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

Data-Driven Techniques for the Prediction of Refrigerant Charge in Heat Pumps through Machine Learning Soft Sensors

N. Fazaa[a,b,c*], A. Schygulla[d], E. Pak[d], E. Schmid[d], F. Faller[d], H. Fugmann[b], H. Madani[c,e], M. Lämmle[d]

[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]Hochschule Offenburg, Faculty of Mechanical and Process Engineering, Badstraße 24, 77652 Offenburg, Germany
[e]Austrian Institute of Technology, Sustainable Thermal Energy Systems, Giefinggasse 4, 1210 Vienna, Austria

Abstract

The transition to natural refrigerants, such as propane (R290), presents new challenges for heat pump development, particularly in optimizing system efficiency while minimizing refrigerant charge. Within the LC150 project, over 6800 steady state operating points were experimentally recorded from 32 heat pump prototypes, with a high-resolution variation of the refrigerant charge. Each data point reflects a half hour average under stabilized conditions, offering a robust dataset for data-driven analysis. This study investigates the use of machine learning (ML) to develop a soft sensor capable of predicting the refrigerant charge based solely on typical operating variables. Seven ML models (k Nearest-Neighbours, Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting, and Multi-Layer Perceptrons (MLPs), a class of feedforward artificial neural networks) were trained and evaluated. A focus was on feature importance and model accuracy, especially for a limited number of input features. When using only five input features, including pressures, temperatures, and valve positions, the highest achieved coefficient of determination (R^2) was 0.93, with Gradient Boosting. When expanded to ten features, the accuracy improved to an R^2 of 0.98 with XGBoost. Notably, the variables identified as most influential by the models correspond to those recognized as important by domain experts, demonstrating a strong alignment between data driven and physics-based understanding. This research was conducted as part of a semester project involving four undergraduate students from Hochschule Offenburg in the "Applied Artificial Intelligence" program in collaboration with Fraunhofer ISE. The project highlights how interdisciplinary collaboration, between AI specialists and heat pump engineers, can accelerate innovation in the development of efficient, charge minimized heat pump systems using natural refrigerants. By leveraging ML for refrigerant charge estimation, this work opens the door to real time diagnostics, enhanced fault detection, and smarter system design, supporting the wider adoption of sustainable heating technologies.

Selection and/or peer-review under the responsibility of the organizers of the 15th IEA Heat Pump Conference 2026.

Keywords: Artificial Intelligence; Propane heat pump; Charge prediction; Soft sensors.

1. Introduction

Artificial intelligence (AI) is increasingly used in engineering applications. For Heat Pumps (HPs), data driven methods allow faster design decisions, improved control strategies, and more reliable performance assessment [1]. Refrigerant charge has a strong influence on system efficiency and safety, especially for propane (R290) HPs that operate with strict charge limits [2]. Reliable charge prediction is therefore important for both laboratory development and field operation. This work investigates how Machine Learning (ML) can be used to estimate refrigerant charge from standard HP measurements.

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

1.1. Motivation

Refrigerant charge affects superheat, subcooling, compressor behaviour, and overall capacity. Manual charging procedures and traditional estimation methods are often slow, uncertain, or dependent on technician experience. Charge limits for R290 increase the need for accurate prediction tools that work across variable operating conditions. A data driven charge estimator can support testing, commissioning, and continuous monitoring without adding new hardware sensors [3].

1.2. State of the art

Existing approaches rely on physical models, refrigerant distribution correlations, or expert-based interpretation of pressure and temperature signals. These methods require detailed knowledge of system volumes and do not always generalize across prototypes or aging conditions [4]. ML can capture nonlinear relations between operational variables, and this study compares several regression models on a unified dataset and examines the influence of feature dimensionality on model accuracy and robustness.

2. Methodology

The data used in this project were generated through a measurement campaign on 32 Refrigeration Circuits (RCs) prototypes. Two datasets were used: one describes each prototype's components, including the condenser, evaporator, compressor and oil names, and pipe diameters and lengths. all components were anonymized using neutral labels (A, B, C, etc.) to ensure confidentiality and consistency across prototypes. The other is a measurement dataset containing time series data recorded during the measurement campaign, such as temperatures, pressures, superheat values and the corresponding refrigerant charges.

2.1. Data Processing and preparation

Preprocessing the measurement data is a necessary step to ensure that all variables are accurate, consistent, and suitable for subsequent modelling. Before any cleaning is performed, the dataset is examined to understand its structure and the behaviour of individual features. Correlations between variables are visualized to identify linear and nonlinear relationships. This analysis supports two goals: selecting appropriate preprocessing actions and informing later decisions on feature selection and model design.

The raw measurements contain occasional errors caused by sensor issues or transient test bench behaviour. Rows with clear measurement faults were addressed through two approaches. In cases where the variable remained close to its expected trend, the faulty points were corrected using a simple linear regression model. Figure 1(A) illustrates this scenario for the refrigerant charge measurement, where some measured data points lie near zero or directly at zero, while the remaining points follow the ideal linear relation. A linear regression model was trained on the correct subset of data, and the learned relationship was used to predict the corrected values for the faulty measurements. Figure 1(B) shows the adjusted dataset, which aligns consistently with the expected trend after correction. When sensor values deviated too strongly from any plausible physical behaviour, the corresponding rows were removed as outliers. The resulting dataset contained around 6800 measurement points. Model development used cross validation with an 80 percent training and 20 percent test split, where the test set was formed by pulling out complete prototypes form the data set.

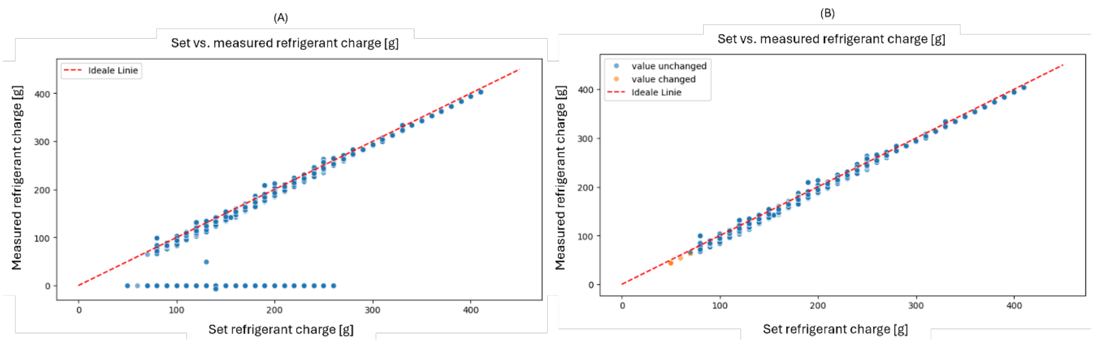

Figure 1:Correction of faulty refrigerant charge measurements. (A) original data (B) corrected data through linear regression

Following correction and filtering, categorical descriptors such as component types or prototype identifiers were transformed using ordinal encoding. This step assigns a consistent numerical ranking to categorical variables, which is required for regression-based machine learning models that cannot operate on raw categorical values. This preprocessing approach ensures that the final dataset is reliable, numerically consistent, and suitable for training and evaluating the machine learning models presented in Section 2.2.

2.2. ML models training

To estimate the refrigerant charge, several supervised regression models were implemented and evaluated. Since the dataset is largely numerical, regression-based methods are the most suitable choice. The models were assessed using the coefficient of determination (R^2) and Root Mean Square Error (RMSE).

The following models were selected to cover a wide range of learning strategies:

- Decision Trees are a hierarchical model that applies a sequence of simple if then rules to divide the feature space into regions with similar target values. This structure allows the model to approximate complex relationships in the data without relying on linear assumptions [5].
- Random Forests extend the Decision Tree approach by training an ensemble of trees using different subsets of the data and features. The final prediction is obtained by averaging across all trees, which reduces variance and mitigates overfitting while improving accuracy [5].
- Support Vector Regression (SVR) applies the core principles of Support Vector Machines (SVM) to regression tasks. By defining a hyperplane in a multidimensional space and using kernel functions, SVR can model both linear and nonlinear relationships with good generalization capability [6].
- Multi-Layer Perceptron (MLP) is a feedforward Neural Network (NN) trained in a supervised manner. Input feature vectors are propagated through one or more hidden layers to produce an output value. Through backpropagation, the network adjusts its weights to learn a nonlinear mapping between inputs and target charge values [7].
- K-Nearest Neighbour Regression is a Machine Learning Model that calculates the distance between points and gives them a prediction based on their local interpolation associated of their nearest neighbours [8].
- Gradient Boosting is an ensemble method that builds multiple weak learners in sequence. Each learner focuses on correcting the residuals of the previous model. This iterative structure allows the ensemble to capture complex dependencies while controlling bias and variance [9].
- XGBoost is an optimized implementation of Gradient Boosting that incorporates regularization, efficient tree construction, and parallelized boosting. These enhancements improve stability, accuracy, and computational performance for large and structured datasets [10].

All models were trained using cross validation, with hyperparameters tuned through structured grid search. This ensured a fair and consistent comparison across methods and supported the selection of the final model used for feature reduction and downstream analysis. Figure 2 shows the process to train the ML models and select the optimal model for this application.

To train the ML model described above, data were collected from a comprehensive 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).

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), and the resulting dataset was used for ML training and validation.

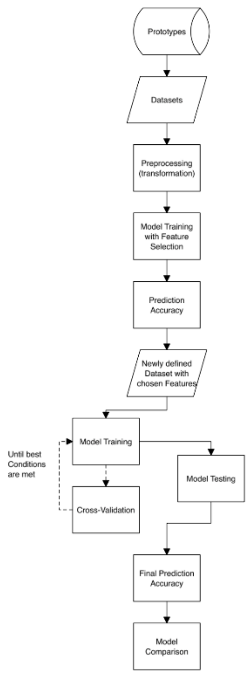

Figure 2: workflow of model training and selection

3. Results

This chapter presents the outcome of the feature selection analysis and the performance comparison of the trained machine learning models. The goal is to understand how feature dimensionality affects charge estimation accuracy and to identify the model that provides the most reliable predictions. Results are shown for both low dimensional and high dimensional feature sets, which were defined through a combination of statistical selection procedures and domain expert assessment.

3.1. Feature selection

Feature selection was conducted to quantify how the number of features influences the regression performance. Figure 3 shows the R^2 score as a function of the number of selected features. The curve increases rapidly at the beginning and then levels off, which indicates that only a limited number of features is necessary to achieve accurate charge prediction. Beyond this point, adding more features increases complexity without meaningful accuracy gains.

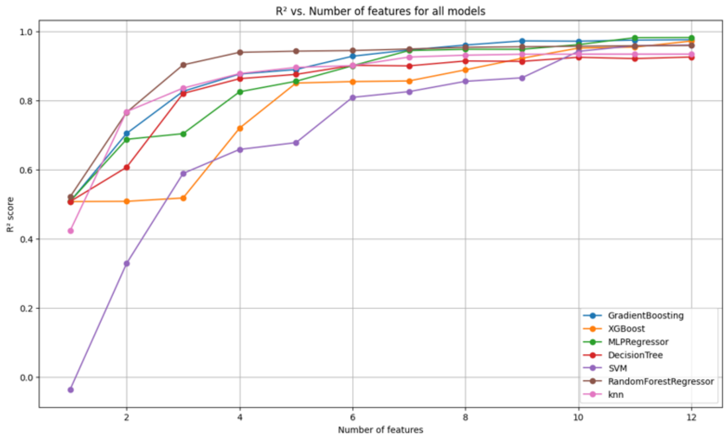

Figure 3: Effect of number of features on the R^2 score for the different ML models

PCA, Recursive Feature Elimination (RFE), and feature importance rankings were used to identify the most relevant inputs. RFE was applied to iteratively remove the least informative features by training a model, ranking all inputs by their contribution, and discarding the weakest ones until the optimal subset remained. Figure 4 shows the RFE results for the Random Forest model. The reduced feature set was then reviewed together with domain experts and only features that appeared consistently in the REF process across multiple ML models and could be obtained in the field with minimal effort were selected for the low dimensional and high dimensional feature sets.

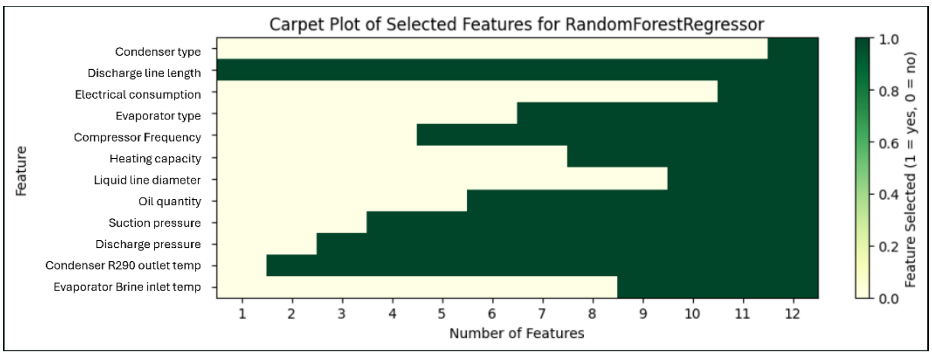

Figure 4: RFE results for Random Forest

Table 1 lists the final features chosen for each set. These variables have direct physical relations to mass flow and enthalpy distribution and are therefore expected to have a strong influence on the target value.

Table 1: Selected features for low and high dimensional sets

Low dimensional set	High dimensional set		
Suction pressure	Suction pressure		Superheat
Discharge pressure	Discharge pressure		Heating capacity
Condenser R290 outlet temp.	Condenser R290 outlet temp.	+	Compressor frequency
Evaporator brine inlet temp.	Evaporator Brine inlet temp.		EEV position
Oil quantity	Oil quantity		Discharge line length

3.2. Model comparison

The models were trained using both the low dimensional and high dimensional feature sets. Figure 5 shows the R² scores and RMSE values for all models, where the left side of the figure corresponds to the low dimensional features and the right side corresponds to the high dimensional ones. For the low dimensional set, the best performing model was Gradient Boosting, followed closely by XGBoost and Random Forest. Their R² scores were 0.93, 0.92, and 0.91, with RMSE values of 15.5 g, 16.8 g, and 17.9 g. For the high dimensional set, XGBoost achieved the highest accuracy with an R² score of 0.98 and an RMSE of 9.18 g. Gradient Boosting and the MLP followed with R² scores of 0.97 and RMSE values of 9.43 g and 9.54 g, respectively.

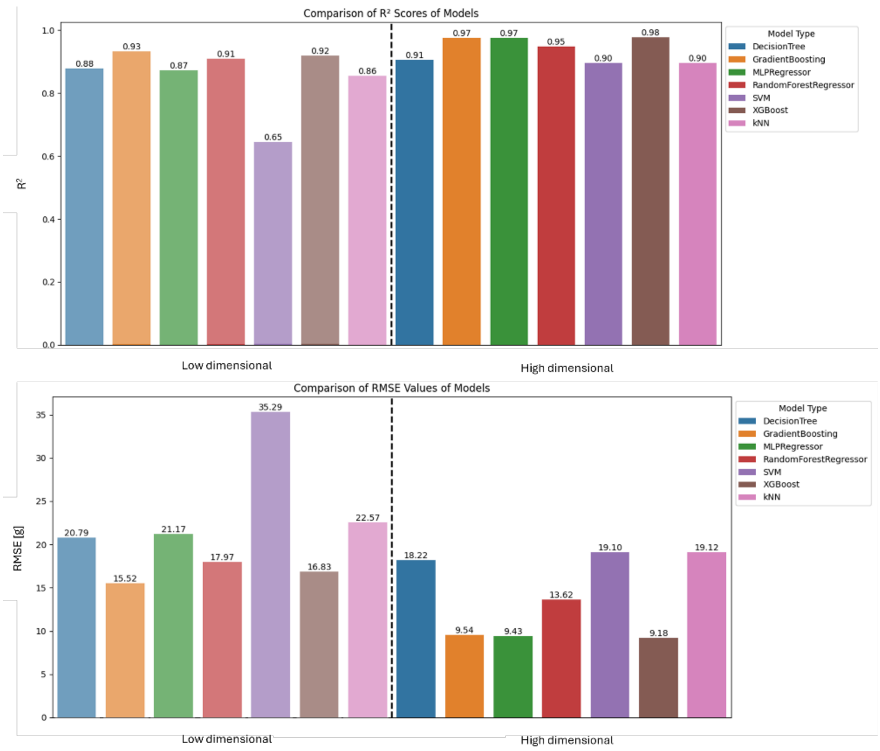

Figure 5: R^2 and RMSE values achieved by the models with low and high dimensional feature sets

As shown in Figure 5, the high dimensional feature set provides higher accuracy. Figure 6 illustrates this more clearly for XGBoost by presenting an overlapping parity plot of the predicted and measured charge values. The faded points represent the low dimensional feature set and form a wider cloud around the ideal parity line, while the filled points represent the high dimensional set and show a much tighter alignment with the ideal relation. This comparison highlights that the high dimensional input improves prediction quality, yet the low dimensional set still offers a stable pattern that may be preferable in situations where data availability, sensor limitations, or long-term robustness are more critical than maximum achievable accuracy.

4. Conclusion

This study shows that machine learning can estimate refrigerant charge with high accuracy using standard heat pump measurements. After comprehensive preprocessing and feature selection, several regression models were evaluated across both low dimensional and high dimensional feature sets. The results confirm that a limited number of physically meaningful variables are sufficient for reliable charge prediction. XGBoost provides the best performance with an R^2 value of 0.98 and an RMSE of 9.18 g, and it shows the most stable residual behaviour. These findings indicate that XGBoost is a strong basis for a soft sensor for propane charge estimation, suitable for system testing, commissioning, and continuous monitoring without additional hardware sensors.

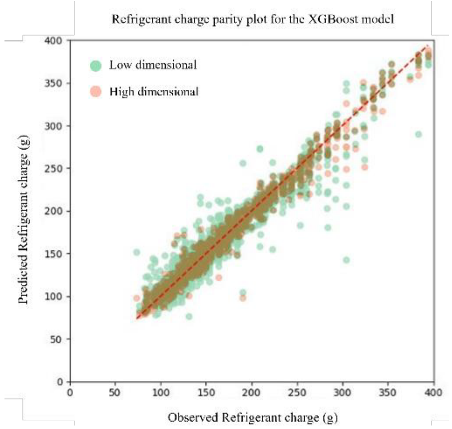

Figure 6: Overlapping parity plot for XGBoost trained with low and high dimensional feature sets

Using AI to predict charge from operational conditions enables real time monitoring, leakage detection, and precise filling during maintenance. Working jointly with AI specialists and heat pump experts also clarifies which features reflect the charge status most strongly, such as superheat and the EEV opening. This approach also helps distinguish features that are physically meaningful from those that only appear important statistically, ensuring that the selected variables remain valid across different prototypes and aging conditions. Despite the promising results, deploying these models in real field environments still faces important challenges. The most significant are the limited availability of stable and well calibrated sensor data in the field and the need for models that remain robust under real operating disturbances including frosting, cycling behaviour, and long-term component degradation. Hence, Practical deployment should include model confidence indicators and automated plausibility checks to flag operating conditions that are outside the training domain. This would prevent overconfident predictions during transients or sensor faults and support safe decisions during commissioning and service. Continued work in this area is required to streamline implementation and fully unlock the potential of AI for the heat pump 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] Shin J-H, Cho Y-H. Machine-Learning-Based Coefficient of Performance Prediction Model for Heat Pump Systems. Applied Sciences 2022; 12(1): 362 [https://doi.org/10.3390/app12010362]
[2] VEIC. Propane Air-to-Water Heat Pump Market Study. Vermont, USA: Vermont Energy Investment Corporation; 2022.
[3] Song Y, Rolando D, Marchante Avellaneda J, Zucker G, Madani H. Data-driven soft sensors targeting heat pump systems. Energy Conversion and Management 2023; 279: 116769 [https://doi.org/10.1016/j.enconman.2023.116769]
[4] Zhang C, Hou B, Li M, *et al.* Refrigerant charge estimation method based on data-physic hybrid-driven model for the fault diagnosis of transcritical CO2 heat pump system. Energy 2024; 309: 133144 [https://doi.org/10.1016/j.energy.2024.133144]
[5] Breiman L. Random forests. Machine Learning 2001; 45(1): 5–32.
[6] Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing 2004; 14(3): 199–222.
[7] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press 2016.
[8] Methods in Biomedical Informatics. Elsevier; 2014.
[9] Friedman JH. Greedy function approximation (a gradient boosting machine). Annals of Statistics 2001; 29(5): 1189–232.
[10] Chen T, Guestrin C. XGBoost (a scalable tree boosting system). In: XGBoost (a scalable tree boosting system); 2016; 785–94.