International Journal of Power Electronics and Drive System (IJPEDS)
Vol. 17, No. 1, March 2026, pp. 765~776
ISSN: 2088-8694, DOI: 10.11591/ijpeds.v17.i1.pp765-776

Application of machine learning for production optimization and predictive maintenance in an iron processing plant

Lakhdari Lahcen, Mohamed Habbab, Alhachemi Moulay Abdellah
Department of Electrical Engineering, Laboratory of Control Analysis and Optimization of Electro-Energetic Systems (CAOSEE), Faculty of Technology, Tahri Mohammed University, Bechar, Algeria

Article Info

Article history:

Received Aug 11, 2025
Revised Jan 5, 2026
Accepted Jan 23, 2026

Keywords:

Iron processing plant
LightGBM
Machine learning
Predictive maintenance
Production optimization

ABSTRACT

The modern metallurgical industry requires advanced solutions for process optimization, cost reduction, and predictive maintenance. This paper proposes a unified simulation-based framework using machine learning (ML) to jointly address production optimization and maintenance prediction in a virtual iron processing environment. Several ML models, including random forest (RF), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), support vector machine (SVM), and k-nearest neighbors (k-NN), were evaluated on synthetic datasets representing production, maintenance, and transport processes. A reproducible methodology was adopted, including preprocessing, time-aware data splitting, and cross-validation to prevent information leakage. Model performance was assessed using F1-score, area under the receiver operating characteristic curve (AUC), and regression metrics. Tree-based models achieved near-perfect classification performance (AUC ≈ 1, precision and recall > 0.99), while light gradient boosting machine (LightGBM) and CatBoost provided the best regression accuracy. Feature importance analysis using SHapley Additive exPlanations (SHAP) identified vibration and temperature as key maintenance indicators. Although based on simulation, the framework is designed for integration with supervisory control and data acquisition (SCADA) and the Industrial Internet of Things (IIoT), supporting real-time industrial deployment and alignment with operational key performance indicators.

Corresponding Author:

Lakhdari Lahcen
Department of Electrical Engineering
Laboratory of Control Analysis and Optimization of Electro-Energetic Systems (CAOSEE)
Faculty of Technology, Tahri Mohammed University
Bechar, Algeria
Email: lakhdari.lahcen@univ-bechar.dz

1. INTRODUCTION

The metallurgical industry is a cornerstone of economic development, supplying critical raw materials such as iron and steel for sectors including construction, automotive manufacturing, infrastructure, and energy. Among the raw materials, iron ore holds a pivotal role as it is the primary feedstock for steel production. The demand for high-quality iron ore is steadily increasing, driven by global urbanization, industrialization, and the transition to modern, energy-efficient infrastructure [1], [2].

In Algeria, the Ghar Djebilet iron ore deposit, located in the Tindouf Province in the southwest of the country, represents one of the largest untapped iron reserves in the world, with estimated deposits exceeding 3.5 billion tonnes. This strategic resource has the potential to transform Algeria into a key player in the global steel industry. The planned development of extraction, processing, and transport infrastructure associated with Ghar Djebilet will not only meet domestic steel demand but also enable significant export opportunities [1].

However, the exploitation of such a large-scale mining and processing project presents complex technical and operational challenges. These include optimizing production parameters, ensuring consistent product quality, and implementing predictive maintenance strategies to minimize equipment downtime. Traditional industrial control systems often rely on static process rules and periodic manual inspections, which may be insufficient in a high-throughput, geographically remote industrial environment [3].

The advent of Industry 4.0 technologies, integrating sensors, real-time data acquisition, advanced analytics, and artificial intelligence (AI), offers unprecedented opportunities to modernize the iron and steel value chain. Within this technological ecosystem, machine learning (ML) has emerged as a powerful approach for extracting actionable insights from large and complex datasets. ML models can identify subtle patterns, adapt to evolving operational conditions, and support data-driven decision-making in both production and maintenance domains [2]–[4].

In this context, the objective of this research is to investigate the application of supervised ML algorithms to optimize production quality and predict equipment failures in an iron processing plant inspired by the operational needs of the Ghar Djebilet project. Given the absence of real operational data from the future facility, synthetic datasets have been generated to simulate realistic industrial scenarios, including production, maintenance, and transport operations [4]–[6].

The contributions of this study are: i) Development of simulation-based datasets representing key operational processes of an iron processing plant; ii) Comparative evaluation of five widely used ML algorithms, random forest (RF), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), categorical boosting (CatBoost), neural network (NN), and RF+NN for classification (predictive maintenance) and regression (production quality); iii) Integration of interpretability techniques (feature importance, SHapley Additive exPlanations (SHAP)) to enhance model transparency and support industrial adoption; and iv) Formulation of a methodological framework adaptable to real industrial datasets for future deployment in the Ghar Djebilet iron processing chain.

In parallel with the national development of the Ghar Djebilet mining project, a dedicated iron-processing plant has been established in the Bechar region of southwestern Algeria. This industrial facility is expected to become a key component in transforming raw extracted ore into semi-finished or finished iron products, while also serving as a technological pilot platform for testing and deploying modern digital solutions in metallurgy. In this study, the Bechar plant is adopted as the reference industrial environment. All simulations and predictive models have been designed in alignment with the anticipated configuration and workflows of this facility. Its proximity to Tahri Mohammed University of Bechar and the active involvement of local academic research in the fields of automation and industrial informatics offer a unique opportunity to bridge the gap between advanced ML research and real-world deployment in a strategic national industrial project [7]–[9].

We address quality stabilization, throughput, and downtime reduction in iron processing, where conventional rule-based control and periodic inspections falter under multivariate non-linearities, drift, and noise. Unlike prior work that isolates a single subproblem and omits rigorous validation and deployability, we propose a unified ML framework over production, maintenance, and transport using simulation-based datasets. We benchmark RF, XGBoost, LightGBM, CatBoost, MLP, SVM, k-NN, and RF+NN, with time-aware splits, cross-validation, latency-aware selection, and SHAP-based interpretation. Results (area under the receiver operating characteristic (ROC) curve (AUC) ≈ 1, precision/recall > 0.99 for classification; coefficient of determination (R^2) > 0.59 for regression with boosters) are linked to key performance indicators (KPI) impacts (downtime, maintenance, throughput in Algerian Dinar (DZD)) and a real-time supervisory control and data acquisition (SCADA)/industrial internet of things (IIoT) deployment path (open platform communications unified architecture (OPC UA)/message queuing telemetry transport (MQTT), open neural network exchange (ONNX), edge inference). This research investigates the application of supervised machine learning algorithms to optimize production quality and predict equipment failures in an iron processing plant. The operational needs of the Ghar Djebilet project inspire the study.

The industrial context of this study is illustrated in Figure 1, which shows the open-pit quarry structure of the Gara Djebilet mining site. The large-scale extraction environment involves heavy equipment, continuous material transport, and dynamic operating conditions, highlighting the importance of predictive maintenance and intelligent production management. Figure 2 presents a representative view of an industrial iron-processing facility. The complex arrangement of processing units and monitoring infrastructures generates heterogeneous operational data streams that can be effectively exploited using machine learning techniques to improve system reliability, efficiency, and decision-making.

Figure 1. Interior view of the Gara Djebilet quarry structures illustrates the open-pit deposit and the mining context

Figure 2. View of robust industrial useful as a schematic illustration

2. REAL-TIME DEPLOYMENT FEASIBILITY SUPERVISORY CONTROL AND DATA ACQUISITION (SCADA)

Although all experiments are conducted in simulation, the approach is compatible with real-time industrial deployment with minor adaptation [10], [11].

a) Integration SCADA
   - Target architecture: PLC/SCADA ⟶ edge gateway (OPC UA or Modbus/TCP client, MQTT broker) ⟶ inference microservice (RF, LightGBM, XGBoost, MLP, or RF+NN) ⟶ alarm feedback to SCADA (OPC UA write) and data historization (InfluxDB/Timescale/PI).
   - Model packaging: ONNX export (Onnxruntime on CPU) or native format (XGBoost/LightGBM). This ensures portability (C++/Python) and startup time < 1 s.

b) Latency budget and sampling rate

End-to-end latency should remain below a fraction of the sampling period:

$$T_{total} = T_{acq} + T_{net} + T_{pre} + T_{inf} + T_{pub}, \text{aim for } T_{total} \leq 0.3 \times (\text{sampling period}) \quad (1)$$

where T_acq = data acquisition time, T_net = network communication time, T_pre = data preprocessing time, T_inf = model inference time, T_pub = publishing time (writing back to SCADA or to databases), and T_total = end-to-end latency of the processing chain. Examples: 10 Hz (100 ms) ⇒ target ≤ 30 ms; 100 Hz (10 ms) ⇒ target ≤ 3 ms. Use incremental/streaming feature updates when windowing (e.g., 1 s window, 100 ms hop) to avoid added delay. Hardware constraints (edge):

- CPU x86/ARM IPC: sufficient for compact RF/GBDT/MLP at 10–100 Hz.
- Light GPU/NPU (Jetson, Intel NPU): useful for larger MLP/CNN or multi-stream batching.
- Memory budget: keep service < 200 MB (model + runtime + buffers). k-NN is RAM-heavy; prefer RF/GBDT/MLP or RF+NN.
- Interfaces: OPC UA/Modbus/TCP for acquisition; MQTT for pub/sub; secure with TLS and RBAC.

c) Order-of-magnitude inference latency (modern edge CPU)
   - Random forest (100–500 trees, depth ≤ 10): 0.1–2 ms/sample
   - LightGBM/XGBoost (depth ≤ 8): 0.2–3 ms
   - Linear SVM: < 0.1–0.5 ms; RBF SVM: 1–10 ms (depends on #SV)
   - MLP (2–3 layers, 64–128): 0.2–2 ms
   - RF+NN (stacking): 1–4 ms

d) Deployment patterns
   - Hard real-time (≤ 10 ms in PLC): avoid complex ML; use rules/linear/shallow trees.
   - Soft real-time (10–200 ms on edge gateway): compact RF/GBDT/MLP; RF+NN is feasible.
   - Offline monitoring (on-prem/cloud): retraining, recalibration, controlled rollouts.

One-sentence takeaway: Packaged in ONNX and executed on an edge gateway via OPC UA/MQTT, compact RF/GBDT/MLP (or RF+NN) typically achieve ≤ 1–5 ms per sample, sufficient for 10–100 Hz operation if preprocessing and networking stay within budget.

3. ECONOMIC AND OPERATIONAL IMPACT INDICATORS

Purpose: Convert model performance (e.g., RF, LightGBM, RF+NN) into DA gains for factory KPIs [12], [13]. The key economic and operational indicators are defined as follows:

- Availability

$$A = \frac{T_{planned} - T_{downtime}}{T_{planned}} \quad (2)$$

- Reliability and maintainability are evaluated using mean time between failures (MTBF) and mean time to repair (MTTR),

$$\text{Overall equipment effectiveness (OEE)} = A \times Performance \times Quality \quad (3)$$

- Monetization Algerian Dinar (DA)
- Downtime saved

$$\text{DA} \approx n_{Failures\ avoided\ (C_{CM} - C_{PM}) + \Delta MTTR.C_{labor/h}} \quad (4)$$

$$\text{DA} = \Delta\text{H}_{\text{down}}.\text{C}_{\text{h}}$$

- Maintenance saved

$$\text{DA} \approx n_{Failures\ avoided\ (C_{CM} - C_{PM}) + \Delta MTTR.C_{labor/h}} \quad (5)$$

- Throughput value

$$\Delta\text{A} = \Delta\text{H}_{\text{down}}/\text{T}_{\text{planned}}, \qquad \Delta\text{OEE} \approx \Delta\text{A}$$

- Nominal production gain

$$\text{DA} = \text{Nominal rate.}\ \ \Delta\text{OEE.}\,\text{H}_{\text{run}}.\,\text{Margin(DA/unit)} \quad (6)$$

Year-1 ROI (DA):

$$\text{ROI} = \frac{\text{DA}_{\text{Downtime}} + \text{DA}_{\text{Maintenace}} + \text{DA}_{\text{Throughput}} - \text{Project}_{\text{Cost(DA)}}}{\text{Project}_{\text{Cost(DA)}}} \quad (7)$$

where A denotes availability; T_planned denotes planned operating time; T_downtime denotes total downtime (planned and unplanned); MTBF denotes mean time between failures; MTTR denotes mean time to repair; OEE denotes overall equipment effectiveness; n_(failures avoided) denotes the number of failures prevented by the ML model; C_CM denotes corrective maintenance cost per failure; C_PM denotes preventive maintenance cost per intervention; ΔMTTR denotes reduction in repair time; C_(labor/h) denotes labor cost per hour; ΔH_down denotes downtime hours avoided; C_h denotes cost per hour of downtime; ΔA denotes improvement in availability; ΔOEE denotes improvement in OEE; H_run denotes total operating hours after improvement; Nominal rate denotes nominal production rate; Margin (DA/unit) denotes profit margin per unit produced; DA denotes economic gain expressed in Algerian Dinars; and ROI denotes return on investment.

4. STANDARDS COMPLIANCE STRATEGY

Goal: Ensure an industrial ML solution that is interoperable, traceable, and auditable by aligning with core standards [14], [15].

a. Reference frameworks to follow
   - International Organization for Standardization – Standard 13374 (ISO-13374) (condition monitoring): Chain functions acquisition → data quality/preprocessing → state detection → health assessment → prognostics → advisory/alarms.
   - International Electrotechnical Commission Standard (IEC) 62264/International Society of Automation Standard 95 (ISA-95) (enterprise–control integration): Place components across levels 0–4 (sensors/programmable logic controller (PLC)/supervisory control and data acquisition (SCADA) → manufacturing execution system (MES)/edge → enterprise resource planning (ERP)/business intelligence (BI) with open platform communications – unified architecture (OPC UA)/message queuing telemetry transport (MQTT) interfaces.

b. Minimum implementation requirements
- Data (governance): Catalog and lineage (sensor → feature → prediction), versioned schemas, quality controls (completeness, freshness, drift), role-based access control (RBAC), transport layer security (TLS), encryption at rest, retention policies, and audit logs.
- Models (ML governance): Model registry (ID/version/data/limits), reproducibility (artifacts & configs), explainability (feature importance/SHAP for random forest (RF)/gradient boosting decision tree (GBDT), documented thresholds with hysteresis.
- Validation (IQ/OQ/PQ): IQ: environments and OPC UA/MQTT connectivity; OQ: offline tests with acceptance criteria (F1/AUC or RMSE/MAE/R^2); and PQ: online pilot (read-only), then canary 5–10% with clear rollback criteria.
- Operations (lightweight machine learning operations (MLOps)): Continuous monitoring (data/metric drift, latency, alert rates), scheduled recalibration/retraining, formal change control (review, regression tests, approval, versioned release).

c. Required deliverables
- Data management plan, IQ/OQ/PQ dossier (protocols & results), model registry, SOPs (alarm procedure, degraded/rollback mode), dashboards, and audit logs.

5. METHODOLOGY

a. Simulation-based data generation

In the absence of historical plant data, we generated three synthetic datasets to emulate realistic scenarios. Three datasets were created:
- Production dataset: Includes furnace temperature, pressure, production rate, energy consumption, and measured iron content (%). These parameters are critical for metallurgical processes, where small deviations can significantly impact final quality [16]–[18].
- Maintenance dataset: Includes vibration level (mm/s), machine temperature (°C), and a binary failure state (0 = normal, 1 = failure). These variables are typical indicators of mechanical and thermal stress in industrial equipment [17].
- Transport dataset: Includes shipment tonnage, transport distance, cost (in DZD), and travel time (hours). Transport logistics represent a major cost factor in large-scale industrial operations [16]–[18].

b. Machine learning models

Five widely used ML algorithms were selected for evaluation:
- Random forest: An ensemble method using multiple decision trees to improve predictive accuracy and control overfitting.
- XGBoost: An efficient gradient boosting algorithm optimized for performance and speed.
- LightGBM: A gradient boosting framework based on decision tree algorithms, designed for high efficiency and low memory usage.
- CatBoost: A gradient boosting algorithm particularly effective with categorical features and reducing overfitting.
- Multilayer perceptron (MLP): A feedforward artificial neural network capable of modeling complex non-linear relationships [14].
- Hybride RF+NN (stacking): The "RF+NN" approach combines a random forest (RF) as a base learner and a multilayer perceptron (MLP) as a meta-model

The global workflow of the proposed Machine learning-based optimization approach is depicted in Figure 3. It outlines the sequential steps from synthetic dataset generation to model evaluation and interpretation, aligning with both predictive maintenance and production quality improvement objectives [19].

c. Evaluation metrics

For classification (predictive maintenance) [20]–[22]:
- Accuracy: proportion of correct predictions

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

- Precision: proportion of correctly predicted positives among all predicted positives

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{9}$$

- Recall: proportion of correctly predicted positives among all actual positives

$$\text{Recall} = \frac{\text{TP}}{\text{TP+ FN}} \tag{10}$$

– F1-score: harmonic mean of precision and recall

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision X Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

– AUC-ROC: area under the receiver operating characteristic curve, indicating the model's discrimination ability.

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\text{d}(\text{FPR}) \tag{12}$$

with

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad , \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

where: TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives).

For regression (quality prediction) [23]–[25]:

– Root mean squared error (RMSE): measures prediction error magnitude

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{yi} - \hat{\text{y}}\text{i})^2} \tag{13}$$

– Mean absolute error (MAE): average of absolute prediction errors

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{14}$$

– Coefficient of determination (R^2): measures the proportion of variance explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{15}$$

With ŷi is the predicted values; yi is real values, and $\bar{y}_i$: is the average of the actual values.

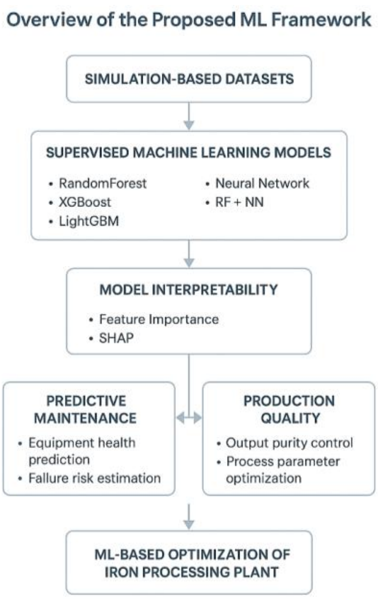

Figure 3. Methodological flowchart of the ML-based industrial optimization framework

d. Model training and validation

Data were split into training (70%) and test (30%) sets. K-fold cross-validation was applied to ensure robustness. SHapley Additive exPlanations (SHAP) analysis was used to interpret the impact of each feature on model predictions [26]–[28]. Figure 4 illustrates the global simulation and Machine learning workflow adopted in this study. The process begins with the generation and preprocessing of synthetic

datasets covering production, maintenance, and transport operations. These datasets are then used to train and evaluate five Machine learning algorithms for classification and regression tasks. The pipeline integrates performance analysis, interpretability techniques (feature importance and SHAP), and visual analytics to extract actionable insights. Finally, all results are compiled in a structured format for inclusion in industrial decision-making and scientific reporting. Data splitting:
– Chronological split to avoid leakage: train/validation/test = 70/15/15 (or 70/30 with CV on the train part).
– Validation: TimeSeriesSplit (k=5) for time-ordered series (production/transport) and Stratified K-Fold (k=5) for binary maintenance labels.
– The test set remains untouched until final reporting.

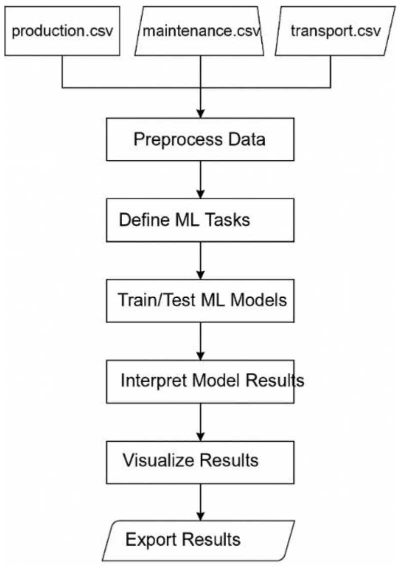

Figure 4. Flowchart of the simulation and machine learning optimization pipeline, from dataset input to result export

6. RESULTS AND DISCUSSION

6.1. Risk of bias from synthetic data and mitigation

Nature of the bias. Synthetic datasets may under-represent certain regimes (transients, rare failures), simplify correlations (non-linearities, cross-interactions), and overestimate stationarity (noise/drift lower than in the field). This can yield optimistic performance (high AUC/F1) and overfitting to simulation assumptions. Mitigation measures (applied/recommended):
- Domain randomization: vary noise amplitudes, sensor gain/offset, operating regimes (throughput, temperatures), and inject outliers/missing values → reduces overfitting to idealized distributions.
- Strict temporal validation: chronological splits + TimeSeriesSplit (prevents leakage) and report mean ± std across folds.
- Sensitivity & stress tests: re-evaluate metrics with ±10–20% changes in sampling rate, window size, and noise intensity.
- Ablations & parsimony: selectively remove variables (vibration and temperature) to verify model dependence; use penalties/early stopping for boosters/MLP.
- Calibration & thresholds: calibration curves and validation-based thresholding (Youden/F1), then lock thresholds before testing.
- Path to real data: offline replay of SCADA tags, then read-only POC (latency/alarms) and canary rollout (5–10%) with rollback criteria.

6.2. Classification results

Key insights: i) Vibration and machine temperature were the top predictors of failure; and ii) SHAP analysis confirmed their strong influence on classification output. Figures 5 to 15 offer an integrated visual assessment of the classification results obtained from all tested models. Figures 5 to 9 display the comparative accuracy, precision, recall, F1-score, and AUC metrics, clearly showing the dominance of tree-based methods (random forest, XGBoost, LightGBM, CatBoost, and RF+NN) over the Neural Network baseline. Figure 10 presents the ROC curves, where ensemble models achieve near-perfect class separation. The confusion matrix in Figure 11 further validates the random forest model's performance, revealing zero false positives and zero false negatives in the simulated dataset. Finally, Figure 12 illustrates the random forest feature-importance profile, identifying vibration level and machine temperature as the most impactful predictors.

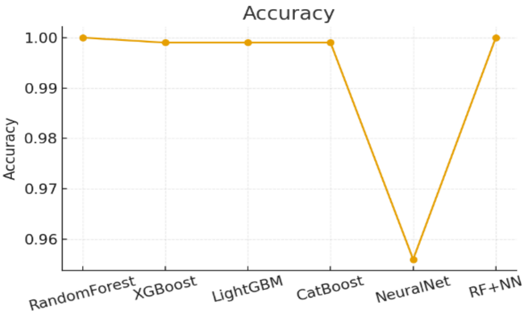

Figure 5. Comparative accuracy

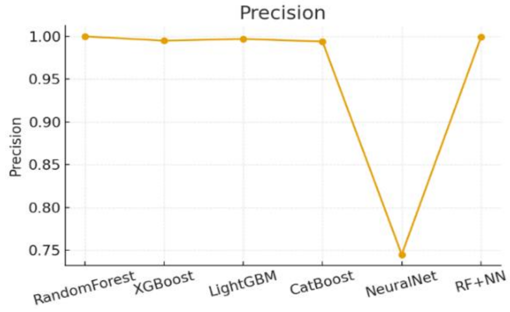

Figure 6. Comparative precision

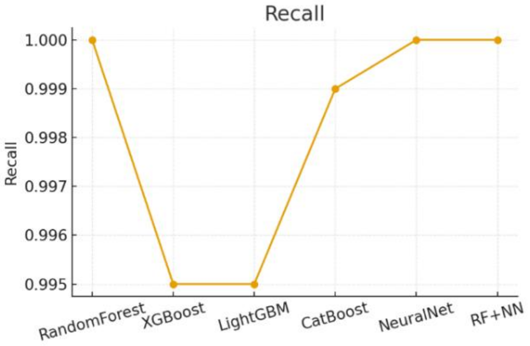

Figure 7. Comparative recall

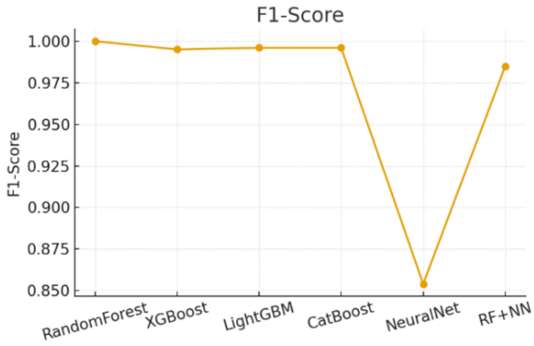

Figure 8. Comparative F1-Score

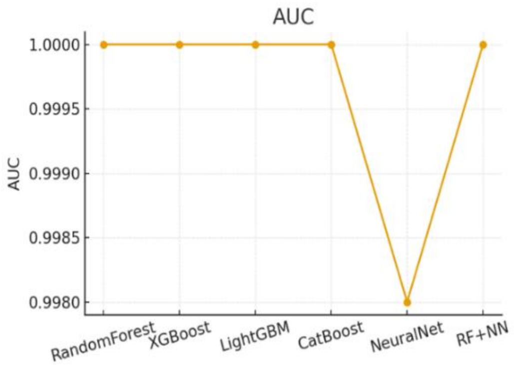

Figure 9. Comparative AUC

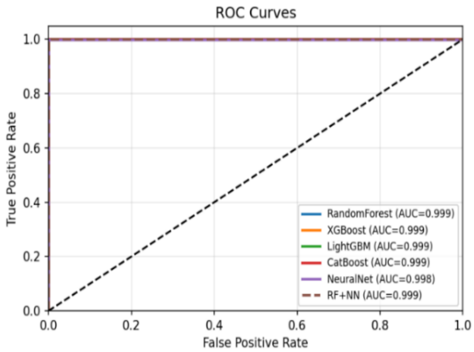

Figure 10. ROC curves for classification models

6.3. Regression results

Figures 13 to 15 present the comparative regression performance of all evaluated models based on RMSE, MAE, and R^2 metrics. The results show that tree-based ensemble methods (random forest, XGBoost, LightGBM, CatBoost) consistently outperform the Neural Network baseline. LightGBM achieves the lowest RMSE and lowest MAE, while also obtaining the highest R^2, indicating better prediction stability and reduced error variability. Conversely, the standalone Neural Network shows the weakest performance across all metrics, confirming its limited suitability for this type of industrial regression task. The hybrid RF+NN model achieves acceptable results but does not surpass the best individual tree-based models, suggesting that hybridization offers only marginal benefit for this regression problem.

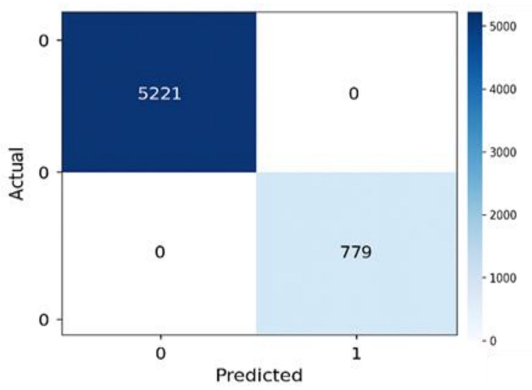

Figure 11. Confusion matrix for random forest

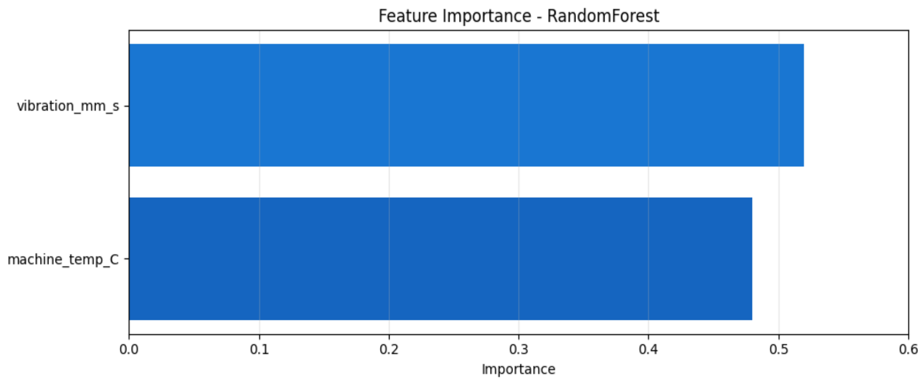

Figure 12. Feature importance random forest

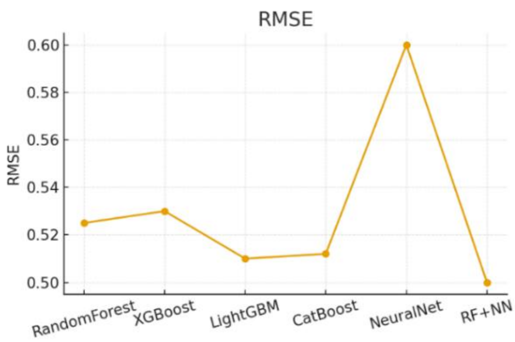

Figure 13. Comparative RMSE

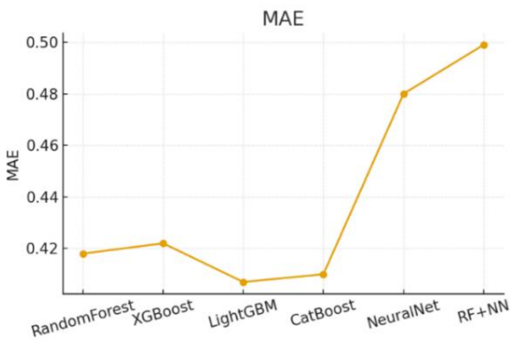

Figure 14. Comparative MAE

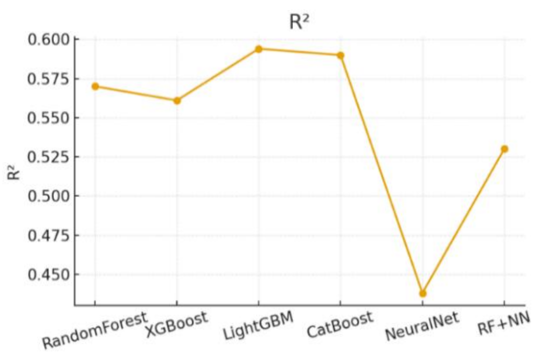

Figure 15. Comparative R^2

7. CONCLUSION

This study demonstrated the potential of machine learning techniques for both predictive maintenance and production quality optimization in the context of an iron processing plant. By using a simulation-based approach, we evaluated six widely adopted ML models, RF, XGBoost, LightGBM, CatBoost, feedforward neural network, and an RF+NN on 3 synthetic datasets representing production, maintenance, and transport processes.

The results of the classification tasks, aimed at predicting equipment failures, revealed that ensemble tree-based models achieved near-perfect predictive performance, with accuracy, precision, recall, and F1-score values exceeding 0.99, and AUC values close to 1. Such performance was explained by the well-defined failure patterns in the simulated maintenance data. In the regression tasks, which focused on predicting iron content as a measure of production quality, gradient boosting algorithms such as LightGBM and CatBoost outperformed other models, achieving the lowest RMSE and MAE values and the highest R^2 scores (above 0.59), even in the presence of simulated measurement noise.

The random forest algorithm achieved the best results due to its robustness to noise, ability to model nonlinear relationships, and effective feature selection, making it well-suited for industrial tabular data. The hybrid RF + MLP model provided only a slight improvement since the random forest already reached near-perfect performance (AUC ≈ 1.0). This marginal difference can be attributed to performance saturation and informational redundancy, as the MLP adds little value when the dataset is already well separated by the RF.

Beyond the numerical performance, the explainability analysis using Feature Importance and SHAP provided valuable insights into the driving factors behind the predictions. For predictive maintenance, vibration levels and machine temperature were identified as the dominant predictors of failure, aligning with real-world engineering knowledge. This transparency is essential for industrial adoption, as it supports trust in the models and facilitates decision-making by engineers and operators.

While the study is based on synthetic data, its methodology is designed to be transferable to real industrial environments. The combination of robust ML algorithms, interpretability methods, and a well-structured evaluation framework can directly support real-world deployments once historical operational data becomes available. The approach remains transferable to industrial settings via SCADA/IIoT integration (OPC UA/MQTT) and ONNX packaging at the edge, with an explicit link between technical metrics and plant KPIs (downtime, maintenance costs, throughput in DZD). The logical next steps are to replay SCADA data offline, run a read-only pilot to calibrate thresholds and latency/drift, execute a limited canary deployment with rollback criteria, and then industrialize (model registry, IQ/OQ/PQ validation, operator SOPs).

8. LIMITATIONS AND FUTURE WORK

a. Limitations of synthetic data

Although this study employs simulation-based datasets, it is important to acknowledge that synthetic data may not fully capture the variability, sensor noise, and rare failure events typically observed in real industrial environments. Such limitations can influence the generalizability of the trained models. Therefore, future work will focus on validating the proposed methodology using real operational data collected from the Bechar iron processing facility to ensure robustness and practical reliability.

b. Human in the loop and operator trust

In addition, incorporating human expertise into the decision-making process remains crucial for the successful industrial deployment of machine learning systems. By integrating explainability techniques such as SHAP and feature importance analysis, the framework enhances transparency and supports operator trust. Future implementations could further explore human-in-the-loop approaches, allowing domain experts to validate model outputs and provide corrective feedback.

c. Broader industrial objectives

Beyond predictive maintenance and production quality optimization, future research directions may extend toward broader objectives such as energy efficiency, process sustainability, and environmental impact reduction. These aspects align closely with current industrial priorities and would enhance the societal and operational relevance of ML-driven optimization systems in manufacturing and mineral processing sectors.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Tahri Mohammed University of Bechar and to the Laboratory of Control, Analysis, and Optimization of Electro-Energetic Systems (CAOSEE) for their institutional and technical support. The authors also acknowledge the valuable contributions of colleagues and experts whose constructive feedback helped improve the quality of this research. Special thanks are extended to the national initiatives for industrial and mining development in Algeria, particularly the strategic project for the exploitation of the Ghar Djebilet iron ore deposit, which inspired the case study framework of this work.

FUNDING INFORMATION

The authors declare that no external funding was received for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Lakhdari Lahcen	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
Habbab Mohamed		✓				✓		✓	✓	✓	✓	✓		
Alhachemi Moulay Abdellah	✓		✓	✓			✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

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

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] J. Jakubowski *et al.*, "Artificial intelligence approaches for predictive maintenance in the steel industry: a survey," *ArXiv*, May 2024, [Online]. Available: http://arxiv.org/abs/2405.12785
[2] L. Cummins *et al.*, "Explainable predictive maintenance: a survey of current methods, challenges and opportunities," *ArXiv*, Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.07871
[3] V. Pruckovskaja *et al.*, "Federated learning for predictive maintenance and quality inspection in industrial applications," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.11101
[4] M. R. Islam, S. Begum, and M. U. Ahmed, "Artificial intelligence in predictive maintenance: a systematic literature review on review papers," in *International Congress and Workshop on Industrial AI and eMaintenance 2023*, 2024, pp. 251–261. doi: 10.1007/978-3-031-39619-9_18.
[5] O. Ogunfowora and H. Najjaran, "Reinforcement and deep reinforcement learning-based solutions for machine maintenance planning, scheduling policies, and optimization," *Journal of Manufacturing Systems*, vol. 70, pp. 244–263, Oct. 2023, doi: 10.1016/j.jmsy.2023.07.014.
[6] A. Meddaoui, M. Hain, and A. Hachmoud, "The benefits of predictive maintenance in manufacturing excellence: a case study to establish reliable methods for predicting failures," *The International Journal of Advanced Manufacturing Technology*, vol. 128, no. 7–8, pp. 3685–3690, Oct. 2023, doi: 10.1007/s00170-023-12086-6.
[7] L. Lahcen and B. Bouchiba, "Fuzzy sliding mode controller for induction machine feed by three level inverter," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 9, no. 1, p. 55, Mar. 2018, doi: 10.11591/ijpeds.v9.i1.pp55-63.
[8] A. D. Scaife, "Improve predictive maintenance through the application of artificial intelligence: A systematic review," *Results in Engineering*, vol. 21, p. 101645, Mar. 2024, doi: 10.1016/j.rineng.2023.101645.
[9] P. Mallioris, E. Aivazidou, and D. Bechtsis, "Predictive maintenance in Industry 4.0: A systematic multi-sector mapping," *CIRP Journal of Manufacturing Science and Technology*, vol. 50, pp. 80–103, Jun. 2024, doi: 10.1016/j.cirpj.2024.02.003.
[10] P. Leitão, A. W. Colombo, and S. Karnouskos, "Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges," *Computers in Industry*, vol. 81, pp. 11–25, Sep. 2016, doi: 10.1016/j.compind.2015.08.004.
[11] I. Sittón-Candanedo, R. S. Alonso, S. Rodríguez-González, J. A. García Coria, and F. De La Prieta, "Edge computing architectures in industry 4.0: a general survey and comparison," in *Advances in Intelligent Systems and Computing*, vol. 81, 2020, pp. 121–131. doi: 10.1007/978-3-030-20055-8_12.
[12] P. Muchiri and L. Pintelon, "Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion," *International Journal of Production Research*, vol. 46, no. 13, pp. 3517–3535, Jul. 2008, doi: 10.1080/00207540601142645.
[13] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, Oct. 2006, doi: 10.1016/j.ymssp.2005.09.012.
[14] M. Wollschlaeger, T. Sauter, and J. Jasperneite, "the future of industrial communication: automation networks in the era of the internet of things and industry 4.0," *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, Mar. 2017, doi: 10.1109/MIE.2017.2649104.
[15] I. O. for Standardization, "ISO 14224:2016 Petroleum, petrochemical and natural gas industries — Collection and exchange of reliability and maintenance data for equipment," 2016.
[16] C. Tsallis, P. Papageorgas, D. Piromalis, and R. A. Munteanu, "Application-wise review of machine learning-based predictive maintenance: trends, challenges, and future directions," *Applied Sciences*, vol. 15, no. 9, p. 4898, Apr. 2025, doi: 10.3390/app15094898.

[17] L. Meitz *et al.*, "A literature review framework and open research challenges for predictive maintenance in industry 4.0," *Computers & Industrial Engineering*, vol. 206, p. 111193, Aug. 2025, doi: 10.1016/j.cie.2025.111193.

[18] K. I. Masani, P. Oza, and S. Agrawal, "Predictive maintenance and monitoring of industrial machine using machine learning," *Scalable Computing: Practice and Experience*, vol. 20, no. 4, pp. 663–668, Dec. 2019, doi: 10.12694/scpe.v20i4.1585.

[19] S. Elkateb, A. Métwalli, A. Shendy, and A. E. B. Abu-Elanien, "Machine learning and IoT – based predictive maintenance approach for industrial applications," *Alexandria Engineering Journal*, vol. 88, pp. 298–309, Feb. 2024, doi: 10.1016/j.aej.2023.12.065.

[20] P. P. Hanzelik, A. Kummer, and J. Abonyi, "Edge-computing and machine-learning-based framework for software sensor development," *Sensors*, vol. 22, no. 11, p. 4268, Jun. 2022, doi: 10.3390/s22114268.

[21] L. Lahcen and H. Mohamed, "Analysis and prediction of solar power plant energy production: a comparative study of machine learning algorithms," in *Lecture Notes in Networks and Systems*, 2025, pp. 536–543. doi: 10.1007/978-3-031-80301-7_58.

[22] H. Wang, W. Zhang, D. Yang, and Y. Xiang, "Deep-learning-enabled predictive maintenance in industrial internet of things: methods, applications, and challenges," *IEEE Systems Journal*, vol. 17, no. 2, pp. 2602–2615, Jun. 2023, doi: 10.1109/JSYST.2022.3193200.

[23] Y. Hua, "Machine learning algorithms for predictive maintenance in industrial systems," *International Journal of Unique and New Updates (IJUNU)*, vol. 3, no. 2, pp. 11–18, 2021, [Online]. Available: https://ijunu.com/index.php/journal/article/view/22/22

[24] J. Dalzochio *et al.*, "Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges," *Computers in Industry*, vol. 123, p. 103298, Dec. 2020, doi: 10.1016/j.compind.2020.103298.

[25] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers & Industrial Engineering*, vol. 150, p. 106889, Dec. 2020, doi: 10.1016/j.cie.2020.106889.

[26] I. ul Hassan, K. Panduru, and J. Walsh, "Predictive maintenance in industry 4.0: A review of data processing methods," *Procedia Computer Science*, vol. 257, pp. 896–903, 2025, doi: 10.1016/j.procs.2025.03.115.

[27] Z. Bao, C. Liu, H. Yang, J. Zhang, and Y. Li, "From theory to industry: A survey of deep learning-enabled bearing fault diagnosis in complex environments," *Engineering Applications of Artificial Intelligence*, vol. 163, p. 113068, Jan. 2026, doi: 10.1016/j.engappai.2025.113068.

[28] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, p. 106024, Nov. 2019, doi: 10.1016/j.cie.2019.106024.

BIOGRAPHIES OF AUTHORS

Lakhdari Lahcen is a professor in the Department of Electrical Engineering at Tahri Mohammed University of Bechar, Algeria. He received his Engineering degree and M.Sc. in Computer Science, and his Ph.D. in Automation and Industrial Computer Science from Tahri Mohammed University, Bechar, Algeria, in 2008, 2011, and 2018, respectively. He is a member of the Laboratory of Control, Analysis, and Optimization of Electro-Energetic Systems (CAOSEE). His research focuses on the application of Artificial Intelligence (AI), Machine Learning (ML), and intelligent automation techniques in industrial systems, with a particular emphasis on production optimization and predictive maintenance in metallurgical and mining industries. He has authored and co-authored several scientific publications in the fields of intelligent control, industrial optimization, and digital transformation of industrial processes. Dr. Lahcen also participates in collaborative research projects aiming to integrate Industry 4.0 technologies into large-scale industrial plants. He can be contacted at email: lakhdari.lahcen@univ-bechar.dz.

Mohamed Habbab was born in 1969 in Relizane, Algeria, received a state engineer degree in electronic engineering in 1994 from the University of Sciences and Technology of Oran (USTO), Algeria, and an M.Sc. degree and a Ph.D. degree in Electrical Engineering from the University Tahri Mohamed, Bechar, Algeria in 2015 and 2018, respectively. He is currently a professor of electrical engineering at the University Tahri Mohamed Bechar, Algeria. His research interests include power electronics, artificial intelligence, and their applications. He can be contacted at email: habbab.mohamed@univ-bechar.dz.

Alhachemi Moulay Abdellah was born on the 2nd of January 1999 in Adrar, Algeria. He received an M.S. degree in Automation and Industrial Computing from Tahri Mohammed University in 2022. He has been a member of the Control Analysis and Optimization of the Electro-Energetic Systems (CAOSEE) Laboratory since 2023. His research interests cover power electronics, control of multilevel converters, artificial intelligence, and renewable energy. He can be contacted at email: alhachemi.abdellah@univ-bechar.dz.