Deviation (RND) threshold based on the deviation-aware metrics. It involves evaluating the RND scores of individual data points in a feature (i) by comparing them to the previously calculated median[i] and MAD[i]. The RND score ($\mathrm{RND_{score}}$[i]) for a given data point (D) in a specific feature (i) is determined by the formula

3.3 Anomaly Detection and Data Balancing

The algorithm initiates the anomaly detection and data balancing phase by iterating through each instance in the dataset. The RND score is calculated for every data point within an instance, and anomalies are identified based on a comparison with the RND threshold and lower/upper median values. Non-anomalous instances are then aggregated into a new dataset termed Non_anomalous_Data.

3.4 Class Imbalance Mitigation

To effectively address the inherent class imbalance in predictive maintenance datasets, the Anomaly Resilient Balancer algorithm strategically balances the distribution of instances within Non_anomalous_Data. Depending on whether the majority class indicates no machine failure (0) or machine failure (1), the algorithm tactically removes random samples from the majority class. It synthesizes new instances for the minority class. The synthesis of instances involves generating values for each feature in the synthetic instance using the Eq. (3):

D_synthetic[j] = r * (UpperMedian[j] – LowerMedian[j]) + LowerMedian[j] (3)

3.5 Dataset Combination and Model Evaluation

In the final steps, the algorithm combines the balanced datasets based on class distribution, creating two datasets (balancedDataset1 and balancedDataset2). The algorithm then evaluates the performance of both balanced datasets using a Support Vector Machine (SVM) classifier because it can handle classification and regression tasks efficiently.

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

This section delves into the empirical findings and discussions that underscore the efficacy of the Anomaly Resilient Balancer algorithm in predictive maintenance. The exploration commences with an in-depth elucidation of the dataset deployed for the experiments.

4.1 Dataset Description:

The AI4I 2020 predictive maintenance dataset, mimicking the operation of a milling machine, serves as the experimental bedrock. Comprising 10,000 data points and 14 features [16], including unique identifiers, product details, and operational parameters, the dataset introduces complexity through five distinct failure modes. These modes encompass tool wear failure, heat dissipation failure, power failure, overstrain failure, and random failures.

4.2 Performance Metrics:

In evaluating the predictive prowess of an Anomaly Resilient Balancer, a suite of performance metrics is employed for a comprehensive assessment. Accuracy, the bedrock metric measuring correctly classified instances, achieved an impressive 93.7960%, underscoring the algorithm's accuracy in identifying equipment failures and sustaining operational reliability. P

4.3 Accuracy Comparison:

A comparative analysis with existing methods reinforces the supremacy of the Anomaly Resilient Balancer. When juxtaposed against methods like DFPAIS, SDFIS, CatBoost, and their combinations, the proposed algorithm surpasses with an accuracy of 93.7960%, as shown in Table 1.

Table 1: Accuracy comparison

Author	Year	Method	Accuracy (%)
Kong et al. [17]	2023	DFPAIS (Data-filling approach based on probability analysis in incomplete soft sets)	83.74
Kong et al. [17]	2023	SDFIS (Simplified approach for data filling in incomplete soft sets)	82.17
Chen et al. [18]	2022	CatBoost (Categorical Boosting)	64.23
Chen et al. [18]	2022	SmoteNC + CatBoost (Synthetic Minority Over-Sampling Technique for Nominal and Continuous)	88.09
Chen et al. [18]	2022	ctGAN + CatBoost (Conditional Tabular Generative Adversarial Network)	87.08
Chen et al. [18]	2022	SmoteNC + ctGAN + CatBoost	88.83
Proposed Method (Anomaly Resilient Balancer)	2023	Anomaly Resilient Balancer	93.7960

5 CONCLUSION AND FUTURE WORK

In conclusion, the Anomaly Resilient Balancer algorithm is a pioneering and efficient solution in predictive maintenance for industrial applications. The algorithm demonstrates exceptional capabilities in identifying equipment failures and ensuring operational reliability through a meticulous exploration of deviation-aware metrics, a dynamic thresholding mechanism, and strategic class imbalance mitigation. The core strengths of the algorithm lie in its ability to intricately balance the interplay between anomaly detection and class imbalance, setting it apart from existing methods. By addressing these challenges in tandem, the Anomaly Resilient Balancer outperforms established approaches in predictive maintenance. It introduces a groundbreaking, integrated solution that sets a new standard for robust machine learning models in real-world applications.

REFERENCE

[1] Pech, M., Vrchota, J., & Bednář, J. (2021). Predictive maintenance and intelligent sensors in smart factories. Sensors, 21(4), 1470.

[2] Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability, 12(19), 8211.

[3] Kamat, P., & Sugandhi, R. (2020). Anomaly detection for predictive maintenance in industry 4.0-A survey. In E3S web of conferences (Vol. 170, p. 02007). EDP Sciences.

[4] Sridhar, S., & Sanagavarapu, S. (2021, September). Handling data imbalance in predictive maintenance for machines using SMOTE-based oversampling. In 2021, the 13th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 44-49). IEEE.

[5] Carrasco, J., López, D., Aguilera-Martos, I., García-Gil, D., Markova, I., Garcia-Barzana, M., ... & Herrera, F. (2021). Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms. Neurocomputing, 462, 440-452.

[6] Xu, H., Sun, Z., Cao, Y., & Bilal, H. (2023). A data-driven intrusion and anomaly detection approach using automated machine learning for the Internet of Things. Soft Computing, 27(19), 14469-14481.

[7] Maleki, S., Maleki, S., & Jennings, N. R. (2021). Unsupervised anomaly detection with LSTM autoencoders using statistical data filtering. Applied Soft Computing, 108, 107443.

[8] Ripan, R. C., Sarker, I. H., Hossain, S. M. M., Anwar, M. M., Nowrozy, R., Hoque, M. M., & Furhad, M. H. (2021). A data-driven heart disease prediction model through K-means clustering-based anomaly detection. SN Computer Science, 2, 1-12.

[9] Liu, X., Ding, Y., Tang, H., & Xiao, F. (2021). A data mining-based framework for identifying daily electricity usage patterns and anomaly detection in building electricity consumption data: Energy and Buildings, 231, 110601.

[10] Perales Gómez, Á. L., Fernández Maimó, L., Huertas Celdrán, A., & García Clemente, F. J. (2020). Madics: A methodology for anomaly detection in industrial control systems. Symmetry, 12(10), 1583.