www.czasopisma.pan.pl PAN www.journals.pan.pl

BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 73(3), 2025, Article number: e154063
DOI: 10.24425/bpasts.2025.154063

ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE

Comparison and optimization of machine learning methods for fault detection in district heating and cooling systems

Mehmet ÇINAR[1], **Emrah ASLAN**[2]*, and **Yıldırım ÖZÜPAK**[3]

[1] Bitlis Eren University, Organized Industrial Zone Vocational School, Electrical Department, Bitlis, Türkiye
[2] Mardin Artuklu University, Faculty of Engineering and Architecture, Department of Computer Engineering, Mardin, Türkiye
[3] Dicle University, Silvan Vocational School, Electrical Department, Diyarbakır, Türkiye

Abstract. In this study, the methods used for the detection of sub-station pollution failures in district heating and cooling (DHC) systems are analyzed. In the study, high, medium, and low-level pollution situations are considered and machine learning methods are applied for the detection of these failures. Random forest, decision tree, logistic regression, and CatBoost regression algorithms are compared within the scope of the analysis. The models are trained to perform fault detection at different pollution levels. To improve the model performance, hyperparameter optimization was performed with random search optimization, and the most appropriate values were selected. The results show that the CatBoost regression algorithm provides the highest accuracy and overall performance compared to other methods. The CatBoost model stood out with an accuracy of 0.9832 and a superior performance. These findings reveal that CatBoost-based approaches provide an effective solution in situations requiring high accuracy, such as contamination detection in DHC systems. The study makes an important contribution as a reliable fault detection solution in industrial applications.

Keywords: pollution detection; grid search optimization; machine learning; DHC.

1. INTRODUCTION

District heating and cooling (DHC) systems play an important role in urban energy management. These systems provide efficient heating and cooling to buildings, increasing energy efficiency and reducing costs. However, the reliability of DHC systems can be compromised by various failures. This necessitates the development of an effective fault detection and diagnosis (FDD) model. Traditional methods, especially manual checks, are insufficient to handle the complexities of modern DHC systems [1].

In recent years, advances in data-driven techniques, especially machine learning (ML) based approaches, offer promising solutions to these problems. However, the lack of comprehensive and high-quality data sets limits the effectiveness of these models [2, 3]. This shortcoming has led researchers to rely on synthetic data sets generated by simulations or open data sources. Among ML algorithms, logistic regression, decision tree, random forest, and CatBoost regression have been successfully applied for fault detection in DHC systems. However, these methods proved to have limitations, especially for subtle problems such as thermal losses [4]. Strategies such as the integration of real-time data with simulation results are under investigation to improve model robustness [5]. Historically, fault detection often relies on reactive methods, which in most cases are inefficient and costly [6]. Today, modern data analytics methods, ML, and deep learning techniques can provide earlier warnings by detecting trends before failures occur [7, 8].

This study compares ML algorithms such as logistic regression, decision tree, random forest, and CatBoost regression for the detection of pollution faults in DHC substations. The dataset used in the study is synthetically generated and represents different pollution levels (high, medium, low). To optimize the performance of the models, hyperparameter adjustments were performed and the random search optimization method was used in this process. The results show that the CatBoost regression algorithm provides the highest accuracy rate (98.32%) and the best overall performance compared to other methods. The main objective of the study is to improve early fault detection in DHC systems, especially to provide solutions to complex problems such as pollution. However, along with the advantages of the CatBoost regression model, challenges such as the need for large data sets and the need for validation with real-world data have also been noted.

This paper contributes to the literature by presenting the applicability of ML for fault detection and its implications for industrial applications. Section 1 provides an overview of the problem, Section 2 reviews the literature related to the study, and Section 3 presents the materials and methods. Section 4 presents the discussion and conclusions, while Section 5 details future work and general conclusions.

*e-mail: emrah.aslan@dicle.edu.tr

Manuscript submitted 2024-10-08, revised 2025-01-14, initially accepted for publication 2025-02-27, published in March 2025.

2. LITERATURE REVIEW

The literature on FDD in DHC systems reveals a growing emphasis on leveraging advanced analytical techniques to address operational inefficiencies. Traditional methods, while effective in identifying obvious failures, often fall short in detecting subtle issues. ML and hybrid approaches have emerged as promising alternatives, offering robust solutions to complex problems. This section synthesizes key studies that highlight both the progress and gaps in this domain, providing a comprehensive background for the current research.

FDD in DHC systems has been a prominent research topic, focusing on identifying and mitigating failures that can disrupt system efficiency. Common failures at the DHC system level include issues related to sensors, actuators, and commissioning, which can often be monitored using available process data. The literature extensively discusses generic methods for detecting sensor-related failures through operational data analysis. For instance, IEA DHC Annex XIII provides a comprehensive overview of typical failures in DHC systems, serving as a foundational framework for fault detection studies [9, 10].

Panday *et al.* [11] explored fault detection methodologies in boilers and combined heat and power systems, identifying critical strategies for improving operational reliability. Similarly, Hundi and Shavari [12] examined inefficiencies in DHC systems, emphasizing the importance of identifying subtle operational anomalies. Gadd and Werner [13] noted that many failures occur at substations, often due to heat load pattern irregularities and poor control mechanisms. These findings are further supported by Buffa *et al.* [14] and Månsson *et al.* [15], who provided in-depth analyses of substation-level failures and detection methods. Leoni *et al.* [16] highlighted that improper configuration of substation valves often results in elevated return temperatures, exacerbating system inefficiencies.

In a study by Bilici and Özdemir, three mathematical models were developed with meteorological data to forecast Turkey's natural gas demand [17]. The scarcity of high-quality data further challenges the development of robust, data-driven fault detection models. Hybrid ML approaches, which integrate real-world data with simulation outputs, have shown considerable potential for diagnosing faults across various energy systems, including DHC [18].

Simulation-based datasets have also played a crucial role in evaluating DHC system failures. For example, a dataset designed for fault detection in DHC systems demonstrated the effectiveness of five ML models, underscoring the importance of integrating simulation results in model development [19]. Bilici *et al.* compared four different meta-heuristic algorithms to forecast Turkey's natural gas demand. In the models trained with 2010–2017 data and tested with 2018-2020 data, the quadratic model of the particle swarm optimization algorithm showed the most successful forecasting performance [20]. Leakage fault detection using ML has achieved significant results, with an accuracy of 85.85%, providing a foundation for further advancements in the field [21].

This study contributes to the existing literature by focusing on contamination detection in DHC substations through a detailed comparative analysis of advanced regression-based ML models. Unlike previous studies that primarily emphasize leakage or actuator failures, this work addresses the underexplored area of fouling-induced contamination. By employing a simulation-based dataset and optimizing model performance using random search optimization, this study achieves a high accuracy rate of 98.32% with the CatBoost regression model. Not only do these findings highlight the applicability of regression models for fault detection but also provide actionable insights for improving system efficiency and reliability in industrial DHC applications.

3. MATERIALS AND METHODS

In this study, different ML algorithms are used for the detection of sub-station pollution faults in DHC systems. logistic regression, decision tree, random forest, and CatBoost regression methods are modeled and optimized to detect faults according to pollution levels. The dataset used in the study is synthetically generated and contains errors representing different pollution concentrations (high, medium, and low) through simulations. The dataset is split into 70% training and 30% test data to measure the performance of the models and evaluate their generalization capabilities. The hyperparameter settings of the models were optimized with the random search optimization method and the best results were obtained for each algorithm. The results are compared using performance metrics such as accuracy, Matthews correlation coefficient, and processing time [22, 23]. This section presents the development process of the proposed model, describing the scope of the study and the details of the applied methods.

3.1. Dataset

The dataset used in this study contains synthetic error data for substation contamination in DHC systems. The dataset was developed within the scope of the International Energy Agency's (IEA) DHC Annex XIII project "Artificial Intelligence Fault Detection and Prediction of Heat Production and Demand in District Heating Networks". The project aims to develop artificial intelligence methods for the prediction of heat demand and production and to evaluate the algorithms used for fault detection. The experiments in the dataset were created through simulations covering 28 days. During this time, failures at different time points were observed and pollution conditions of different intensities (high, medium, and low) were integrated into the model. Failures were simulated through scenarios that could occur suddenly or gradually. Fault intensities can be interpreted in different ways depending on the simulation model used. This dataset provides a valuable resource for the development of new approaches for fault detection in DHC systems.

The dataset used in the study is divided into two parts for model training and testing, 70% training and 30% testing. The training data is used to complete the learning process of the models, while the test data is used to evaluate the overall performance of the models. This approach is preferred as a standard method to measure the generalization capacity of the models and to objectively evaluate their performance. The basic statistical information of the dataset, the names of the input variables, and their respective descriptions are presented in Table 1. This information provides a detailed overview of the scope of the study and the structure of the dataset.

Table 1
Basic statistical information about the data used in the dataset

Variable	Explanation	Min	Max
BC	Id of the boundary conditions used for this experiment	0	12
F1_type	Type of fault used in the experiment	0	1
F1_start	Start time of the fault, in hours	0	671
F1_stop	Stop time of the fault, in hours	0	672
F1_init	İnitial intensity of the fault, in the range [0–1]	0	1
F1_final	Final intensity of the fault, in the range [0–1]	0	1

The heating and cooling sources, storage units, and flow to residential and commercial buildings are shown in the simplified block diagram of a DHC system in Fig. 1. The diagram summarises the operational framework of the system, which serves as the basis for the fault detection methods discussed in this paper.

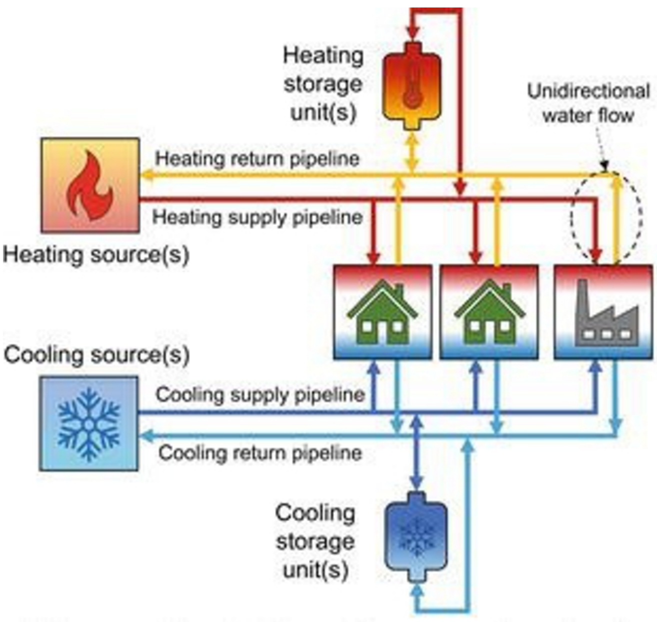

Fig. 1. DHC system block diagram [24]

3.2. Logistic regression

Logistic regression is a statistical ML method used for classification problems. The method uses a sigmoid function to estimate the probability that the dependent variable belongs to a certain category. The model produces a finite output (for example, 0 or 1) using a linear combination of independent variables. The predicted probabilities are compressed by the sigmoid function between 0 and 1 and classified according to a threshold value (usually 0.5). The basic mathematical expression of a logistic regression model

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}, \tag{1}$$

where $h_\theta(x)$ is the probability that an observation belongs to a class, θ is the model parameters, and x is the values of the independent variables. The loss function is defined as negative log-likelihood in logistic regression and is written as

$$J_\theta = \frac{-1}{m}\sum_{i=1}^{m}\Big[y^{(i)}\log\Big(h_\theta(x^{(i)})\Big) + (1-y^{(i)})\log\Big(1-h_\theta(x^{(i)})\Big)\Big], \tag{2}$$

This function is used to optimize the model parameters.

3.3. Decision tree

Decision tree is an algorithm that performs classification or regression tasks by branching the data. At each node, the data is divided and subdivided according to a specific feature and threshold value. This process is optimized using a measure of data purity (e.g., *Gini* coefficient or information gain).

The *Gini* coefficient used as a purity measure of a decision tree is calculated as

$$Gini = 1 - \sum_{i=1}^{n} p_i^2, \tag{3}$$

where p_i is the probability of belonging to the i-th class.

Decision tree models provide fast learning but often tend to overfit. Therefore, the complexity of the model is controlled by methods such as pruning.

3.4. Random forest

Random forest is an ensemble version of decision tree algorithms. It creates multiple decision trees and combines the results of these trees for classification or regression tasks [25]. Each tree is trained on a random sampling of the dataset and a subset of features. In classification problems, the vote of each tree is taken to determine the final class; in regression, the trees are averaged. The basic mathematical expression of random forest is

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x), \tag{4}$$

where B is the total number of trees and $f_b(x)$ is the prediction result of the b-th tree.

3.5. CatBoost regression

CatBoost is a gradient-boosting algorithm optimized for categorical data. Unlike other boosting algorithms, it reduces the need for data preprocessing by automatically coding categorical variables. It also uses an innovative ordered boosting mechanism to avoid overfitting.

The main goal of CatBoost is to generate successive forecasts using decision trees to minimize the loss function. The gradient boosting procedure of the model can be described as

$$F_{m+1}(x) = F_m(x) + \gamma h_m(x), \tag{5}$$

where $F_m(x)$ are the predictions of the model at the m-th iteration, $h_m(x)$ are the weak students generated by the decision tree, and γ is the learning.

3.6. Random search optimization

Random search optimization is a widely used technique for hyperparameter optimization that offers a simple approach. This method is an optimization process in which hyperparameters are randomly selected in various ranges of values. Basically, a specific range is defined for each hyperparameter and random values are selected from this range. The performance of the model is then evaluated for each selected combination [26, 27].

This method is particularly preferred when the hyperparameter space is very large or when other optimization techniques are costly and complex to implement. One of the advantages of random search is that the probability of reaching the global optimum is higher because the parameters are chosen randomly, so the probability of getting stuck in a local minimum is lower.

Compared to methods such as grid search, random search can achieve better results with a much smaller number of trials. This is because grid search tests all possible values of each parameter, while random search takes only a few random samples, which can be more time-efficient. Mathematically, random search operates on a set of hyperparameters θ and selects a random combination of hyperparameters,

$$\theta^* = \arg\max_{\theta \in \Delta} f(\theta), \tag{6}$$

where θ represents a randomly chosen combination of hyperparameters in the hypothesis space, Δ is the space containing all possible hyperparameter combinations, $f(\theta)$ is an objective function that measures the performance of the model, usually accuracy, error rate, or some other performance metric. θ^* denotes the hyperparameter combination that achieves the best performance. As a result, random search optimization often provides more efficient hyperparameter optimization and can achieve good results in a shorter time, avoiding high computational costs.

3.7. Generation of synthetic failure dataset

To develop the dataset, potential failure modes were first identified through failure modes, effects, and criticality analysis (FMECA). Various failure scenarios were then modeled based on these identified modes. The simulations were designed to replicate different fault profiles, including both step and ramp failures, under a variety of conditions, such as fluctuating outdoor temperatures and varying heat demand. Additionally, critical parameters, such as the onset time of faults and their severity levels, were incorporated into the simulation process.

The dataset was intentionally diversified to mimic real-world conditions by randomly selecting the fault onset times and the associated severities. Each simulation instance recorded essential variables, ensuring that the dataset was structured and formatted appropriately for training ML models used in FDD tasks. This diversity in the data helps in capturing the complexity and variability found in actual system operations.

In the simulations, failure profiles were modeled as either step or ramp types. A step profile represents a sudden onset of a fault, whereas a ramp profile simulates a gradual development of the fault over time. The failure dynamics were further defined by the following parameters:

Start time (t_0): The moment when the fault begins, initiating the fault condition.

End time (t_x): This parameter is specific to ramp profiles, marking the point in time when the fault reaches its maximum severity.

Start intensity (v_0): The intensity of the fault at the moment of onset, which always starts at zero.

Final severity (v_x): The final severity level of the fault, ranging from 0 to 1, where 1 indicates the maximum severity of the fault.

These parameters were used to construct various fault scenarios within the simulations, ensuring a comprehensive dataset that could be used for different types of fault detection tasks. The severity of faults in the dataset was expressed on a scale from 0 to 1, where 0 represents the absence of any fault, and 1 signifies the highest possible severity level for that fault type. This scale of severity is model-dependent and varies depending on the fault type simulated.

Figure 2 illustrates the fault appearance profiles used in the simulation process. The left side of the figure depicts the ramp profile for faults that gradually increase in severity, while the right side shows the step profile for faults that occur abruptly. These profiles serve as the basis for generating synthetic fault data used in the ML models for detection and diagnosis.

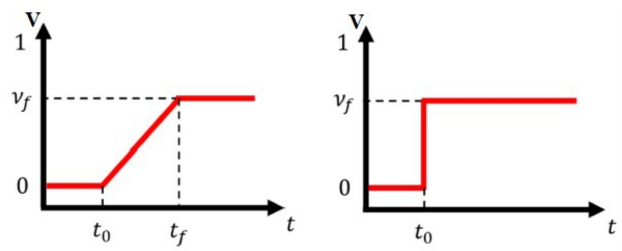

Fig. 2. Fault appearance profiles used in the simulation (left: ramp for advancing faults, right: step for abrupt faults) [19]

4. RESULTS AND DISCUSSION

In this study, four different regression methods are applied to detect pollution failures in DHC systems. These methods are logistic regression, decision tree, random forest, and CatBoost regression. The dataset used consists of synthetic fault data generated by Modelica simulations. In the dataset, high, medium, and low pollution scenarios were created and these scenarios were divided into 70% training and 30% testing data to test the accuracy of the model. While the training data supported the learning process of the model, the test data was used to objectively evaluate the performance of the models.

In this study, four regression models were utilized for fault detection in DHC systems: logistic regression, decision tree, random forest, and CatBoost regression. Each of these models has distinct advantages and disadvantages, depending on the type of problem and the data used. Below, a summary of the advantages and disadvantages of each model is provided, followed by Table 2 summarizing these aspects for clarity.

Table 2
Advantages and disadvantages of regression models

Regression model	Advantages	Disadvantages
Logistic regression	– Simple and easy to implement – Efficient for binary classification – Provides probabilistic output	– Struggles with nonlinear relationships – Underperforms with highly correlated features
Decision tree	– Intuitive and easy to interpret – Can handle both categorical and continuous data – Captures nonlinear relationships	– Prone to overfitting – Sensitive to small changes in data – Less robust in noisy data
Random forest	– Combines multiple decision trees for better accuracy – Robust and less prone to overfitting – Can handle large datasets	– Computationally expensive – Difficult to interpret due to many trees
CatBoost regression	– Optimized for categorical data – Excellent handling of imbalanced datasets – Minimal data preprocessing required	– Computationally intensive – Requires more memory and resources than simpler models

To improve the performance of the model, the hyperparameters of each regression model were optimized using the random search optimization algorithm. This method aims to identify the best model parameters by making random choices in the hyperparameter space. The optimization process was performed to obtain the optimal results by evaluating different combinations of hyperparameters for each model. Table 3 shows the hyperparameters set for each regression model and their values. This method was used as an effective strategy to improve the overall accuracy of the model and achieve better performance.

The hyperparameter ranges used in our study were determined by considering studies in the literature and the characteristics of the methods. These ranges aim to comprehensively address critical variables that may affect model performance. The random search optimization method scans these ranges and determines the optimal combinations that improve model performance. This process aims to minimize bias in parameter selection and improve the generalisability of the model.

Table 3
Hyperparameters tuned for regression models using random search optimization

Regression model	Hyperparameter	Optimized value
Logistic regression	C	Tuned
	Solver	lbfgs
	Max iterations	1000
Decision tree	Max depth	10
	Min samples split	2
	Min samples leaf	1
	Max features	None
Random forest	Number of estimators	100
	Max depth	10
	Min samples split	2
	Min samples leaf	1
CatBoost regression	Learning rate	0.1
	Depth	6
	L2 regularization term	3
	Iterations	1000

In this study, four different regression models are used for fault detection in DHC systems: logistic regression, decision tree, random forest, and CatBoost regression. The performance of the models was evaluated using key metrics such as accuracy, Matthews correlation coefficient (MCC), and elapsed time. Table 4 clearly shows the performance of each model on these metrics and these results reflect the capacity of the models to detect contamination failures in DHC systems.

Table 4
Performance metrics for the regression algorithms

Algorithms	Accuracy	Matthews corr. coef.	Elapsed time (seconds)
Logistic regression	0.8723	0.7286	14.35
Decision tree	0.9051	0.8125	32.15
Random forest	0.9506	0.9012	60.25
CatBoost regression	0.9832	0.9664	112.50

First, the highest performance in terms of accuracy came from the CatBoost regression model, which achieved the best results with an accuracy of 98.32%. This high accuracy rate demonstrates the success of CatBoost in capturing patterns in complex datasets. The random forest model ranked second with an accuracy of 95.06%, while the decision tree and logistic regression models performed lower with 90.51% and 87.23% accuracy rates respectively. The lower accuracy of the decision tree may be due to its tendency to overlearn (overfitting). Logistic regression, on the other hand, is a simpler model and did not model the non-linear relationships in the dataset well enough.

Matthews correlation coefficient (MCC) is a metric that measures the accuracy of each model as well as the balance of the classification, providing a more accurate performance assessment in imbalanced data sets. The CatBoost regression model scored the highest among all models with an MCC value of 0.9664. Random forest ranked second with an MCC value of 0.9012, while decision tree and logistic regression performed lower with MCC values of 0.8125 and 0.7286, respectively. These results show that CatBoost is successful in correctly classifying both positive and negative classes.

The processing times of the four regression models used in the study were analyzed to shed light on the practical use of the algorithms in different application scenarios. The fastest model was logistic regression with a processing time of 14.35 seconds. This was followed by the decision tree at 32.15 seconds and the random forest at 60.25 seconds. The CatBoost regression model, which has the highest accuracy rate, had a processing time of 112.50 seconds.

The long processing time of CatBoost regression reflects the gradient-boosting structure of the algorithm and the large number of trees used. This shows that while faster models, such as logistic regression or random forest, may be preferred in applications requiring a quick response, CatBoost is a superior option when high accuracy is required.

Figure 3 compares the performance of various ML algorithms in terms of accuracy, the Matthews correlation coefficient (MCC), and processing time. In terms of accuracy, CatBoost regression achieved the highest performance with an accuracy rate of 98.32%, followed by random forest (95.06%), decision tree (90.51%), and logistic regression (87.23%). These results indicate that CatBoost regression provides superior classification performance, making it the most suitable algorithm for this dataset.

The Matthews correlation coefficient (MCC) values further emphasize the robustness of CatBoost regression, as it obtained the highest MCC value (0.9664), indicating a well-balanced classification in terms of true positives and true negatives. Random forest (0.9012) and decision tree (0.8125) displayed moderate performance, while logistic regression had the lowest MCC value (0.7286). The high MCC value of CatBoost regression underscores its effectiveness in minimizing false positives and negatives compared to other algorithms.

In terms of computational cost, the processing time results reveal that logistic regression is the fastest algorithm (14.35 seconds), followed by decision tree (32.15 seconds) and random forest (60.25 seconds). CatBoost regression, despite its high accuracy and MCC, is the slowest algorithm with a processing time of 112.50 seconds. These findings suggest that while CatBoost regression excels in performance metrics, faster algorithms like logistic regression may be more suitable for applications requiring quick results.

The Confusion matrix in Fig. 4 is used to evaluate the prediction performance of the model. According to the results, the model performed very well with an accuracy rate of 98.32%. No errors were made in the prediction of the negative class (False Positive: 0), while 979 samples were classified as false negative in the positive class (False Negative). The Precision value calculated for the positive class was 100%, indicating that all positive predictions were correct. The Recall value was 95.74%, indicating that the positive class was correctly detected at a high rate. The F1-Score was 97.81%, indicating a balanced performance between Precision and Recall. These results prove that the model provides a reliable solution, especially in fault detection, and is suitable for industrial applications.

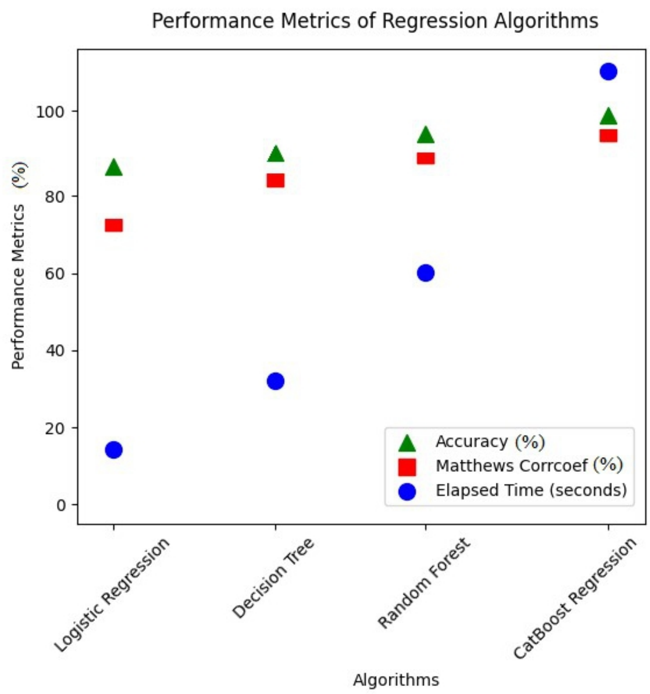

Fig. 3. Comparison of accuracy performance of ML algorithms

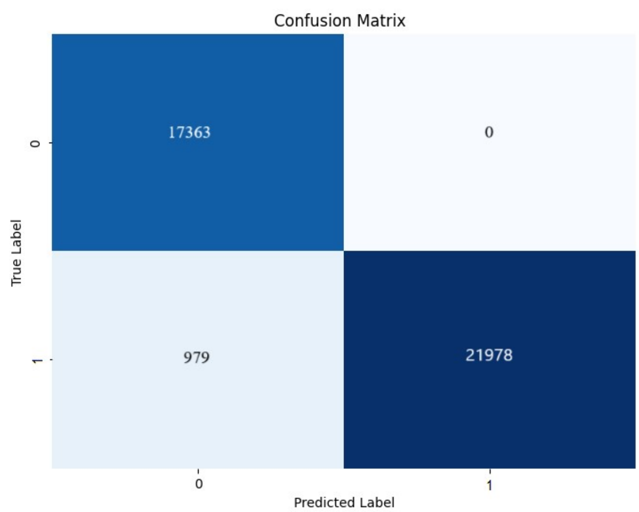

Fig. 4. Confusion matrix

In Fig. 5, the performance of the CatBoost regression model at different pollution levels (75% very high, 20% and 11% medium, and 5% low) is analyzed in terms of fault detection probability and UA (heat transfer coefficient) values. At an extremely high pollution level (75%), with a sharp drop in the UA value, the model showed a high accuracy rate, detecting the

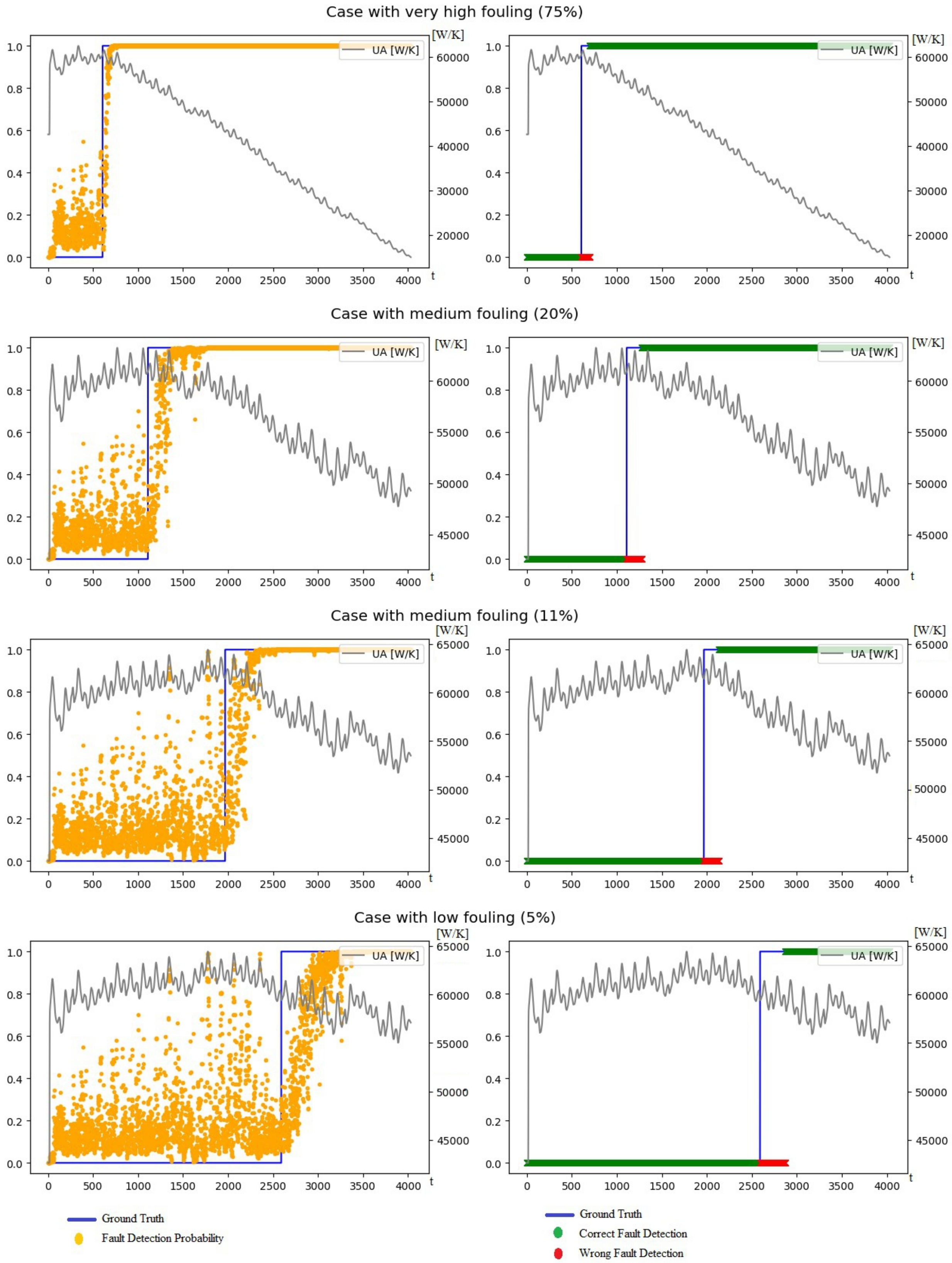

Fig. 5. Performance of the CatBoost regression model

fault quickly and accurately. In this case, there were no false detections, demonstrating the reliability of the model at this level. At medium pollution levels (20% and 11%), the change in the UA value was slower, but the model successfully detected the fault. However, slight delays in detection time were observed at medium levels, indicating the effect of the rate of pollution increase on the model. At the low pollution level (5%), although the change in the UA value was quite slow, the CatBoost model correctly detected the fault and performed its task with minimal false detections. In general, the CatBoost regression model is characterized by fast and accurate detection, especially at high pollution levels, and adaptive and reliable performance at low and medium levels. These findings confirm that the model offers an effective solution for critical applications such as contamination detection in DHC systems.

The ground truth expression in Fig. 5 refers to the reference values that reflect the reality of the labeled data in the dataset. These values are used to ensure that anomalies or normal conditions occurring in the system are correctly classified.

5. CONCLUSIONS

This study was conducted using various regression models, namely logistic regression, decision tree, random forest, and CatBoost regression models for fault detection in DHC systems. The performance of the models is evaluated by key metrics such as accuracy, MCC, and processing time. CatBoost regression was the most successful model with the highest accuracy (98.32%) and MCC (0.9664). Random forest performed well with high accuracy (95.06%) and balanced MCC (0.9012), offering a good balance between high accuracy and computational efficiency. Although CatBoost provides the best results, considering the high processing time (112.50 seconds), random forest may be a more efficient alternative for real-time applications where time constraints are important. The decision tree and logistic regression models, although faster, achieved lower results in terms of accuracy and MCC, and are therefore less suitable for complex fault detection tasks. In conclusion, although CatBoost regression is the most accurate model for fault detection in DHC systems, it should be considered in terms of processing time. Random forest offers a strong alternative in terms of performance and computational efficiency, while decision tree and logistic regression may be preferable for simpler tasks where speed is of the utmost importance. In future work, strategies can be developed to improve computational performance without compromising model accuracy.

REFERENCES

[1] F. Zhang, N. Saeed, and P. Sadeghian, "Deep learning in fault detection and diagnosis of building HVAC systems: A systematic review with meta-analysis," *Energy AI*, vol. 12, p. 100235, 2023, doi: 10.1016/j.egyai.2023.100235.

[2] M. Vallee, T. Wissocq, Y. Gaoua, and N. Lamaison, "Generation and evaluation of a synthetic dataset to improve fault detection in district heating and cooling systems," *Energy*, vol. 283, p. 128387, 2023, doi: 10.1016/j.energy.2023.128387.

[3] A. Rafati and H.R. Shaker, "Predictive maintenance of district heating networks: A comprehensive review of methods and challenges," *Therm. Sci. Eng. Progress*, vol. 53, p. 102722, 2024, doi: 10.1016/j.tsep.2024.102722.

[4] C. Hermans, J. Al Koussa, T. Van Oevelen, and D. Vanhoudt, "Fault detection for district heating substations: Beyond three-sigma approaches," *Smart Energy*, vol. 16, p. 100159, 2024, doi: 10.1016/j.segy.2024.100159.

[5] Q. Shi, M. Liu, S. Zhang, R. Zheng, and X. Lan, "Multi-Agent Path Finding Method Based on Evolutionary Reinforcement Learning," *2024 43rd Chinese Control Conference (CCC)*, Kunming, China, 2024, pp. 5728–5733, doi: 10.23919/CCC63176.2024.10661475.

[6] F.L.O. Martins *et al.*, "Multiple Fault Diagnosis in a Wind Turbine Gearbox with Autoencoder Data Augmentation and KPCA Dimension Reduction," *J. Nondestruct. Eval.*, vol. 43, p. 114, 2024, doi: 10.1007/s10921-024-01131-3.

[7] F. Alpsalaz and M.S. Mamiş, "Detection of Arc Faults in Transformer Windings via Transient Signal Analysis," *Appl. Sci.*, vol. 14, p. 9335, 2024, doi: 10.3390/app14209335.

[8] Y. Zhou, S. Zheng, and J.L.M. Hensen, "Machine learning-based digital district heating/cooling with renewable integrations and advanced low-carbon transition," *Renew. Sustain. Energy Rev.*, vol. 199, p. 114466, 2024, doi: 10.1016/j.rser.2024.114466.

[9] M. Vallee, T. Wissocq, Y. Gaoua, and N. Lamaison, "Generation and evaluation of a synthetic dataset to improve fault detection in district heating and cooling systems," *Energy*, vol. 283, p. 128387, 2023, doi: 10.1016/j.energy.2023.128387.

[10] X. Yang, Q. Zhao, Y. Wang, and K. Cheng, "Fault signal reconstruction for multi-sensors in gas turbine control systems based on prior knowledge from time series representation," *Energy*, vol. 262, p. 124996, 2023, doi: 10.1016/j.energy.2022.124996.

[11] R. Panday, N. Indrawan, L.J. Shadle, and R.W. Vesel, "Leak detection in a subcritical boiler," *Appl. Therm. Eng.*, vol. 185, p. 116371, 2021, doi: 10.1016/j.applthermaleng.2020.116371.

[12] P. Hundi and R. Shahsavari, "Comparative studies among machine learning models for performance estimation and health monitoring of thermal power plants," *Appl. Energy*, vol. 265, p. 114775, 2020, doi: 10.1016/j.apenergy.2020.114775.

[13] H. Gadd and S. Werner, "Fault detection in district heating substations," *Appl. Energy*, vol. 157, pp. 51–59, 2015, doi: 10.1016/j.apenergy.2015.07.061.

[14] S. Buffa *et al.*, "Advanced control and fault detection strategies for district heating and cooling systems—a review," *Appl. Sci.*, vol. 11, no. 1, p. 455, 2021, doi: 10.3390/app11010455.

[15] S. Månsson *et al.*, "Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities," *Energy*, vol. 180, pp. 163–174, 2019, doi: 10.1016/j.energy.2019.04.220.

[16] P. Leoni, R. Geyer, and R.-R. Schmidt, "Developing innovative business models for reducing return temperatures in district heating systems: Approach and first results," *Energy*, vol. 195, p. 116963, 2020, doi: 10.1016/j.energy.2020.116963.

[17] Z. Bilici, D. Özdemir, and H. Temurtaş, "Comparative analysis of metaheuristic algorithms for natural gas demand forecasting based on meteorological indicators," *J. Eng. Res.*, vol. 11, no. 3, pp. 259–265, 2023. doi: 10.1016/j.jer.2023.100127.

[18] A.J. Patil, R. Naresh, R.K. Jarial, and H. Malik, "Optimized synthetic data integration with transformer's DGA data for improved ML-based fault identification," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 32, no. 1, pp. 598–607, 2025, doi: 10.1109/TDEI.2024.3421915.

[19] M. Vallee, T. Wissocq, Y. Gaoua, and N. Lamaison, "Generation and evaluation of a synthetic dataset to improve fault detection in district heating and cooling systems," *Energy*, vol. 283, p. 128387, 2023, doi: 10.1016/j.energy.2023.128387.

[20] Z. Bilici and D. Özdemir, "Comparative analysis of metaheuristic optimization algorithms for natural gas demand forecasting based on meteorological parameters," *J. Fac. Eng. Archit. Gazi Univ.*, vol. 38, no. 2, pp. 1153–1168, 2022. doi: 10.17341/gazimmfd.1014788.

[21] P. Xue *et al.*, "Machine learning-based leakage fault detection for district heating networks," *Energy Build.*, vol. 223, p. 110161, 2020, doi: 10.1016/j.enbuild.2020.110161.

[22] S. Månsson *et al.*, "A machine learning approach to fault detection in district heating substations," *Energy Procedia*, vol. 149, pp. 226–235, 2018, doi: 10.1016/j.egypro.2018.08.187.

[23] E. Aslan and Y. Özüpak, "Detection of road extraction from satellite images with deep learning method," *Cluster Comput.*, vol. 28, p. 72, 2025, doi: 10.1007/s10586-024-04880-y.

[24] I. De la Cruz and C.E. Ugalde-Loo, "District Heating and Cooling Systems," in *Microgrids and Local Energy Systems*, UK: IntechOpen Limited, 2021. doi: 10.5772/intechopen.99740.

[25] E. Aslan, "Prediction and comparative analysis of emissions from gas turbines using random search optimization and different machine learning-based algorithms," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 72, no. 6, p. e151956, 2024, doi: 10.24425/bpasts.2024.151956.

[26] B. Ozdemir, E. Aslan, and I. Pacal, "Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection," *IEEE Access*, vol. 13, pp. 27050–27069, 2025, doi: 10.1109/ACCESS.2025.3539122.

[27] Y. Özüpak, "Machine learning-based fault detection in transmission lines: A comparative study with random search optimization," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 73, no. 2, p. e153229, 2025, doi: 10.24425/bpasts.2025.153229.