15TH IEA HEAT PUMP CONFERENCE
MAY 26–29, 2026 | HOFBURG
VIENNA 2026
DECARBONISATION THROUGH INNOVATION

Leveraging AI tools along the heat pump value chain from component design to operation supervision

N. Fazaa[a,b,c*], L. Heinen[a,b], H. Fugmann[b], L. Schnabel[b], N. Réhault[b], H. Madani[c,d], A. Velte-Schäfer[a,b]

[a]University Freiburg, Department of Sustainable Systems Engineering (INATECH), Emmy-Noether-Straße 2, 79110 Freiburg, Germany
[b]Fraunhofer Institute for Solar Energy Systems ISE, Heidenhofstraße 2, 79110 Freiburg, Germany
[c]KTH Royal Institute of Technology, Department of Energy Technology, Brinellvägen 68, 10044 Stockholm, Sweden
[d]Austrian Institute of Technology, Sustainable Thermal Energy Systems, Giefinggasse 4, 1210 Vienna, Austria

Abstract

Artificial Intelligence (AI) provides engineers with powerful tools. We showcase applications of AI tools along the heat pump value chain. The development of efficient heat pump circuits with minimized refrigerant charges requires complex trade-offs between component selection and overall system performance. By using Machine Learning (ML) models trained on a large dataset from a heat pump development project, a design optimization tool to assist engineers in identifying the most optimal component combinations is developed and evaluated. The tool's main goal is to recommend component combinations that achieve the highest possible efficiency (COP: Coefficient of Performance) with the lowest refrigerant charge for a given heat capacity and to reduce development times. On building energy systems (BES) design level, reliable annual simulations of variable speed heat pumps require a detailed performance map (> 100 operation points) of the heat pump. In early stages of heat pump and BES design, available component or equipment data is often limited. By applying ML models including deep learning tools to predict heat pump performance out of a minimal dataset, the number of required measurement points can be drastically reduced. We explore, evaluate, and show how to leverage these tools for an efficient BES design process. At the operational stage of the heat pump lifecycle, large amounts of sensor data are generated continuously by building automation systems. However, this data is often unstructured, noisy, and lacks explicit labels, posing significant challenges for direct ML application. We address this by developing a modular data-driven monitoring framework using task-specific ML models for fault detection, operational state classification, and performance evaluation. A central element is the integration of expert feedback to refine model outputs and validate anomalies. This human-in-the-loop strategy enhances model transparency, which is essential in the BES context, where explainability and operational relevance are critical for deployment.
© HPC2026.
Selection and/or peer-review under the responsibility of the organizers of the 15th IEA Heat Pump Conference 2026.

Keywords: Artificial Intelligence; Heat pump Design Optimization; Performance Maps; Fault Detection and Diagnosis;

1. Introduction

Artificial intelligence (AI) has become a powerful tool used in many areas. In this work we show how it can accelerate Heat Pump (HP) development, enabling data-driven design, optimization, and performance enhancement across the entire product lifecycle. Chapter 2 discusses the design phase, AI can be used to predict system performance, and recommend optimal component combinations that balance efficiency, cost, and refrigerant charge. Chapter 3 shows how during planning and integration, AI can create HP performance maps using fewer operational points, support load forecasting, system sizing, and smart control strategies. Chapter 4 shows AI use in the monitoring stage, AI-powered analytics process large volumes of operational data to detect anomalies, predict faults, and continuously improve system reliability and performance. Together, these

*Corresponding author. Tel.: +49 174 922 4891
E-mail address: nader.suhail.qadri.fazaa@ise.fraunhofer.de

applications accelerate innovation and enhances overall system efficiency. Figure 1 shows the complete product development process, which highlights the key phases where AI can provide valuable support.

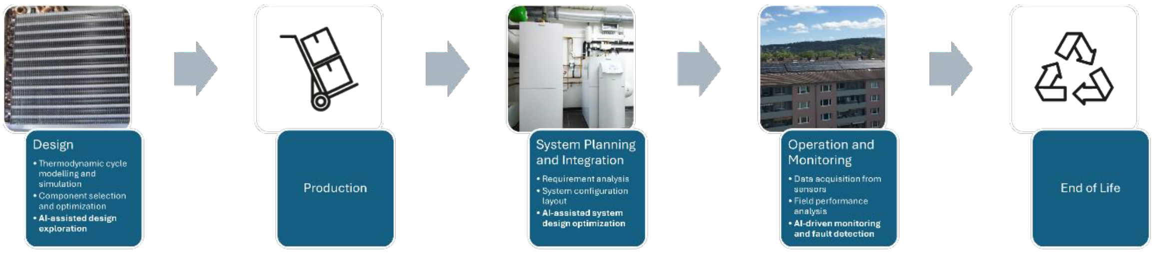

Figure 1: Heat pump development and integration workflow

2. AI in HP Design

The transition to low Global Warming Potential (GWP) refrigerants has positioned propane (R290) as a leading natural alternative for residential and commercial HPs. Propane combines excellent thermodynamic performance with a low GWP of 3. However, propane is highly flammable thus it is necessary to minimize the refrigerant charge to avoid additional safety measures. This directly affects achievable heating capacity and efficiency [1]. Therefore, optimizing component combinations to minimize refrigerant charge while maintaining high system performance has become a central design challenge for HPs.

2.1. Using ML in propane HP design

Designing efficient HP circuits with minimized refrigerant charge involves complex trade-offs between component selection and system performance. Conventional design workflows depend on iterative simulations and prototype testing, which are time-consuming and resource intensive. AI, particularly Machine Learning (ML), enables a faster and more systematic exploration of design possibilities by learning from experimental and simulated data.

In this work, a ML model was developed to estimate the heating capacity and Coefficient Of Performance (COP) of propane HPs based on component and operating parameters. The model supports engineers in evaluating design configurations quickly and identifying combinations that offer high efficiency under varying conditions. In a related study, a ML was trained to predict the total refrigerant charge, complementing the present work. The next step is to build a component database and train the model for automatic component selection, linking performance prediction with design optimization.

2.2. Data collection and pre-processing

To train the ML model described above, data were collected from a comprehensive propane HP measurement campaign that included 32 prototypes. Measurements were performed under various operating conditions, and only steady-state points were considered. Each record includes refrigerant-side parameters (e.g. pressures, temperatures) and system-level indicators (e.g. compressor speed, heating capacity).

With the help of domain experts, the following features were selected as the most relevant for model training: compressor type, compressor part-load ratio, superheating, refrigerant charge, brine inlet temperature, and water outlet temperature. Outliers and incomplete records were removed using interquartile range filtering and physical plausibility checks, for example, enforcing an energy balance within ±10 %. Feature selection was guided by thermodynamic understanding of the vapor-compression cycle. Highly correlated variables were reduced using Principal Component Analysis (PCA). The processed dataset had approximately 6800 measurements and was used for ML training and validation with a 75%-25% split.

2.3. Goal of the ML model and its selection

The gradient boosting model, implemented here as a multioutput regressor, demonstrated strong predictive accuracy for both COP and heating capacity, with an average R^2 score of 0.8721 and an RMSE of 0.6221. Figure 2 presents the parity plots comparing predicted and measured values, showing that most data points lie close to the ideal 1:1 line, confirming the model's ability to generalize across various prototypes and operating conditions. These results provide valuable engineering insights by linking physical understanding of HPs with data-driven analysis. The approach demonstrates how ML models can capture nonlinear dependencies and assist in evaluating design choices more efficiently than traditional simulation-based methods.

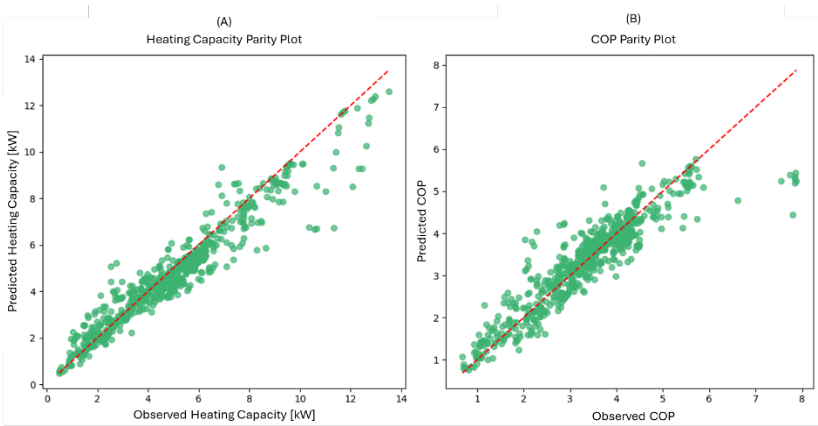

Figure 2: (A) Parity plot of predicted heating capacity (kW) vs. measured heating capacity in kW.
(B) Parity plot of predicted COP vs the measured COP

3. AI in HP Planning

The performance map of heat pumps is required to perform annual simulations in the BES design process. Different approaches were presented in literature to model the performance map to minimize the number of required operation points. Cheng et al. [2] present a polynomial model that requires at least 20 operation points. Belman-Flores et al. [3] use relatively small Artificial Neural Networks (ANN) to model the performance map of variable-speed HPs. Both the polynomial models and the small ANNs have the inherent drawback of being valid for exactly one HP – for each new variant a new specific set of coefficients needs to be determined, or the ANN needs to be retrained. This issue could potentially be tackled with Tabular Foundation Models (TFMs) as recently presented by Hollmann et al. [4] because TFMs are deep neural networks being pre-trained on a multitude of different general datasets already and require a relatively small amount of problem-specific training data.

3.1. HP performance maps and prediction quality indicators

The typical structure of an air-sourced HP performance map is composed of the following quantities: the supply temperature of the heating system $T_{wtr,out}$, the outdoor air temperature T_{air}, the Part-Load Ratio (PLR), the COP, and the heating capacity $\dot{Q}_h$. The first three quantities are the feature columns, the latter two quantities are to be predicted. For identification of each HP in a dataset with performance maps of different HPs the nominal heating capacity (HP 'Size') is added as a fourth feature.

PLR is defined as the ratio of the actual heating capacity $\dot{Q}_h$ to the maximum heating capacity under the same temperature conditions. COP is defined as ratio of heating capacity and electrical power consumption P_{el}. The Hoval Belaria HP family [5] is chosen for the evaluation, however, the approach can also be applied across HPs different manufacturers. For each of the four HP variants in this family a dataset of 126 operating points is available from the datasheet. The prediction quality is evaluated in terms of Root Mean Square Deviation (RMSD), Coefficient of Variance (CV), and the R^2 score.

3.2. Training strategy

The training data set is mainly composed of approximately 80 % of available datasheet points of the 15 kW HP variant. Additionally, for each of the 8 kW and the 24 kW HP variants, 12 datasheet points are part of the training data set which is less than 10 % of the available datasheet points of these variants. Only a single point of the 13 kW HP variant is part of the training data set. We use this training strategy to assess whether the TFM can learn from the characteristics of one HP performance map to predict the other three, using only a very small training dataset.

3.3. Results and Discussion

The results of the prediction of the heating capacity at maximum compressor speed with the TFM "TabPFN" are shown in Figure 3 along with the datasheet values for various supply temperatures at air temperature of 2 °C. The qualitative comparison in Figure 3 (a) shows that the TFM prediction matches the datasheet values of all variants with only small deviations. The quantitative comparison of the prediction quality indicators CV and R^2 score in Figure 3 (b) shows that the whole performance map of the 13 kW variant can be predicted with a CV less than 10 % and an R^2 score of > 0.95. Also in terms of COP the obtained CV value is less than 10 %. For having only a single point in the training data set this is a remarkable accurate prediction. Increasing the number of training data points leads to an even better prediction quality as the CV values and the R^2 values for the 8 kW variant and the 24 kW variant show.

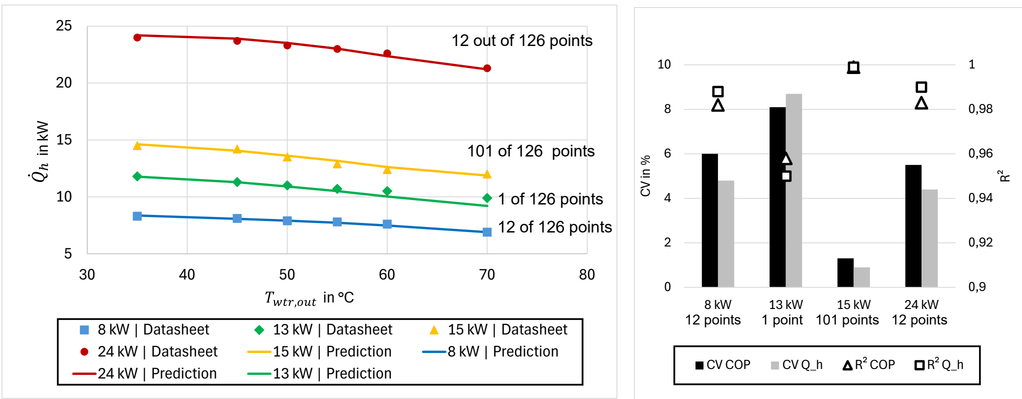

Figure 3 – (a) Comparison of predicted heating capacity (max. compressor speed) with datasheet values for the four HP variants at air temperature of 2 °C, (b) Influence of number of training points on prediction quality of heating capacity.

In literature the application of various ML approaches to HP performance maps is discussed. For instance, Chesser et al. [6] applied Random Forest (RF) and Multiple Linear Regression (MLR) to predict the performance map of a HP. Thus, we applied Gradient Boost, Random Forest, and Linear Regression to our HP performance maps. With the same training dataset, RMSD and CV values of all evaluated ML approaches are at least twice as high as shown in Figure 3 (b), which emphasizes the superior prediction quality of TabPFN.

4. AI in HP Operation

Buildings with HP systems generate an increasing volume of sensor data via their energy management infrastructure. This trend has accelerated over the past few decades, creating many opportunities for continuous system monitoring. Effective monitoring is essential to guarantee the HP's optimal performance and detect anomalies at an early stage [7]. Traditionally, fault detection relies on manually defined rule sets tailored to each installation, a process that is time-consuming and error prone. By contrast, data-driven approaches offer greater adaptability and reduced setup effort, as they can learn directly from the operational data stream. Despite its potential, the straightforward application of ML techniques is hindered by several challenges. Sensor data is usually unstructured and noisy and labelled training data is scarce, because failures are either undetected or eliminated. Furthermore, the residential sector shows a wide variety of HP configurations, which limits the transferability of trained models. Any automated detection framework must finally allow for human intervention and interpretability, since identified faults require verification and corrective action by qualified personnel [8].

4.1. Proposed Fault Detection and Diagnosis Framework

To address these challenges, we propose a decentralized, human-in-the-loop fault detection and diagnosis framework. An expert oversees the automated alarms generated by modular ML-modules deployed on highly specialized subsystems. Whenever a module raises an alert or requests feedback, the expert analyses the case and provides a label, thereby contributing to the continuous enrichment of the training data. This feedback loop scales fault detection across large portfolios and yields high-quality annotations that adapt the modules to the specific characteristics of each system. Thanks to the modular design, components developed for one installation can be partially transferred to others with overlapping characteristics. Each ML-module implements the 'COMETH' [9] principle through a pair of complementary binary classifiers calibrated for different sensitivity levels. During normal operation, the two classifiers compare their outputs; any discrepancy triggers a feedback request, prompting expert intervention and label collection. Over time, this mechanism reduces the frequency of requests while the ML methods adapt with more training data. We use a decision tree classifier [10] to improve interpretability and a DBSCAN-based anomaly detector [11] to provide high sensitivity.

4.2. Sensor Error Detection Study

We evaluated an important aspect of the proposed framework using a dataset collected over two months at five-minute intervals from a residential air-to-water HP energy management system. Our study focused exclusively on temperature readings from the heat pumps internal sensors. One ML-module was trained using the condenser and evaporator return and supply temperatures ($T_{cond,sup}$, $T_{cond,ret}$, $T_{evap,sup}$, $T_{evap,ret}$) as well as the incoming air temperature ($T_{air,sup}$) and the water temperature supplied to the heating system ($T_{wat,sup}$). Three synthetic fault events were introduced by resetting the $T_{cond,ret}$ and $T_{evap,sup}$ sensors to zero at predetermined times. All other measurements were treated as nominal, providing a controlled environment in which to assess model performance and feedback evolution during application (see Figure 4).

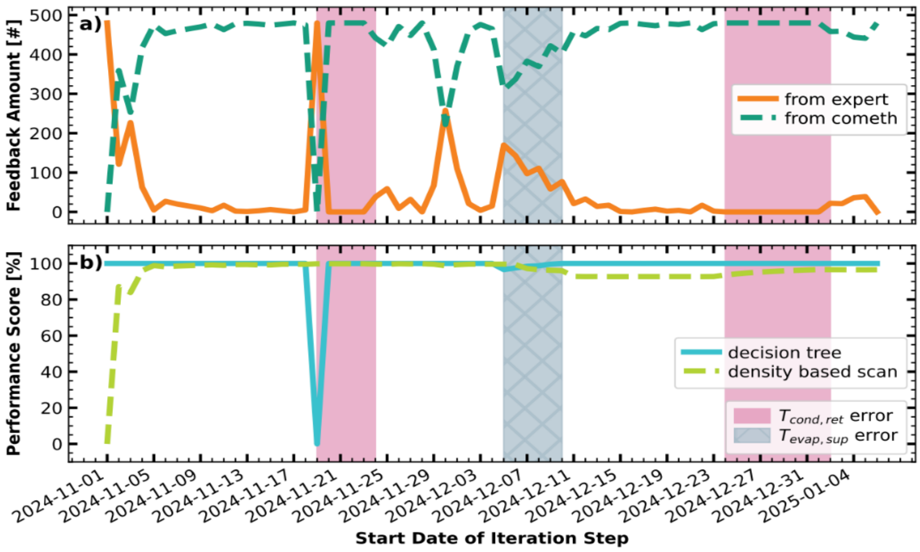

Figure 4: Iterative application of a single ML module on the prepared dataset. Part a) shows the amount of timesteps per iteration step that were labelled automatically by the cometh module or manually by an expert. Part b) shows the F_1-score performance of the individual ML methods during the application. The locations of the synthetic faults in the dataset are highlighted with vertical boxes.

Initially, both classifiers operate without prior training, providing default predictions. The models are retrained iteratively using one-day data batches to incorporate newly acquired data labels. As the framework learns the initial nominal regimes, the volume of feedback requests declines markedly, but whenever a new faulty or nominal operating mode occurs the amount of feedback requests increases. The expert labels from the answers to these requests enable the models to learn from the data and generalize to future anomalies. A second instance of the $T_{evap,sup}$ error does not require an increased amount of feedback as the models recognize its symptoms from the previous occurrence. By the end of the evaluation period, the module has adapted to the behavior of the system, showing that this approach effectively mitigates the data quality and labeling challenges. Ongoing work aims to extend this framework by incorporating alternative learning algorithms, evaluating cross-system transferability, and pretraining modules on simulation data to proactively address common fault scenarios.

5. Conclusion

In conclusion, implementing AI across the design, planning, and operational stages of HP development can help engineers to explore a wider range of configurations and operating strategies potentially at lower costs and shorter development times. While this study focuses on a qualitative evaluation, it demonstrates practical pathways to integrate AI into the HP value chain and highlights its potential to accelerate design iterations, reduce the need for extensive measurements, and enhance system performance by maximizing the use of available data. To achieve real value, different AI and ML approaches must be tailored to each phase of the HP lifecycle, from component design to system optimization and monitoring. Nevertheless, several challenges remain, most notably the limited availability of high-quality data and the lack of standardized data structures across components, complete HP systems, and building-level integrations. Continued efforts in this field are essential to streamline AI implementation and fully unlock its potential in the HP industry.

Acknowledgements

We gratefully acknowledge financial support of the German Federal Ministry for Economic Affairs and Energy (BMWE) due to an enactment of the German Bundestag under grant number 03EN4046 (LCR290).

References

[1] Timo Methler, Hannes Fugmann, Clemens Dankwerth, Christian Sonner, Katharina Morawietz, Lena Schnabel. Performance and safety analysis of charge reduced brine to water heat pumps using R290. In: 14th IEA Heat Pump Conference, May 15-18 2023, Chicago, US “Heat Pumps – Resilient and Efficient” Conference Proceedings - Full Papers: Performance and safety analysis of charge reduced brine to water heat pumps using R290. Heat Pump Centre 2023.

[2] Cheng L, Braun JE, Horton WT. A methodology for mapping the performance of variable-speed residential cooling equipment using load-based testing. International Journal of Refrigeration 2021; 132: 133–44 [https://doi.org/10.1016/j.ijrefrig.2021.09.001]

[3] Belman-Flores JM, Ledesma SE, Garcia MG, Ruiz J, Rodríguez-Muñoz JL. Analysis of a variable speed vapor compression system using artificial neural networks. Expert Systems with Applications 2013; 40(11): 4362–9 [https://doi.org/10.1016/j.eswa.2013.01.015]

[4] Hollmann N, Müller S, Purucker L*, et al.* Accurate predictions on small data with a tabular foundation model. Nature 2025; 637(8045): 319–26 [https://doi.org/10.1038/s41586-024-08328-6][PMID: 39780007]

[5] Hoval Deutschland. Technologisch überlegene Raumklima-Lösungen; 2026 [cited 2026 January 12] Available from: URL: https://www.hoval.de/de_DE/.

[6] Chesser M, Lyons P, O'Reilly P, Carroll P. Air source heat pump in-situ performance. Energy and Buildings 2021; 251: 111365 [https://doi.org/10.1016/j.enbuild.2021.111365]

[7] Barandier P, Marques Cardoso AJ. A review of fault diagnostics in heat pumps systems. Applied Thermal Engineering 2023; 228: 120454 [https://doi.org/10.1016/j.applthermaleng.2023.120454]

[8] Leite D, Andrade E, Rativa D, Maciel AMA. Fault Detection and Diagnosis in Industry 4.0: A Review on Challenges and Opportunities. Sensors (Basel) 2024; 25(1) [https://doi.org/10.3390/s25010060][PMID: 39796851]

[9] Benndorf GA, Wystrcil D, Réhault N. A fault detection system based on two complementary methods and continuous updates. IFAC-PapersOnLine 2018; 51(24): 353–8 [https://doi.org/10.1016/j.ifacol.2018.09.601]

[10] Loh W-Y. Classification and regression trees. WIREs Data Min & Knowl 2011; 1(1): 14–23 [https://doi.org/10.1002/widm.8]

[11] Ester M, Kriegel H-P, Sander J, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise; 1996.