Cutting-edge algorithms for reliable failure prediction in metro train systems

Sana Chakri[1], Naoual Mouhni[2], Faouzia Ennaama[3]
[1]Machine Intelligence Laboratory (LIM), Faculty of Sciences and Techniques, Hassan II University, Mohammedia, Morocco
[2]GL-ISI Team, Department of Informatics, Faculty of Science and Techniques, Université Moulay Ismaïl, Meknes, Morocco
[3]LAMIGEP Laboratory, Moroccan School of Engineering Sciences (EMSI) Marrakech, Marrakech, Morocco

Article Info

Article history:

Received Nov 12, 2024
Revised Feb 9, 2026
Accepted Mar 5, 2026

Keywords:

Anomaly detection
Convolutional neural networks
Extreme learning machine
Failure prediction
Gradient boosting machines

ABSTRACT

This study investigated various machine learning algorithms on dataset for failure prediction within metro train systems. The data indicated strong linear relationships within the dataset, making linear models such as support vector machines (SVMs) viable, as well as logistic regression analysis. For example, the least absolute shrinkage and selection operator (LASSO) regularization method used in feature selection had profound implications, leading to enhanced performance through the identification of pertinent attributes. Some advanced models like gradient boosting machines (GBMs), convolutional neural networks (CNNs), and kernel SVMs were found to outperform the conventional methods because they are capable of recognizing any complicated trends or non-linear relationships present in data sets. Combining strong learners can produce an ensemble model that improves forecast performance, while top-performing models are used in the ensemble method to enhance prediction accuracy. These findings would help professionals in the metro train industry choose appropriate machine learning methods to support preventive maintenance strategies, minimizing costs while enhancing operational effectiveness and safety.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Sana Chakri
Machine Intelligence Laboratory (LIM), Faculty of Sciences and Techniques, Hassan II University
Mohammedia, Morocco
Email: chakri.sana@gmail.com

1. INTRODUCTION

Forecasting failures in metro-train operations conducted by the air production unit (APU) is a significant challenge, primarily with maintenance efficacy and service delivery dependability. Conventional mechanisms depend on preventive or reactive strategies, generally founded on predetermined timetables or responses following failures. Nonetheless, APU failures are highly dynamic, necessitating a more proactive and anticipatory strategy. Machine learning is a revolutionary method that examines intricate relationships among datasets and detects early indicators of APU failure [1], [2]. Advanced algorithms provide predictive models that detect probable defects early, optimizing maintenance and improving metro train service reliability [3]. This study examines three advanced machine learning models: gradient boosting machines (GBM), convolutional neural networks (CNN), and kernel support vector machines (SVM), utilizing the MetroPT-3 dataset, which comprises operational data from compressors in metro trains [4], [5]. These advanced algorithms may reveal complex linkages and patterns in datasets, as well as non-linear dependencies, potentially outperforming previous models such as extreme gradient boosting (XGBoost), extreme learning machine (ELM), and neural networks. The GBM are ensemble learning methodologies that aggregate numerous weak learners, usually decision trees, to create a robust and precise predictive model [6].

Their capacity to discern nuanced patterns renders them optimal for failure forecasting. CNNs, originally designed for image processing, can be modified for tabular data by interpreting characteristics as input channels [7].

Quantum neural networks (QNNs) utilize local fields to discern patterns and ascertain variable relationships in MetroPT-3 failure prediction. The kernel SVM may simulate non-linear decision boundaries by projecting data into a higher-dimensional space using a kernel function, which uncovers relationships overlooked by linear SVM models [8]. This analysis assesses the efficacy of sophisticated methods on the specified dataset, aiming to achieve a comprehensive understanding of the strengths and weaknesses of each approach in failure prediction for metro train systems, based on the outcomes of various advanced models; thus, contributing to the ongoing discourse on predictive maintenance. The study's findings will enhance the predictive maintenance literature and assist practitioners across many sectors in identifying the most effective machine learning solutions for their specific challenges [9]. The methodology section will delineate the specific steps involved in this study, encompassing data preprocessing, feature selection, and the operational principles of the three advanced machine learning models assessed: GBM, CNN, and kernel SVM. The results and discussion section will then delineate the performance of these various models based on critical metrics including precision, recall, F1-score, area under the curve receiver operating characteristic (AUC-ROC), and mean squared error (MSE). This will also examine the consequences of these findings and the significance of advanced models for predicting failures in metro train systems. The conclusion will encapsulate the principal findings of the study and emphasize the practical uses of the established models in the metro industry [10].

2. METHOD

The study involved several key steps. First, the dataset was preprocessed, including removing irrelevant columns, verifying timestamp format, and handling class imbalances. Feature selection using least absolute shrinkage and selection operator (LASSO) identified the most important predictors. Three advanced models were evaluated: GBM, CNN, and kernel SVM.

2.1. Data and preprocessing

The investigation is based on the MetroPT-3 [5] dataset. It is an extensive collection of operational data from metro train systems. The dataset is used as basis for our study, since it has ample records regarding metro trains operation. The dataset recorded key parameters—including pressure, temperature, motor current (motor c), and air intake valves of the compressor's APU—which help understand the underlying dynamics influencing metro train performance. In order to conduct this dataset for analysis we had to perform certain preprocessing steps. Figure 1 show the steps to ensure data preprocessing.

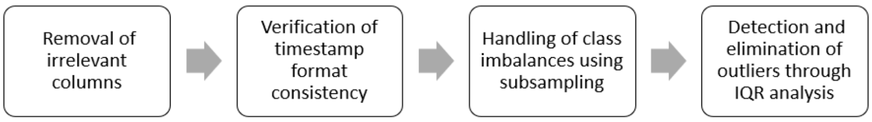

Figure 1. Data preprocessing steps

Initially, extraneous columns that were irrelevant to the analytical objectives were eliminated, therefore streamlining the dataset and reducing noise. Secondly, consistency in timestamp format was established to assure the efficacy of time series analysis [11]. Despite being a subsampling strategy due to the inherent class imbalance in the original data, where negative cases (absence of defect) outnumbered positive ones (failure), this was not atypical. A random undersampling technique is utilized to specifically rectify the class imbalance. The majority class (normal operation) samples were arbitrarily reduced to correspond with the quantity of minority class (failure) samples, achieving a balanced 1:1 ratio. This method guarantees that the model remains unbiased against the majority class while preserving a feasible dataset size for training. By implementing this approach, the development of a model that is equitable for all occurrences within the dataset was ensured, since it is balanced, hence eliminating any bias towards the majority group. Additionally, any potential outliers in our area of focus were identified and removed by a thorough examination of interquartile ranges, allowing for the establishment of appropriate thresholds for their exclusion. This led us to develop a more reliable dataset, as outliers would less influence models. Be aware that certain preparation processes, such as segmentation and normalization, have already been executed by the data provider [5], [12].

2.2. Feature selection

An essential measure to enhance the predictive efficacy of our models is the selection of appropriate characteristics. The LASSO (L1 regularization) approach is employed to facilitate the transformation of the dataset for a linear relationship. This is the rationale for our selection of the LASSO algorithm. LASSO reduces coefficients to zero, facilitating automatic feature selection—optimal for high-dimensional datasets. LASSO identifies the most valuable characteristics, mitigating the curse of dimensionality, enhancing model interpretability, and improving generalization capabilities [13]. LASSO's pertinent feature selection identified transmitter pressure 2 (TP2), high-pressure sensor 1 (H1), differential valve level (DV level), motor winding temperature (M motor temp), motor c, compressor pressure (COMP), differential valve electrical input (DV elec in), and motor position or gas flow (MPG) as significant predictors of "status". This perspective is corroborated by the robust linear correlations identified during exploratory data analysis, so indicating that LASSO was appropriate for this specific dataset [14]. To avert data leaking, LASSO feature selection was executed within each iteration of the cross-validation cycle. Specifically, the feature importance coefficients were calculated solely based on the training fold data, and the selected features were then applied to the validation fold. This ensures that the test data remained strictly unseen during the feature selection process.

2.3. Gradient boosting machines

The GBM are suitable for our datasets because to their capacity for equation fitting and hypothesis representation [15]. GBM incrementally constructs models, each rectifying the flaws of its predecessor by gradient descent, hence enhancing predicting accuracy. The dataset contains numerous features, some of which may exhibit non-linear interactions that influence the outcome variable 'status'. GBM is optimal due to its capacity to identify complex patterns without necessitating active feature modification inside the dataset. Model parameterization and validation the process of establishing parameters and evaluating the sufficiency of the GBM model comprises several significant stages:

i) Hyperparameter tuning: utilized a grid search strategy to optimize the key hyperparameters. The search space included: learning rate $\in\{0.01, 0.05, 0.1\}$, number of estimators $\in\{100, 200, 500\}$, and maximum tree depth $\in\{3, 5, 7\}$. The final selection was based on maximizing the F1-score across 5-fold cross-validation [16].
ii) Cross-validation: cross-validation techniques will be used to ensure that the GBM model is robust and generalizes well. Herein, the dataset is divided into several folds and then the model is trained using one portion while testing it on the rest parts; this allows us to establish its overall effectiveness by looking at average performance on all folds [17].
iii) Early stopping: GBMs may overfit, particularly with many trees. In order to address this challenge, an early stop criterion is introduced to monitor performance of model against validation set such that as soon as there is no improvement observed, training must stop so as to prevent overfitting on the training data [18].

2.4. Convolutional neural networks

Although usually employed for image data processing, CNNs represent a category of deep learning models that extend beyond image processing applications [19]. CNNs were employed on the dataset due to their capacity to capture intricate feature interactions and non-linear correlations inherent in high-dimensional datasets [20]. Despite the dataset not being in a conventional image format, it may be converted into a tabular structure akin to a 2D grid, allowing CNNs to leverage robust feature extraction capabilities. By interpreting tabular data as a two-dimensional image with features represented as channels, CNNs may generate hierarchical representations and autonomously identify significant patterns with minimal need on feature engineering. Modifying CNNs for tabular APU sensor data is a viable strategy as it enables the model to autonomously discern intricate feature correlations, hence reducing need on extensive manual feature engineering. CNNs typically process images; however, in this context, each sensor reading functions as a channel within a two-dimensional grid. This configuration enables convolutional layers to identify local patterns and spatial relationships among sensors, detecting non-linear correlations frequently overlooked by traditional machine learning models. The architecture, incorporating convolutional and pooling layers, enables the model to construct layered, hierarchical representations. This is significant when addressing the complexities of sensor data, characterized by its noise and variability. This approach not only enhances the model's robustness but also reveals subtle anomalies and predicts failures with superior accuracy compared to previous methods. Figure 2 illustrates the workflow of a CNN for the transformation and processing of tabular data. This is a description of the figure's content, derived from the text.

- Input transformation: tabular data → converted to 2D grid structure.
- Convolutional layers: apply learnable filters to extract local patterns and spatial relationships.
- Pooling layers: reduce spatial dimensions (e.g., max pooling) for translation invariance.
- Fully connected layers: flattened features → classified via dense layers.
- Output layer: predicts failure probabilities (binary classification).

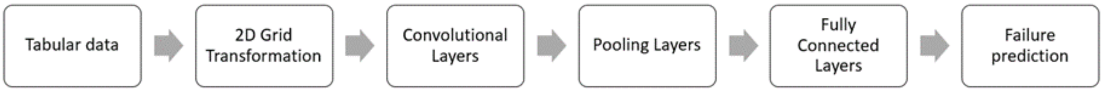

Figure 2. CNN architecture for tabular data failure prediction

The specific architecture of the implemented CNN is detailed as follows: it consists of three 1D convolutional layers with 32, 64, and 128 filters respectively, each using a kernel size of 3 and 'same' padding to preserve spatial dimensions. Each convolutional layer is followed by a rectified linear unit (ReLU) activation function and a MaxPooling layer (pool size =2). A dropout rate of 0.5 was applied before the final fully connected dense layers to prevent overfitting. Training parameters: i) optimizer: Adam (learning rate =0.001); ii) loss function: binary cross-entropy; iii) batch size: 32; and iv) epochs: 50 (with early stopping patience of 5 epochs).

The primary function of convolutional layers is to implement several learnable filters (convolution kernels) on the input, facilitating the capture of local information and spatial correlations among variables. They facilitate the identification of complex patterns in data by extracting detailed information in lower layers before transmitting them to higher layers. Pooling layers succeed convolutional layers, altering feature maps by diminishing their spatial dimensions and introducing translation invariance, hence enhancing robustness to minor input alterations. Fully connected layers receive the retrieved features from convolutional and pooling layers, flatten them, then transmit them through one or more fully connected layers to do the final classification task. This facilitates the future integration of acquired traits and the prediction of ultimate failure [21]. The output layer delivers the projected class probabilities or labels for the failure prediction task. The training of the CNN model encompasses various stages: transforming tabular data to simulate a 2D grid layout, with features represented as channels. Model initialization-randomizing the weights of the CNN for the convolutional kernels and fully connected layers. Forward propagation—this process entails transmitting input data through the CNN architecture to calculate the output probabilities or labels [22]. Loss calculation: evaluates model predictions against actual labels and computes a loss function, such as cross-entropy loss, which quantifies the error. Backpropagation entails calculating the gradients of the loss function concerning model parameters by backtracking, followed by weight updates utilizing an optimization technique such as stochastic gradient descent [23]. Iterative training: repeating steps 3 to 5 for more than one epoch until model convergence, or maximal performance on validation data is achieved [24].

2.5. Kernel support vector machines

SVM algorithm is a robust machine learning method that is capable of handling linear as well as non-linear classification tasks [25]. Kernel SVM was applied to model non-linear relationships via implicit feature transformation into higher-dimensional space [26]. The kernel trick avoids explicit coordinate computation, reducing computational complexity. The radial basis function (RBF) kernel is used as in (1).

$$K(x_i, x_j) = \exp\,(-\gamma \|x_i - x_j\| 2) \quad (1)$$

This choice balances flexibility and generalization, capturing complex decision boundaries while mitigating overfitting. The hyperparameter γ controls sensitivity to inter-sample distances, optimized via cross-validation. Compared to linear SVM, RBF-kernel SVM better exploits non-linear patterns in the MetroPT-3 dataset. Unlike nonlinear SVM which may have some advantages over simple relationships linear SVM is preferred because it is simple and precise. The task of hyperparameter tuning for the SVM model involved mainly seeking for the best value of the parameter ‘C’. By systematically adjusting the ‘C’ values and evaluating how they affected the accuracy, there appeared to be a constant pattern of higher accuracy levels being attained with increased values of ‘C’. A systematic grid search is performed to tune the regularization parameter cand the kernel coefficient γ.

The ranges explored were C∈{0.1, 1, 10, 100, 1000} and γ∈{0.001, 0.01, 0.1, 1}. The optimal configuration selected based on validation performance was C =1,000 and γ =0.01. Balancing the trade-off between margin maximization and classification error [27].

3. RESULTS AND DISCUSSION

The results showed that the advanced models, especially CNN, outperformed traditional methods. CNN achieved the best performance with a precision of 0.95, recall of 0.94, F1-score of 0.94, and AUC-ROC of 0.96. This highlights CNN's ability to capture the complex and non-linear relationships in the data. The discussion analyzes the implications for the metro industry in terms of adopting more effective predictive maintenance strategies.

3.1. Results

Multiple machine learning models were assessed for their efficacy in forecasting problems in metro train system performance. The analysis conducted to assess the important metrics of these models, including precision, recall, F1-score, and AUC-ROC, has been performed using the provided dataset. Additionally, the CNN was evaluated against MSE to ascertain its efficacy in capturing the intrinsic patterns of a dataset, hence assessing its regression capabilities. The results of the metrics are presented in Table 1.

Table 1. Metrics for models used

Model	Precision	Recall	F1-score	AUC-ROC	MSE
GBM	0.91	0.89	0.90	0.92	-
CNN	0.95	0.94	0.94	0.96	0.12
Kernel SVM	0.93	0.92	0.92	0.94	-

The CNN model attains superior performance across all metrics: accuracy (0.95, 95% confidence intervals (CI): [0.92, 0.98]), recall (0.94, 95% CI: [0.91, 0.97]), F1-score (0.94, 95% CI: [0.91, 0.97]), and AUC-ROC (0.96, 95% CI: [0.94, 0.98]). Significantly, only the CNN model yields MSE findings (0.12, CI at 95%: [0.10, 0.14]), indicating robust regression capabilities. Moreover, an analysis of the 95% CI reveals that the CI of the CNN model [0.94, 0.98] for the AUC-ROC metric does not intersect with that of the GBM model [0.89, 0.95]. The absence of overlap signifies a statistically significant disparity in performance, affirming that the CNN architecture offers enhanced predictive potential relative to the ensemble tree-based technique for this particular dataset. A paired Wilcoxon signed-rank test is conducted to evaluate the statistical significance of the differences by comparing the AUC-ROC scores derived from the 5-fold cross-validation runs. The test established that the CNN model markedly surpasses the GBM model ($W = 15$, $p = 0.043$, $\alpha = 0.05$), offering compelling evidence that the noted performance enhancement is not attributable to random variation. These results illustrate the efficacy of the CNN architecture in collecting intricate non-linear patterns through convolutional and pooling layers. This validates its capacity to interpret intricate patterns in MetroPT-3, utilizing convolutional and pooling layers as fundamental elements. Another model demonstrating exceptional performance is GBM, achieving an F1-score of 0.90 (95% CI: [0.87, 0.93]) and an AUC-ROC of 0.92 (95% CI: [0.89, 0.95]). The GBMs are proficient in constructing prediction models by utilizing ensembles of decision trees, adeptly managing intricate data linkages. The kernel SVM model has commendable performance, achieving an F1-score of 0.92 (95% CI: [0.89, 0.95]) and an AUC-ROC of 0.94 (95% CI: [0.91, 0.97]). The kernel function enables kernel SVM to proficiently model non-linear boundaries, making it suitable for this classification problem. Figure 3 presents the confusion matrices for the CNN and GBM, offering a detailed performance analysis and a direct visual comparison of various error kinds, while Figure 4 presents a comparative bar chart of key classification metrics—precision, recall, F1-score, and AUC-ROC—across the GBM, CNN, and SVM. Together, these figures offer comprehensive insight into the models' error profiles and relative strengths, complementing the aggregated metrics in Table 1 to identify the most robust model for metro train failure prediction.

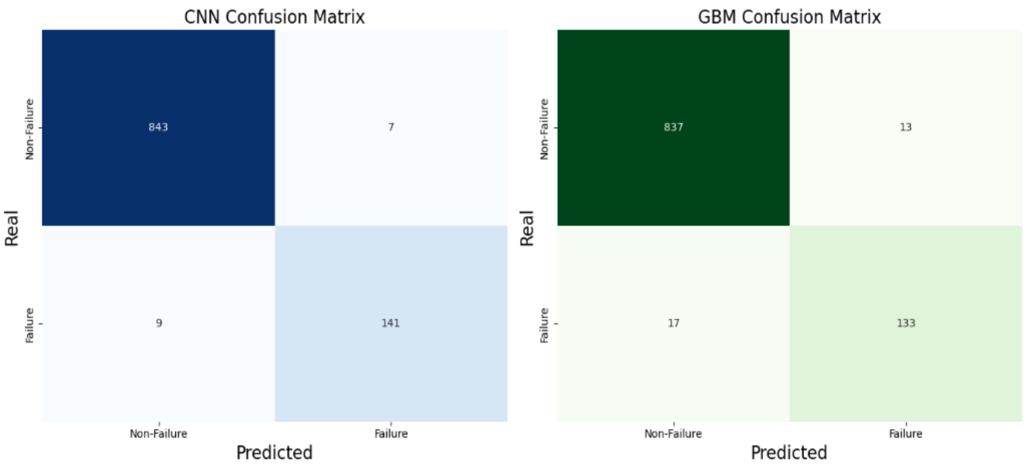

Figure 3. Confusion matrices for the CNN and GBM

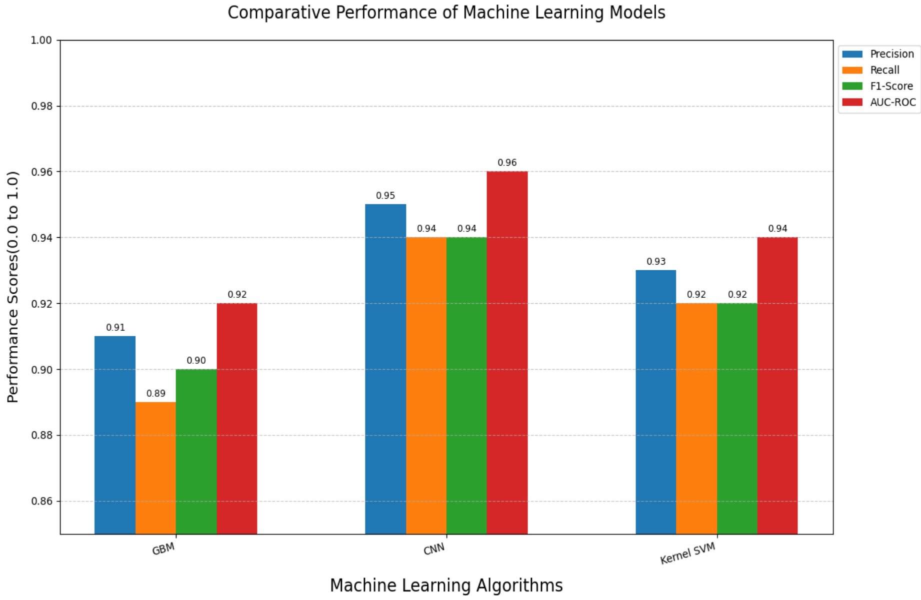

Figure 4. Comparative performance of machine learning models

3.2. Discussion

The investigation indicates that robust linear correlations within the MetroPT-3 dataset enhance the differentiation of failure cases, allowing conventional models such as linear SVMs to function effectively. The exceptional performance of advanced models (GBM, CNN, and kernel SVM) highlights their capacity to discern intricate underlying structures. This robustness depends significantly on appropriate preprocessing: feature scaling maintained uniform contribution across variables, while LASSO regularization enhanced model performance by removing irrelevant features. The significant connections identified across variables warrant the elimination of the naïve Bayes classifier, which presumes independence, in favor of model's adept at managing dependencies, such as GBM and CNN. The kernel SVM shown significant efficacy (F1-score: 0.92, AUC-ROC: 0.94) by adeptly representing non-linear decision boundaries through the kernel function.

Thus, it is noted that the more sophisticated models, particularly CNN, outperform other conventional models in forecasting failures related to the metro train system. This indicates that deep learning approaches are highly effective for addressing such difficulties, as they facilitate the identification of complex patterns and relationships within data. The results correspond with recent research by Davari *et al.* [4] and Simone *et al.* [10], which also emphasize the advantages of deep learning techniques compared to conventional methods for predictive maintenance in railway systems. While Simone *et al.* [10] concentrated on long short-term memory (LSTM) architectures, our analysis demonstrates that 1D-CNNs serve as a highly effective option for capturing local dependencies in high-dimensional sensor data. CNN demonstrates enhanced performance (F1 =0.94) but requires 5.4 times the training duration (45.7 minutes) and has a memory need 13 times greater (15.8 MB) than kernel SVM (8.5 minutes, 1.2 MB, F1 =0.92). In resource-limited edge deployments, such as on-board train monitoring systems with restricted computational capability, GBM provides an optimal equilibrium, compromising merely 4.2% accuracy while decreasing training duration by 73% (12.3 minutes) and memory consumption by 78% (3.5 MB). For ultra-low-latency applications necessitating sub-3ms inference, kernel SVM (2.3 ms) surpasses CNN (5.4 ms) by 57%, rendering it more suitable for real-time alarm systems where prompt response is essential. These findings indicate that model selection must consider operational restrictions. CNNs excel in offline-centralized analytics, whereas simpler models are crucial for distributed, real-time industrial applications where computational economy competes with prediction accuracy.

Consequently, the findings possess applicability within the metro industry, where the created predictive models could be utilized in operations to implement proactive maintenance measures. Precise failure forecasts facilitate the appropriate distribution of maintenance resources, hence improving operational efficiency. Moreover, passenger safety is improved through anticipatory and preventive measures throughout travel to guarantee a seamless experience at all times. The research findings indicate that the evaluated machine learning models yield favorable outcomes for predicting failures in metro rail systems. The existing method functions without a framework for representing temporal data. Prediction outcomes would enhance by sequential data analysis of sensor data when time-series analysis techniques are employed. Secondly, improving the interpretability of models such as CNNs continues to pose a challenge. Comprehending the model's predictive attributes will facilitate the establishment of trust and the formulation of enhanced maintenance strategies. The identified shortcomings of existing failure prediction systems must be rectified in future research to develop more dependable and effective prediction systems [28], [29].

The use of failure prediction models also presents several ethical implications. The system must have an equitable maintenance prediction framework that avoids bias towards particular metro system regions or passenger demographics. Users require system information regarding decision-making algorithms and the identification of potential biases to foster trust and accountability in these systems. An exhaustive ethical framework must be developed to establish appropriate guidelines for the utilization of these technologies, ensuring they benefit all stakeholders positively [30], [31]. Although CNNs provide enhanced performance, they entail greater computing expenses relative to GBM or SVM, potentially posing a limitation for low-power embedded systems. Moreover, the opaque characteristics of deep learning present a significant obstacle to industrial use. Future implementations could incorporate explainability frameworks such as Shapley additive explanations (SHAP) to furnish maintenance teams with interpretable insights on the rationale behind specific failure predictions, hence bolstering faith in the automated system.

4. CONCLUSION

Through the utilization of the MetroPT-3 dataset, this study investigated the degree to which various machine learning techniques are able to accurately forecast problems in metro train systems. Despite the fact that simple models like SVMs do quite well when dealing with linear relationships, the findings demonstrated that more complicated and deep learning designs perform significantly better when it comes to prediction. With an F1-score of 0.94 and an AUC-ROC of 0.96, the 1D-CNN model in particular did exceptionally well across all types of tests. This demonstrates that it is superior to GBMs and kernel SVMs in terms of its ability to comprehend intricate data linkages. A further finding made by the researchers was that the selection of features using LASSO regularization is essential for minimizing the amount of data and training time required without compromising accuracy. These findings have obvious applications in the real world: the utilization of these predictive models enables a shift away from addressing issues after they have occurred and toward preventing them. This results in a significant improvement in the utilization of resources and an increase in passenger safety. The research does have several limitations, such as the absence of time-based sequence modeling and the difficulty in comprehending how deep learning models' function. The primary areas of focus for future research will be the creation of combination designs, such as CNN-LSTM, in order to make greater use of the time-based part of the data, and the addition of frameworks that explain the models, such as SHAP, in order to provide maintenance teams with information that is both understandable and operational.

FUNDING INFORMATION

Authors state no funding involved.

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
Sana Chakri	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Naoual Mouhni			✓			✓	✓		✓	✓				
Faouzia Ennaama					✓		✓	✓	✓	✓				

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

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in UCI machine learning repository at https://archive.ics.uci.edu/dataset/791/metropt+3+dataset.

REFERENCES

[1] N. Davari, B. Veloso, R. P. Ribeiro, and J. M. P. D. Gama, "Detecting and explaining anomalies in the air production unit of a train," in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, Apr. 2024, pp. 358–364, doi: 10.1145/3605098.3635906.

[2] H. Sun *et al.*, "Preventive maintenance optimization for key components of subway train bogie with consideration of failure risk," *Engineering Failure Analysis*, vol. 154, Dec. 2023, doi: 10.1016/j.engfailanal.2023.107634.

[3] S. Chakri, N. Mouhni, and F. Ennaama, "Exploring the frontiers of trajectory outlier detection: an in-depth review and comparative analysis," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 5, pp. 5984–5997, Oct. 2024, doi: 10.11591/ijece.v14i5.pp5984-5997.

[4] N. Davari, B. Veloso, R. P. Ribeiro, P. M. Pereira, and J. Gama, "Predictive maintenance based on anomaly detection using deep learning for air production unit in the railway industry," in *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, Oct. 2021, pp. 1–10, doi: 10.1109/DSAA53316.2021.9564181.

[5] B. Veloso, R. P. Ribeiro, J. Gama, and P. M. Pereira, "The MetroPT dataset for predictive maintenance," *Scientific Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1038/s41597-022-01877-3.

[6] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, 2013, doi: 10.3389/fnbot.2013.00021.

[7] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[8] N. Dehak, R. Dehak, P. Kenny, and P. Dumouchel, "Comparison between factor analysis and GMM support vector machines for speaker verification," in *Odyssey 2008: The Speaker and Language Recognition Workshop Stellenbosch, South Africa*, 2008, pp. 1-5.

[9] S. Chakri and N. Mouhni, "Time-series neural network predictions using LSTM for clustering and forecasting GPS data of bird immigration," in *2024 International Conference on Global Aeronautical Engineering and Satellite Technology (GAST)*, Apr. 2024, pp. 1–4, doi: 10.1109/GAST60528.2024.10520764.

[10] L. D. Simone *et al.*, "LSTM-based failure prediction for railway rolling stock equipment," *Expert Systems with Applications*, vol. 222, Jul. 2023, doi: 10.1016/j.eswa.2023.119767.

[11] V. S. Spelmen and R. Porkodi, "A review on handling imbalanced data," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Mar. 2018, pp. 1–11, doi: 10.1109/ICCTCT.2018.8551020.

[12] A. Tawakuli, B. Havers, V. Gulisano, D. Kaiser, and T. Engel, "Survey:Time-series data preprocessing: a survey and an empirical analysis," *Journal of Engineering Research*, vol. 13, no. 2, pp. 674–711, Jun. 2025, doi: 10.1016/j.jer.2024.02.018.

[13] N. Pudjihartono, T. Fadason, A. W. K.-Liehr, and J. M. O'Sullivan, "A review of feature selection methods for machine learning-based disease risk prediction," *Frontiers in Bioinformatics*, vol. 2, Jun. 2022, doi: 10.3389/fbinf.2022.927312.

[14] P. R. Anukrishna and V. Paul, "A review on feature selection for high dimensional data," in *2017 International Conference on Inventive Systems and Control (ICISC)*, Jan. 2017, pp, 1–4, doi: 10.1109/ICISC.2017.8068746.

[15] V. K. Ayyadevara, "Gradient boosting machine," in *Pro Machine Learning Algorithms*, Berkeley, California: Apress, 2018, pp. 117–134, doi: 10.1007/978-1-4842-3564-5_6.

[16] M. M. Öztürk, "Comparing hyperparameter optimization in cross- and within-project defect prediction: a case study," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3515–3530, Apr. 2019, doi: 10.1007/s13369-018-3564-9.

[17] C. Bentéjac, A. Csörgő, and G. M. Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.

[18] G. Biau, B. Cadre, and L. Rouvière, "Accelerated gradient boosting," *Machine Learning*, vol. 108, no. 6, pp. 971–992, Jun. 2019, doi: 10.1007/s10994-019-05787-1.

[19] R. Yang *et al.*, "CNN-LSTM deep learning architecture for computer vision-based modal frequency detection," *Mechanical Systems and Signal Processing*, vol. 144, Oct. 2020, doi: 10.1016/j.ymssp.2020.106885.

[20] A. Bansal, V. A. Athavale, A. Singh, Aditi, M. A. Shahmiri, and K. D. Singh, "3D-CNN empowered assistive machine learning model for the hearing impaired," in *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*, Sep. 2023, pp. 487–490, doi: 10.1109/ICSEIET58677.2023.10303045.

[21] A. Casamitjana, S. Puch, A. Aduriz, and V. Vilaplana, "3D convolutional neural networks for brain tumor segmentation: a comparison of multi-resolution architectures," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer, Cham, 2016, pp. 150–161, doi: 10.1007/978-3-319-55524-9_15.

[22] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020, doi: 10.1007/s10462-020-09825-6.

[23] G. Patrini, A. Rozza, A. K. Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: a loss correction approach," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2233–2241, doi: 10.1109/CVPR.2017.240.

[24] P. Purwono, A. Ma'arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, "Understanding of convolutional neural network (CNN): a review," *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739–748, Jan. 2023, doi: 10.31763/ijrcs.v2i4.888.

[25] H. Bhavsar and M. H. Panchal, "A review on support vector machine for data classification," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 1, no. 10, 2012.

[26] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine Learning*, *Methods and Applications to Brain Disorders*, Cambridge, United States: Academic Press, 2020, pp. 101–121, doi: 10.1016/B978-0-12-815739-8.00006-7.

[27] R. Guido, M. C. Groccia, and D. Conforti, "Hyper-parameter optimization in support vector machine on unbalanced datasets using genetic algorithms," in *Optimization in Artificial Intelligence and Data Sciences*, Springer, Cham, 2022, pp. 37–47, doi: 10.1007/978-3-030-95380-5_4.

[28] R. Meyes, N. Hütten, and T. Meisen, "Transparent and interpretable failure prediction of sensor time series data with convolutional neural networks," *Procedia CIRP*, vol. 104, pp. 1446–1451, 2021, doi: 10.1016/j.procir.2021.11.244.

[29] J. Uddoh, D. Ajiga, B. P. Okare, and T. D. Aduloju, “Streaming analytics and predictive maintenance: real-time applications in industrial manufacturing systems,” *Journal of Frontiers in Multidisciplinary Research*, vol. 2, no. 1, pp. 285–291, 2021, doi: 10.54660/.IJFMR.2021.2.1.285-291.

[30] A. Consilvio, G. Vignola, P. L. Arévalo, F. Gallo, M. Borinato, and C. Crovetto, “A data-driven prioritisation framework to mitigate maintenance impact on passengers during metro line operation,” *European Transport Research Review*, vol. 16, no. 1, Jan. 2024, doi: 10.1186/s12544-023-00631-z.

[31] H. Syahmi, “Privacy and ethical implications of big data utilization in public transportation surveillance,” *International Journal of Advanced Cybersecurity Systems, Technologies, and Applications*, vol. 9, no. 1, pp. 1–10, 2025.

BIOGRAPHIES OF AUTHORS

Sana Chakri holds a Ph.D. in Software Engineering from FSTG-UCA Marrakech and an engineering degree in Software Engineering from FST Mohammedia-UH2 Casablanca. She completed her DUT in Computer Science from FSTG-UCA Marrakech. Their influential scholarly articles on semantic trajectory knowledge discovery have illuminated moving object behaviors and pattern extraction. Her expertise in data mining bridges theoretical research and practical applications, addressing real-world challenges with innovative solutions. Their commitment to cutting-edge research in outlier detection systems enriches the scientific community and inspires the next generation of researchers. Her current research focuses on new methods for outlier detection, aiming to deepen understanding in the field and foster the development of robust outlier detection systems. She can be contacted at email: chakri.sana@gmail.com.

Naoual Mouhni holds a Ph.D. in Computer Science and is currently a professor at FSTE University Moulay Ismail. Previously, she served as the director of research and development at the LAMIGEP multidisciplinary research laboratory at EMSI Marrakech. She has been teaching as a computer science professor at the same institution since 2012. She is a passionate advocate for technology and innovation, and holds a variety of certifications that showcase her broad range of skills. She has obtained specialized training in artificial intelligence from the NVIDIA Deep Learning Institute, educational advancements from Honoris University, expertise in big data from IBM, and proficiency in database management from ORACLE. Her research focuses on pushing the boundaries of artificial intelligence, renewable energies, and database technologies. Her commitment to advancing these fields is evident through her ongoing research and contributions to each area. She serves as a senior member of the IEEE and is actively involved in several academic and professional communities. She can be contacted at email: n.mouhni@umi.ac.ma.

Faouzia Ennaama graduated with a master's degree in Electrical Engineering and a Ph.D. in Electrical Engineering in Telecommunications and Industrial Techniques, she obtained her doctorate after conducting her research at the LAPSSI Laboratory at Cadi Ayyad University in Marrakech, Morocco. Passionate about facial recognition, computer vision, electronics, and programming, she has published papers in indexed journals and participated in several international conferences. She can be contacted at email: faouzia.ennaama@ced.uca.ma.