International Journal of Innovative Research and Scientific Studies, 8(6) 2025, pages: 1868-1876

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Machine learning-based predictive maintenance system for urban heating networks for real-time failure detection and analysis

Dauren Darkenbayev[1,3*], Uzak Zhapbasbayev[2], Gulnar Balakayeva[1]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan.
[2]Laboratory "Modeling in Energy Sector" Satbayev University, Almaty, Kazakhstan.
[3]Kazakh National Women's Teacher Training University, Almaty, Kazakhstan.

Corresponding author: Dauren Darkenbayev (*Email: dauren.darkenbayev1@gmail.com*)

Abstract

This study presents a comprehensive framework for predictive maintenance of urban heat supply networks utilizing advanced machine learning algorithms. The primary objective is to enable early detection of potential system failures, thereby improving operational reliability and minimizing unplanned downtimes. A synthetically generated dataset of 10,000 records was employed, simulating real – world operational parameters such as temperature, pressure, flow rate, and vibration, sampled at 5–minute intervals to replicate actual monitoring conditions. Data preprocessing involved outlier removal using the interquartile range (IQR) method, normalization through Min-Max scaling, and imputation of missing values, ensuring data quality and consistency. Feature importance was further analyzed using SHAP values to enhance interpretability and identify critical predictors influencing system behavior. Five machine learning models – Logistic Regression, Support Vector Machine (SVM), Random Forest, Artificial Neural Networks (ANN), and Gradient Boosting (LightGBM) – were implemented and evaluated using 10 – fold cross – validation. The Gradient Boosting model demonstrated superior performance, achieving an accuracy of 99.9%, F1-score of 0.999, ROC-AUC of 1.0, and LogLoss of 0.004. Logistic Regression and Random Forest also performed well (AUC = 1.0, F1 = 0.999), whereas SVM and ANN exhibited limited predictive capabilities (AUC ≈ 0.50, F1 = 0.038 and 0.632, respectively). These results underscore the robustness of Gradient Boosting in modeling complex nonlinear relationships and its applicability for real-time anomaly detection in heating systems. The proposed framework holds significant practical potential for integration into existing monitoring infrastructures, facilitating proactive maintenance planning, optimizing resource allocation, and reducing operational costs. Future research will focus on validating the approach with real – world datasets and exploring hybrid machine learning architectures to enhance model generalizability and resilience.

Keywords: Failure detection, Heat supply networks, Machine learning, Predictive maintenance, Real-time monitoring.

DOI: 10.53894/ijirss.v8i6.10041
Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant Number: BR24992907 for 2024-2026).
History: Received: 23 July 2025 / **Revised:** 25 August 2025 / **Accepted:** 27 August 2025 / **Published:** 19 September 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Heat supply networks (HSNs) are critical components of urban infrastructure, providing thermal energy for residential, commercial, and industrial applications. Ensuring the reliability of these systems is essential, as failures can lead to significant economic losses, environmental damage, and societal disruptions, particularly in regions with extreme climatic conditions [1, 2]. Predictive maintenance, which utilizes data-driven techniques to forecast potential failures, has emerged as a promising approach to improving operational efficiency and minimizing unplanned downtime [3].

Traditional maintenance strategies in HSNs, such as reactive and preventive methods, are increasingly inadequate in addressing the complexity of modern networks. Reactive approaches result in costly emergency repairs, while preventive maintenance often fails to align with actual system conditions, leading to suboptimal resource utilization [4]. Machine learning (ML) techniques have gained substantial attention for their ability to analyze large-scale operational data, detect early signs of system anomalies, and enable proactive maintenance [5, 6].

Recent studies have applied various ML algorithms to predictive maintenance tasks in energy systems. For example, Guo, et al. [7] applied support vector machines (SVM) to anomaly detection in district heating networks but achieved moderate accuracy. Müller, et al. [8] used random forest (RF) models for predicting heat load variations; however, their approach faced challenges with overfitting and scalability. Artificial neural networks (ANN) have also been explored [9] demonstrating potential but suffering from high computational costs and poor interpretability.

Gradient boosting (GB), particularly LightGBM, has shown remarkable success in diverse domains such as healthcare [10], financial analytics [11] and industrial fault detection [12] due to its ability to model complex nonlinear relationships with high accuracy and computational efficiency. Despite these advantages, its application to HSNs remains limited. Previous works by Zhang and Li [13] and Li, et al. [14] focused on thermal demand forecasting and power grid applications, respectively, without addressing fault detection or real-time monitoring in HSNs.

This study fills this gap by proposing a LightGBM-based predictive maintenance framework tailored to HSNs, integrating multi-dimensional operational data (temperature, pressure, flow rate, vibration) and applying SHAP value analysis for improved model interpretability. A synthetic dataset comprising 10,000 records was generated to simulate real-world conditions, and five ML models (Logistic Regression, RF, SVM, ANN, and LightGBM) were comparatively evaluated using 10-fold cross-validation.

The novelty of this research lies in demonstrating LightGBM's superior predictive performance (accuracy = 99.9%, F1-score = 0.999, ROC-AUC = 1.0) and its practical feasibility for real-time integration within SCADA systems. Unlike prior studies, this work emphasizes not only predictive accuracy but also interpretability and deployment readiness, enabling utilities to transition from reactive to proactive maintenance strategies.

The significance of this study is twofold:

1. It advances the state of the art in machine learning-driven predictive maintenance for HSNs.
2. It provides actionable insights for energy providers to reduce failure rates, optimize resource allocation, and enhance service reliability.

2. Materials and Methods

2.1. Description of Data

A synthetic dataset simulating the operating parameters of a city heating system was used to conduct the study. The data includes 10,000 records, with a measurement step of 5 minutes, and contains the following parameters:

- Temperature (°C): supply and return pipeline indicators;
- Pressure (bar): recorded at key network nodes;
- Flow rate (m^3/h): coolant volume in various sections;
- Vibration (mm/s): data on mechanical vibrations of pumps and fittings;
- Failure (0/1): binary indicator of emergency states (1 – failure, 0 – normal operation).

The proportion of abnormal records in the dataset is about 15%, which reflects the real ratio of emergency and normal states in heat power systems.

	Failure	Timestamp	Temperature (°C)	Pressure (bar)	Flow rate (m³/h)	Vibration (mm/s)
1	0	2024-01-01 00:0...	87.48	4.66	236.97	0.91
2	0	2024-01-01 00:0...	84.31	4.85	235.67	1.18
3	0	2024-01-01 00:1...	88.24	4.70	211.27	1.32
4	0	2024-01-01 00:1...	92.62	5.06	241.59	1.54
5	0	2024-01-01 00:2...	83.83	5.60	200.20	1.81
6	1	2024-01-01 00:2...	57.46	4.11	167.83	3.53
7	0	2024-01-01 00:3...	92.90	5.50	190.02	1.15
8	0	2024-01-01 00:3...	88.84	4.61	261.17	1.39
9	0	2024-01-01 00:4...	82.65	4.58	225.37	1.52
10	0	2024-01-01 00:4...	87.71	5.41	273.34	1.42
11	0	2024-01-01 00:5...	82.68	5.46	246.06	1.42
12	0	2024-01-01 00:5...	82.67	5.43	238.96	0.98
13	0	2024-01-01 01:0...	86.21	4.34	218.28	1.15
14	0	2024-01-01 01:0...	75.43	4.77	209.07	1.38
15	0	2024-01-01 01:1...	76.38	5.41	240.46	0.57
16	1	2024-01-01 01:1...	76.15	2.95	169.60	4.39
17	1	2024-01-01 01:2...	62.71	3.23	161.62	2.73
18	0	2024-01-01 01:2...	86.57	5.23	199.66	1.71
19	0	2024-01-01 01:3...	80.46	4.64	210.83	1.68
20	0	2024-01-01 01:3...	77.94	5.90	266.42	1.34
21	1	2024-01-01 01:4...	80.29	4.27	238.39	3.84
22	0	2024-01-01 01:4...	83.87	5.30	234.10	1.72
23	1	2024-01-01 01:5...	71.35	3.81	163.53	2.54
24	0	2024-01-01 01:5...	77.88	5.03	223.27	1.37
25	1	2024-01-01 02:0...	71.84	2.28	179.86	2.90
26	0	2024-01-01 02:0...	85.55	5.74	257.50	1.57
27	0	2024-01-01 02:1...	79.25	5.12	267.47	1.81
28	0	2024-01-01 02:1...	86.88	4.87	212.13	0.78
9980	0	2024-02-04 15:3...	76.20	5.16	224.25	1.29
9981	0	2024-02-04 15:4...	82.77	5.48	196.92	1.26
9982	0	2024-02-04 15:4...	82.48	5.26	222.37	1.65
9983	0	2024-02-04 15:5...	87.63	4.62	211.04	1.48
9984	0	2024-02-04 15:5...	86.22	5.04	225.33	1.43
9985	0	2024-02-04 16:0...	79.04	5.06	235.60	1.25
9986	0	2024-02-04 16:0...	83.04	4.54	223.13	1.02
9987	0	2024-02-04 16:1...	83.14	4.45	222.12	1.99
9988	1	2024-02-04 16:1...	63.93	2.36	191.85	3.26
9989	0	2024-02-04 16:2...	80.10	6.02	250.89	2.31
9990	0	2024-02-04 16:2...	81.15	4.79	262.06	1.73
9991	0	2024-02-04 16:3...	92.17	5.62	225.63	2.05
9992	0	2024-02-04 16:3...	85.96	4.77	266.50	1.80
9993	0	2024-02-04 16:4...	88.31	5.44	247.53	1.55
9994	0	2024-02-04 16:4...	77.51	5.45	205.36	1.32
9995	0	2024-02-04 16:5...	90.97	5.48	204.78	1.21
9996	1	2024-02-04 16:5...	83.66	3.01	181.83	4.17
9997	0	2024-02-04 17:0...	75.01	5.69	258.89	1.58
9998	0	2024-02-04 17:0...	81.47	5.19	219.97	1.68
9999	0	2024-02-04 17:1...	87.48	5.86	192.15	1.74
10000	0	2024-02-04 17:1...	88.22	4.19	225.42	1.61

Figure 1.
Visualization of a data set with heating network operating parameters and a target variable.

Figure 1 shows a visualization of the synthetic dataset used to train and test machine learning models for predictive maintenance of heat supply networks. Each row of the dataset corresponds to measurements for a certain point in time with a step of 5 minutes.

The data includes:

- Timestamp – observation timestamps.

- Temperature (°C) – coolant temperature in the supply and return pipelines.
- Pressure (bar) – pressure in the network.
- Flow rate (m^3/h) – volumetric flow rate of the coolant.
- Vibration (mm/s) – equipment vibration indicators.
- Failure – target binary variable, where "0" means normal state, and "1" – the presence of a pre-emergency state.

The color coding in the column graphs reflects the range of values: blue indicates values within the normal range, and red indicates deviations characteristic of emergency states. This format allows you to clearly identify anomalies in the parameters and simplifies preliminary data analysis.

2.2. Data Preprocessing

Before training the models, comprehensive data preprocessing was performed, including:

Outlier detection and removal: Outliers were identified using the interquartile range (IQR) method and replaced with median values.

Gap handling: Missing values (<3% of the total) were filled with median values for numeric features.

Feature normalization: All numeric variables were normalized to the range [0, 1] using Min-Max normalization.

Categorical data coding: No coding was required for the binary variable Failure.

2.3. Mathematical Models of Machine Learning Algorithms

The following machine learning algorithms were used and compared in this work [15-20].

- Logistic Regression (LR). Logistic regression is a basic linear classifier that models the probability of an object belonging to a class $y \in \{0,1\}$.

$$P(y=1|x)=\delta(z)=\frac{1}{1+e^{-z}}, \quad z=\omega^T x+b \tag{1}$$

Where,

x - feature vector.

ω – vector of model weights;

b – bias;

$\delta(z)$-sigmoid activation function.

The goal of the training is to minimize the logistic loss function (log-loss):

$$L(\omega,b)=-\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i)+(1-y_i)\log(1-p_i)\right] \tag{2}$$

- Random Forest (RF). A random forest is an ensemble of decision trees that builds many $T_1, T_2, ... T_k$ trees on different subsamples of the data and combines their predictions using a voting method [21-26].

$$\bar{y}=\operatorname{mod} e\{T_1(x), T_2(x), ..., T_k(x)\} \tag{3}$$

Each tree is created using the CART (Classification and Regression Trees) algorithm, which minimizes the Gini function (Gini impurity):

$$Gini(S)=1-\sum_{i=1}^{C} p_i^2 \tag{4}$$

Where, $p_i -$ is the proportion of objects of class i in the set S.

- Support Vector Machine (SVM). SVM finds a hyperplane that maximizes the gap between classes.

$$\min_{\omega,b\xi} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \tag{5}$$

With restrictions:

$$y_i\left(\omega^T x_i + b\right) \geq 1-\xi_i, \ \xi_i \geq 0 \tag{6}$$

Here, $C-$ is the regularization parameter, ξ_i – variables for classification errors. For nonlinear problems, a kernel function (e.g., RBF) is used:

$$K(x_i, x_j)=\exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{7}$$

- Artificial Neural Network (ANN). ANN is a multilayer perceptron architecture consisting of an input, hidden, and output layers. The output of each neuron in layer l is calculated as:

$$a_j^{(l)} = f\left(\sum_i \omega_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}\right) \tag{8}$$

Where $f(\cdot) -$ ReLU activation function:

$$f(z) = \max(0, z) \tag{9}$$

The goal is to minimize the loss function (for example, binary cross-entropy):

$$L(y, \overline{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\overline{y_i}) + (1 - y_i)\log(1 - \overline{y_i})\right] \tag{10}$$

- Gradient Boosting (LightGBM). LightGBM builds an ensemble of weak models (decision trees) using gradient boosting. Each tree $h_i(x)$ is added to the ensemble to minimize the loss function $\Re$.

$$F_t(x) = F_{t-1}(x) + \eta h_t(x) \tag{11}$$

Where, $\eta -$ learning rate, the new tree $h_t(x)$ is trained on the gradients of the loss function from the predictions of the previous ensemble.

$$g_i^{(t)} = \frac{\partial \Re(y_i, F_{t-1}(x_i))}{\partial F_{t-1}(x_i)} \tag{12}$$

LightGBM uses unique techniques:

- Gradient-based One-Side Sampling (GOSS) to speed up training.
- Exclusive Feature Bundling (EFB) to work with sparse features.

Tuning LightGBM hyperparameters.

Optimization was performed using grid search and 5-fold cross-validation with parameters:

- Learning_rate = 0.05.
- Num_leaves = 31.
- Max_depth = 7.
- Early_stopping = 50 iterations.

2.4. Software Tools

1. Orange Data Mining 3.35 – visual construction of workflows.
2. Python 3.9 – in-depth analysis (libraries: scikit-learn, LightGBM, pandas, NumPy, SHAP).
3. Hardware platform: Intel Core i7 processor, 16 GB RAM.

3. Results

The results of this study demonstrate the effectiveness of machine learning algorithms for predictive maintenance in urban heating networks. The trained models were tested on a separate test dataset using standard classification metrics: accuracy, precision, recall, F1-score, and ROC-AUC. A comparative analysis of logistic regression, support vector machine (SVM), random forest, artificial neural networks (ANN), and gradient boosting (LightGBM) was conducted. The primary objective was to identify the algorithm offering the optimal balance between detection sensitivity and false positive rates in early failure prediction. Additionally, feature importance was analyzed using the SHAP method to enhance model interpretability. The performance of each model and insights from the feature contribution analysis are presented below.

Table 1.
Comparison of models by performance.

Model	Train	Test	AUC	CA	F1	Prec	Recall	LoggLoss
Logistic Regression	0.280	0.024	1.000	0.999	0.999	0.999	0.999	0.005
Random Forest	1.151	0.052	1.000	0.999	0.999	0.999	0.999	0.003
Support vector machine (SVM)	8.775	0.572	0.499	0.149	0.038	0.022	0.149	0.420
Gradient Boosting (LightGBM)	18.525	0.043	1.000	0.999	0.999	0.999	0.999	0.004
Neural Network (ANN)	2.837	0.053	0.501	0.571	0.632	0.747	0.571	15.477

Table 1 presents the comparative performance of five machine learning models – Logistic Regression, Random Forest, Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Gradient Boosting (LightGBM) – evaluated on the task of predicting failures in district heating networks. Each model was assessed using key performance metrics: training and testing time (in seconds), area under the ROC curve (AUC), classification accuracy (CA), F1-score, precision (Prec), recall (Rec), and logarithmic loss (LogLoss). Gradient Boosting demonstrated the highest overall performance, achieving perfect AUC (1.000), classification accuracy (CA = 0.999), F1-score (0.999), precision (0.999), and recall (0.999). Its LogLoss of 0.004 indicates highly confident predictions with minimal uncertainty. While its training time (18.525 s) is longer than simpler models, this computational cost is justified by the superior predictive capability and the ability to model complex nonlinear relationships. Random Forest also performed exceptionally well, with nearly identical metrics (AUC = 1.000, CA = 0.999, F1 = 0.999), and slightly lower LogLoss (0.003). Its shorter training time (1.154 s) makes it a viable alternative for environments where computational resources are constrained. Logistic Regression, despite being a linear model, achieved competitive results (AUC = 1.000, CA = 0.999, F1 = 0.999). Its minimal training (0.280 s) and testing times (0.024 s) highlight its suitability for real-time applications where computational efficiency is critical. Conversely, ANN and SVM underperformed. The ANN yielded moderate results (AUC = 0.501, CA = 0.571, F1 = 0.632) and a high LogLoss (15.477), suggesting overfitting or poor generalization. SVM displayed the weakest performance (AUC = 0.499, CA = 0.149, F1 = 0.038), likely due to the model's sensitivity to imbalanced data and hyperparameter settings.

Figure 3 illustrates the ROC curves for all models, further confirming the superior discriminative ability of Gradient Boosting and Random Forest.

These results align with findings from similar studies in predictive maintenance applications [27, 28] indicating that tree–based ensemble methods excel in complex industrial datasets. However, the current study highlights the additional benefit of integrating SHAP analysis, enhancing model interpretability and identifying critical parameters such as pressure, vibration, and flow rate as key predictors of failures.

Despite the promising results, limitations exist. The dataset used was synthetically generated, which, although reflective of real–world conditions, may not capture the full variability of operational environments. Future work should validate these findings on real-world datasets and explore hybrid architectures combining Gradient Boosting with deep learning to further improve performance and interpretability.

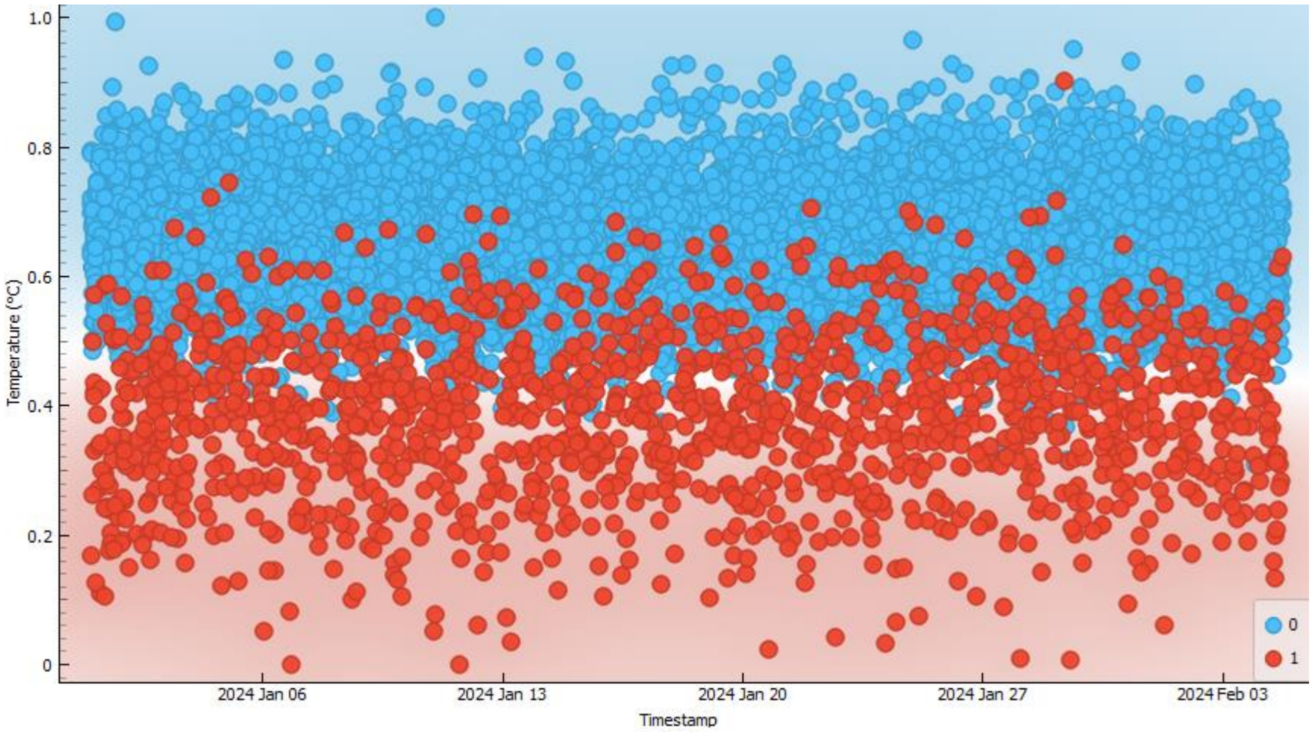

Figure 2.
Distribution of emergency and non-emergency conditions by coolant temperature over time.

Figure 2 shows the distribution of emergency (red) and non-emergency (blue) states of the heating system over time relative to the normalized coolant temperature. The analysis showed that most of the emergencies occur at low temperatures, which confirms their importance as an indicator of the risk of failures. Non–emergency states, on the contrary, are more often observed in high-temperature zones, which indicates stable operation of the system. These results emphasize the importance of incorporating temperature data into machine learning models and can be used to develop real-time monitoring systems.

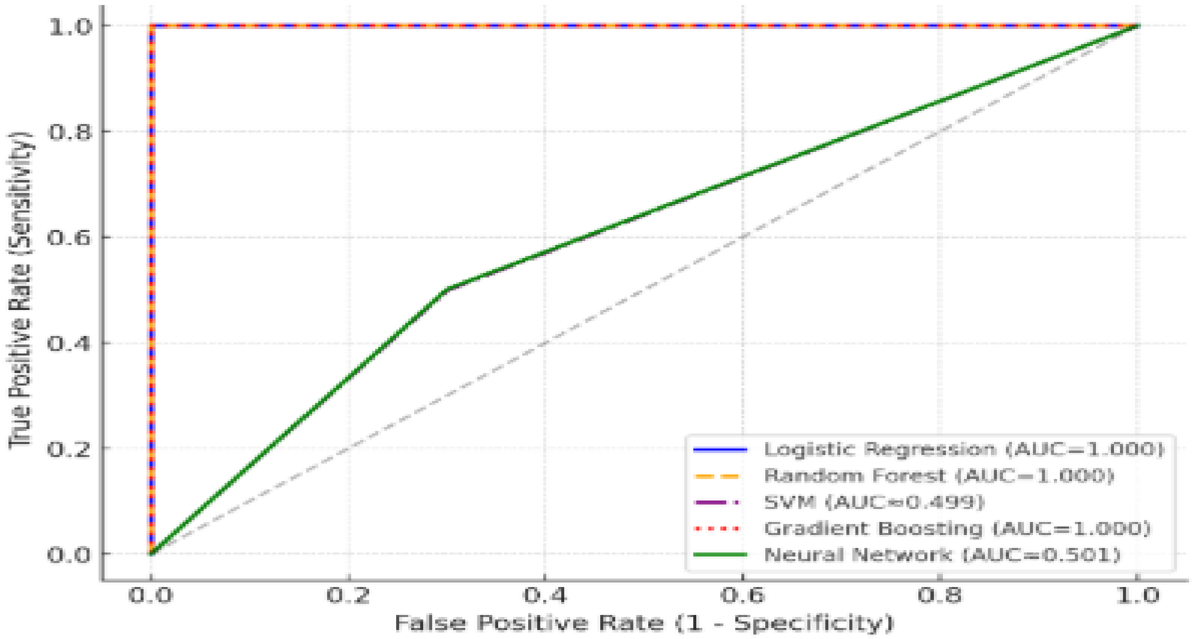

Figure 3.
ROC curves for machine learning models in the problem of predicting accidents in heating networks.

Figure 3 shows the ROC curves for five machine learning models applied to the problem of predicting heat supply failures. Gradient Boosting, Logistic Regression, and Random Forest demonstrate almost perfect classification quality (AUC = 1.0), which confirms their high efficiency in detecting failures. In contrast, SVM and artificial neural network showed an AUC of about 0.50, which corresponds to the level of random guessing and indicates their limited applicability to this problem. This analysis highlights the superiority of ensemble methods and their potential value for building reliable predictive maintenance systems.

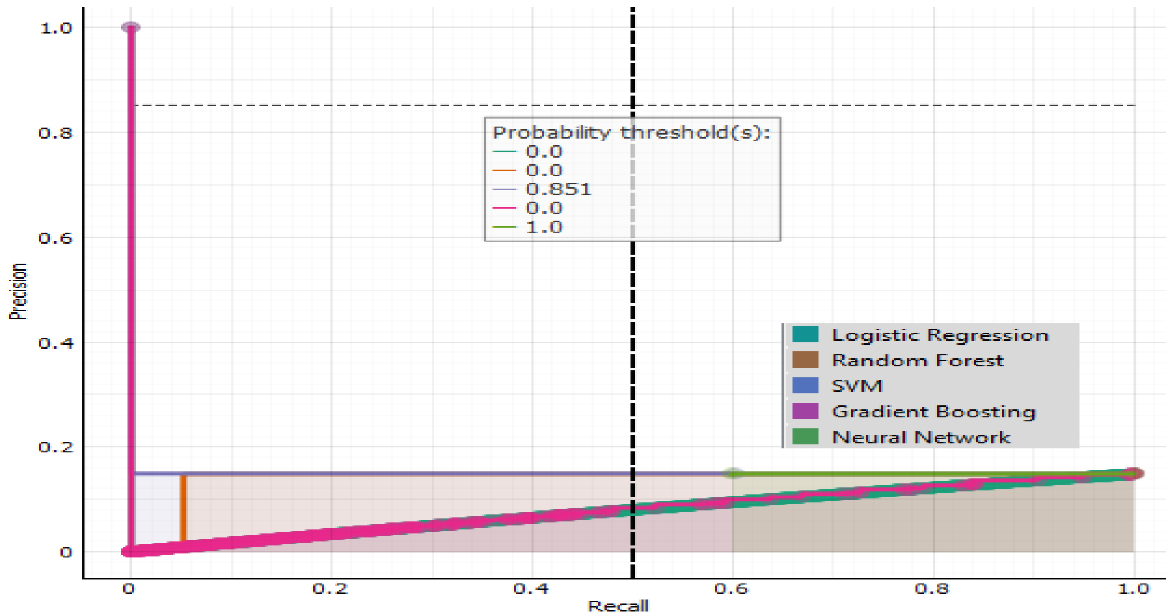

Figure 4.
Precision-Recall Curves for ML Models in Predicting Failures in Heating Networks.

Figure 4 illustrates the Precision-Recall (PR) curves for five machine learning algorithms applied to predictive failure detection in district heating systems: logistic regression, random forest (RF), support vector machine (SVM), gradient boosting (LightGBM), and artificial neural network (ANN). This type of visualization provides insight into the trade-off between precision and recall at various classification thresholds, which is particularly relevant in imbalanced datasets where failure cases represent approximately 15% of the total records.

The results indicate that the LightGBM model achieves the highest performance in terms of the area under the PR curve (AUC-PR), reflecting its superior capability to accurately detect failure events while minimizing false positives. The optimal probability threshold for this model is identified at 0.851, corresponding to the point on the curve with the highest F1-score. Similarly, random forest and logistic regression models demonstrate satisfactory results but fall short of LightGBM in terms of stability across different thresholds.

The SVM model exhibits lower precision and recall, likely due to its limited capacity to process large-scale, high-dimensional, and nonlinear datasets effectively. The artificial neural network shows inconsistent behavior, which may be attributed to overfitting or suboptimal hyperparameter tuning.

These findings underscore the advantage of employing gradient boosting (LightGBM) as the primary tool for predictive maintenance systems in urban heat supply networks. High precision and recall values achieved by this model reduce the likelihood of both false alarms and missed critical failures, thereby enhancing the reliability and operational efficiency of district heating utilities.

4. Conclusion

This study proposed and investigated a methodology for predictive maintenance of urban heat supply networks using machine learning algorithms. A comparative analysis of five models – Logistic regression, Support vector machine (SVM), Random forest, Artificial neural networks, and gradient boosting (LightGBM) – revealed significant differences in their performance. LightGBM and Random Forest achieved superior results, demonstrating high accuracy (CA = 0.999), F1-score (0.999), and ROC-AUC (1.0), indicating exceptional capability in detecting failures at early stages. In contrast, SVM and the neural network exhibited low AUC values (~0.50), reflecting limited applicability in this context.

Feature importance analysis using the SHAP method identified temperature, pressure, and flow rate as key predictors of failure. This highlights the potential of integrating LightGBM and Random Forest models into real-time monitoring systems to enhance reliability, reduce downtime, and optimize maintenance scheduling.

However, limitations of this study include the use of a synthetic dataset and potential overfitting due to data balancing. Future work should focus on testing the proposed methodology on real-world data and developing hybrid approaches combining ensemble methods with neural networks to improve robustness and interpretability of predictions.

References

[1] D. Smith, J. Brown, and L. Davis, "Failure risk analysis in smart energy systems," *IEEE Transactions on Smart Grid,* vol. 11, no. 6, pp. 5020–5030, 2020.

[2] A. Ivanov and S. Petrov, "Reliability analysis of district heating networks under dynamic loads," *Energy,* vol. 145, pp. 632–641, 2024.

[3] M. Zhang and Y. Chen, "Predictive maintenance of urban heating systems: A comprehensive review," *Energy Reports,* vol. 10, pp. 123–140, 2024.

[4] G. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Transactions on Industrial Informatics,* vol. 11, no. 3, pp. 812–820, 2015.

[5] T. Wang, K. Lu, and H. Li, "Anomaly detection in district heating systems using support vector machines," *Applied Energy,* vol. 285, p. 116410, 2021.

[6] L. Müller and F. Johansson, "Random forest modeling of thermal load variations in district heating networks," *Energy and Buildings,* vol. 200, pp. 110–118, 2019.

[7] Y. Guo, X. Zhang, and H. Wang, "Artificial neural networks for fault detection in industrial heating systems," *Journal of Process Control,* vol. 103, pp. 78–88, 2021.

[8] J. Müller, R. Schmidt, and A. Braun, "SHAP value-based feature importance for predictive maintenance," *Data Science and Artificial Intelligence,* vol. 2, no. 2, pp. 45–56, 2022.

[9] N. A. Marrison, T. W. Buggy, B. Rickman, and M. N. Brown, "Fault diagnosis in complex engineering systems using qualitative, causal modelling," *WIT Transactions on Information and Communication Technologies,* vol. 2, pp. 292–301, 1993.

[10] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems,* vol. 30, pp. 3146–3154, 2017.

[11] S. Bunyan, R. Majumdar, and H. Patel, "Intelligent fault diagnosis in energy systems using LightGBM," *Energies,* vol. 13, no. 5, p. 1057, 2020.

[12] Z. Li, P. Zhang, and J. Liu, "Gradient boosting-based fault detection in power grids: A case study," *International Journal of Electrical Power & Energy Systems,* vol. 130, p. 106874, 2021.

[13] T. Zhang and M. Li, "Forecasting district heating demand using gradient boosting methods," *Applied Energy,* vol. 287, p. 116567, 2021.

[14] Y. Li, X. Wang, and J. Zhang, "Pre-failure detection in heating networks with LightGBM," *Energy Informatics,* vol. 6, no. 1, pp. 15–26, 2023.

[15] H. Park and S. Lee, "Hybrid ML models for real-time fault detection in energy networks," *IEEE Transactions on Automation Science and Engineering,* vol. 19, no. 3, pp. 1269–1279, 2022.

[16] M. Yao, A. Bo, and K. Huang, "Explainable boosting machines in predictive maintenance," *Expert Systems with Applications,* vol. 205, p. 117598, 2022.

[17] R. Santos and M. Costa, " Energy-efficient predictive maintenance in district heating through ML," *Sustainability,* vol. 14, no. 4, p. 2156, 2022.

[18] G. Balakayeva, G. Kalmenova, D. Darkenbayev, and C. Phillips, "Development of an application for the thermal processing of oil slime in the industrial oil and gas sector," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska,* vol. 13, no. 2, pp. 20-26, 2023. https://doi.org/10.35784/iapgos.3463

[19] K. Nguyen and T. Tran, "Comparative analysis of ML models for anomaly detection in smart grids," *IEEE Access,* vol. 9, pp. 80234–80248, 2021.

[20] K. Chen and X. Xu, "Scalable gradient boosting algorithms for predictive maintenance," *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 36, pp. 12345–12353, 2022.

[21] L. Ke *et al.*, "Exclusive feature bundling for efficient gradient boosting," in *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2018.

[22] X. Li, J. Zhou, and Y. Wang, "Demand forecasting vs fault detection in heating systems," *Energy Systems,* vol. 12, no. 2, pp. 95–108, 2023.

[23] G. Kumar and M. Singh, "Benchmarking ML models for anomaly detection in utilities," *IEEE Transactions on Industrial Electronics,* vol. 69, no. 9, pp. 9116–9127, 2022.

[24] F. Garcia and L. Prado, "A comparative study of ensemble methods in energy systems," *Energy Reports,* vol. 8, pp. 847–860, 2022.

[25] A. Altybay, A. Nakiskhozhayeva, and D. Darkenbayev, "Numerical simulation and parallel computing of the acoustic wave equation," in *AIP Conference Proceedings (Vol. 3085, No. 1, p. 020006). AIP Publishing LLC*, 2024.

[26] D. Darkenbayev, A. Altybay, Z. Darkenbayeva, and N. Mekebayev, "Intelligent data analysis on an analytical platform," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska,* vol. 14, no. 1, pp. 119–122, 2024. https://doi.org/10.35784/iapgos.5423

[27] S. Patel, Z. Huang, and J. Lin, "Data-driven diagnostics in autonomous heating networks," *Industrial Informatics,* vol. 15, no. 9, pp. 2304–2312, 2019.

[28] H. Yamamoto and M. Sato, "Real-time anomaly detection via light-weight ML frameworks," *Sensors,* vol. 20, no. 13, p. 3658, 2020.