2024 11th International Conference on Internet of Things: Systems, Management and Security (IOTSMS) | 979-8-3503-6650-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/IOTSMS62296.2024.10710259

Fault Detection On Heat Pump Operational Data Using Machine Learning Algorithms

Md Mahbubur Rahman[1,2]
[1]Department of Computer Science and Media Technology
Malmö University
Malmö, 20506, Sweden
[2]Internet of Things and People Research Centre
Malmö University
Malmö, 20506, Sweden
tanmay1007031@gmail.com

Reza Malekian[1,2]
[1]Department of Computer Science and Media Technology
Malmö University
Malmö, 20506, Sweden
[2]Internet of Things and People Research Centre
Malmö University
Malmö, 20506, Sweden
reza.malekian@mau.se

Vilhelm Akerstroem
Engineering Software
Robert Bosch GmbH
Lund, Sweden
Vilhelm.Akerstroem@se.bosch.com

***Abstract*—Heat pumps, being complex systems, are susceptible to various malfunctions. By harnessing contemporary IoT technologies, these devices continuously transmit data which enables monitoring, maintenance, and efficiency. This study focuses on identifying compressor short duration cycles as faults through supervised machine learning algorithms such as XGBoost, Random Forest, SVM, and k-NN. Data preprocessing and labeling were conducted using extensive logged data from heat pump systems, addressing issues like high dimensionality, data sparsity, and temporal dependencies. The methodology included feature engineering, interpolation of missing data, and downsampling for compressor short duration cycles. Supervised machine learning models were applied to classify these short duration cycles. Among the models, XGBoost achieved the highest accuracy and F1-scores, effectively distinguishing between normal and fault conditions. The findings highlight the potential of machine learning to enhance predictive maintenance and operational efficiency in heat pumps.**

***Index Terms*—Heat Pumps, Fault Detection, Compressor Short Duration Cycles, Supervised Machine Learning, Internet of Things**

I. Introduction

The European Union (EU) has implemented regulations aimed at reducing greenhouse gas emissions in order to safeguard natural resources and human health. With the EU's target of achieving carbon neutrality by 2050, there is a pressing need to decarbonize various energy sectors, including heating and cooling, power generation, and transportation [1] The heating and cooling sector, which currently accounts for half of the EU's energy consumption [2] and heavily relies on fossil fuels, will play a pivotal role in transitioning towards low-carbon energy sources. To achieve the national and international energy policies, it is therefore vital to introduce sustainable and efficient renewable-based energy systems into near-zero energy buildings.

Heat pumps represent an innovative and efficient technology to reduce emissions in the heating sector by utilizing electricity instead of fossil fuels [2]. Heat pumps transform electrical power to heat with high efficiency and thus can substitute the usage of fossil fuels for heating purposes. In recent years, the heat pump market has experienced significant growth, with an annual average growth rate of 12% in the EU since 2015 [6]. According to the International Energy Agency (IEA) [3], the global installation of heat pumps is projected to more than triple by 2030 compared to the levels seen in 2020 and has the potential to reduce emissions by at least 500 million tons by 2030 which is equivalent to the annual emission produced by all cars currently operated in Europe.

Today, IoT-connected devices are more present than ever before in vast array of application domains. IoT is characterized by the pervasive deployment of smart and heterogeneous devices (e.g., sensors, actuators, RFIDs) interconnected through the Internet for direct communications without human intervention. According to Transforma Insights [7], currently there are approximately 18.1 billion IoT devices and by 2033 the number of deployed IoT devices is expected to reach 39.4 billion. As a result, the amount of data being generated by IoT devices is expect to reach 79.4 zettabytes by 2025. Industrial applications are now using IoT devices extensively resulting in Industry 4.0 [8]. This shift is facilitated by the internet, industrial assets, cloud technology, cognitive computing and artificial intelligence. The increasing use of connected devices is driving the need for efficient analyses of the generated data in order to make informed decision.

Heat pump is no exception. Typical heat pump systems utilize an array of sensors to monitor the cooling or heating process as well as to monitor both the indoor and outdoor environments [5]. By providing heat pump systems with internet connectivity and smart sensors, manufacturers have access to large quantities of procedurally stored logged information which is crucial for predictive maintenance and system improvement [9]. Inefficient maintenance of heat pumps leads to significant operational inefficiencies and can shorten the system's lifespan due to unnecessary activations such as short compressor runs. Fault detection techniques can mitigate energy losses by 40%, maintenance costs, and service disruptions, ensuring optimal performance and energy efficiency in heating and cooling applications [10].

This paper investigates the detection of compressor short duration cycles in heat pumps using supervised machine learning algorithms. The primary contributions of this research are:

- Development of a comprehensive data preprocessing pipeline addressing high dimensionality, sparsity, and temporal dependencies.
- Implementation and comparison of several supervised machine learning models, including XGBoost, Random Forest, SVM, and k-NN, for fault detection.
- Demonstration of the effectiveness of these models through extensive experimentation on real-world heat pump data.

This research continues previous work on error detection in heat pump operations, conducted in collaboration with Robert Bosch GmbH in Lund, Sweden. [11]

The remainder of this paper is organized as follows: Section II outlines the research objectives and questions. Section III reviews related work in fault detection for heat pumps. Section IV provides the theoretical background. Section V describes the methodology, including data preprocessing and machine learning techniques. Section VI presents the results and discussion. Finally, Section VII concludes the paper with a summary of findings and suggestions for future work

II. Research Objectives and Questions

A. Research Objectives

The primary objective of this research is to implement effective fault detection techniques for the analysis of heat pump operational data. Specifically, this study focuses on detecting compressor short cycling and comparing the efficacy of different fault detection models.

B. Research Questions

- **RQ1:** What are the key characteristics present in large-scale logged data from heat pump systems?
- **RQ2:** Which fault detection models best identify compressor short cycling in heat pumps?

III. Related Work

The critical role of heat pump systems in modern infrastructure underlines the importance of effective fault detection methods. Although numerous studies have focused on fault detection and diagnosis within HVAC systems, particularly in chiller and air handling unit systems [12], there is a notable absence of prior research specifically addressing data-driven fault detection for heat pump operational data. Müller et al. [13] investigated transferring machine learning models for fault detection from experimental to real-world office building heat pump data. Initial satisfactory results on experimental data did not translate well to real-world settings. Li et al. [14] introduced data-driven models for Air Source Heat Pump performance evaluation and anomaly detection using real water heating system data. It demonstrates no generic model that can work for all heat pump and suggests broader datasets for broader insights.

A comparative analysis [15] of supervised learning algorithms for fault detection in air-to-air heat pumps showed k-nearest Neighbors (k-NN) achieving the highest performance with over 99% accuracy, while other methods also exceeded 90%. Another study [16] on wind turbine fault detection using random forests (RF) and extreme gradient boosting (XGBoost) found these methods robust and superior to support vector machines (SVM). Ebrahimifakhar et al. [17] compared nine algorithms for rooftop unit fault diagnosis, with SVM achieving the highest accuracy at 96.2%.

Sulaiman et al. [18] evaluated the impact of various faults on the coefficient of performance (COP) and tested deep learning, SVM, and MLP for fault classification, with MLP achieving the highest accuracy and precision. Ebrahimifakhar et al. [19] found logistic regression most accurate (93.6%) for diagnosing rooftop unit faults compared to k-NN and random forest. W. Yao et al. [20] evaluates tree-based ensemble learning methods RF, XGBoost, LightGBM and multivariate control charts (Hotelling's T2, MCUSUM, MEWMA) for chiller fault detection. The LightGBM-MEWMA method effectively identifies seven common chiller faults with an 88.71% detection rate and 82.3% diagnosis accuracy. Despite these advancements, there remains a significant gap in research specifically targeting data-driven fault detection in heat pump systems. The lack of targeted research on time-dependent features and the need for diverse, real-world datasets highlight the importance for further investigation.

This research aims to explore and implement fault detection techniques for heat pump operational data. This work has the potential to significantly impact the energy sector by optimizing heat pump performance, reducing maintenance costs, and contributing to the transition towards sustainable energy sources.

IV. Theoretical Background

A. Short Cycling in Heat Pump Compressors

Short cycling is a common situation in heat pump where compressor running duration is less than what it should be. This short cycling can cause many problems. Firstly, it requires a lot of wear and tear on the components of the compressor. Every time the compressor starts, it experiences a power surge which eventually makes the mechanical parts and electrical system stressed [21]. This repeated strain can cause the compressor to fail prematurely that can lead to expensive repairs or replacement. Moreover, short cycling has negative effect on the overall efficiency of the heat pump system. This system is designed in a way that supposed to operate longer cycles to provide the most efficiency. The system consumes more energy to achieve the desired heating or cooling state when it turns on and off too often, which eventually lead to increased utility bills [22].The consequences of short cycling extend beyond energy consumption and mechanical wear. The comfort levels within the conditioned space can also be adversely affected. A heat pump that is short cycling may not run long enough to distribute conditioned air evenly throughout the space [23]. This can lead to uneven temperatures, with some areas being too warm and others too cold. Additionally, short cycle can cause fluctuations in humidity levels, further compromising indoor comfort [24]. Addressing short cycling is crucial for maintaining the longevity, efficiency, and performance of heat pump systems. Identifying the root causes of short cycling is essential for implementing effective solutions, ultimately leading to cost savings and extended equipment life

B. Random Forest

Ensemble learning algorithms, such as random forest (RF), bagging, and boosting outperform single classifiers in accuracy and noise resilience [25] [26]. An RF comprises multiple classifiers, each voting for the most frequent class of the input vector x:

$$\text{majority vote}\{\hat{C}_b(x)\}_{b=1}^{B},$$

where $\hat{C}_b(x)$ is the b-th tree's prediction. RFs enhance tree diversity by using bootstrap aggregating (bagging) to create different training data subsets [25], increasing stability and classification accuracy [27]. Research [27] [28] shows RFs are less sensitive to noise and overtraining compared to boosting methods.

Each bagging subset contains about two-thirds of the dataset, with the remaining third forming the out-of-bag (OOB) subset for performance evaluation. The OOB misclassification ratio estimates generalization error, which decreases as the number of trees increases, preventing overfitting [27]. RFs select the best split from a random subset of features at each node, reducing correlation and error, and grow trees without pruning, enhancing computational efficiency. The Gini Index, used for selecting the best split, measures the impurity of an element with respect to other classes:

$$\text{Gini Index} = \sum_{j \neq i} f(C_i|T) \cdot f(C_j|T),$$

where $f(C_i|T)$ is the probability of a case belonging to class C_i. RFs assess feature importance by calculating the decrease in accuracy through oob error estimation and the decrease in the Gini Index when an input variable is switched [28].

C. Extreme Gradient Boosting (XGBoost)

XGBoost, short for "Extreme Gradient Boosting," is an advanced gradient boosting algorithm based on the Gradient Boosting Decision Tree (GBDT) framework, introduced by Tianqi Chen and Carlos Guestrin in 2016 [29]. The core principle of XGBoost is to build models sequentially, where each new model corrects the errors of its predecessors using gradient descent to minimize a specified loss function [30].

In the training process, the jth Decision Tree (DT) is dependent on the previous $j-1$ trees. The relationship between the predictive variable x and the residuals $(y - \hat{y}_{j-1})$ is fitted by the predicted function $f_j(x, \theta_j)$ of the jth DT, described as:

$$\begin{aligned}
\hat{y}_0 &= 0 \\
\hat{y}_1 &= \hat{y}_0 + f_1(x, \theta_1) \\
\hat{y}_2 &= \hat{y}_1 + f_2(x, \theta_2) \\
&\vdots \\
\hat{y}_T &= \hat{y}_{T-1} + f_T(x, \theta_T)
\end{aligned}$$

where T is the number of boosting trees and θ_j represents the structure of the jth DT.

The optimization objective of XGBoost is:

$$\min \text{obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{T} \Omega(f_j)$$

where n is the number of training samples, $l(y_i, \hat{y}_i)$ is the loss function, and $\Omega(f_j)$ is the regularization term for the jth DT, representing the complexity of the tree. The goal is to minimize both the sample loss and the model complexity.

D. Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a robust classifier that seeks to find the optimal hyperplane by maximizing the margin between classes. As the margin increases, the generalization error of the SVM decreases. Consider a dataset with N training samples, each represented as $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i$ denotes the feature vector in the input space and y_i is the class label, which can be either positive $(+1)$ or negative (-1).

Let $\mathbf{z}_i$ represent the feature space vector, mapped from the input space to a high-dimensional feature space via the function Φ. The hyperplane in this feature space can be defined as:

$$\mathbf{z}_i = \Phi(\mathbf{x}_i) \quad (1)$$

$$\mathbf{w} \cdot \mathbf{z} + b = 0 \quad (2)$$

Here, $\mathbf{w}$ is the vector that determines the orientation of the hyperplane, and b is the bias term. The data is considered linearly separable if there exist values for $\mathbf{w}$ and b such that:

$$\mathbf{w} \cdot \mathbf{z}_i + b \geq +1 \quad \text{when} \quad y_i = +1 \quad (3)$$

$$\mathbf{w} \cdot \mathbf{z}_i + b \leq -1 \quad \text{when} \quad y_i = -1 \quad (4)$$

To handle cases where the data is not linearly separable, the constraints (3) and (4) can be modified by introducing non-negative slack variables ξ_i:

$$\mathbf{w} \cdot \mathbf{z}_i + b \geq 1 - \xi_i \quad (5)$$

The ξ_i values are non-zero for data points that do not meet the conditions in (3) or (4).

The SVM constructs the optimal hyperplane by minimizing the following error function:

$$\frac{1}{2}\mathbf{w}^2 + C\sum_{i=1}^{N}\xi_i \quad (6)$$

subject to the constraints in (5). Here, C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error. The minimization of the error function (6) involves the use of Lagrange multipliers and the Kuhn-Tucker conditions from optimization theory. The non-zero coefficients from the Lagrange multipliers correspond to the support vectors. [32]

E. k-NN

The K-Nearest Neighbor (k-NN) algorithm classifies data based on the closest 'K' neighbors using distance metrics like Euclidean, Manhattan, Chebyshev, and Hamming. The Euclidean distance formula is:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(ar(x_i)_k - ar(x_j)_k)^2}$$

where x_i and x_j represent different data points, and $ar(x)$ denotes the attributes of these points. k-NN, an instance-based and lazy learning algorithm, requires no training phase and makes no assumptions about data distribution, enhancing its versatility. Data normalization is crucial for accurate distance measurement and classification.

V. METHODOLOGY

This section introduces the methodology applied during this research. The methodology based on CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, which offers a structured method to data mining and analytics. Our research approach combines both quantitative and qualitative, constituting a mixed-method approach where The quantitative parts of the research consist of experimental testing and comparative analysis of model performance. The qualitative aspects involve a thorough literature review and the exploration of concepts and patterns revealed through an in-depth examination of the domain and research topic.

A. The CRISP-DM Process

This study utilized the CRISP-DM methodology, a widely used framework for managing data mining projects [31]. Figure 1 provides a summary of the CRISP-DM process. Typically, this approach is divided into six key stages, outlined as follows

1) **Business Understanding:** Define business requirements and objectives to create a project plan.
2) **Data Understanding:** Explore data to comprehend its limitations and possibilities.
3) **Data Preparation:** Prepare data through techniques like cleaning, feature engineering, and integration.
4) **Modeling:** Test multiple models and parameters to find optimal configurations.
5) **Evaluation:** Systematically evaluate models to ensure they meet project requirements.
6) **Deployment:** Plan, monitor, and maintain the model(s) and finalize the project.

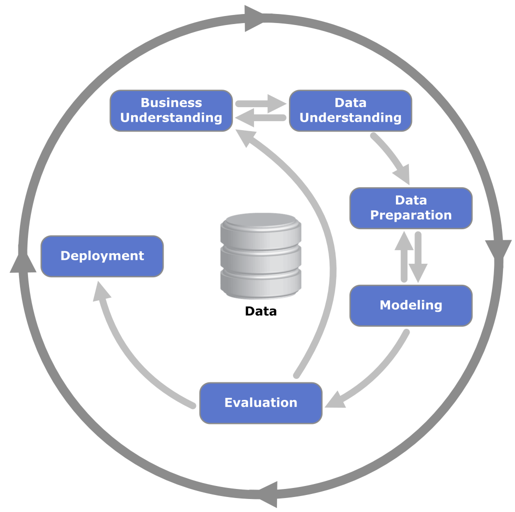

Fig. 1: CRISP-dm

This research follows steps 2-5 in the CRISP-DM model, discarding the business understanding and deployment steps as they are outside the scope. The contributions of this methodology include the comprehensive processing of industrial datasets, the application of machine learning algorithms for fault detection, and the evaluation of model performance using various metrics.

B. Data Understanding

This section explores into the second phase of the CRISP-DM process, focusing on comprehending the data. The datasets, sourced from Robert Bosch AB, which includes operational information on heat pumps from different models and manufacturers across Europe. These time-series datasets include parameter data, command data, control data, and monitor data, recorded as log files. Each heat pump system's dataset comprises roughly 200 to 300 features, with new observations recorded around six times per minute. Even if no changes occur, data is still reported every four minutes, resulting in a sparse dataset predominantly filled with empty values.

Data has been collected over a range from 100 days up to 2 years, resulting in millions of observations per heat pump. The features are not consistent across all models, with only about 50% being common to all models. The features are of various types, including binary values indicating state shifts (e.g., On/Off), tertiary values for three-state shifts, and static contextual information about the heat pump. Numerical features include momentary values from sensors and accumulated values over time, such as total energy consumption. This understanding of the data's structure, volume, and characteristics is crucial for the subsequent data preparation phase.

C. Data Preparation

This section presents the execution of the third stage of the CRISP-DM process, namely, data preparation.

1) Data Extraction and Consolidation: Data for each heat pump is provided in compressed files containing monthly data. These files are extracted to access daily data, which is then converted for easier handling. Finally, the daily data is consolidated into a single file for analysis.

2) Data Interpolation: Heat pump datasets are sparse, with values reported only upon changes. Missing values are handled using forward fill, where each missing value is filled with the most recent observed value. For initial missing values, back fill is used, filling them with the first observed subsequent value, ensuring no leading missing values remain.

3) Feature Engineering : Due to variability in feature counts (100 to 200+ features) across datasets, 85 common features were identified as a baseline. The filter method was used to remove low variance features, which are unlikely to carry meaningful information for predictive modeling, thereby reducing noise and dimensionality. At the end of this step, the number of features varied across datasets, typically ranging from 36 to 55 features, depending on each dataset's specific characteristics. To capture temporal dependencies, feature lagging was applied, creating lagged versions of features to incorporate historical data.

4) Data Normalisation: Normalization ensures scale consistency, crucial for fault detection methods that rely on distance measures. Z-score normalization has been used which standardizes features to have a mean of 0 and a standard deviation of 1.

5) Data Labeling: In this research, data is labeled to identify short operational periods of a compressor. Periods where the compressor is ON for less than 10 minutes are labeled as faults, indicating potential issues. The process involves segmenting data into continuous ON periods and labeling those shorter than 10 minutes as faults. This threshold of 10 minutes was chosen based on insights and recommendations from heat pump domain experts with extensive experience in the operational characteristics and maintenance of heat pumps. This approach creates a dataset highlighting abnormal compressor behavior, which is essential for training models to detect and predict faults.

6) Downsampling: To address dataset imbalance, where instances of short compressor life cycles are few, downsampling is achieved by defining a 12-minute window before and after each compressor's short duration periods. This approach increases the representation of anomalous periods, enhancing the model's ability to learn and identify rare events. Figure 2 illustrates the process of labeling data points and performing downsampling.

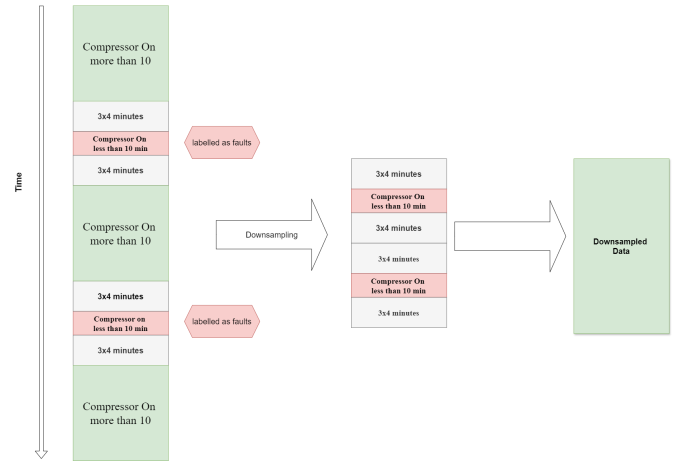

Fig. 2: Data Labeling and Downsampling

7) Data Splitting: Maintaining chronological order, the data is split into training (80%) and testing (20%) sets without shuffling. This preserves temporal dependencies, which is crucial for accurate time series modeling.

D. Modeling

This section describes the modeling step of the CRISP-DM process, focusing on the implementation of various supervised machine learning algorithms to identify the best-performing models. The key steps are:

Algorithm Selection: We chose Random Forest, Balanced Random Forest, XGBoost, Support Vector Machine (SVM), and K-Nearest Neighbors (k-NN) based on their proven effectiveness across a variety of classification tasks, their ability to handle imbalanced datasets, and their performance in terms of accuracy, interpretability, and computational efficiency.

Parameter Configuration: Each algorithm was configured with specific parameters based on own preliminary experiments. Table 1 lists the specific parameters used:

TABLE I: Model Configurations

Algorithm	Parameters
Random Forest	n_estimators=100, criterion='gini', max_depth=2, min_samples_split=2, reg_alpha=0, reg_lambda=1
Balanced Random Forest	n_estimators=100, criterion='gini', max_depth=2, min_samples_split=2, class_weight='balanced_subsample'
XGBoost	n_estimators=100,learning_rate=0.3, max_depth=6
SVM	kernel='linear', C=1.0, class_weight='balanced'
K-Nearest Neighbors (k-NN)	n_neighbors=2, weights='uniform', algorithm='auto', leaf_size=30, p=2

E. Evaluation

This section presents the different metrics utilized to evaluate the results of the models.

1) Classification Terminology: When evaluating fault detection , the problem can be considered as a binary classification task where faults are represented as 1 (Positive), and normal points as 0 (Negative). In binary classification, predictions are categorized into four types:

- **True Positives (TP)**: Faults correctly identified as Faults.
- **True Negatives (TN)**: Normal points correctly identified as normal.
- **False Positives (FP)**: Normal points incorrectly classified as Faults.
- **False Negatives (FN)**: Faults incorrectly classified as normal.

TABLE II: Confusion Matrix

		Predicted Class	
		Negative	Positive
Actual Class	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

2) Evaluation Metrics: In this section, the metrics that were used to evaluate the performance of the fault detection models are presented.

Accuracy: Measures the overall correctness of predictions

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision: Measures the accuracy of positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall: Measures the model's ability to detect all actual faults.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

F1-score: Combines precision and recall into a single metric.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Macro F1-score: Combines precision and recall into a single metric.

$$\text{Macro F1} = \frac{1}{N}\sum_{i=1}^{N} F1_i \tag{5}$$

Process diagram has been shown in figure 3, which is followed during this research.

VI. Result and Discussion

The performance of five different classification models—Random Forest, Balanced Random Forest, XGBoost, SVM, and k-NN was evaluated across various datasets. This discussion focuses on the performance of these models on five selected datasets.The datasets are referred to by their arbitrary IDs: H1, H2, H3, H4, and H5. In the context of a heat pump system, high precision is crucial as it reduces the incidence of false alarms, thereby saving costs associated with unnecessary maintenance. Conversely, high recall is vital for ensuring that most faults are detected, enhancing the system's reliability and

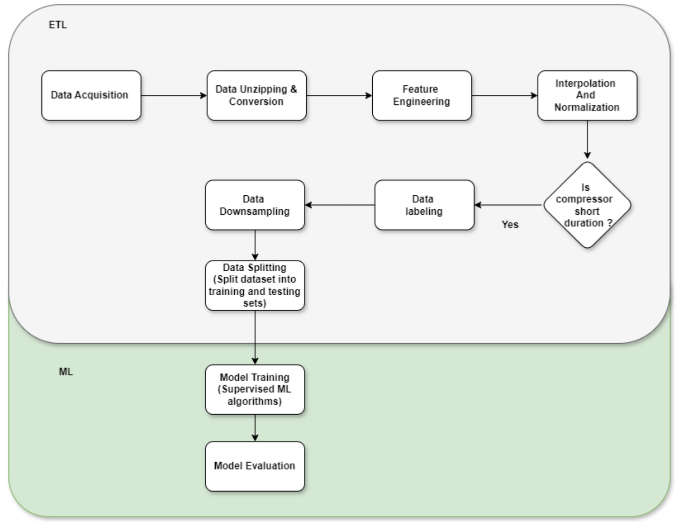

Fig. 3: Process diagram showcasing the fault detection process, from the initial data loading to the final evaluation step

safety. The F1-Score, which combines precision and recall, provides a balanced measure of the overall effectiveness of the fault detection system. The evaluation metrics used in this analysis include accuracy, F1-score for both classes (Normal and Fault), and the macro F1-score. The detailed results for each dataset are summarized below in Tables III-VII.

TABLE III: Model Performance Metrics for Dataset ID: H1

Model	**Accuracy**	**F1-Score (Class 0)**	**F1-Score (Class 1)**	**Macro F1-Score**
Random Forest	0.90	0.94	0.61	0.78
Balanced Random Forest	0.92	0.96	0.76	0.86
XGBoost	0.95	0.97	0.80	0.89
SVM	0.72	0.82	0.44	0.63
k-NN	0.91	0.95	0.68	0.82

TABLE IV: Model Performance Metrics for Dataset ID: H2

Model	**Accuracy**	**F1-Score (Class 0)**	**F1-Score (Class 1)**	**Macro F1-Score**
Random Forest	0.87	0.92	0.74	0.83
Balanced Random Forest	0.89	0.93	0.80	0.87
XGBoost	0.90	0.94	0.80	0.87
SVM	0.72	0.77	0.64	0.71
k-NN	0.91	0.94	0.82	0.88

XGBoost consistently produced the highest accuracy and F1-scores across all datasets, performing well in handling both majority (Normal) and minority (Fault) classes. Balanced Random Forest improved minority class performance significantly over standard Random Forest, demonstrating higher F1-scores for positive class (faults) and better macro F1-scores, making it suitable for imbalanced datasets. SVM showed variable

TABLE V: Model Performance Metrics for Dataset ID: H3

Model	Accuracy	F1-Score (Class 0)	F1-Score (Class 1)	Macro F1-Score
Random Forest	0.95	0.97	0.80	0.89
Balanced Random Forest	0.95	0.97	0.81	0.89
XGBoost	0.97	0.98	0.86	0.92
SVM	0.82	0.89	0.57	0.73
k-NN	0.91	0.95	0.60	0.77

TABLE VI: Model Performance Metrics for Dataset ID: H4

Model	Accuracy	F1-Score (Class 0)	F1-Score (Class 1)	Macro F1-Score
Random Forest	0.89	0.93	0.76	0.84
Balanced Random Forest	0.89	0.93	0.73	0.83
XGBoost	0.89	0.93	0.74	0.83
SVM	0.91	0.94	0.83	0.89
k-NN	0.88	0.93	0.73	0.83

performance, with high precision for the majority (Normal) class but lower recall for the minority (Fault) class, indicating it may not be ideal for imbalanced datasets without additional balancing techniques. k-NN maintained a good balance between accuracy and F1-scores, though it did not reach the performance levels of XGBoost and Balanced Random Forest, making it a reliable but not exceptional choice.

Figure 4 shows the confusion matrix for the dataset H3, where XGBoost correctly identifies 97.96% of the normal instances and 86.41% of the faults, with only 2.04% false positives and 13.59% false negatives, demonstrating its good performance among five models.

VII. Conclusion

This research aimed to develop fault detection techniques specifically for compressor short duration cycles. To guide the study, two research questions were formulated. Answering the first research question: Analysis of large-scale logged data from heat pump systems revealed several key characteristics essential for understanding system performance. Due to the practice of logging only state changes, significant data sparsity was encountered, necessitating advanced data handling techniques to complete the time series data before model training. The dataset's high dimensionality, with 200 to 300 features per system, includes both categorical and numerical data such as operational states and sensor readings. Temporal dependencies played a crucial role, with data collected over 100 to 600 days and influenced by weather conditions. This understanding guided the data preparation steps, which included feature engineering and lagging for temporal dependency, interpolation of missing values, and data downsampling by implementing a 12-minute window for compressor short durations. Most importantly, the focus was on data labeling for supervised machine learning algorithms.

TABLE VII: Model Performance Metrics for Dataset ID: H5

Model	Accuracy	F1-Score (Class 0)	F1-Score (Class 1)	Macro F1-Score
Random Forest	0.97	0.98	0.94	0.96
Balanced Random Forest	0.98	0.99	0.97	0.98
XGBoost	0.99	0.99	0.98	0.98
SVM	0.98	0.99	0.96	0.97
k-NN	0.95	0.97	0.90	0.93

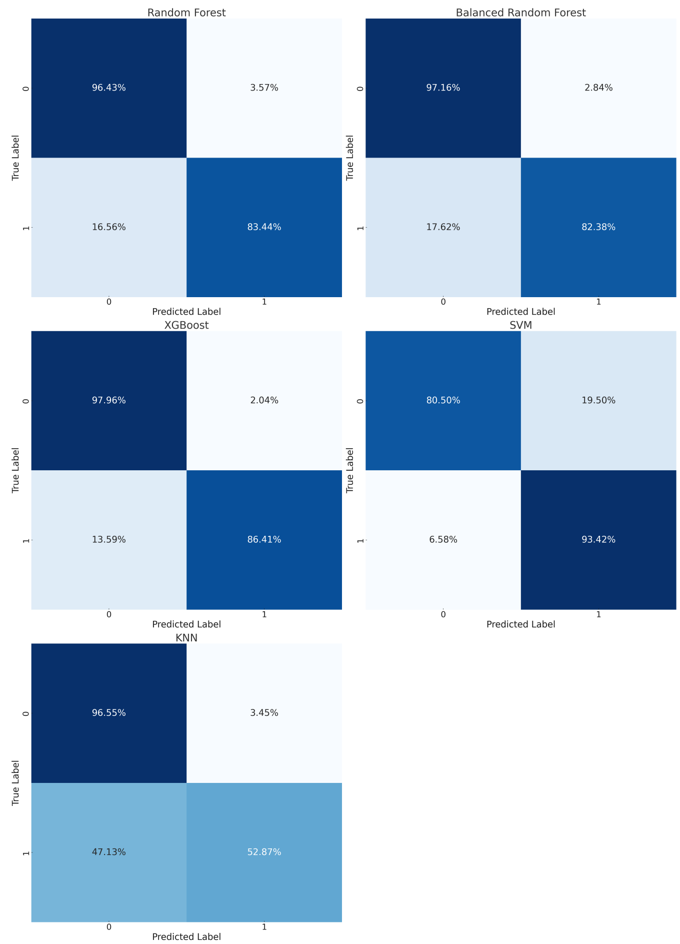

Fig. 4: Confusion Matrix of models on dataset H3

Answering the second research question: We implemented five machine learning models. Among them, XGBoost consistently achieved the highest accuracy and balanced F1-scores, effectively handling both normal and fault classes. Balanced Random Forest also performed well, especially with imbalanced datasets. SVM showed high precision for normal classes but had lower recall for faults, indicating a need for additional balancing techniques. k-NN maintained a good balance between accuracy and F1-scores but did not outperform XGBoost and Balanced Random Forest. Deploying robust fault detection systems using these models can significantly improve the reliability and efficiency of heat pump operations, leading to better maintenance strategies and reduced operational costs.

This research demonstrates that supervised machine learning algorithms, particularly XGBoost, are effective in detecting compressor short cycling in heat pumps. Implementing such models can significantly enhance predictive maintenance and operational efficiency. Future work should focus on developing generalized models that can be applied across different datasets and exploring advanced data balancing techniques to further improve fault detection accuracy.

VIII. Acknowledgment

This research was conducted in collaboration with Robert Bosch GmbH, Lund, Sweden. The authors gratefully acknowledge Michael Wommer and Yuting Tan from Robert Bosch GmbH for their valuable contributions to this paper. This work is partially supported by the Internet of Things and People Research Centre at Malmö University.

References

[1] European Commission, "Heating and Cooling," accessed Mar. 21, 2024. [Online]. Available: urlhttps://energy.ec.europa.eu/topics/energy-efficiency/heating-and-cooling_en

[2] European Environment Agency, "Decarbonising Heating and Cooling in Europe," accessed Jul. 26, 2024. [Online]. Available: urlhttps://www.eea.europa.eu/publications/decarbonisation-heating-and-cooling

[3] International Energy Agency, "The Future of Heat Pumps," accessed Jul. 26, 2024. [Online]. Available: urlhttps://www.iea.org/reports/the-future-of-heat-pumps

[4] M. Wahl, T. Droscher, J. Sprey, and A. Moser, "Modelling of Heat Pump Load Profiles for Grid Expansion Planning," in *Proc. 2018 53rd Int. Univ. Power Eng. Conf. (UPEC)*, 2018, pp. 1-6, doi: 10.1109/UPEC.2018.8541958.

[5] Y. H. V. Lun and S. L. D. Tung, *Heat Pump Configuration*. Springer International Publishing, 2020.

[6] IEA, *The Future of Heat Pumps*, License: CC BY 4.0, IEA, 2022. [Online]. Available: urlhttps://www.iea.org/reports/the-future-of-heat-pumps

[7] Transforma Insights, "Current IoT Forecast Highlights," Transforma Insights, 2024. [Online]. Available: urlhttps://transformainsights.com/research/forecast/highlights

[8] S. Vaidya, P. Ambad, and S. Bhosle, "Industry 4.0 – A Glimpse," *Procedia Manufacturing*, vol. 20, pp. 233-238, 2018.

[9] A. Blázquez-García, A. Conde, U. Mori, and J. Lozano, "A Review on Outlier/Anomaly Detection in Time Series Data," *ACM Computing Surveys*, vol. 54, Apr. 2021, pp. 1-33, doi: 10.1145/3444690.

[10] I. Bellanco, E. Fuentes, M. Vallès, and J. Salom, "A review of the fault behavior of heat pumps and measurements, detection and diagnosis methods including virtual sensors," *Journal of Building Engineering*, vol. 39, Jan. 2021, p. 102254, doi: 10.1016/j.jobe.2021.102254.

[11] M. Epsteins and F. Forsström, "Finding Known and Novel Errors in Heat Pumps Using Unsupervised ML," LU-CS-EX, Master Thesis, Department of computer science lth, Lund University, 2024.

[12] Y. Li and Z. O'Neill, "A critical review of fault modeling of HVAC systems in buildings," *Building Simulation*, vol. 11, no. 5, pp. 953–975, Oct. 2018, doi: 10.1007/s12273-018-0458-4.

[13] G. Bode, S. Thul, M. Baranski, and D. Mueller, "Real-world application of machine-learning-based fault detection trained with experimental data," *Energy*, vol. 198, Mar. 2020, p. 117323, doi: 10.1016/j.energy.2020.117323.

[14] W.-T. Li, N. U. Hassan, F. Khan, C. Yuen, and Y. M. Keow, "Data Driven Model for Performance Evaluation and Anomaly Detection in Integrated Air Source Heat Pump Operation," in *Proc. 2019 IEEE Int. Conf. Ind. Technol. (ICIT)*, 2019, pp. 1280-1285, doi: 10.1109/ICIT.2019.8755022.

[15] P. Barandier, M. Mendes, and A. J. M. Cardoso, "Comparative analysis of four classification algorithms for fault detection of heat pumps," *Energy and Buildings*, vol. 316, 2024, Art. no. 114342, doi: 10.1016/j.enbuild.2024.114342. [Online].

[16] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost," *IEEE Access*, vol. 6, pp. 21020-21031, 2018, doi: 10.1109/ACCESS.2018.2818678.

[17] A. Ebrahimifakhar, A. Kabirikopaei, and D. Yuill, "Data-driven fault detection and diagnosis for packaged rooftop units using statistical machine learning classification methods," *Energy and Buildings*, vol. 225, Art. no. 110318, 2020, doi: 10.1016/j.enbuild.2020.110318.

[18] N. A. Sulaiman, P. Abdullah, H. Abdullah, and M. N. Shah, "Fault detection for air conditioning system using machine learning," *International Journal of Artificial Intelligence*, vol. 9, no. 1, pp. 109–116, 2020, doi: 10.11591/ijai.v9.i1.pp109-116.

[19] A. Ebrahimifakhar and D. Yuill, "Application of Machine Learning Classification Methods in Fault Detection and Diagnosis of Rooftop Units," 2021.

[20] W. Yao, D. Li, and L. Gao, "Fault detection and diagnosis using tree-based ensemble learning methods and multivariate control charts for centrifugal chillers," *Journal of Building Engineering*, vol. 51, p. 104243, 2022, doi: 10.1016/j.jobe.2022.104243.

[21] S. Ilic, C. Bullard, and P. Hrnjak, "Effect of shorter compressor on/off cycle times on A/C system performance," Air Conditioning and Refrigeration Center, CR-43, 2001.

[22] M. Dongellini and G. L. Morini, "On-off cycling losses of reversible air-to-water heat pump systems as a function of the unit power modulation capacity," Energy Conversion and Management, vol. 196, pp. 966-978, 2019.

[23] Fire & Ice, "What Is Short Cycling in HVAC and Why Is It a Problem?," Accessed: 2024-05-26. [Online]. Available: https://indoortemp.com/resources/short-cyclinghvac-problem.

[24] C. Roe, "How Long Should AC Stay Off Between Cycles," Precision Heating & Air, 15-Nov-2023. [Online]. Available: https://www.precision-hvac.com/how-long-should-ac-stay-off-between-cycles/. [Accessed: 14-Jun-2024]..

[25] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, pp. 123–140, 1996.

[26] T. G. Dietterich, "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization," Machine Learning, vol. 40, no. 2, pp. 139–157, Aug. 2000, doi: 10.1023/A:1007607513941. [Online]. Available: https://doi.org/10.1023/A:1007607513941.

[27] J. C.-W. Chan and D. Paelinckx, "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," Remote Sensing of Environment, vol. 112, no. 6, pp. 2999–3011, 2008, doi: https://doi.org/10.1016/j.rse.2008.02.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425708000679.

[28] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," Remote Sensing of Environment, vol. 86, no. 4, pp. 554–565, 2003, doi: https://doi.org/10.1016/S0034-4257(03)00132-9. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425703001329.

[29] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[30] C. Bentejac, A. Csorgo, and G. Martinez-Munoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, pp. 1937-1967, 2021.

[31] R. Wirth and J. Hipp. "CRISP-DM: Towards a standard process model for data mining". In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining (Jan. 2000).

[32] E. Vaiciukynas, A. Gelzinis, and A. Verikas, "Exploring similarity-based classification of larynx disorders from human voice," *Speech Communication*, vol. 54, no. 5, pp. 601-610, 2012, doi: 10.1016/j.specom.2011.04.004.