BIG DATA ANALYTICS FOR EFFICIENT PREDICTIVE MAINTENANCE FOR INTELLIGENT AND SUSTAINABLE MANUFACTURING WITH OPTIMIZED ADABOOST

Shahida T[1*], Farhat Jummani[1], Ramakrishna Kumar[2]

Abstract: Predictive maintenance has gained popularity in intelligent and sustainable manufacturing because it may increase uptime, eliminate unscheduled downtime, and maximize resource use. This research proposes an improved AdaBoost algorithm and big data analytics for intelligent manufacturing system predictive maintenance. First, the proposed system gathers and preprocesses massive industrial data from sensors, IoT devices, and other sources. The meta-algorithm AdaBoost improves mediocre learners' performance, predicting machinery failure and degradation. Algorithmic optimization adjusts hyperparameters like iterations and learning rate to balance model accuracy with processing efficiency. The proposed model improves on previous work with 0.972 accuracy, 0.977 precision, 0.972 recall, and 0.974 F1-score. An optimized AdaBoost-enabled predictive maintenance framework manages big data analytics complexity in manufacturing in a scalable, cost-effective, and smart way. The framework uses AdaBoost's potential and optimization approaches to improve predictive maintenance tactics, intelligent manufacturing, and industrial sustainability.

Keywords: Optimized AdaBoost; Big Data Analytics; Predictive Maintenance; Intelligent Manufacturing; Sustainable Manufacturing; Quantization Techniques

[1]Department of Computer Science & Engineering, Shri Jagdish Prasad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan – 333001
[2]Department of Electrical and Communications Engineering, College of Engineering, National University, Muscat. Sultanate of Oman; Email; ramakrishnakumar@nu.edu.om

[*]**Corresponding Author:** Shahida T
[*]Department of Computer Science & Engineering, Shri Jagdish Prasad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan – 333001

DOI: 10.53555/ecb/2022.11.02.068

1. Introduction

Big Data analytics can transform industrial operations into smart, eco-friendly systems. Companies may enhance their manufacturing processes, gain new insights, and make better decisions by evaluating huge volumes of data across the production lifecycle. Intelligent manufacturing relies on sensors in machines and other equipment to monitor temperature, pressure, vibration, and more. All supply chain data must be integrated to optimize procurement, inventory management, and logistics. Sensor data can warn workers of equipment failures, enabling predictive maintenance [1]. This keeps machinery working smoothly, reduces repair costs, and extends their lifespan. Big data analytics can monitor the manufacturing process and analyze data to ensure high-quality goods [2]. Quality issues may be monitored and fixed rapidly. Monitoring and analyzing energy data may help factories save energy. Improvements in energy efficiency benefit more than just cost savings. Analyzing data from multiple production stages might identify inefficient actions and bottlenecks [3]. This data aids workflow optimization, resource utilization, and productivity. Big Data analytics helps manufacturers optimize supply chains by revealing demand projections, inventory management, and supplier performance. This may reduce waste and boost productivity [4]. Businesses can tailor products by analysing customer data and preferences, improving manufacturing efficiency and waste reduction. Big Data analytics can track and evaluate industrial process environmental impacts. For firms to accomplish sustainability goals, emissions, waste, and other issues must be monitored. Rising data quantities make data security and regulatory compliance more vital. Big Data analytics can be used to improve security and ensure manufacturing processes comply with industry standards. By analyzing data from several sources, manufacturers may continuously improve their processes, products, and green efforts. Big Data analytics promotes efficiency, decision-making, and sustainability by making production systems more adaptable and resilient [5].

Big Data analytics-based predictive maintenance aids intelligent and sustainable production. Predictive maintenance uses sensor data to determine when machinery needs repairs. Machinery faults can be predicted with predictive maintenance. Proactive problem solving can reduce unscheduled downtime and boost productivity for manufacturers. Manufacturers can arrange maintenance during planned downtime using historical and real-time data to avoid production schedule delays and maximize resource use. Early problem discovery and repair extend machine life. Avoiding catastrophic malfunctions can extend product life, reduce unnecessary equipment replacements, and increase sustainability [6]. Predictive maintenance reduces maintenance costs by preventing costly emergency repairs and decreasing time-based maintenance. It optimizes resource distribution, saving money over time. Well-maintained equipment performs better. Predictive maintenance can save power use and enhance the environment by keeping machines running efficiently. Manufacturers can better manage replacement component supply by anticipating maintenance needs. This saves the needless expense of stockpiling spare parts while ensuring emergency access to vital components. Big Data analytics allows real-time equipment health tracking. When conditions change, alerts are sent, allowing for quick action. Comparing historical and present data helps understand equipment performance. Data-driven insights help manufacturers choose preventative maintenance, budget allocation, and equipment upgrades. IoT and sensor data power predictive maintenance. These devices collect data in real time, enabling equipment health monitoring. Combine sensor data with Big Data analytics to improve predictions. Zero waste and low environmental impact are important to sustainable production. Predictive maintenance avoids unscheduled repairs and equipment disposal, helping the environment. Smart, eco-friendly production relies on big data analytics' predictive maintenance. Lowering waste and energy consumption improves operating efficiency, lowers costs, increases equipment lifespan, and promotes environmental sustainability [7].

Problem Formulation

When creating a predictive maintenance challenge for intelligent and sustainable manufacturing, goals, variables, restrictions, and criteria for mathematical models or data-driven approaches must be defined. A Big Data analytics-based predictive maintenance problem structure is as follows:

Objective:

1. Reduce Downtime and Production Loss:

- Develop a model to forecast equipment failures in advance to reduce unplanned downtime and production losses.

2. Optimize Maintenance Costs:
-Create a maintenance schedule approach that strikes a good balance between the expenses of maintenance operations (people, components, and downtime) and the benefits of averting breakdowns.

3. Maximize Equipment Reliability and Performance:
- Make sure the predictive maintenance strategy aids in improving the machinery used in production.

4. Improve Sustainability:
- Reduce waste, energy use, and the frequency of equipment replacements to improve sustainability, which is a key aspect of any long-term plan with a positive influence on the environment.

Variables

Equipment Health Indicators
- Define variables that describe the state of manufacturing equipment based on sensor data, previous maintenance records, and other pertinent information.

Maintenance Decision Variables
- Determine decision variables include scheduling, resource allocation, and the kind of maintenance to be performed (preventative, corrective, or predictive).

Constraints:

1. Resource Constraints:
- Consider restrictions on maintenance resources, including labor, spare parts inventories, and maintenance personnel availability.

2. Production Constraints:
- Plan maintenance when there will be the least impact on production so that goals may be met.

3. Regulatory and Safety Compliance:
- When organizing and carrying out maintenance tasks, it is important to follow all applicable regulations and safety protocols.

Criteria

1. Accuracy of Predictions
- Assess how well predictive models do in predicting when pieces of equipment will break down. Accuracy may be measured by recall and F1 score.

2. Cost-Benefit Analysis
- Calculate the monetary effect of the predictive maintenance plan by weighing the expenses of maintenance, downtime, and possible savings.

3. Sustainability Metrics
- Incorporate the assessment criteria. These include advances in energy efficiency, waste reduction, and the environmental effect of maintenance operations.

4. Equipment dependability Metrics:
- Measure the dependability and performance of equipment by examining key performance indicators (KPIs) related to uptime, mean time between failures (MTBF), and overall equipment effectiveness (OEE) [8].

Data Requirements:

In order to do predictive maintenance, the following data is required:
1. Sensor Data:
- Specify types of sensor data (temperature, vibration, pressure, etc.) that will be required.

2. Historical Maintenance Records
- Use these records to train models and find trends in equipment breakdowns.

3. External Factors
- Think about how things outside of your control, like the weather or changes in demand, might affect the condition of your equipment and how often you need to service it [9].

This formulation of the predictive maintenance problem allows manufacturers to develop a systematic and all-encompassing plan for incorporating Big Data analytics into their operations in order to achieve more intelligent and environmentally friendly outcomes. This structure lays the groundwork for creating mathematical models, ML algorithms, and optimisation techniques to handle targeted problems and obtain desired outputs [10].

Research Contribution

There are following research contribution as below:
- This paper optimised AdaBoost algorithm for automotive predictive maintenance.
- By distributing calculations over multiple processors or nodes, the predictive maintenance model may efficiently evaluate huge volumes of sensor and IoT data.

- By recognizing and selecting relevant attributes, the model can better capture essential patterns and correlations that enable predictive maintenance.
- The proposed method reduces disruptions and extends equipment life, which improves resource efficiency and reduces environmental impact.
- Adopting advanced analytics, machine learning, and optimization technologies improves industry efficiency and competitiveness..

Paper organization

The remainder of the article is structured as follows: A quick summary of the many literature evaluations already presented on the topic is provided in Section II. The research approach is covered in Section III. The research's findings are presented in Section IV. Potential applications are discussed in Section V. The paper is ultimately concluded in Section VI.

2. Related work

Scherer et al. [11] described how an HDaaS platform solution employing EMC® Isilon®, Pivotal® Hadoop Distribution (HD), and VMware vSphere Big Data Extensions may help distribute Big Data analytics by optimizing resource use and administration.

Luckow et al. note that Hadoop has several applications, including the automotive industry [12]. Hadoop has created a varied ecosystem, including databases. Questions like "What kinds of applications and data sets would work well with Hadoop?" inspired this essay. How can a multi-tenant Hadoop cluster support many frameworks and tools? Do these programs fit relational database management structures? The question is how to secure corporate demands.

Bracke et al. [13] calculate product fleet risk likelihood using a multivariate investigation of product failure behavior and consumer product consumption profile. An automobile case study employing a synthetic data set that includes genuine implications of typical field failure behavior and vehicle fleet utilization patterns demonstrates the technique theoretically and practically.

Data analytics and intelligent manufacturing help solve Vater et al.'s problems [14]. Prescriptive analytics may enhance manufacturing output. This essay begins with a detailed analysis of production prescriptive analytics. This report also underlines the need and suggests future investigation.

Singh et al. [15] discuss massive data challenges. The paper explains big data's technical foundations. This article shows how MapReduce, a background data mining technology, works.

Wen-Xin et al. estimate the functional area, partition it quantitatively, analyze the geographical pattern qualitatively, and evaluate the division's precision [16]. Results show that the Kappa coefficient for categorizing functional land in Xi'an's core urban region is 0.748, indicating 79.26% accuracy. A more rational functional land structure permits dynamic updating and fine-grained function division in the research area.

Pavithra et al. [17] examines massive data development and examination. This paper also briefly discusses the pros and cons of applying this article's Big Data analytics ideas in each subject. Large dataset analysis methods in various real-world scenarios are also covered.

Gupta et al. [18] asserted that R&M has helped all automakers, dealers, drivers, and insurance. Today, a new technology is rapidly changing R&M processes and applications. AI has a ripple impact in the auto industry.

Rahman et al. [19] proposed open central VHMS and a taxonomy using IoE and machine learning. Finally, this idea's outcome affects the auto sector. It may motivate the researcher to develop a centralized, intelligent, and secure vehicle condition diagnosis system to help this industry meet Industry 4.0 norms.

Jayender et al. [20] study the compatibility of Big data, IOT analytics, and ERP to construct an intelligent decision-making support system for the Automotive Supply Chain as an alternative to ERP. This study presents a framework for an autonomous intelligent system that uses AI to recognize statistical models in SCM processes.

Huang et al. [21] emphasize our interdisciplinary effort to develop a comprehensive car dataset from various internet sources and formats. The collection includes 899 car models with 1.4 million images, model specifications, and UK sales statistics from over a decade. In addition to our philosophy, technical details, and data format, we present three basic case studies to demonstrate the use of our data for business studies and applications.

Lourens et al. [22] demonstrate how these technologies are used in important automotive value chain processes. We use future use cases to show how these breakthroughs can impact the industry.

Using a novel application of K-means clustering, Li et al. [23] separated vehicle risk into 30 categories, which served as a baseline for the construction of a vehicle model risk assessment system in China.

Combine elements from numerous industries, including cloud computing. Zhang et al. [24] found SOM's pattern-selection skills in vast data effective for attribute optimization and clustering. A car customization case study shows how the SOM can add clusters as it learns more about customer needs. To enable Industry 4.0, the self-organizing tool includes many smart design attributes.

Research gaps

While there is much potential in utilizing big data analytics for predictive maintenance in intelligent and sustainable production, many unanswered questions remain. Filling in these spaces will make these systems more useful and efficient. Some major unanswered questions in this field are as follows:

- Manufacturing environments and equipment can change quickly. Predictive maintenance solutions rarely dynamically adapt to these alterations. More work is needed to build adaptive algorithms that learn from experience and update their models in real time.
- Several manufacturing processes generate real-time data. Streaming data may overwhelm current predictive maintenance models. Researchers require real-time analytics systems that can interpret high-velocity data streams to produce accurate predictions.
- Edge computing reduces latency and strain on centralized systems by processing data closer to the source. Manufacturing predictive maintenance is expanding and could benefit from edge computing and big data analytics research.
- Manufacturing data comes from many sources and is presented in many formats. Sensors, IoT devices, and archival records may be difficult to integrate and analyze. Future research should focus on methods for combining and interpreting heterogeneous data sources.
- Predictive maintenance models must provide forecasts and uncertainty or confidence levels. Research on uncertainty and confidence evaluation can improve predictive maintenance models.
- Machine learning models, especially predictive maintenance models, are often "black boxes" without interpretability. Additional research is needed to make these models more interpretable and understandable, especially in circumstances where human operators must trust and act on predictions.
- Most predictive maintenance algorithms offer minimal future knowledge. Research into extending the prediction horizon can help proactive maintenance techniques foresee equipment degradation and breakdowns over time.
- The cost-benefit analysis of big data analytics for predictive maintenance is crucial. Future research should include implementation costs, maintenance savings, and manufacturing efficiency gains in a cost-benefit analysis.
- Industrial data confidentiality becomes more important as interconnected manufacturing systems become common. Data security, privacy, and secure communication methods need more predictive maintenance research..

The completion of these studies will not only advance our theoretical understanding of big data analytics in predictive maintenance, but will also provide real-world applications for the implementation of sustainable and intelligent manufacturing systems.

3. Material and Method

Dataset

These statistics come from manufacturing company production equipment. Data helps prevent costly equipment failures by predicting repair needs. Manually tracking maintenance becomes harder as firms grow. It planned predictive maintenance using sensor data [25]. Sensor data is utilized to schedule preventative maintenance. Below are features or columns.

- UDI (Unique Device Identifier)
- Product ID
- Type: Categorized as Low, medium and high.
- Air Temperature.
- Process Temperature.
- Rotational Speed.
- Torque
- Tool Wear
- Target (Machine Failure)
- Failure Type

Fig 1 demonstrate the data distribution as below

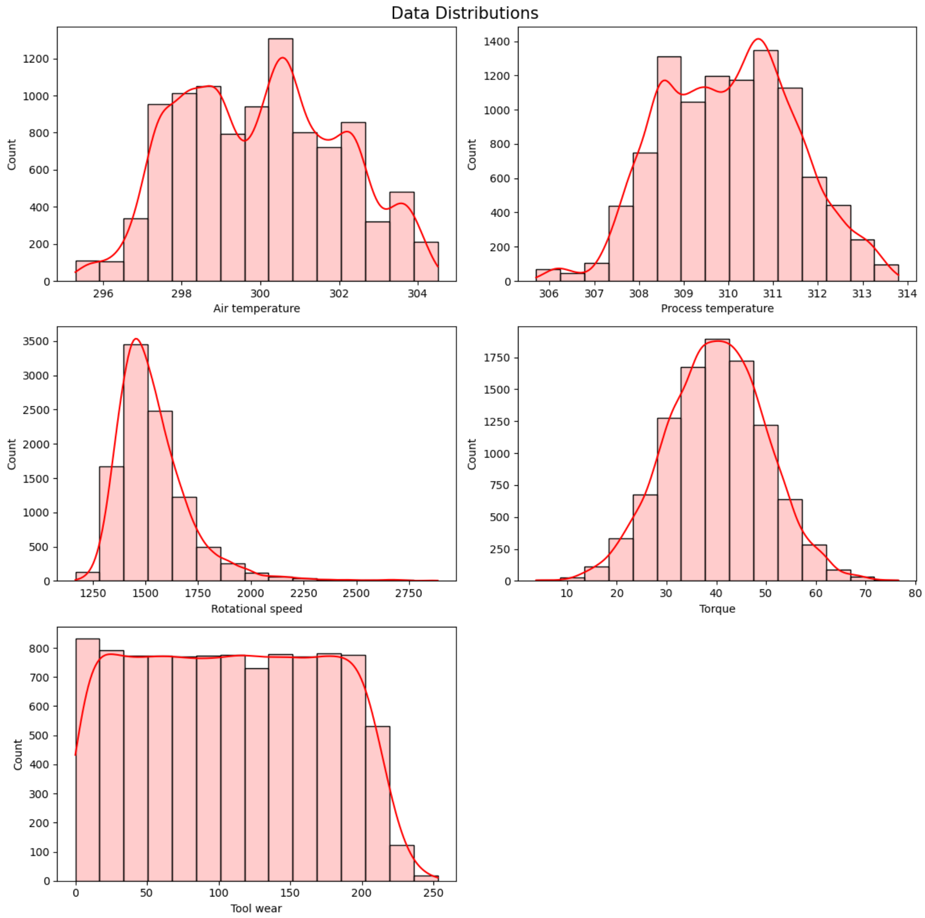

Fig. 1 Data distribution

Figure 2 shows how this data is evaluated to predict future maintenance needs. The company will know when to repair devices, reducing costly downtime [26].

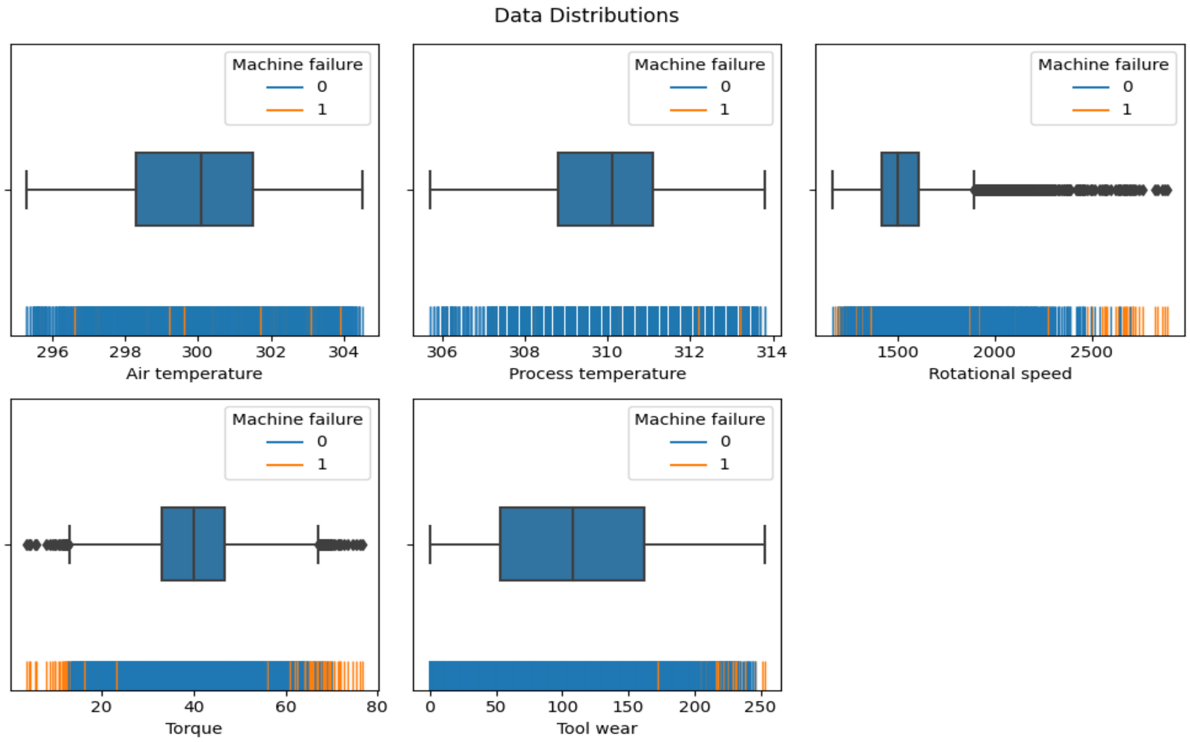

Fig. 2 Data distribution for target variable

As businesses grow in size and complexity, keeping up with routine maintenance becomes increasingly difficult shown in figure 3.

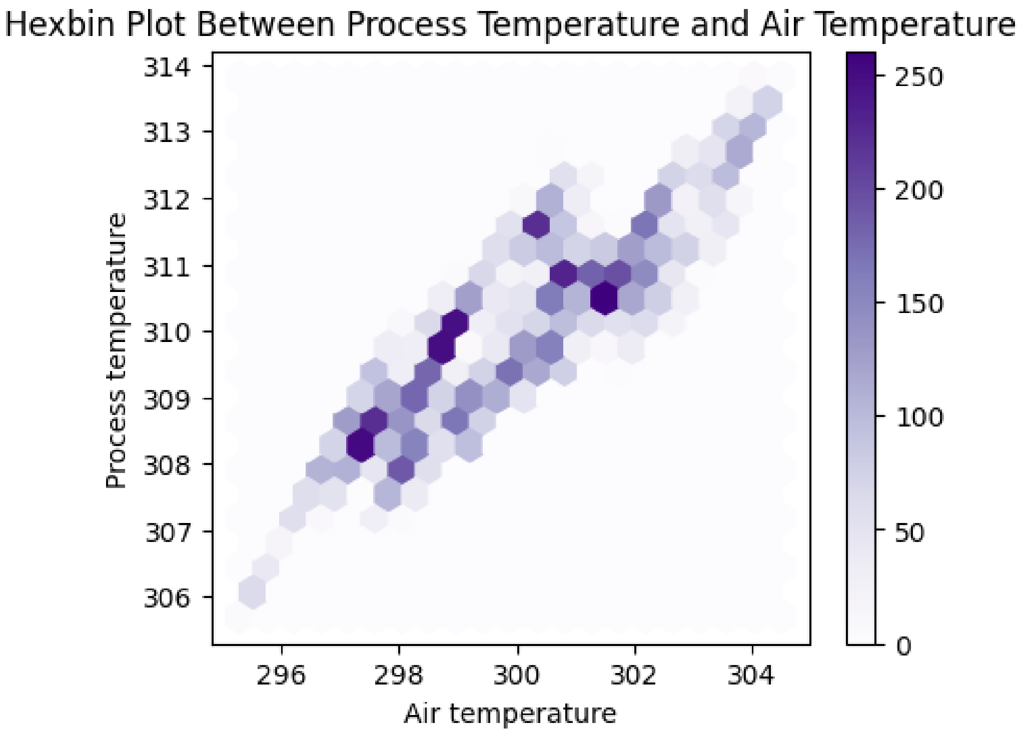

Fig. 3 Hexbin Plot for features

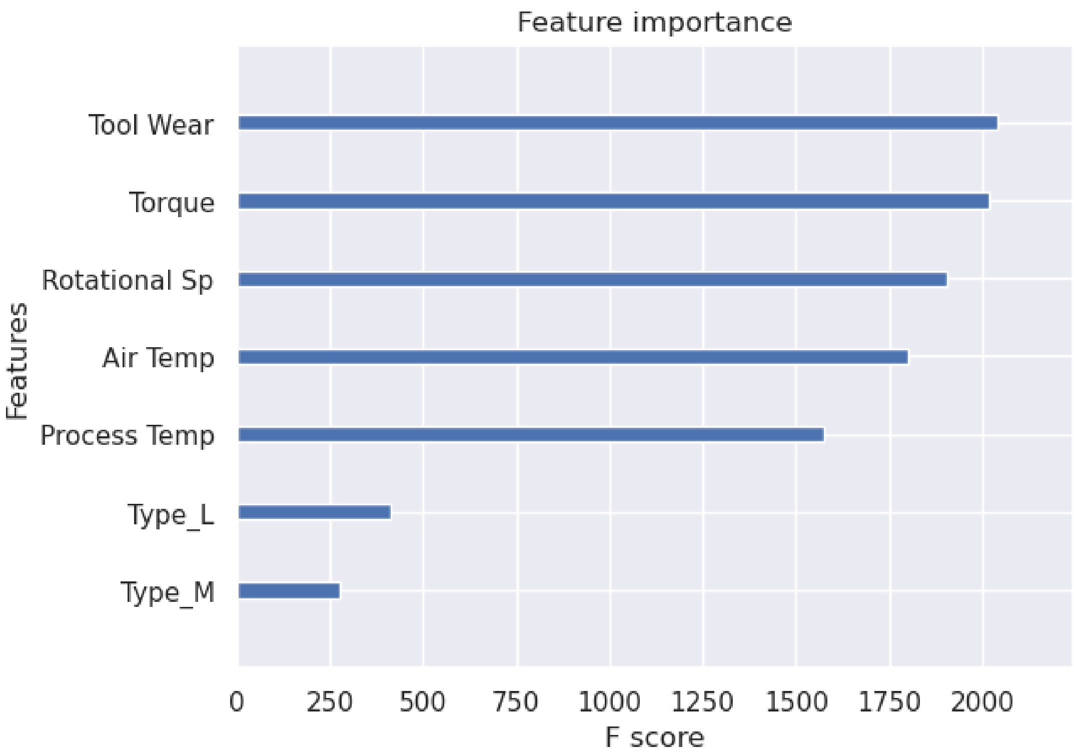

Fig. 4 Feature importance

Figure 4 shows the relevance of features. Ensemble learning, which includes AdaBoost, is a strong machine learning paradigm. AdaBoost, developed by Yoav Freund and Robert Schapire in 1996, has been widely utilized to improve weak learners and create a more accurate and dependable ensemble model [27]. Figure 5 shows feature SHAP values.

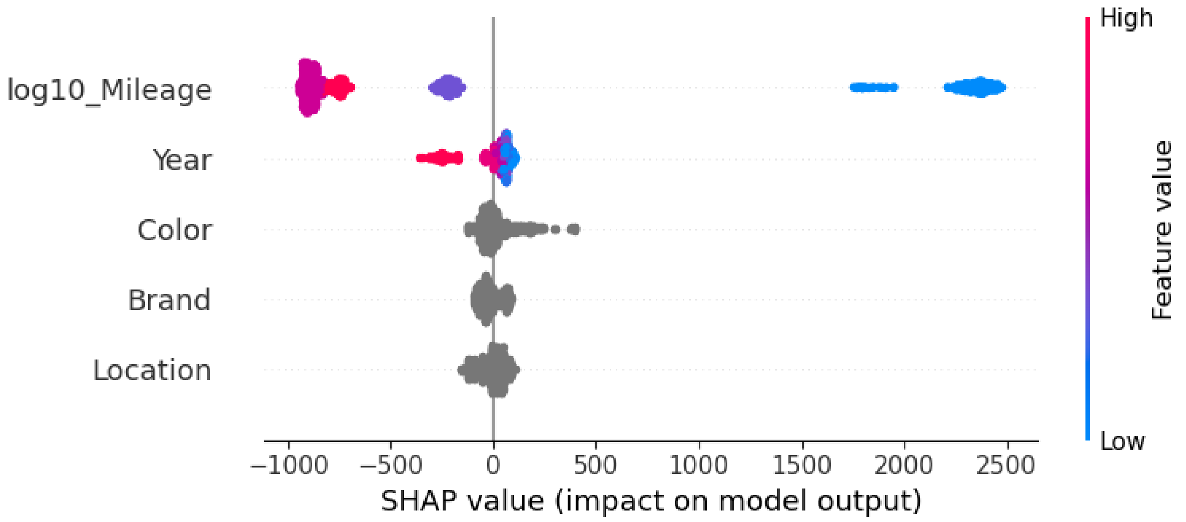

Fig. 5 SHAP value for Feature value

This section will explain AdaBoost's fundamentals, inner workings, and real-world applications to demonstrate its importance to machine learning. AdaBoost uses boosting to improve a model's accuracy by giving poorly labeled examples more weight. To create a powerful and accurate classifier, the algorithm integrates the outputs of several weak learners, often basic models better than random chance. AdaBoost's adaptability comes from its capacity to dynamically adjust training case weights to prioritize accurate classification of previously misclassified samples [28-32].

AdaBoost iterates throughout training. Each iteration, a weak learner is taught on the dataset, and the method weights misclassified events more. The model performs better overall because this adaptive weighting directs weak learners' attention to tough cases. An accuracy-based weighting technique combines weak learners into one model [33-37]. AdaBoost employs weighted voting to aggregate ineffective learner predictions. Learners' training results influence their importance. Better learner precision increases aggregate forecast weight. This ensemble strategy reduces overfitting and improves model generalization.

Due to its various benefits, AdaBoost is widely used. First, its simplicity aids comprehension and acceptance. AdaBoost's interoperability with several base classifiers increases model diversity. AdaBoost also overfits less than individual classifiers, making it useful in situations with limited training data. Many situations can employ AdaBoost [38-40]. AdaBoost is used in computer vision for face, object, and image segmentation. Bioinformatics uses it to classify proteins and analyze gene expression. AdaBoost has also excelled in text categorization and fraud detection, where accurate projections are essential. AdaBoost works well in many situations, but it has drawbacks. Data noise and outliers may reduce algorithm performance. Base classifier and algorithm iterations affect efficiency [40-42]. AdaBoost shows how weak learners can be merged to create a robust and accurate classifier via ensemble learning. AdaBoost's adaptability, simplicity, and efficacy make it a machine-learning staple. As machine learning technology advances, AdaBoost remains an important technique that improves model quality in various circumstances.

Algorithm 1: Adaboosting

Step 1. Set each training instance's initial sample weight, w_i, to (1/N).
Step 2. For all values of t from 1 to T for step 3-6.
Step 3. Develop a simple classifier h_t using the weighed data.
Step 4. Find the weak classifier's epsilon_t error.
Step 5. The weak classifier's alpha_t weight has to be determined.
Step 6. Change w_i of samples to reflect how well h_t is doing.
Step 7. Combine weak classifiers into a strong classifier

AdaBoost optimization demands careful feature engineering, algorithmic judgments, and computer resource efficiency. With sufficient hyperparameter tweaking, noise attention, and parallelization, AdaBoost can be a powerful tool for a number of machine learning tasks. Even when optimization algorithms evolve, AdaBoost may be tailored to a wide range of datasets. AdaBoost relies significantly on its base classifier. Classifiers that are easy to develop and computationally efficient will better represent your data. Decision stumps, shallow decision trees with one decision node and two leaf nodes, often work.

The learning rate and iterations (T) are hyperparameters that greatly impact AdaBoost. A systematic search or grid search or Bayesian optimization can find hyperparameter values that maximize model accuracy. Quantizing model parameters reduces inference time and memory. This increase is crucial for AdaBoost models in resource-constrained environments.

```
# Import necessary libraries
import numpy as np
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
# Define the AdaBoost algorithm with Quantization and Hyperparameter Tuning
class QuantizedAdaBoost:
    def __init__(self, n_iterations=50, learning_rate=1.0, base_classifier=None,
quantization_bits=8):
        self.n_iterations = n_iterations
        self.learning_rate = learning_rate
        self.base_classifier = base_classifier or DecisionTreeClassifier(max_depth=1)
        self.quantization_bits = quantization_bits
        self.models = []
        self.alphas = []
    def quantize_weights(self, weights):
        # Implement weight quantization logic here (e.g., rounding to specified number of bits)
        quantized_weights = ...
        return quantized_weights
    def fit(self, X, y):
        # Initialize sample weights
        sample_weights = np.ones(len(X)) / len(X)
        for t in range(self.n_iterations):
            # Train a weak classifier
            weak_classifier = self.base_classifier.fit(X, y, sample_weight=sample_weights)
            # Calculate the error of the weak classifier
            predictions = weak_classifier.predict(X)
            error = np.sum(sample_weights * (predictions != y)) / np.sum(sample_weights)
            # Calculate the weight of the weak classifier
            alpha = self.learning_rate * np.log((1 - error) / error)
            self.alphas.append(alpha)
            # Update sample weights
            sample_weights *= np.exp(-alpha * y * predictions)
            sample_weights /= np.sum(sample_weights)
            # Quantize the weights
            quantized_weights = self.quantize_weights(sample_weights)
            # Store the weak classifier and its quantized weights
            self.models.append((weak_classifier, quantized_weights))
    def predict(self, X):
        # Make predictions using the final ensemble model
        final_predictions = np.zeros(len(X))
        for model, alpha in zip(self.models, self.alphas):
            weak_classifier, quantized_weights = model
            predictions = weak_classifier.predict(X)
            final_predictions += alpha * predictions
        # Convert final predictions to binary (e.g., using sign function)
        final_predictions = np.sign(final_predictions)
        return final_predictions
# Example usage:
# Instantiate QuantizedAdaBoost with desired hyperparameters
```

```
adaboost_model = QuantizedAdaBoost(n_iterations=50, learning_rate=0.1, quantization_bits=4)
# Fit the model to training data
adaboost_model.fit(X_train, y_train)
# Make predictions on test data
predictions = adaboost_model.predict(X_test)
# Evaluate accuracy
accuracy = accuracy_score(y_test, predictions)
print(f"Accuracy: {accuracy}")
```

4. Results

The Optimized AdaBoost model improves predictive accuracy, helping manufacturers detect and prevent equipment breakdowns, stabilizing production. The hatmap with proposed model results is in figure 6.

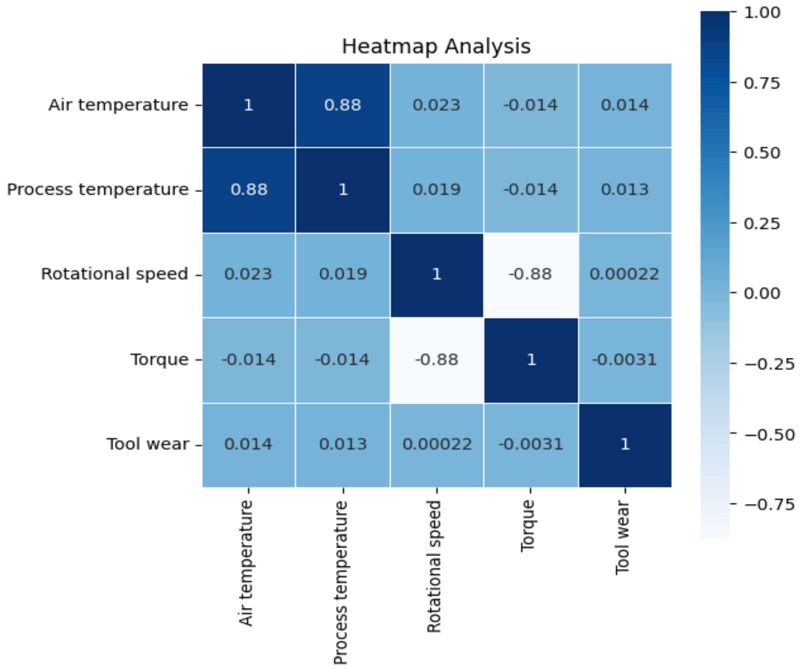

Fig. 6 Confusion matrix

Figure 7 compares the number of successfully anticipated occurrences to the total instances to show predictive maintenance model accuracy. Find the accuracy rate by dividing the number of correct diagnoses by the total number of correct and false positives. The model's predictive power improves with accuracy and precision. The authors calculate the fraction of correct diagnoses over right and faulty diagnoses. To record every machine breakdown, a high recall rate is needed. Get a balanced model evaluation with the F1 score, a harmonic mean of accuracy and recall. The suggested model achieves 0.972 accuracy, 0.977 precision, 0.972 recall, and 0.974 F1-score.

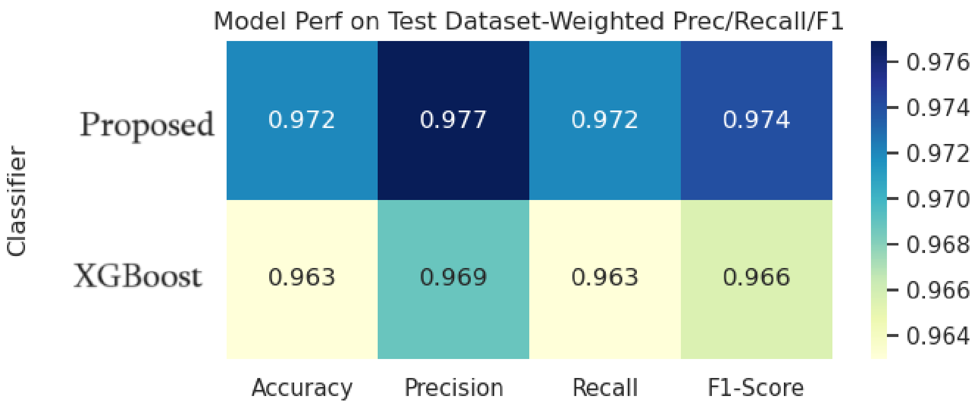

Fig. 7 Result analysis

An Optimized AdaBoost model for predictive maintenance in intelligent and sustainable manufacturing must be studied for effectiveness, economy, and practicality. This study's findings support the model's feasibility and its continued development and refinement to meet industrial demands.

Discussion

Optimization of AdaBoost requires establishing appropriate hyperparameters, such as iterations and learning rate. Adjusting to production data's quirks improves the algorithm's efficiency. Model parameter quantization affects memory efficiency, especially in resource-constrained production. Discuss quantization methods and their effects on memory and accuracy. Debate should center on how the approach handles enormous data scalability challenges. Parallelization disperses computations, allowing the model to manage large datasets and real-time data streams. Forecasts are crucial to predictive maintenance, and real-time data helps. Discussing the model's ability to assess high-velocity data streams and provide insights for proactive decision-making is crucial. The study's usefulness in boosting green production should be highlighted. The manufacturing ecosystem benefits from sustainability and resource efficiency when equipment lifespan, downtime, and maintenance schedules are optimized. Discussing integration with current systems, user usability, and industrial process compatibility is crucial. Predictive maintenance solutions must overcome barriers to implementation in manufacturing. The proposed method focuses on user input and stakeholder interaction. The strategy's user-centricity should be considered in light of manufacturing process participants' needs and aims. Compare the proposed method against standard predictive maintenance methods and other machine learning algorithms to assess its performance. Discuss the benefits and unique characteristics of the upgraded AdaBoost-enabled solution. Understanding the solution's boundaries and limits requires discussion. This includes data quality, model interpretability, and implementation challenges. Expanding the model's flexibility to accommodate for more nuanced industrial environments, studying other integration approaches, and solving new field challenges are possible research directions. Given the growing relevance of data-driven industrial decisions, trust in the installed predictive maintenance system is crucial, making ethical, data protection, and responsible AI usage discussions necessary.

5. Conclusion and Future scope

Optimization, parallelization, and quantization make the model efficient and scalable for processing enormous datasets and real-time data streams with few resources. Real-time analytics lets you make informed decisions and prevent unwanted downtime. Green manufacturing supports predictive maintenance because it maximizes equipment uptime, reduces downtime, and conserves resources. The report emphasizes practical deployment issues notwithstanding the challenges of adopting predictive maintenance in real-world production environments. The memory-efficient, scalable concept is suitable for industrial use. Customers and other stakeholders must be consulted during production. The solution's success depends on customer satisfaction, integration ease, and operational requirements. Continuous improvement approaches including retraining the model with new data and adapting to changing production conditions ensure the predictive maintenance system's long-term utility and relevance. It will take time and effort to improve the model's dynamic adaptation to changing manufacturing conditions. Integration with edge computing and advanced sensor technologies should be studied to improve the model's efficiency and real-time capabilities. Future research should make the model more comprehensible and transparent so users and stakeholders can trust it. Future research must incorporate cyber security considerations into predictive maintenance model design and deployment to ensure safety and reliability as industrial systems become more networked. AdaBoost with big data analytics for predictive maintenance can solve intelligent and sustainable manufacturing concerns. As manufacturing evolves, complex algorithms and analytics will ensure industrial processes' resilience, efficiency, and sustainability.

DECLARATIONS:

Availability of data and materials: Publicly available dataset was analyzed in this study.

Competing Interest: The authors declare no competing interests.

Funding: No funding was received for this work.

Author Contributions: ST were responsible for Validation, Software, Data Curation, and Writing - Original Draft. FJ were responsible for Conceptualization, Writing - Original Draft. YKS was responsible for Writing - Original Draft, Visualization. ST was responsible for Writing - Review & Editing. FJ were responsible for Formal Analysis. YKS was responsible for Writing -

Original Draft, Resources, Supervision. The author(s) read and approved the final manuscript.

Ethics approval and consent to participate statement: Not applicable.

Consent for The Publication: Not applicable.

References

1. Berges, C., Bird, J., Shroff, M. D., Rongen, R., & Smith, C. (2021). Data analytics and machine learning: Root-cause problem-solving approach to prevent yield loss and quality issues in semiconductor industry for automotive applications. Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA, 2021-Septe. https://doi.org/10.1109/IPFA53173.2021.9617238
2. Itoh, M., Yokoyama, D., Toyoda, M., & Kitsuregawa, M. (2015). A System for visual exploration of caution spots from vehicle recorder data. 2015 IEEE Conference on Visual Analytics Science and Technology, VAST 2015 - Proceedings, 199–200. https://doi.org/10.1109/VAST.2015.7347677
3. Saldivar, A. A. F., Goh, C., Chen, W. N., & Li, Y. (2016). Self-organizing tool for smart design with predictive customer needs and wants to realize Industry 4.0. 2016 IEEE Congress on Evolutionary Computation, CEC 2016, 5317–5324. https://doi.org/10.1109/CEC.2016.7748366
4. Kljaic, Z., Skorput, P., & Amin, N. (2016). The challenge of cellular cooperative ITS services based on 5G communications technology. 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2016 - Proceedings, 587–594. https://doi.org/10.1109/MIPRO.2016.7522210
5. Bakkar, M., & Alazab, A. (2019). Designing security intelligent agent for petrol theft prevention. Proceedings - 2019 Cybersecurity and Cyberforensics Conference, CCC 2019, Ccc, 123–128. https://doi.org/10.1109/CCC.2019.00006
6. Lamb, J., & Godbole, N. S. (2019). Smart Energy Efficiency for a Sustainable World. 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019, July 2018, 0855–0861. https://doi.org/10.1109/UEMCON47517.2019.8992941
7. Shunkai, W., Yi, W., Hongyu, N., & Fan, Z. (2022). Research on urban intelligent network industry index based on entropy weight method and big data analysis. Proceedings - 2022 International Conference on Data Analytics, Computing and Artificial Intelligence, ICDACAI 2022, 161–168. https://doi.org/10.1109/ICDACAI57211.2022.00040
8. Aydin, I., Sevi, M., Gungoren, G., & Irez, H. C. (2022). Signal Synchronization of Traffic Lights Using Reinforcement Learning. 2022 International Conference on Data Analytics for Business and Industry, ICDABI 2022, 103–108. https://doi.org/10.1109/ICDABI56818.2022.10041559
9. Ferrell, U. D., & Anderegg, A. H. A. (2022). Validation of Assurance Case for Dynamic Systems. AIAA/IEEE Digital Avionics Systems Conference - Proceedings, 2022-Septe, 1–11. https://doi.org/10.1109/DASC55683.2022.9925731
10. Cui, S., Li, L., Tang, Y., & Li, C. (2021). Exploring the Diversity of Alliance Portfolio and Firm Performance Based on the QCA Method. 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2021, 2019, 541–546. https://doi.org/10.1109/ICCCBDA51879.2021.9442558
11. Scherer, V., & Kaponig, B. (2013). EMC Hadoop as a service solution for use cases in the automotive industry. 2013 International Conference on Connected Vehicles and Expo, ICCVE 2013 - Proceedings, 488–493. https://doi.org/10.1109/ICCVE.2013.6799842
12. Luckow, A., Kennedy, K., Manhardt, F., Djerekarov, E., Vorster, B., & Apon, A. (2015). Automotive big data: Applications, workloads and infrastructures. Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015, 1201–1210. https://doi.org/10.1109/BigData.2015.7363874
13. Bracke, S., Lücker, A., & Sochacki, S. (2016). Reliability analysis regarding product fleets in use phase: Multivariate cluster analytics and risk prognosis based on operating data. International Conference on Control, Decision and Information Technologies, CoDIT 2016, 210–215. https://doi.org/10.1109/CoDIT.2016.7593562
14. Vater, J., Harscheidt, L., & Knoll, A. (2019). Smart Manufacturing with Prescriptive Analytics A review of the current status and future work. Proceedings of 2019 8Th International Conference on Industrial Technology and Management (Icitm 2019), 224–228.

15. Singh, S., & Jagdev, G. (2020). Execution of Big Data Analytics in Automotive Industry using Hortonworks Sandbox. Indo - Taiwan 2nd International Conference on Computing, Analytics and Networks, Indo-Taiwan ICAN 2020 - Proceedings, 158–163. https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181314
16. Wen-Xin, S., & Yun, G. (2020). Identification and Analysis of Urban Functional Areas Based on VGI Data. 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2020, 408–414. https://doi.org/10.1109/ICCCBDA49378.2020.9095657
17. Pavithra., N., & Manasa., C. M. (2021). Big Data Analytics Tools: A Comparative Study. CSITSS 2021 - 2021 5th International Conference on Computational Systems and Information Technology for Sustainable Solutions, Proceedings, 1–6. https://doi.org/10.1109/CSITSS54238.2021.9683711
18. Gupta, S., Amaba, B., McMahon, M., & Gupta, K. (2021). The Evolution of Artificial Intelligence in the Automotive Industry. Proceedings - Annual Reliability and Maintainability Symposium, 2021-May, 1–7. https://doi.org/10.1109/RAMS48097.2021.9605795
19. Rahman, M. A., Rahim, M. A., Rahman, M. M., Moustafa, N., Razzak, I., Ahmad, T., & Patwary, M. N. (2022). A Secure and Intelligent Framework for Vehicle Health Monitoring Exploiting Big-Data Analytics. IEEE Transactions on Intelligent Transportation Systems, 23(10), 19727–19742. https://doi.org/10.1109/TITS.2021.3138255
20. Jayender, P., & Kundu, G. K. (2022). Big data, IOT, ERP interoperability - An Intelligent SCM Decision System. Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2022, 549–555. https://doi.org/10.1109/ICAIS53314.2022.9742745
21. Huang, J., Chen, B., Luo, L., Yue, S., & Ounis, I. (2022). DVM-CAR: A Large-Scale Automotive Dataset for Visual Marketing Research and Applications. Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022, 4140–4147. https://doi.org/10.1109/BigData55660.2022.10020634
22. Lourens, M., Sharma, S., Pulugu, R., Gehlot, A., Manoharan, G., & Kapila, D. (2023). Machine learning-based predictive analytics and big data in the automotive sector. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 1043–1048. https://doi.org/10.1109/icacite57410.2023.10182665
23. Li, P., Xu, B., & Xue, B. (2023). Research on Vehicle Model Risk Rating Based on GLM Model and K-Means Clustering Algorithm for Car Insurance Pricing Scenario. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 134–137. https://doi.org/10.1109/ICCCBDA56900.2023.10154859.
24. Zhang, J., & Zheng, B. (2023). Finite Element Analysis and Optimization Design of the Spiral Groove Brake Drum. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 228–232. https://doi.org/10.1109/ICCCBDA56900.2023.10154645
25. Souksavanh, V., & Liu, Y. (2020). NVH Data Analytics and Its Application in Vehicle Rating. 2020 IEEE 7th International Conference on Industrial Engineering and Applications, ICIEA 2020, 287–292. https://doi.org/10.1109/ICIEA49774.2020.9101968
26. Pang, H., Liu, P., Wang, S., Wang, Z., & Zhang, Z. (2020). Usage Pattern Analytics of Fuel Cell Vehicle Based on Big Data Analysis. 2020 10th International Conference on Power and Energy Systems, ICPES 2020, 373–378. https://doi.org/10.1109/ICPES51309.2020.9349670
27. Kantert, J., & Nolting, M. (2021). How to integrate with real cars - Minimizing lead time at volkswagen. Proceedings - International Conference on Software Engineering, 358–359. https://doi.org/10.1109/ICSE-SEIP52600.2021.00045
28. Lilhore UK, Manoharan P, Simaiya S, Alroobaea R, Alsafyani M, Baqasah AM, Dalal S, Sharma A, Raahemifar K. HIDM: Hybrid Intrusion Detection Model for Industry 4.0 Networks Using an Optimized CNN-LSTM with Transfer Learning. Sensors. 2023; 23(18):7856. https://doi.org/10.3390/s23187856
29. Uguroglu, E. (2021). Near-Real Time Quality Prediction in a Plastic Injection Molding Process Using Apache Spark. Proceedings - 2021 International Symposium on Computer Science and Intelligent Controls, ISCSIC 2021,

284–290. https://doi.org/10.1109/ISCSIC54682.2021.00059

30. Nair, J. P., & Vijaya, M. S. (2021). Predictive Models for River Water Quality using Machine Learning and Big Data Techniques-A Survey. Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021, 1747–1753. https://doi.org/10.1109/ICAIS50930.2021.9395832
31. Gireesh Babu, C. N., Chandrashekhara, K. T., Verma, J., & Thungamani, M. (2021). Real time alert system to prevent Car Accident. 2021 International Conference on Forensics, Analytics, Big Data, Security, FABS 2021, 1, 1–4. https://doi.org/10.1109/FABS52071.2021.9702559
32. Zhou, J., Guo, Y., Huang, H., Li, R., & Gan, Y. (2021). The Potential Customer's Background of the Chinese Electric Vehicle Market Base on Big Data. Proceedings - 2021 International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2021, 263–268. https://doi.org/10.1109/CAIBDA53561.2021.00062
33. Dalal, S., Seth, B., & Radulescu, M. (2023). Driving Technologies of Industry 5.0 in the Medical Field. In Digitalization, Sustainable Development, and Industry 5.0: An Organizational Model for Twin Transitions (pp. 267-292). Emerald Publishing Limited.
34. Dalal, S., Lilhore, U. K., Simaiya, S., Sharma, A., Jaglan, V., Kumar, M., ... & Rana, A. K. (2023). Original Research Article A Blockchain-based secure Internet of Medical Things framework for smart healthcare. Journal of Autonomous Intelligence, 6(3).
35. Lilhore, U. K., Dalal, S., Faujdar, N., Margala, M., Chakrabarti, P., Chakrabarti, T., ... & Velmurugan, H. (2023). Hybrid CNN-LSTM model with efficient hyperparameter tuning for prediction of Parkinson's disease. Scientific Reports, 13(1), 14605.
36. Dalal, S., Lilhore, U. K., Simaiya, S., Jaglan, V., Mohan, A., Ahuja, S., ... & Chakrabarti, P. (2023). Original Research Article A precise coronary artery disease prediction using Boosted C5. 0 decision tree model. Journal of Autonomous Intelligence, 6(3).
37. Zhang, B., & Zhang, F. (2022). Analysis and Optimization of Communication Strategy of New Energy Vehicles at Home and Abroad Based on Data Mining. Proceedings - 2022 6th Annual International Conference on Data Science and Business Analytics, ICDSBA 2022, 614–618. https://doi.org/10.1109/ICDSBA57203.2022.00025
38. Lilhore, U.K., Simaiya, S., Dalal, S. et al. A smart waste classification model using hybrid CNN-LSTM with transfer learning for sustainable environment. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-16677-z
39. Dalal, S., Lilhore, U. K., Manoharan, P., Rani, U., Dahan, F., Hajjej, F., ... & Raahemifar, K. (2023). An Efficient Brain Tumor Segmentation Method Based on Adaptive Moving Self-Organizing Map and Fuzzy K-Mean Clustering. Sensors, 23(18), 7816.
40. Zhao, B., Zhang, J., Yuan, D., Yang, X., & Zhang, Y. (2022). Correlation Analysis of Public Welfare Activities and Brand Marketing Activities of Car Enterprises Based on Cloud Computing. Proceedings - 2022 6th Annual International Conference on Data Science and Business Analytics, ICDSBA 2022, 527–531. https://doi.org/10.1109/ICDSBA57203.2022.00111
41. Dalal, S., Lilhore, U.K., Foujdar, N. et al. Next-generation cyber attack prediction for IoT systems: leveraging multi-class SVM and optimized CHAID decision tree. J Cloud Comp 12, 137 (2023). https://doi.org/10.1186/s13677-023-00517-4
42. Zheng, B., & Yao, C. (2023). Automobile Profession Responds to the Development of Automobile Intelligence. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 419–423. https://doi.org/10.1109/ICCCBDA56900.2023.10154651