Hybrid Detection for Heat Pump Contamination Using Physics-Informed Machine Learning

Ahmed Qarqour[1], Gernot Heisenberg[2], Sahil-Jai Arora[3], Drazen Martinovic[4]

[1,3,4]*Bosch Thermotechnik GmbH, 73243 Wernau (Neckar), Germany*

[1,2]*Institute of Information Science, Technical University of Applied Sciences Cologne, 50678 Cologne, Germany*

Ahmed.qarqour@de.bosch.com

ABSTRACT

The deployment of residential heat pump systems is a key enabler of the decarbonization of the heating sector. However, their long-term reliability remains a barrier to sustained performance and user acceptance. A major degradation driver is water contamination within the hydraulic circuit, which leads to fouling, scaling, and corrosion of components such as plate heat exchangers – ultimately reducing efficiency and shortening system lifetime. Although installation procedures and operational filtration measures, including magnetite filtration, aim to reduce particle accumulation, continuous condition-based monitoring of component degradation remains limited. To address the scarcity of real-world failure data for training predictive models, this paper proposes a physics-informed, data-prior approach that combines physical knowledge with machine learning. Instead of embedding physics into the model architecture or loss functions, the approach incorporates it at the data level by generating labeled healthy and faulty scenarios through a physics-based laboratory setup. This enables the model to learn degradation patterns grounded in physical behavior, supporting early fault detection and producing outputs that remain interpretable and plausible for domain experts. The approach is demonstrated on a plate heat exchanger contamination use case. A design-of-experiments campaign in a climate chamber generated labeled data representing healthy, moderately contaminated, and severely contaminated states. A Random Forest classifier achieved consistent cross-validation performance (AUC ≈ 0.98) with low variance across folds. Precision–recall analysis revealed robust early fault detection, with average precision values of approximately 0.96 for moderate contamination and 0.97 for severe contamination. Cumulative gain and lift analyses indicated that inspecting the top 20–40 % of systems ranked by model risk can identify 80–100 % of the contaminated cases, supporting efficient maintenance prioritization. Model-derived feature importance was assessed using Gini importance and subsequently validated through expert review, enabling interpretable failure logic for condition-based maintenance strategies. The results demonstrate that combining physically grounded data, supervised machine learning, and explainable diagnostics provides a transferable hybrid approach for interpretable reliability assessment and condition-based monitoring beyond the specific case.

1. INTRODUCTION

Residential heat pump systems are increasingly deployed as part of the transition toward low-carbon heating (International Energy Agency, 2022). At the same time, ensuring their long-term operational reliability remains a central engineering challenge. Persistent degradation processes concerns – particularly those linked to operational degradation – can compromise sustained system performance and increase operational risk (Qarqour et al., 2024). Water contamination in the hydraulic circuit is a significant degradation factor, as it directly causes scaling, fouling, and corrosion in components such as plate heat exchangers. This degradation not only reduces thermal efficiency but also shortens the system's useful life, underlining the need for effective monitoring and predictive maintenance strategies (Tovazhnyanskyy et al., 2007; Ardsomang et al., 2013).

Reliability modeling offers a range of approaches to assess degradation in engineering systems. These are typically categorized into three types: physics-based, data-driven, and hybrid models (Lei et al., 2018). Each differs in how it balances physical understanding, data requirements, and model interpretability. Heat pump systems involve multiple interacting components, often sourced from different suppliers. The level of available data and domain knowledge varies across components, making the selection of a suitable

modeling approach particularly challenging (Qarqour et al., 2025).

In previous work, we addressed this modeling selection challenge by developing a structured decision framework that links key system indicators – such as physical knowledge and data availability – to appropriate reliability modeling strategies (Qarqour et al., 2025). When applied to plate heat exchanger degradation, this framework indicated hybrid modeling as the most suitable approach. However, the practical implementation and validation of the identified hybrid approach for plate heat exchanger degradation remained open and are addressed in this work.

Hybrid modeling approaches differ in the way physical knowledge is incorporated into the learning process. This physics-informed learning has been classified by Karniadakis et al. (2021) into three types: data priors (physics enters via data), inductive biases (via model structure), and learning biases (via loss functions). Embedding physics in this way supports the generation of physically plausible patterns, enabling validation against real-world behavior and facilitating explainable diagnostics for condition-based monitoring (Wu et al., 2024).

Fouling in plate heat exchangers is a widespread degradation mechanism that reduces efficiency and shortens component lifetime – even under proper water treatment (Ardsomang et al., 2013). While both physics-based and data-driven approaches exist to model fouling, they face limitations in real-world applications (Lei et al., 2018; Ardsomang et al., 2013). Physics-based methods often fail to capture dynamic system behavior, while purely data-driven ones require extensive sensor data and often lack interpretability (Hou et al., 2025; Meng et al., 2025). As a result, a systematic and transferable approach that combines physical understanding with data-driven monitoring remains missing for residential heat pump systems. This motivates the hybrid approach proposed in this work.

To address the identified gap, this work proposes a hybrid approach that integrates physical understanding at the data level. In line with the data-prior category described above, the approach incorporates physical knowledge at the data level by generating labeled healthy and faulty scenarios through a controlled laboratory setup. This design of experiments campaign enables the machine learning model to learn degradation patterns grounded in physical system behavior. As a result, the outputs remain interpretable, physically plausible, and suitable for early fault detection. The approach is demonstrated on a plate heat exchanger use case, where contamination scenarios are experimentally varied to emulate realistic fouling conditions for model training and validation.

In contrast to conventional diagnostic strategies that often separate physical system understanding from data-driven evaluation, the proposed approach combines controlled degradation generation, interpretable machine learning, and operationally relevant monitoring concepts within a unified diagnostic workflow. By systematically translating physically defined degradation mechanisms into data-driven diagnostics, the approach establishes a bridge between domain knowledge and scalable condition-based monitoring. This foundation supports the reliable deployment of interpretable machine learning models in residential heat pump systems and related engineering applications. It also provides a transparent basis for supervised degradation modeling in settings where real-world failure data are scarce.

2. BACKGROUND AND RELATED WORK

In residential heat pump systems, plate heat exchangers are essential for transferring heat between the refrigerant circuit and the water or brine loop (Jnod Energy, 2026). Their compact design, high heat transfer efficiency, and low temperature approach make them well-suited as evaporators or condensers (Munnangi et al., 2026). These properties contribute to heat pump's ability to operate efficiently across a range of conditions. As the primary thermal interface, the reliability of PHEs is crucial for maintaining system performance, minimizing energy consumption, and ensuring stable long-term operation (Kapustenko et al., 2023).

Fouling is a common and persistent issue in both industrial and residential heat exchangers. It leads to reduced heat transfer efficiency, increased energy consumption, and shorter component lifetimes (Kapustenko et al., 2023). Even when water quality is managed during installation, contamination can still occur during operation, especially in systems with limited filtration or harsh water conditions (Virginia Heat Transfer, 2026). In residential PHEs, fouling often builds up gradually and silently, making it a critical degradation mechanism that is difficult to detect (Berce et al., 2021). This underlines the need for reliable monitoring strategies to assess fouling severity and its impact on system performance.

Fouling in plate heat exchangers has traditionally been addressed using empirical correlations or fixed fouling factors applied in design calculations. In addition, indirect performance indicators derived from measured temperatures and flow rates have been used to infer degradation effects (Kapustenko et al., 2023). In practice, monitoring often relies on estimating heat transfer effectiveness or overall heat transfer coefficients based on thermodynamic balances (Romanowicz et al., 2023). While these methods are straightforward to implement, they depend on strong assumptions regarding fluid properties, flow regimes, and sensor accuracy. In practice, these assumptions often do not hold under real operating conditions (Kapustenko et al., 2023). Consequently, they often fail to capture the complex, time-varying progression of fouling observed in practical field deployments (Patil et al., 2022).

To overcome these limitations, physics-based approaches have been proposed. Alhuthali et al. (2022) developed a dynamic plate heat exchanger model integrating protein denaturation kinetics with a fouling deposition formulation and improved parameter adaptability using dimensional analysis and symbolic regression. Ikonen et al. (2023) introduced a monitoring framework that combines thermodynamic modeling, state estimation, and vibration-based sensing to estimate heat transfer degradation in real time. Although these approaches enhance modeling accuracy and operational insight, they depend on detailed system-specific calibration, comprehensive physical knowledge, and, in some cases, dedicated instrumentation.

Data-driven approaches have gained attention for their ability to detect fouling patterns directly from operational measurements without requiring explicit thermodynamic modeling (Soomro et al., 2026). Kouidri et al. (2025) developed cascaded forward and recurrent neural network models to estimate fouling resistance using temperature and flow data, achieving high predictive accuracy. Similarly, Sundar et al. (2020) developed a deep learning model to predict fouling thermal resistance in a cross-flow heat exchanger using operational data, including temperatures and flow rates, and demonstrated its capability to capture nonlinear relationships between operating conditions and fouling levels. These approaches enable automated detection and prognostic assessment from field data. However, despite their predictive capability, such models often act as black boxes and provide limited physical interpretability for engineers and operators (Soomro et al., 2026).

To address these limitations, recent studies advocate hybrid modeling approaches that integrate physical knowledge with data-driven learning to enhance interpretability, robustness, and deployment readiness in practical applications (Wu et al., 2024). Tang et al. (2025) present a learning-bias hybrid approach by proposing a physics-informed LSTM (PI-LSTM) framework for dynamic heat exchanger modeling. In their work, thermodynamic constraints are incorporated directly into the neural network loss function, enforcing energy conservation during training. By coupling physical residuals with temporal sequence learning, the model reduced data dependency and improved generalization compared to purely data-driven architectures. However, the framework focused on dynamic system response prediction rather than degradation classification or reliability assessment.

In contrast, Hou et al. (2025) adopted a data-prior hybrid strategy by combining simulation-based data generation with an LSTM network to track degradation and forecast key thermal metrics. Their framework linked physically defined fouling thresholds with real-time monitoring and demonstrated that embedding degradation knowledge at the data level can improve predictive consistency under varying operating conditions. However, the approach remained centered on component-level performance forecasting and did not explicitly account for broader system interactions or varying operational regimes. Similarly, Sansana et al. (2024) integrated mechanistic feature engineering with machine learning to forecast fouling-related key performance indicators. In their framework, physics-based surrogate variables derived from heat transfer principles are constructed prior to model training, thereby enhancing interpretability and maintenance relevance. While effective for long-term forecasting in an industrial heat exchanger, the approach was likewise limited to a single component and did not explicitly address system-level interactions or transferability across different application domains.

Although hybrid approaches have demonstrated improved fouling detection and performance forecasting in plate heat exchangers, their applicability to residential heat pump systems remains limited. Existing studies predominantly focus on isolated component behavior and emphasize thermal performance prediction rather than degradation classification for reliability assessment. Furthermore, system-level interactions under varying operating modes are rarely considered. The influence of seasonal conditions and operational variability on degradation detection is therefore insufficiently addressed. Consequently, the transferability of current hybrid solutions to real-world residential heat pump environments remains constrained. To address this gap, this work introduces a physics-informed, data-prior approach that links component-level degradation to system-level behavior, enabling interpretable and transferable monitoring in residential heat pump systems.

3. HYBRID RELIABILITY MODELING APPROACH

For the detection of component degradation in a physically meaningful and scalable way, this work introduces a hybrid approach that links system behavior to machine learning-based monitoring. The approach follows a physics-informed learning strategy in which physical knowledge is incorporated at the data level through experimentally defined degradation states. These states form the basis for supervised training under controlled operating conditions. The approach consists of four stages that translate physical degradation behavior into interpretable, data-driven diagnostics: (1) degradation design, (2) data generation, (3) model development, and (4) evaluation. The overall objective is to support condition-based monitoring by enabling early fault detection, providing interpretable degradation patterns, and supporting maintenance prioritization. These stages are summarized in Figure 1 and described sequentially within this section.

The initial stage (Degradation Design) defines the physical setup and control strategy required to enable structured data generation for degradation monitoring. the proposed approach follows a physics-informed strategy in which degradation is deliberately induced based on known failure mechanisms. This enables the definition of targeted design

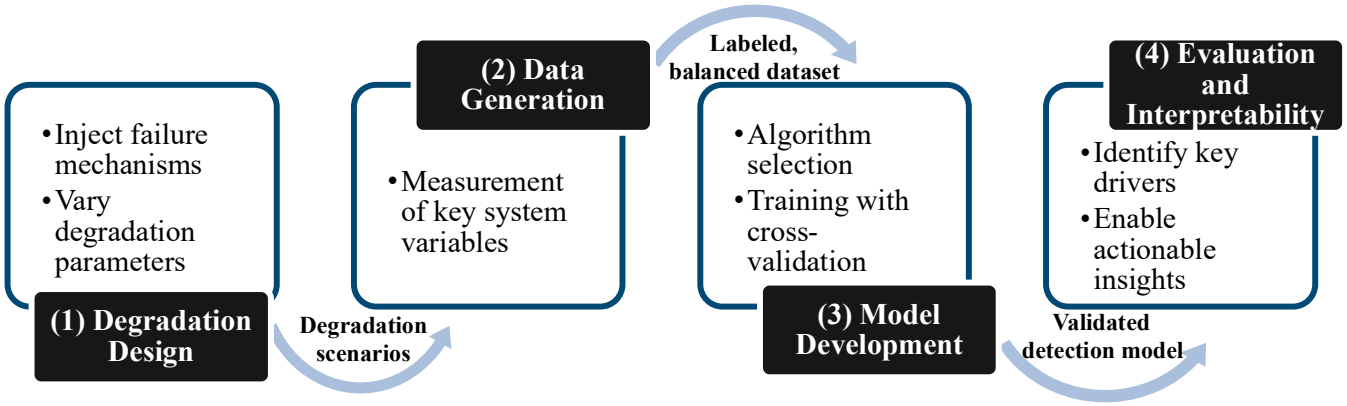

Figure 1. Overview of the proposed hybrid approach for component-level degradation detection and monitoring. The workflow illustrates the four-stage process linking controlled degradation design, structured data generation, model development, and interpretable evaluation.

parameters, which are then actively manipulated to simulate the onset and progression of component-level degradation. Relevant operational and environmental variables – such as flow conditions and temperature profiles – are systematically varied to reflect their effect on system behavior under different degradation levels. To support learning of meaningful and generalizable patterns, the degradation process is designed to be reproducible and clearly structured. Healthy and degraded states are separated based on observable system responses, enabling consistent labeling and robust model training. This setup forms the basis for interpretable, condition-based diagnostics in the subsequent stages.

The following stage 2 (Data Generation) aims to generate structured, labeled datasets that capture system behavior under defined degradation conditions. To achieve this, the experimental setup must be equipped with domain-relevant sensors for measuring variables such as temperature, pressure, and flow rate. This ensures accurate and repeatable acquisition of key system variables. These variables should be selected in consultation with domain experts to ensure they capture both degradation-relevant influences and observable effects. Each operating scenario is labeled based on experimentally defined degradation levels that produce observable shifts in thermal and hydraulic system response, thereby ensuring physically grounded class boundaries. The controlled environment reduces confounding effects typically present in field data, such as ambiguous boundary conditions. To support balanced learning, class distributions must be actively shaped to provide sufficient representation of all degradation levels. The resulting dataset is then preprocessed into a standardized format, including systematic labeling, handling of missing values, and verification of class balance. By linking controlled degradation states to measurable system responses, this stage establishes a robust foundation for supervised model training and subsequent interpretability in condition-based monitoring applications.

The objective of Stage 3 (Model Development) is to convert the labeled dataset into a predictive model capable of detecting component-level degradation. The modeling procedure is designed to ensure both high predictive performance and interpretability, thereby supporting integration into condition-based monitoring frameworks. The process begins with the selection of an appropriate machine learning algorithm. Following algorithm selection, the training is performed using k-fold cross-validation to assess generalization across varying operating conditions.

A successful training procedure is characterized by stable validation performance and low inter-fold variance, indicating that the model captures consistent degradation patterns across varying validation splits. Upon completion of this stage, the approach proceeds to model evaluation and interpretability – establishing links between predictive outputs and physically meaningful system behavior.

The fourth stage (Evaluation and Interpretability) assesses the model's predictive performance and ensures that its outputs can be meaningfully interpreted in the context of system behavior. Evaluation begins with standard performance metrics such as precision–recall, which quantify the model's ability to distinguish between healthy and degraded states - including early-stage faults. To further assess practical usefulness, cumulative gain and lift analyses are applied. These metrics evaluate how well the model can prioritize systems at highest risk, enabling targeted maintenance strategies based on ranked predictions. Beyond predictive performance, model interpretability is explicitly addressed. Feature importance measures – such as Gini impurity–based importance or Shapley values – are computed to identify which input variables drive the model's decisions. These results are then examined in relation to known physical system behavior, enabling validation of the learned patterns from an engineering perspective. This step ensures that the model does not function as a black box but provides transparent and physically meaningful diagnostics. Together,

performance evaluation and interpretability assessment establish the model's suitability for reliable condition-based monitoring and informed maintenance decision support.

The following section applies the described approach to a residential heat pump system, focusing on plate heat exchanger degradation as a representative use case. The case study demonstrates the practical implementation of the modeling stages and provides empirical evaluation of the proposed approach.

4. PLATE HEAT EXCHANGER CASE STUDY

To demonstrate the practical implementation and validation of the proposed approach, this section applies it to a plate heat exchanger (PHE) within a residential heat pump system. The component is prioritized through a prior assessment framework (Qarqour et al., 2025) and selected due to its relevance for hybrid reliability modeling. The subsequent sections present the systematic application of the approach, progressing from system characterization and degradation design to model development and interpretation. This use case demonstrates how experimentally derived knowledge can be translated into reliable prognostics and interpretable diagnostics suitable for field deployment.

4.1. System Context

The investigated use case involves a residential air-to-water heat pump system installed in an outdoor environment and tasked with providing space heating and domestic hot water. The system is integrated into a building infrastructure that includes hydronic distribution, domestic water storage, and user-specific thermal demands. Its operation is influenced by dynamic external factors such as ambient temperature fluctuations, water chemistry, installation quality, and building-specific usage profiles (Qarqour et al., 2024). The core system consists of several interacting components and functional modules, including a PHE that serves as the condenser. The PHE facilitates thermal energy transfer from the refrigerant loop to the water circuit, and its performance is critical for achieving the desired heating output and overall system efficiency. Given its exposure to variable flow rates, fluctuating return temperatures, and long-term water-side interactions, the PHE is particularly vulnerable to degradation mechanisms such as fouling and scaling (Kapustenko et al., 2023).

To support efficient operation and enable basic fault detection, the system is equipped with a range of sensors distributed across the refrigerant and water circuits, as well as the surrounding environment (Brudermueller et al., 2025). These include pressure and temperature sensors distributed along the refrigerant circuit (e.g., compressor suction line, condenser inlet and outlet), as well as pressure and flow sensors on the water side and ambient air temperature sensors. The collected measurements are continuously processed by the system controller to regulate setpoints, determine operating modes, and trigger safety protocols through predefined error codes (Qarqour et al., 2024). Sensor configurations and placements are sometimes adapted in the field to address site-specific issues or improve diagnostic coverage. However, the effects of degradation – such as scaling or partial blockage - do not manifest uniformly. Depending on its location and severity, degradation can influence multiple parameters across subsystems, leading to complex, nonlinear interactions. As a result, predefined error codes – typically designed for control and protection mechanisms rather than early fault diagnosis – are often insufficient for reliable detection. These challenges motivate modeling approaches that leverage system behavior and contextual sensor information to identify degradation patterns more effectively. The resulting data – spanning thermal, hydraulic, and operational variables – forms a high-resolution stream that enables the development of advanced condition-based monitoring strategies.

The following section corresponds to the first stage of the proposed approach for hybrid reliability model development. It outlines the experimental setup used to emulate PHE degradation under controlled conditions and generate representative data for model training.

4.2. Experimental Setup

This section demonstrates the implementation of Stage 1 (Degradation Design) of the proposed approach, which focuses on replicating fault conditions in a controlled laboratory environment. In this use case, the objective is to simulate fouling-related contamination of the PHE, specifically on the water side of the component. In real-world installations, suspended particles originating from internal plumbing systems can accumulate on the plate surfaces over time, leading to a gradual reduction of the effective heat transfer area. The resulting fouling layer acts as an additional thermal resistance on the plate surface, limiting convective heat exchange between the fluid and the wall. As a result, the overall heat transfer performance of the exchanger decreases, impairing energy transfer between the refrigerant loop and the water circuit.

In addition to increased thermal resistance, fouling can also alter the flow behavior within the PHE. To emulate these coupled degradation effects in a controlled and repeatable manner, the experimental setup focuses on reproducing the contamination-induced reduction in effective flow and heat-transfer area. To achieve this, a sleeve is inserted at the water inlet of the PHE to create restricted flow distribution across defined portions of the internal plate structure. Different sleeve lengths are used to generate predefined degradation configurations with progressively reduced active plate participation. The experimental design includes three degradation levels, defined in consultation with domain experts: a clean reference case (0% blockage), a moderately contaminated state (approximately 50 % blockage), and a

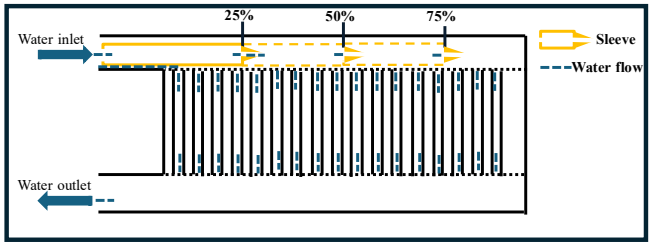

Figure 2. Schematic representation of the plate heat exchanger (PHE) blockage setup illustrating the sleeve position at the water inlet and the resulting reduction of the active flow area.

severely contaminated state (approximately 75% blockage). These states are selected based on observed effects on flow rate and thermal behavior and physically correspond to reduced active plate area inside the exchanger. The principle of this degradation emulation is illustrated in Figure 2, which shows the sleeve position within the flow path and the resulting reduction of the active water-side flow area in the PHE. The exchanger operates in a counterflow arrangement between the water and refrigerant liquid. While the setup does not directly reproduce microscopic fouling layer formation on the plate surfaces, it emulates the resulting thermohydraulic effects associated with contamination-induced flow restriction and reduced effective heat-transfer utilization.

To ensure that degradation effects could be isolated from external influences, all three contamination states are tested under identical environmental and operational conditions. These include a standardized heating curve, a fixed domestic hot water draw profile, and outdoor temperature variations d erived from a DIN-standard annual profile. This setup allows each experiment to simulate a full seasonal cycle (spring, summer, autumn, winter) while maintaining consistent external load factors across all system states. Each contamination state is operated for the same duration to ensure comparability and balanced data generation across classes. The result is a physically meaningful degradation scenario with clearly separable system behavior, forming the foundation for the subsequent data generation stage. The system is installed in a programmable climate chamber and operated using a hardware-in-the-loop setup to ensure reproducible execution of temperature and load profiles.

The effectiveness of the simulated fouling is confirmed by consistent shifts in system behavior across degradation states. Specifically, compressor start frequency, heating cycle duration, and pump energy consumption all increased progressively from the healthy to the severely contaminated condition. These trends align with physical expectations – reduced heat transfer through the PHE increases thermal resistance, requiring longer and more frequent operation to meet setpoints, and demanding higher flow effort from the water-side pump. Figure 3 illustrates these effects in a comparative bar chart, highlighting the average values of

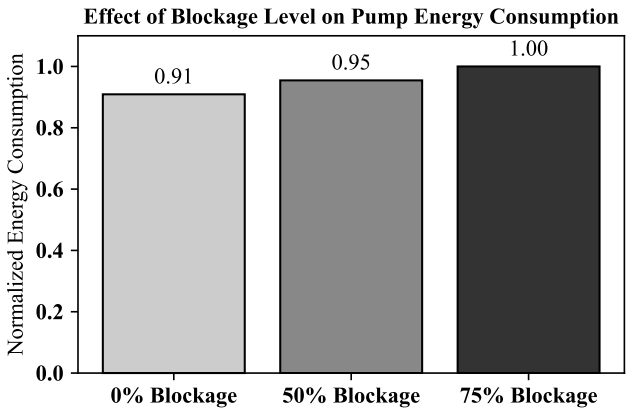

Figure 3. Normalized pump energy consumption across the three contamination levels (0%, 50%, and 75% blockage), showing the average values measured during the experiments.

selected parameters across the three states. These results are reviewed and validated by domain experts, who confirmed the physical plausibility of the trends and their diagnostic relevance.

4.3. Collected Dataset

This section corresponds to Stage 2 (Data Generation) of the proposed approach. Based on the experimental design in Stage 1, a labeled dataset is generated to represent three system states: healthy, moderately contaminated, and severely contaminated. The heat pump system is operated in a climate chamber across a simulated annual cycle (spring, summer, autumn, winter) for each contamination level, ensuring a balanced and representative data distribution. Sensor data is recorded at 0.03 Hz (5 samples per minute), capturing thermal, hydraulic, and operational dynamics.

In total, the experiment yields approximately 45,000 samples, approximately equally distributed across the three system states to avoid class imbalance during model training and to enable controlled comparison of physically interpretable degradation patterns across the defined contamination levels.

Figure 4 illustrates the refrigerant and water circuits of the heat pump system, which interact through the PHE as the thermodynamic interface between both domains, together with the distribution of the integrated sensor system across these circuits. The dataset includes readings from temperature sensors (e.g., refrigerant inlet/outlet of the PHE, water supply/return), pressure sensors on both sides, and flow meters on the water loop. Additional sensors monitored ambient air temperature and key operating signals from the controller. This high-resolution multi-domain data captures the complex system response to degradation and forms the foundation for robust classification.

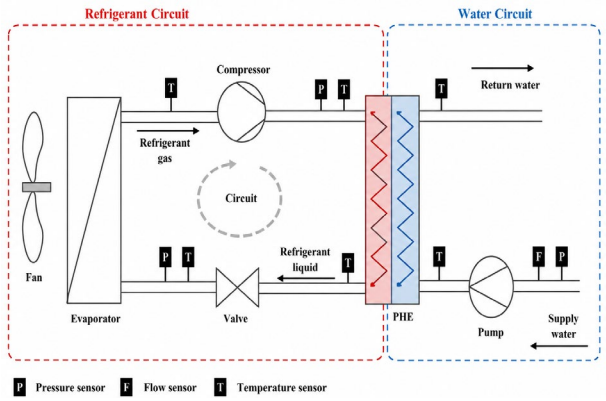

Figure 4. Simplified schematic of the heat pump system indicating the refrigerant and water circuits and the locations of the temperature (T), pressure (P), and flow (F) sensors used for data acquisition.

To ensure data quality, a standardized preprocessing procedure is applied. This includes time alignment of multi-sensor streams and removal of corrupted or missing entries. High-frequency sensor fluctuations, arising from measurement resolution and control dynamics, are smoothed using a moving average filter. Short signal gaps result from the BUS-based data transmission logic, where values are updated only upon change. These gaps are imputed using a zero-order hold strategy. Longer gaps are excluded to avoid introducing bias. Importantly, no normalization or scaling is performed. The selected features represent physically meaningful system variables with consistent measurement units and comparable operating ranges, such that preserving their original scale does not distort the inherent data structure. Retaining physical units supports interpretability and consistency with domain expertise. The resulting dataset maintains the temporal resolution and operational variability required for reliable degradation detection.

4.4. Model Training and Performance

To implement Stage 3 (Model Development) of the proposed approach, a supervised learning model is developed to detect and classify the degradation state of the PHE. The modeling task is framed as a multiclass classification problem, assigning each system instance to one of the three discrete classes: healthy, moderately contaminated, or severely contaminated. This structure is enabled by fully labeled experimental data, generated under controlled conditions with clearly defined physical criteria for each degradation level. Given the discrete definition of the degradation states in the experimental setup, classification is selected over regression. A continuous degradation index is not pursued, as the laboratory design does not provide a physically measurable continuous degradation variable. Instead, the objective was to generate interpretable condition labels that align with practical maintenance decision categories.

To identify the most suitable algorithm, a diverse set of classification models is evaluated. These included linear models (Logistic Regression), kernel-based models (Support Vector Classifier, Naive Bayes), and tree-based models (Decision Tree, Random Forest, Light Gradient Boosting Machine), selected as widely adopted and representative implementations of their respective learning paradigms. This range of models is chosen to cover fundamentally different model assumptions, each offering distinct capabilities in handling nonlinearity, feature interactions, and interpretability. All models are trained on the same preprocessed dataset derived from the controlled degradation experiments, ensuring consistent feature distributions and balanced class representation. A stratified five-fold cross-validation scheme is employed to obtain robust generalization estimates, with each sample used once for validation and class proportions preserved across folds. To benchmark classification performance, receiver operating characteristic (ROC) curves are generated for each model, serving as a robust tool for identifying the most effective model.

To rigorously evaluate model performance and select the most suitable algorithm, multiple complementary metrics are employed: AUC, balanced accuracy, F1 score, and average precision. AUC served as the primary indicator of discriminative ability across the three degradation classes, capturing the trade-off between true and false positive rates. Balanced accuracy ensured fair assessment under class imbalance, while the F1 score reflected the trade-off between precision and recall. Average precision provides insight into the model's reliability in detecting rare fault cases. Model validation is conducted using stratified five-fold cross-validation, preserving class proportions within each fold to

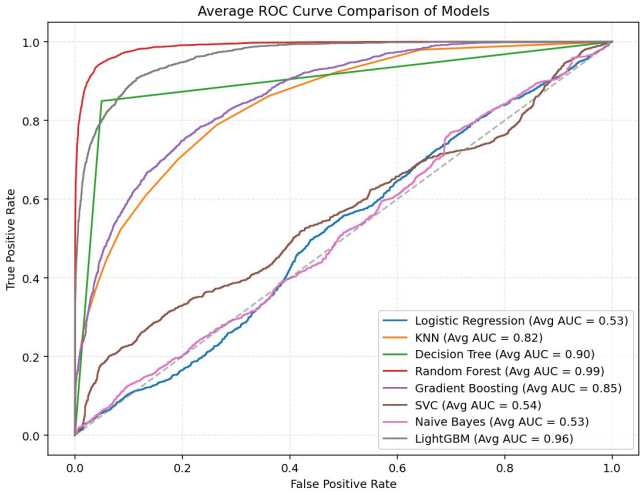

Figure 5. ROC curves of the evaluated classification models for multiclass degradation detection. Tree-based models achieve higher AUC values compared to linear and kernel-based approaches, with the Random Forest classifier exhibiting the steepest curve.

ensure comparability. Performance metrics exhibit high consistency across folds, with variation below 1% – underscoring the robustness of the evaluation procedure. Based on these outcomes, model comparison is further supported by ROC curve analysis using the same labeled dataset from the controlled degradation experiments. As shown in Figure 5, tree-based models outperform linear and kernel-based approaches due to two main characteristics of the system: nonlinear feature interactions and the presence of distinct operating regimes within the feature space. Linear models impose a single global decision boundary and therefore cannot adequately represent coupled thermal-hydraulic interactions. Kernel-based models capture nonlinearities through global transformations of the input space but still rely on a unified decision structure. In contrast, tree-based models recursively partition the feature space and learn localized, regime-specific decision rules, which better reflect the heterogeneous operating behavior of the heat pump system. Among them, the Random Forest classifier achieves the steepest ROC curve and highest AUC (~0.98), along with an F1 score of 0.95 – making it the final model selected for downstream reliability diagnostics. The selected Random Forest model consists of an ensemble of 200 decision trees trained on bootstrapped subsets of the training data. Final class predictions are obtained by aggregating the tree-level votes, which improves robustness against overfitting and supports reliable multiclass classification across heterogeneous operating conditions.

4.5. Interpretation and Field Relevance

To evaluate the model's suitability for field deployment, the precision–recall performance is analyzed for both degradation classes. These metrics are particularly relevant in maintenance applications, where high precision minimizes false alarms, and high recall enables early and reliable fault detection. This trade-off directly influences the effectiveness of condition-based service strategies. As shown in Figure 6, the Random Forest achieves an average precision (AP) of 0.96 for severe contamination, maintaining stable precision up to approximately 90 % recall. The slight drop beyond this point reflects the model's effort to capture borderline cases. For moderate contamination, the AP reaches 0.97 with similarly consistent behavior. Together, these results confirm the model's ability to detect faults accurately at early stages, supporting reliable condition-based maintenance in practical deployments. To further assess the model's utility for field deployment, cumulative gain and lift charts are evaluated. These metrics show how well the model ranks faulty systems, which is essential for prioritizing inspections under resource constraints.

As shown in Figure 7, inspecting the top 20 % of ranked instances captures around 80 % of severely contaminated systems. This is 4× improvement over random selection. A similar effect is observed for moderately contaminated cases. Most faults are concentrated in the top prediction band. These

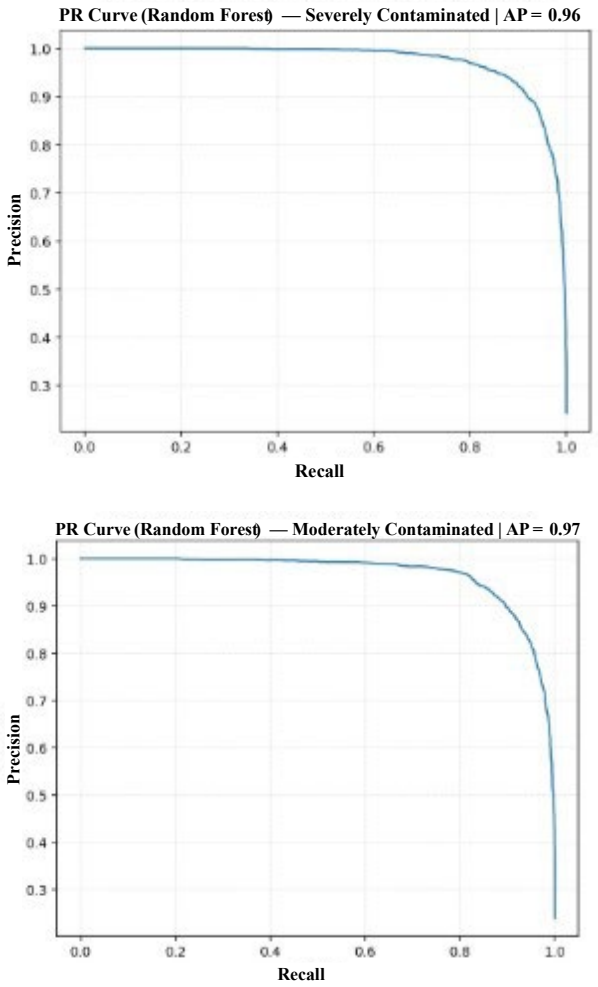

Figure 6. Precision–recall curves of the Random Forest classifier for moderate and severe contamination classes.

results confirm that the model supports targeted inspection and early intervention. This enables more efficient resource allocation and improves the effectiveness of condition-based maintenance.

To interpret the model's decision logic and assess its operational relevance, the feature importance is analyzed using the Gini impurity index from the Random Forest classifier. Feature importance is quantified based on the mean decrease in impurity, reflecting how strongly each variable contributes to reducing class uncertainty across the decision trees. Results show that refrigerant-side variables – particularly the PHE outlet temperature – provide the highest discriminative contribution for predicting degradation states within the controlled experimental setup. Water-side parameters such as flow rate and pressure also contribute to degradation detection, indicating that degradation effects are reflected across both hydraulic and thermodynamic system behavior. The observed relevance of refrigerant-side variables suggests that contamination-induced changes in heat-transfer behavior can provide complementary diagnostic

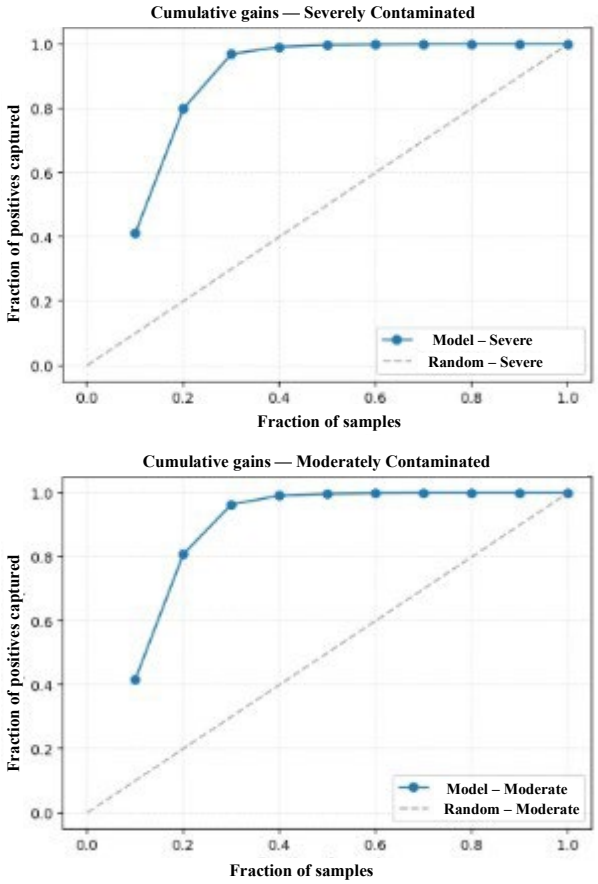

Figure 7. Cumulative gains curves of the Random Forest classifier for the severe and moderate contamination classes, showing the fraction of detected faults as a function of the inspected sample fraction.

information beyond conventional water-side measurements alone. These findings are validated through expert feedback: technicians note that current diagnostics tend to focus on water side measurements and welcome the identification of previously overlooked indicators with higher diagnostic value. To support practical deployment, association rule analysis is applied to the trained model, yielding simple if–then patterns that consistently signal degradation. These interpretable rules provide a transparent, low-complexity interface between advanced modeling outputs and existing monitoring infrastructure. This enables more reliable, data-driven fault detection in the field.

5. DISCUSSION AND FUTURE WORK

The results indicate that physically defined degradation mechanisms can be systematically translated into interpretable diagnostic models. This is achieved by embedding domain knowledge at the data generation stage through controlled and reproducible degradation scenarios, which establish clearly defined system states. These structured degradation regimes allow supervised learning to capture patterns that are directly linked to known physical mechanisms. This structural consistency is reinforced by the staged design of the approach – from degradation design to evaluation and interpretability – ensures traceability between physical cause, observed system response, and diagnostic output. In this way, the approach addresses a central challenge in condition monitoring: aligning predictive performance with engineering interpretability and domain validation.

The observed performance patterns suggest that the controlled degradation design generates structurally distinct and physically meaningful operating regimes. Although the degradation is introduced through the water-side flow path, its effect propagates through the coupled thermodynamic interaction between the water and refrigerant circuits inside the PHE. As a result, refrigerant-side variables become highly sensitive indicators of altered heat-transfer behavior and contribute strongly to degradation-state discrimination within the controlled experimental setup. The experimentally defined degradation states represent controlled reductions in active plate participation induced through the sleeve-based flow restriction described in Section 4.2, rather than direct percentages of real fouling coverage across the full PHE surface. As illustrated by the progressive shifts in compressor behavior, pump energy consumption, and heating-cycle characteristics shown in Figure 3, these laboratory-defined operating regimes were intentionally selected to generate physically separable and diagnostically interpretable system responses under reproducible laboratory conditions. This does not imply that hydraulic variables are irrelevant for degradation detection. Instead, the results indicate that refrigerant-side measurements provide complementary diagnostic information beyond conventional hydraulic indicators such as flow rate or pressure. This observation contrasts with conventional diagnostic practices, which typically emphasize water-side measurements alone. The results therefore highlight the diagnostic value of integrating refrigerant-side measurements for earlier and more reliable detection of contamination-related degradation due to their strong sensitivity to altered heat-transfer behavior. This interpretation is further supported by the stable precision–recall behavior, which indicates the presence of well-separated degradation states rather than marginally overlapping states. As a result, detection thresholds can be adjusted according to operational priorities without an abrupt loss of reliability. This flexibility is particularly relevant for balancing early fault identification against unnecessary service interventions. In practical deployments, intervention thresholds would likely be calibrated at earlier degradation stages according to operational and maintenance requirements. In addition, the pronounced lift performance indicates that degradation risk can be concentrated within a limited subset of systems, enabling prioritized inspection and structured maintenance planning across large populations of installed systems.

These findings have direct implications for the design of condition-based monitoring strategies in residential heat pump systems. The ability to detect early-stage fouling enables maintenance actions to be initiated before severe efficiency losses or system interruptions occur. Beyond individual fault detection, the integration of risk-based ranking supports structured maintenance planning across distributed installations where routine inspection is not feasible. The use of interpretable features and rule-based explanations further facilitates transparent fault communication and supports integration into existing service workflows. In contrast to conventional diagnostics that rely primarily on static thresholds or isolated water-side indicators, the demonstrated relevance of refrigerant-side variables suggests a shift toward behavior-based and physically informed monitoring concepts**.** This shift also has implications for monitoring system design, indicating that thermodynamic indicators from the refrigerant circuit should be considered alongside conventional hydraulic measurements when developing sensor strategies for heat pump diagnostics. At the same time, real field installations are expected to be dominated by healthy operating conditions. Therefore, the balanced dataset used in this study should be interpreted as a controlled basis for learning physically separable degradation patterns rather than as a direct representation of field class distributions. In practical deployment, these diagnostic patterns could support broader monitoring workflows in which non-normal operating behavior is first identified before detailed degradation classification and maintenance prioritization are applied.

Beyond the specific use case of the plate heat exchanger, the proposed approach demonstrates high potential for generalization across other degradation-prone components. Overall, the results demonstrate that physics-informed data generation under controlled operating conditions can support the development of interpretable supervised diagnostic models even in settings with limited real-world failure data. The generalizability of the approach lies less in transferring a specific trained model and more in transferring the structured approach used to derive physically interpretable diagnostic relationships from controlled degradation behavior. Core transferable elements include physically motivated degradation design, controlled generation of labeled operating regimes, integration of domain knowledge into data generation, and interpretable evaluation of sensor-response relationships. Key prerequisites include access to interpretable sensor signals, the ability to define degradation-relevant control parameters, and sufficient labeling through experimental or historical data. These conditions are particularly relevant in residential and industrial heating, ventilation, and air conditioning (HVAC) applications, where components like circulation pumps, filters, or expansion valves exhibit similarly observable degradation dynamics. In addition, the approach offers transferable value to other domains with physics-governed degradation processes. The use of physics-informed training data improves transferability by embedding real-world operational behavior into the learning process, allowing the resulting models to maintain diagnostic relevance even under varying field conditions.

While the approach shows promising results, it is important to recognize several underlying limitations. First, the degradation scenarios are simulated under controlled laboratory conditions. While this ensures interpretability and reproducibility, it may not fully capture the variability and measurement-related disturbances present in real-world systems, including sensor drift, installation-dependent flow imbalances, and irregular user-driven load profiles. Moreover, while certain degradation processes – such as contamination in the plate heat exchanger – can be deliberately induced and monitored until failure, others cannot be easily simulated or may not exhibit clear observable transitions. This limits the approach's direct applicability to components with accessible and controllable fault mechanisms. Second, the approach relies on domain knowledge to define degradation thresholds and relevant features, which introduces expert bias and limits automation. Third, although Random Forest proved effective in this context, the approach does not yet incorporate temporal dynamics or explicitly model gradual degradation trends, which may be essential for long-term reliability forecasting. Future work should therefore investigate data-driven approaches based on field data to capture time-dependent degradation behavior under real operating conditions. Such models could characterize degradation trajectories over time and support component-level reliability assessment, including risk evolution and potential remaining useful life (RUL) estimation. In addition, this data-driven perspective is particularly relevant for components whose degradation mechanisms cannot be reproducibly induced or clearly labeled in controlled laboratory experiments. Finally, beyond these modeling aspects, current interpretability relies on global importance measures and association rules, which may overlook instance-specific effects. Addressing these limitations through field deployment, integration of time-series models, local interpretability techniques, and investigation of embedded AI deployment constraints offers promising directions for future research.

6. CONCLUSION

This work establishes an approach that combines physically grounded degradation design with data-driven modeling to enable interpretable component-level diagnostics in residential heat pump systems. Applied to a plate heat exchanger as a representative use case, the approach integrates controlled degradation design, physics-informed data generation, and supervised machine learning to support condition-based monitoring. Each stage of the approach is designed to maintain alignment with physical system behavior, enabling the generation of structured, interpretable

data and the training of robust classifiers. The approach demonstrates that meaningful degradation patterns can be learned from operationally relevant signals and translated into actionable diagnostic insights. Feature importance analysis and association rules provide insights into underlying mechanisms and informed sensor placement and fault prioritization strategies. The approach's ability to support early detection, targeted inspection, and transparent decision-making makes it well-suited for real-world deployment. While developed for the PHE, the approach is generalizable to other critical components in residential or industrial systems. Its structured design, grounded in both engineering knowledge and data-driven modeling, offers a scalable pathway toward predictive reliability diagnostics across domains.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the valuable collaboration and support received throughout this research. Special thanks are extended to the Automation and Control Systems Group at Fraunhofer IIS and the engineering and data analytics teams of the Bosch Home Comfort Group for their contributions during the development and validation of the proposed approach. This work forms part of a doctoral research project conducted within a collaborative program between Bosch Home Comfort Group and the Technical University of Applied Sciences Cologne, with the shared goal of advancing reliability modeling in sustainable heating technologies.

REFERENCES

Alhuthali, S., Delaplace, G., Macchietto, S., & Bouvier, L. (2022). Whey protein fouling prediction in plate heat exchanger by combining dynamic modelling, dimensional analysis, and symbolic regression. Food and Bioproducts Processing, 134, 163–180. https://doi.org/10.1016/j.fbp.2022.05.009

Ardsomang, T., Hines, J. W., & Upadhyaya, B. R. (2013). Heat exchanger fouling and estimation of remaining useful life. In Proceedings of the Annual Conference of the Prognostics and Health Management Society 2013.

Berce, J., Zupančič, M., Može, M., & Golobič, I. (2021). A review of crystallization fouling in heat exchangers. *Processes*, 9(8), 1356. https://doi.org/10.3390/pr9081356

Brudermüller, T., Potthoff, U., & Fleisch, E. (2025). Estimation of energy efficiency of heat pumps in residential buildings using real operation data. Nature Communications, 16, 2834. https://doi.org/10.1038/s41467-025-58014-y

Hou, G., Zhang, X., Li, Y., & Wang, H. (2025). Application of machine learning algorithms in real-time health status monitoring of plate heat exchangers. Applied Thermal Engineering, Article S0735-1933(25)00234-9. https://doi.org/10.1016/j.applthermaleng.2025.120456

Ikonen, E., Liukkonen, M., Hansen, A. H., Edelborg, M., Kjos, O., Selek, I., & Kettunen, A. (2023). Fouling monitoring in a circulating fluidized bed boiler using direct and indirect model-based analytics. Fuel, 346, 128341. https://doi.org/10.1016/j.fuel.2023.128341

International Energy Agency (IEA). (2022). The future of heat pumps. Paris, France: Author. https://www.iea.org/reports/the-future-of-heat-pumps

Jnod Energy. (2025, August 28). *How does a plate heat exchanger work in a heat pump?* Retrieved January 26, 2026, from https://www.jnodenergy.com/how-does-a-plate-heat-exchanger-work-in-a-heat-pump/

Kapustenko, P., Klemeš, J. J., & Arsenyeva, O. (2023). Plate heat exchangers fouling mitigation effects in heating of water solutions: A review. *Renewable and Sustainable Energy Reviews*, 179, 113283. https://doi.org/10.1016/j.rser.2023.113283

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3, 422–440. https://doi.org/10.1038/s42254-021-00314-5

Kouidri, I., Dahmani, A., Furizal, F., Ma'arif, A., Mostfa, A. A., Amrane, A., Mouni, L., & Sharkawy, A.-N. (2025). Artificial intelligence-based techniques for fouling resistance estimation of shell and tube heat exchanger: Cascaded forward and recurrent models. *Eng*, 6(5), 85. https://doi.org/10.3390/eng6050085

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 104, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.024

Meng, C., Griesemer, S., Cao, D., Seo, S., Yan, L. (2025). When physics meets machine learning: A survey of physics-informed machine learning. Machine Learning in Computational Science and Engineering, 1, Article 20. https://doi.org/10.1007/s44379-025-00016-0

Munnangi, A., Mohamed Arshath, S., Karthikeyan, C., Muthamizhi, K., & Praveenkumar, V. (2026). Plate heat exchangers and their versatile applications — Artificial neural networks. In R. K. Arya, G. D. Verros, & J. P. Davim (Eds.), *Smart heat transfer and thermal management* (Woodhead Publishing Reviews: Mechanical Engineering Series, pp. 273–296). Woodhead Publishing. https://doi.org/10.1016/B978-0-443-33881-6.00014-9

Patil, P., Srinivasan, B., & Srinivasan, R. (2022). Monitoring fouling in heat exchangers under temperature control based on excess thermal and hydraulic loads. *Chemical Engineering Research and Design*, 181, 41–54. https://doi.org/10.1016/j.cherd.2022.02.032

Qarqour, A., Arora, S. J., Heisenberg, G., Rabe, M., & Kleinert, T. (2024). Utilizing data analysis for optimized determination of the current operational state of heating systems. In Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KDIR) (pp. 200–209). SciTePress. https://doi.org/10.5220/0012876200003838

Qarqour, A., Arora, S. J., Heisenberg, G., Rabe, M., & Kleinert, T. (2025). Towards systematic reliability assessment: A multi-criteria decision framework for modeling heat pump systems. In Proceedings of the Asia Pacific Conference of the Prognostics and Health Management Society 2025, 5(1). https://doi.org/10.36001/phmap.2025.v5i1.4463

Romanowicz, T., Taler, J., Jaremkiewicz, M., & Sobota, T. (2023). Determination of heat transfer correlations for fluids flowing through plate heat exchangers needed for online monitoring of district heat exchanger fouling. *Energies*, 16(17), 6264. https://doi.org/10.3390/en16176264

Sansana, J., Rendall, R., Castillo, I., de Bruijne, L., Huggins, J., Phillips, A., & Reis, M. S. (2024). Hybrid approach for advanced monitoring and forecasting of fouling with application to an ethylene oxide plant. *Industrial & Engineering Chemistry Research*, 63(24), 10666–10676. https://doi.org/10.1021/acs.iecr.4c00298

Soomro, A. W., Mat Kiah, M. L., Md Noor, R., Newaz Kazi, S., Shaikh, K., Khan, W. A., & Ali, I. (2026). Artificial intelligence in industrial heat exchanger fouling prediction: A 20-year systematic review of AI, ML, and DL approaches. *ICT Express*, 12(1), 92–110. https://doi.org/10.1016/j.icte.2025.12.003

Sundar, S., & Rajagopal, R. (2020). Fouling modeling and prediction approach for heat exchangers using deep learning. International Journal of Heat and Mass Transfer, 159, 120112. https://doi.org/10.1016/j.ijheatmasstransfer.2020.120112

Tovazhnyanskyy, L., Sherstyuk, V., Kapustenko, P., Khavin, G., Perevertaylenko, A., Boldyryev, S., & Garev, A. (2007). Plate heat exchangers for environmentally friendly heat pumps. Chemical Engineering Transactions, 12, 213–217. https://doi.org/10.3303/CET0712037

Virginia Heat Transfer. (2021, October 8). *Plate exchanger: Fouling vs scaling*. Retrieved January 26, 2026, from https://vaheat.com/?p=147

Wu, Y., Sicard, B., & Gadsden, S. A. (2024). Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. *Expert Systems with Applications*, 255(Part C), 124678. https://doi.org/10.1016/j.eswa.2024.124678