Automatic Differentiation of Variable and Fixed Speed Heat Pumps With Smart Meter Data

Tobias Brudermueller
ETH Zurich

Florian Wirth
University of St. Gallen

Andreas Weigert
University of Bamberg

Thorsten Staake
University of Bamberg; ETH Zurich

***Abstract*—With the increasing prevalence of heat pumps in private households, the need for optimization is growing. At the same time, the growing number of active smart electricity meters generates data that can be used for remote monitoring. In this paper, we focus on the automatic differentiation between fixed speed and variable speed heat pumps using smart meter data. This distinction is relevant because it is necessary for evaluating the state or cycling behavior of a heat pump. In addition, identifying fixed speed heat pumps is important because they are known to be the less efficient systems and therefore may be preferred targets in energy efficiency or replacement campaigns. Our methods are applied to electricity data from 171 Swiss households with a resolution of 15 minutes. In this setting, a K-Nearest Neighbor model achieves a mean AUC of 0.976 compared to 0.5 of a biased random guess model.**

***Index Terms*—smart meter data, heat pump, machine learning, variable speed, fixed speed, modulation, residential, inverter**

I. Introduction

By the end of 2020, the European Heat Pump Association (EHPA) estimated a total of 14.86 million installed heat pumps in 21 European countries, which is an increase of 7.4% compared to the previous year [1]. The European Union (EU) plans on electrically heating 40% of all residential and 65% of all commercial buildings by 2030 [2]. Hence, increasing heat pumps distribution is part of the EU's carbon mitigation plan. Other regions of the world follow the same trend [3]–[6]. Thus, heat pumps will play a major role in heat and cooling decarbonization [7], [8].

However, many market-available and already installed heat pumps are not connected to the internet. Cases of connected heat pumps are often only pilot projects [9]. There is an "adherence to proprietary solutions" because "from the perspective of some suppliers, provider-specific solutions offer a higher potential for customer retention than services based on interoperable platforms" [10]. Even if the digitalization of heat pumps leads to the growth of the heat pump market in the future [11], the lack of connectivity and IOT-readiness of already installed heat pumps will remain for the next decades. On the other hand, the number of deployed smart meters is constantly increasing. In 2019, the European Commission expected that by 2020 almost 43% of European consumers would have a smart meter installed [12]. At this point in time, already seven European Union states reached a roll-out rate of 80% and some started with the second-generation roll-out [12].

In our work, we want to make use of both trends and target towards exploiting smart meter data to monitor heat pumps. We especially hope to identify systems with high energy saving potential. Several works already demonstrated that the existence of heat pump installations can be predicted with smart meter data [13]–[15] but there is a lack of work in categorizing heat pumps by their modulation capability [16]. However, this information is a valuable insight for remote monitoring services and energy saving campaigns. As shown in [17], using pre-selection criteria to identify candidates for a heat pump inspection campaign can lead to an average energy saving effect of 15.2%. Therefore, we want to build on previous work and take it one step further. Assuming the existence of a heat pump installation to be known, we want to distinguish variable speed and fixed speed heat pumps from smart meter data only and analyze influential factors for the classification results. The differences between variable and fixed speed heat pumps and the reasons why this distinction is relevant are briefly described in the following. More details can also be found in [18].

Variable speed heat pumps (also called inverter driven heat pumps) react to changes in heat demand by adjusting the rotation speed of the compressor. Accordingly, the electric power consumed depends on the compressor speed [19]. Variable speed heat pumps are frequently also called (capacity) modulated heat pumps. On the other hand, **fixed speed heat pumps** can only run the compressor on one fixed speed and thus, on a single electric power level. Accordingly, the system reacts to changes in heat demand by switching on and off only. Fixed speed heat pumps are often also referred to as fixed output, single speed or on-off heat pumps. The differentiation between variable speed and fixed speed heat pumps and their correct identification is highly relevant in the context of remote monitoring for a few reasons:

1) The existence of a fixed speed heat pump indicates that a potentially old heating system or less-efficient one is present which might need special monitoring [18]–[24]. Variable speed heat pumps are known to be more efficient than fixed speed heat pumps because they reduce the so-called "cyclic losses" [24]. Hence, they perform less on-off-cycles. This is beneficial because a significant amount of energy is lost during the start-up transient of a heat pump (i.e., switching on and off), which impacts the overall performance [24].

2) Closely connected to the previous point is that different types of heat pumps lead to different cyclic behavior. This also causes different observable patterns in the smart meter data as it will be explained in Section III-D.
3) The information about the modulation capability of a heat pump is preliminary information for follow-up analyses and is necessary to derive correct evaluations of a heat pump's state and cyclic behavior. For example, the annual operating hours and the number of start-up transients can serve as evaluation parameter for fixed speed heat pumps, but not necessarily for variable speed heat pumps.
4) Lastly, a differentiation between modulating and non-modulating heat pumps is also commercially attractive. Fixed speed heat pumps will most likely be the first to be replaced in the future due to older age and lower average efficiency, which is why their owners might be interested in replacement offers. In addition, variable speed heat pumps have a potential for cross-selling. The advantages of modulating heat pumps over on-off systems are especially significant when there is a high need for modulation, for example, in the context of a variable electricity supply [19]. Therefore, a variable speed heat pump owner could be offered extensions to the existing system. This could be either a digital service for optimal control [25]–[28] or the integration of photovoltaic systems [29].

In this paper, we use real-world smart meter data from 171 Swiss households to differentiate between variable speed and fixed speed heat pumps. A K-nearest neighbor model can correctly classify the heat pump type with an AUC of 0.976 on a single week of data. Consequently, we can derive the information in an automated manner even if the heat pumps are not connected to the internet. The remainder of the paper is structured as follows: First, we present related work, followed by a description of the data set and the process of generating ground truth data. Then, we explain the feature engineering process, the evaluation scheme and corresponding results.

II. Related Work

A recent study [30] provides an overview of the research field of smart meter data analytics and categorizes existing work into sub-groups. Here, especially "load profiling" as reviewed in [31]–[33] is closely linked to our work. It "refers to the classification of load curves or consumers according to electricity consumption behaviors" [30]. Hence, the link to the present study is that a classification task is applied to smart meter data. However, it mostly focuses on categorizing user types in an unsupervised learning setting.

Other studies focus on identifying household characteristics from smart meter data formulated as classification task. For example, the work of [15] uses smart meter data to "recognize 19 household classes related to 11 household characteristics (e.g., electric heating, size of dwelling) with an accuracy of up to 95%". Similar work can be found in [34]–[37]. Likewise, the area of "non-intrusive load monitoring" aims at "decomposing a mains electricity measurement into each of its constituent individual appliances" [38]. Different techniques in this respect are reviewed in [39]–[42]. However, it needs to be distinguished between work that targets towards the identification of existent appliances and work that goes one step further by trying to decompose the corresponding patterns.

For the latter, a few studies with focus on heating systems can be highlighted. In study [43], a Bayesian model is used to disaggregate the electrical heating component in an unsupervised manner. The algorithm is evaluated on 676 households with an observation period of at least 6 months and 30-minute resolution. The authors report that the mean (over all the households) relative RMSE error is 16.6%. Kouzelis et al. [44] provide a clustering-based approach to separate flexible from non-flexible consumers. The goal is to estimate the residential heat pump consumption in a probabilistic way. However, the proposed method does not further differentiate between different types of flexible loads. Similarly, the work presented in [45] can disaggregate electrical load into space and domestic hot water heating. It covers residences with an electrical resistance heating system and smart meter data with 5-minute resolution.

While the above-mentioned studies focus on heating systems in general, a few studies aim to detect heat pumps from smart meter data. In [13], smart meter data with 15-minute resolution of 397 households is enriched with weather data and geospatial data to classify if a heat pump is present or not. Additionally, the presented algorithms are used to predict a heat pump's age ($<$10, $\geq$10, $\geq$20 years) and thermal reservoir (ground or air source). The authors report AUC-scores between 0.73 and 0.86. Similarly, in [14], the daily electricity consumption data of households is used to identify heat pump installations. Depending on the type of features being used, the authors report an F1-score between 0.785 and 0.864. Lastly, one study [17] uses electricity consumption data to pre-select heat pumps with a large saving potential for energy efficiency campaigns. The authors estimate that using pre-selection criteria can lead to an average saving effect of 15.2% per inspected heat pump. In short, there is existent work which covers the detection and pre-selection of heat pumps from smart meter data. Nonetheless, to the best of the authors' knowledge, there is a lack of work to distinguish heat pump installations by their modulation capability.

III. Methods

A. The Data Set

In this paper, we use smart meter data from 171 different households in Switzerland. The time series measurements have a resolution of 15 minutes with each value representing the electrical energy consumption in kWh in the last quarter of an hour. The data covers eight years [from 2012 to 2020] with on average 4.6 years of data for each household. Most of the data (77.8%) comes from single family houses. For the rest (22.2%), the building type is either a multi-family house or unknown. All households have one smart meter that measures only the heat pump's electricity consumption and one smart meter that measures any other appliance. Hence, we obtain the values of an aggregated smart meter by adding the values of the two smart meters at every timestamp (Figure 1). In Section IV, we evaluate our methods on both (the separate heat pump smart meter and the aggregated one) to analyze potential differences in performance. Therefore, we also test whether our approach works in cases where only a single smart meter is installed.

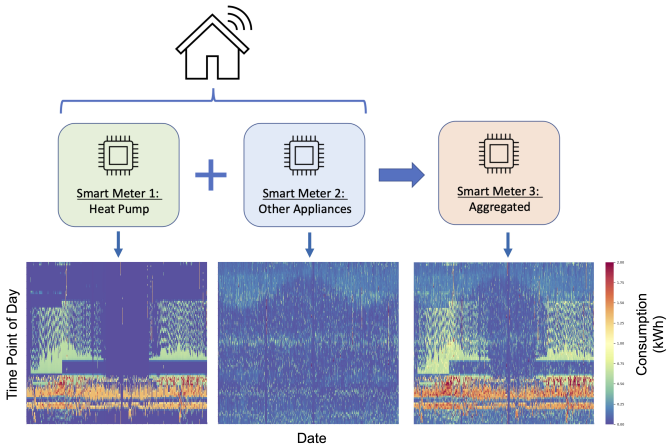

Fig. 1. Overview of smart meters. One smart meter measures only the heat pump's consumption, while the other measures all other appliances. When added, they form a third type of smart meter – an aggregated one. The heat maps cover a single year and are shown with a fixed scale for valid comparisons. Therefore, the color encoding corresponds to the electricity consumption in kWh (blue refers to 0.0 kWh and red to 2.0 kWh).

B. Generating Ground Truth Data

To generate ground truth data for classifying if the smart meter data of each household corresponds to a variable speed or fixed speed heat pump, two energy consultants labeled it. The targeted label is binary: *fixed vs. variable speed heat pump*. Both energy consultants are experts in the field. They have 10+ years of experience in on-site heat pump consultations and optimization. The data was labeled with a labeling tool, which showed the heat pump's smart meter data (without other appliances) of 2019 of each household. The labels were derived from visualizations of the data as heat maps and histograms.

The energy consultants labeled the images separately in individual sessions. Initially, we asked them to label data of 226 households. However, in 14 cases (6.2%), one of the experts marked to be unsure and therefore did not provide a label. There were mainly two reasons for this: In some cases, there was not enough data available for the evaluation. In other observations, the times during which the heat pump was switched on were too short for a proper labeling (e.g., only 30 minutes on-times, i.e., just two readings in a row). All these cases were excluded from our analyses. For the remaining 212 cases, the Fleiss' Kappa measure [46] of 0.52 to evaluate the inter-coder reliability can be interpreted as a moderate agreement according to [47]. There was disagreement between the labelers in 41 of 212 cases (19.3%). In Section III-D, we describe why there can be cases that are difficult to judge. To ensure a high quality of labels, these cases of disagreement were dropped. Thus, only the remaining 171 households are included in the following analyses. For these, the energy consultants chose the same label independently of each other. The final distribution of the labels is as follows: 130 variable speed heat pumps (75.9%) and 41 fixed speed heat pumps (24.1%). We are aware that both energy consultants could have labeled some observations wrongly. However, this is a natural problem when working with real-world data, as a recent study shows that 4-6% of the observations in manually labeled, well-known benchmarking data sets in the machine learning community are mislabeled [48].

C. Final Data Set Creation

Machine learning approaches need enough training data to generalize, which is why we want to increase the number of observations. Further, we want to account for differences in patterns each year and potential influence of weather. Therefore, we apply the following approach. A single household can have multiple years of smart meter data. For each household, we analyze the smart meter data multiple times in different windows. The window sizes can cover a single year, a single month, or a single week. Especially, analyzing a single week is of interest to us because it is privacy preserving and only a small amount of data is needed. We treat the time series of a household in one window as one separate observation. For example, when the window size covers a single year, we consider a household with two years of data as two separate observations. For the weekly and monthly windows, we only examine the months October to February because this covers the main heating period. Additionally, we only use time series with more than 85% completeness. In total this makes 764 yearly, 4'435 monthly and 16'498 weekly observations across all households. We assume that the heat pump type of each household did not change over the years. Therefore, we copy the binary label of each household to each observation referring to the same household.

D. Feature Engineering

In the following, we want to describe the behavioral differences between the two types of heat pumps in features that we extract from the time series in each window. For the monthly and weekly window size, we use 52 features. For the yearly window size, we use 77 features because we additionally consider more statistical features that refer only to the winter period. Before explaining the derived features, we first want to outline the underlying assumptions.

For a variable speed heat pump, we assume that the different compressor speeds lead to a wider spread in the distribution of electrical energy consumption values than for a fixed speed heat pump. We suppose that for a well-planned modulating heat pump the electricity consumption follows a normal distribution centering around the middle of the possible range. This would correspond to the compressor not constantly running on maximum speed and would mean that the heat pump could react to both a lower and higher heat demand. On the other hand, we assume that the electricity consumption of on-off heat pumps is almost constant, and that the distribution has a clearly dominant peak.

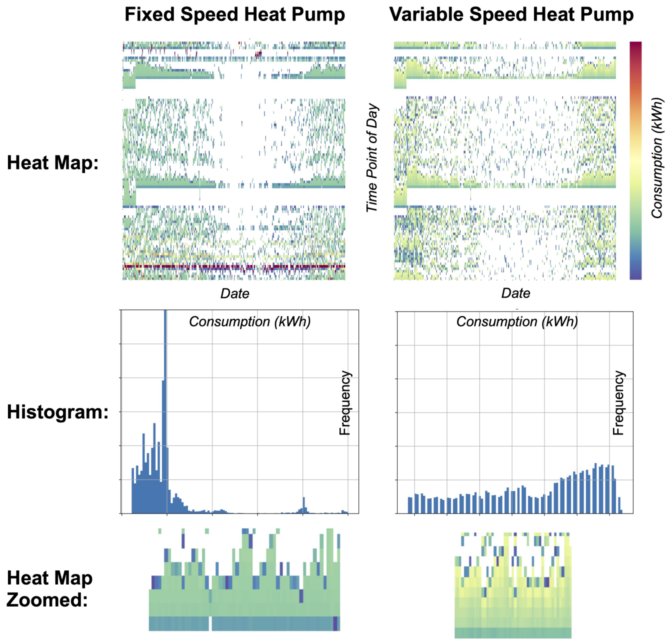

Fig. 2. Visualizations of the smart meter data of two heat pumps - one labeled as *fixed speed* and the other as *variable speed*.

A heat pump can switch on or off at any time within a 15-minute measurement interval. The measured kWh-value then corresponds to the fraction of 15 minutes the device is switched on. Hence, the measurement values in this case can be lower than other measurements although the heat pump might have a constant electricity consumption. In the following, we refer to these as "fractional readings". We suspect that the time points of on and off operations spread equally across a 15-minute time interval and that it causes an equal distribution at the lower left bound of a corresponding histogram. Figure 2 shows visualizations of smart meter data labeled as fixed speed and variable speed heat pump. We observe differences in the shapes of the histograms and the color gradients of the heat maps. For the variable speed heat pump, we see a higher variety in colors with steady transitions referring to the heat pump modulation.

There are edge-cases where the distinction is difficult. The electricity consumption and related pattern are closely linked to the installed size of a heat pump and the load it needs to supply. For example, an undersized variable speed heat pump would need to constantly run on its maximum level. In return, a highly oversized variable speed heat pump would frequently run on its lowest power level. In both cases the variable speed heat pump would almost behave like a fixed speed heat pump with its classical on-off cycles. At least, wrongly predicting a fixed speed heat pump in this case would still identify a system that needs special monitoring or has a high energy saving potential.

TABLE I
DESCRIPTION OF FEATURES (HISTOGRAM BIN SIZE 100).

FEATURE CATEGORY	DESCRIPTION
s_[...]	STATISTICAL FEATURES: MAXIMUM, MINIMUM, MEDIAN, STANDARD DEVIATION, MODE, KURTOSIS, SKEW, VARIANCE. FOR YEARLY FEATURES ADDITIONALLY CALCULATED WITH ONLY CONSIDERING DAYS IN WINTER AND DAY TIME VS. NIGHT TIME.
r_[...]	RATIOS OF CONSUMPTION: MEAN TO MAXIMUM AND MEAN TO MINIMUM.
t_num_above_[...]	NUMBER OF VALUES, WHERE THE CONSUMPTION EXCEEDS OR EQUALS A GIVEN THRESHOLD (0.125 KWH, 0.25 KWH, 0.5 KWH, MEAN, MIN, MAX).
h_sum_norm_diff_>[...]	SUM OF ABSOLUTE DIFFERENCES OF CONSECUTIVE KWH-VALUES. ONLY CONSIDERING VALUES WHERE THE ABSOLUTE DIFFERENCE IS GREATER THAN A THRESHOLD (0.1, 0.2, 0.3, 0.4, 0.5). (NORMALIZED BY THE NUMBER OF ALL ABSOLUTE DIFFERENCES.)
h_share_diff_>[...]	SHARE OF SUM OF ABSOLUTE DIFFERENCES OF CONSECUTIVE KWH-VALUES. ONLY CONSIDERING VALUES WHERE THE ABSOLUTE DIFFERENCE IS GREATER THAN A THRESHOLD (0.1, 0.2, 0.3, 0.4, 0.5). (NORMALIZED BY THE NUMBER OF ALL ABSOLUTE DIFFERENCES.)
h_share_at_max_plus[...]_bin	SHARE OF VALUES THAT FALL INTO THE LARGEST HISTOGRAM BIN OR NEIGHBORING BINS (LEFT AND RIGHT).
h_slope_[...]	ABSOLUTE DIFFERENCES OF LARGEST BIN AND AVERAGE OF ITS NEIGHBORING BINS (LEFT AND RIGHT).
h_num_bins_>[...]%	NUMBER OF BINS THAT EXCEED A GIVEN THRESHOLD (1-7) IN TERMS OF SHARE OF VALUES THEY CONTAIN.
h_centered_window_[...]	STATISTICAL FEATURES IN A WINDOW OF +- STANDARD DEVIATION CENTERED AROUND THE DISTRIBUTION MODE: STANDARD DEVIATION, SKEW, KURTOSIS, SHARE OF VALUES FALLING INTO LEFT AND RIGHT SIDE OF MODE.
h_mean_readings_per_cycle	MEAN NUMBER OF NON-ZERO-READINGS IN A ROW (I.E., AVERAGE ON-TIME).
h_unique_vals_rel_to_max	NUMBER OF UNIQUE CONSUMPTION VALUES DIVIDED BY THE MAXIMUM CONSUMPTION. (CONSUMPTION ROUNDED TO TWO DECIMALS.)

We provide a complete descriptive list of feature categories in Table I and and want to present a few exemplarily. For variable speed heat pumps, we expect to observe more non-zero absolute differences of consecutive kWh-readings due to the modulation. Additionally, we expect variable speed heat pumps to have longer cycles and fewer on-off transients. Therefore, we derive the features *h_sum_norm_diff_[. . .]* and *h_share_diff_[. . .]* which can be seen as a description of the distribution of absolute differences of consecutive measurements. The feature *h_mean_readings_per_cycle* denotes the mean number of consecutive non-zero measurements. Hence, it accounts for the average cycle-length. We assume that the "peakier" the distribution of measurements, the more likely is the observation to represent a fixed speed heat pump. Consequently, we assume that in this case a few histogram bins in immediate adjacency cover most measurements. Therefore, the features *h_share_at_[. . .]*, *h_slope_[. . .]* and *h_unique_vals_rel_to_max* describe how high the share of values is that fall into the largest bin (histogram mode) or the directly neighboring bins. Similarly, the *h_num_bins_[. . .]*-

features count the number of bins that contain a certain share of measurement values.

E. Classification Procedure

We phrase the problem as a binary classification with the classes *fixed speed heat pump (1)* and *variable speed heat pump (0)*. Then, we use the derived features to train and test the following machine learning-based classification models: K-Nearest Neighbor (KNN), Random Forest, Decision Tree, Logistic Regression, Naïve Bayes, and Support Vector Machine. For testing, we perform 5-fold cross validation – hence, we use five different random states for the evaluation. Each classification model is evaluated by averaging the test results across all random states. We perform the following steps for each random state:

1) **Data split and normalization:** First, we split the data into 80% training data and 20% test data. Then, we normalize the data by applying standard scaling to the training and test data (removing mean plus scaling to unit variance).
2) **Grid search:** For each classifier we perform a grid search on the training data to find the optimal hyper-parameters for each model. It builds on a 5-fold cross validation and the choice of parameters is evaluated in terms of the AUC-score.
3) **Training and testing:** With the chosen hyper-parameters, each classifier is trained on the whole training data and applied to the test data.

IV. Results

In the following, we report the mean-scores (and standard deviations) of the previously described models and of a biased random guess model, which constantly predicts the label of the most frequent class (*variable speed heat pump*). Therefore, the biased random guess model accounts for class imbalances and forms the strongest baseline possible. We evaluate each possible combination of smart meter type and window size (Table II). For the window size of a single year, we show the performance scores of all models. The K-Nearest Neighbor model performs best, followed by the random forest. For these two models, we also report the results when using a window size of a single month and single week. Figure 3 additionally shows the mean ROC-curves of all models for aggregated smart meters and a window size of a single week vs. single year. All models perform better than a biased random guess. However, the Naïve Bayes model performs worst. The KNN model constantly performs best and forms a robust estimator for all settings since it constantly achieves an AUC above 0.95.

A. Recursive Feature Elimination

In the following, we test the robustness of our methods by evaluating if the performance is maintained when using less features. This would be beneficial in terms of scalability and cost of computation. Here, we want to focus on the most relevant use-case in practice: using single weeks of data from aggregated smart meters. Therefore, we apply the *Recursive*

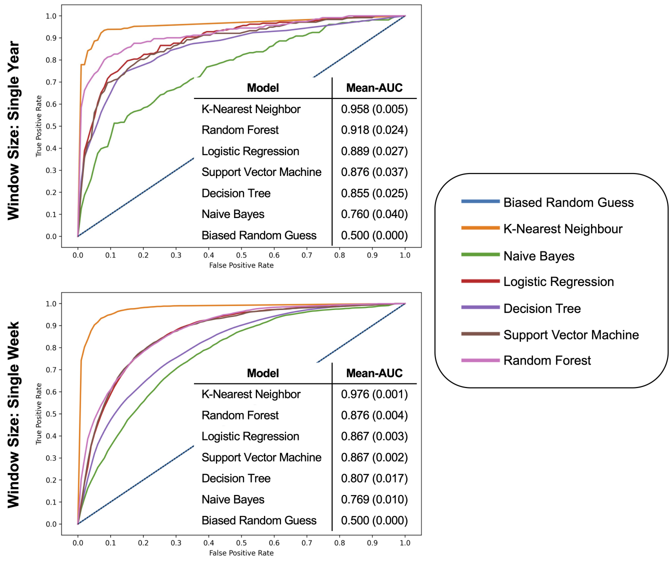

Fig. 3. Overview of the mean-ROC curves of each model when using aggregated smart meters and a window size of a single week vs. a single year.

Feature Elimination (RFE) algorithm to this setting, where iteratively the least important features are dismissed while training the classifier [49]. We use an adapted version as implemented in [50] that "includes a cross-validation loop to find the optimal number of features" [49]. The implementation provided in [50] cannot be applied to the KNN model. However, the RFE is typically used in combination with random forests. In our case we also choose the random forest model because other than the KNN it delivers solid results but also has more room for improvement (AUC of 0.876). Hence, we can evaluate if the random forest trained with less features even outperforms the one trained with more features. We apply the RFE algorithm with a 5-fold cross validation to the random forest with the best hyper-parameters found in the grid search (see process described in Section III-E). The RFE yields that 7 features (13.4%) can be removed. Now we re-train and test the random forest only on the remaining 45 features with a 5-fold cross validation. We use the same data set splits as before in the previous evaluation (Section III-E). Table III shows that our methods are robust and stable. With less features, the random forest performs almost equally well (AUC of 0.874 vs. 0.876) as the same model that was trained on all features.

V. Discussion

Up to this point, we can summarize that for the given problem the K-Nearest Neighbor model outperforms all other classification models. With this model, we achieve good and stable performance scores for all settings. In the most relevant use-case in practice (single week of data from an aggregated smart meter), the KNN achieves a mean AUC of 0.976. When choosing a random forest model instead, the performance decreases by approximately 10% (mean AUC of 0.876). However, also here the results are stable. The performance remains almost the same (mean AUC of 0.874) when decreasing the number of features with a recursive

TABLE II
TABLE OF RESULTS. FOR THE WINDOW SIZE OF A SINGLE YEAR, WE REPORT THE PERFORMANCE SCORES OF ALL MODELS, WHILE FOR A SINGLE MONTH AND WEEK WE ONLY SHOW THE TWO BEST MODELS. THE HIGHEST PERFORMANCE IN EACH CATEGORY IS MARKED IN BOLD.

WINDOW SIZE	SMART METER TYPE	SEPARATE (HEAT PUMP ONLY)					AGGREGATED (HEAT PUMP & OTHER APPLIANCES)				
	SCORE	AUC	ACCURACY	PRECISION	RECALL	F1-SCORE	AUC	ACCURACY	PRECISION	RECALL	F1-SCORE
SINGLE YEAR	K-NEAREST NEIGHBOR	**0.965** (0.004)	**0.928** (0.027)	**0.932** (0.026)	**0.898** (0.03)	**0.912** (0.028)	**0.958** (0.005)	**0.915** (0.023)	**0.906** (0.025)	**0.888** (0.027)	**0.896** (0.024)
	RANDOM FOREST	0.962 (0.011)	0.859 (0.025)	0.894 (0.027)	0.772 (0.018)	0.804 (0.021)	0.918 (0.024)	0.851 (0.036)	0.900 (0.023)	0.756 (0.039)	0.788 (0.044)
	LOGISTIC REGRESSION	0.940 (0.006)	0.886 (0.021)	0.869 (0.013)	0.855 (0.022)	0.861 (0.018)	0.889 (0.027)	0.847 (0.036)	0.835 (0.029)	0.788 (0.042)	0.803 (0.039)
	SUPPORT VECTOR MACHINE	0.930 (0.013)	0.876 (0.022)	0.862 (0.019)	0.834 (0.029)	0.845 (0.024)	0.876 (0.037)	0.844 (0.041)	0.832 (0.039)	0.781 (0.044)	0.799 (0.044)
	DECISION TREE	0.866 (0.029)	0.843 (0.027)	0.813 (0.033)	0.807 (0.034)	0.808 (0.032)	0.855 (0.025)	0.827 (0.028)	0.798 (0.036)	0.779 (0.049)	0.784 (0.045)
	NAÏVE BAYES	0.856 (0.032)	0.582 (0.028)	0.675 (0.037)	0.682 (0.019)	0.579 (0.032)	0.760 (0.040)	0.510 (0.027)	0.627 (0.030)	0.623 (0.035)	0.507 (0.029)
	BIASED RANDOM GUESS	0.500 (0.000)	0.702 (0.048)	0.351 (0.024)	0.500 (0.000)	0.412 (0.016)	0.500 (0.000)	0.702 (0.048)	0.351 (0.024)	0.500 (0.000)	0.412 (0.016)
SINGLE MONTH	K-NEAREST NEIGHBOR	**0.977** (0.004)	**0.945** (0.007)	**0.936** (0.009)	**0.916** (0.010)	**0.925** (0.009)	**0.966** (0.003)	**0.919** (0.009)	**0.901** (0.013)	**0.884** (0.009)	**0.892** (0.010)
	RANDOM FOREST	0.929 (0.009)	0.846 (0.014)	0.851 (0.023)	0.724 (0.017)	0.758 (0.020)	0.879 (0.007)	0.798 (0.010)	0.811 (0.023)	0.622 (0.013)	0.638 (0.018)
	BIASED RANDOM GUESS	0.500 (0.000)	0.745 (0.007)	0.372 (0.004)	0.500 (0.000)	0.427 (0.002)	0.500 (0.000)	0.745 (0.007)	0.372 (0.004)	0.500 (0.000)	0.427 (0.002)
SINGLE WEEK	K-NEAREST NEIGHBOR	**0.988** (0.002)	**0.966** (0.005)	**0.958** (0.006)	**0.949** (0.007)	**0.953** (0.006)	**0.976** (0.001)	**0.935** (0.003)	**0.918** (0.003)	**0.907** (0.004)	**0.912** (0.002)
	RANDOM FOREST	0.925 (0.007)	0.847 (0.009)	0.863 (0.013)	0.713 (0.011)	0.749 (0.012)	0.876 (0.004)	0.796 (0.011)	0.831 (0.006)	0.602 (0.016)	0.611 (0.025)
	BIASED RANDOM GUESS	0.500 (0.000)	0.751 (0.010)	0.375 (0.005)	0.500 (0.000)	0.429 (0.003)	0.500 (0.000)	0.751 (0.010)	0.375 (0.005)	0.500 (0.000)	0.429 (0.003)

TABLE III
RANDOM FOREST WITH AND WITHOUT RECURSIVE FEATURE ELIMINATION. WINDOW SIZE: SINGLE WEEK; SMART METER TYPE: AGGREGATED.

SCORE	AUC	ACCURACY	PRECISION	RECALL	F1-SCORE
RF (WITHOUT RFE)	**0.876** (0.004)	**0.796** (0.011)	**0.831** (0.006)	**0.602** (0.016)	**0.611** (0.025)
RF (WITH RFE)	0.874 (0.006)	0.794 (0.013)	0.822 (0.010)	0.599 (0.022)	0.606 (0.035)
BIASED RANDOM GUESS	0.500 (0.000)	0.751 (0.01)	0.375 (0.005)	0.500 (0.000)	0.429 (0.003)

feature elimination algorithm by approximately 13.4%. When we evaluate the influence of the smart meter type, we can observe the following: Using a separate smart meter can increase the performance up to 10%. Only the KNN is not affected by the difference in smart meter type because as mentioned earlier it performs well for all settings. Similar behavior occurs when evaluating the influence of the window size. Apart from the KNN, the behavior is as expected: the more data considered, the better. However, the difference is not substantial (in the range of 1-5%). We can conclude that a single week of data is enough to achieve a good performance in distinguishing variable speed from fixed speed heat pumps. Surprisingly, for the KNN model, the results for single weeks of data are even better than for single years. Our findings are relevant for three reasons: First, using a small amount of data is in accordance with the principle of data reduction and data economy of the European data protection law. Second, our algorithms apply to households where a smart meter was installed recently. Third, when the heat pump does not need to be measured by a separate smart meter, it reduces costs and also covers households with a single smart meter.

A. Limitations and Future Work

A limitation of the work presented in this paper is that we gained the ground truth data in a preliminary labeling step (as described in Section III-B). Hence, the heat pump labels were derived from the smart meter data. Therefore, future work should cover a verification with additional meta data about the heat pump installations. Additionally, we want to evaluate our methods with data from more households for different geographical conditions and by considering local weather. Further experiments can extract less features and smaller window sizes to explore the limits and improve scalability.

VI. CONCLUSION

We demonstrate that real-world smart meter data with 15-minute resolution can be used to distinguish variable speed heat pumps from fixed speed heat pumps. We further show that our results are robust to different settings and that only a small number of features need to be computed. A single week of data is sufficient for a K-Nearest Neighbor model to achieve a mean AUC of 0.976. The differentiation of heat pumps' modulation capability is of a high practical use because fixed speed heat pumps are known to be less efficient systems [18]–[24]. Hence, our work can help to identify potentially inefficient heat pumps. The differentiation is important for heat pump replacement or energy efficiency campaigns, and to correctly evaluate a heat pump's cyclic behavior. Future work can build on our findings and extract the type of heat pump as preliminary information for methods in predictive maintenance.

REFERENCES

[1] E. European Heat Pump Association, "European heat pump market and statistics report 2021," accessed: 2022 May 24. [Online]. Available: https://www.ehpa.org/market-data/market-report-2021/

[2] E. Commission, "Communication from the commission to the European parliament, the council, the European economic and social committee and the committe of the regions - powering a climate-neutral economy: An EU strategy for energy system integration," accessed: 2022 May 24. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2020:299:FIN

[3] M. Lapsa, G. Khowailed, K. Sikes, and V. Baxter, "The US residential heat pump market, a decade after "the crisis"," *12thIEA Heat Pump-Conference*, vol. 2017, p. 11, 2017.

[4] T. Nishimura, "Heat pumps—status and trends in Asia and the Pacific," *International Journal of Refrigeration*, vol. 25, no. 4, pp. 405–413, 2002.

[5] K. Saito, "Latest heat pump technologies in Japan," in *12th IEA Heat Pump Conference*, 2017, pp. 3–4.

[6] H. Zhao, Y. Gao, and Z. Song, "Strategic outlook of heat pump development in China," in *12th IEA Heat Pump Conference 2017*, 2017, pp. 1–5.

[7] M. H. Abbasi, B. Abdullah, M. W. Ahmad, A. Rostami, and J. Cullen, "Heat transition in the European building sector: Overview of the heat decarbonisation practices through heat pump technology," *Sustainable Energy Technologies and Assessments*, vol. 48, p. 101630, 2021.

[8] A. S. Gaur, D. Z. Fitiwi, and J. Curtis, "Heat pumps and our low-carbon future: A comprehensive review," *Energy Research & Social Science*, vol. 71, p. 101764, 2021.

[9] L. Sugden, "Smart grids create new opportunities for heat pumps," *REHVA J*, pp. 20–22, 2012.

[10] F. Fleischle and M. Kaniut, "Gutachten digitalisierung der energiewende topthema 1: Verbraucher, digitalisierung und geschäftsmodelle," *Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie; Bundesministeriums für Wirtschaft und Energie: Berlin, Germany*, accessed: 2022 May 24. [Online]. Available: https://www.bmwk.de/Redaktion/DE/Publikationen/Studien/digitalisierung-der-energiewende-thema-1.html

[11] T. Nowak and P. Westring, "Growing for good? the European heat pump market-status and outlook," in *European Heat Pump Conference, Rotterdam*, 2017.

[12] European Commission, *Benchmarking smart metering deployment in the EU-28: final report*. Publications Office, accessed: 2022 May 24. [Online]. Available: https://data.europa.eu/doi/10.2833/492070

[13] A. Weigert, K. Hopf, N. Weinig, and T. Staake, "Detection of heat pumps from smart meter and open data," *Energy Informatics*, vol. 3, no. 1, pp. 1–14, 2020.

[14] H. Fei, Y. Kim, S. Sahu, M. Naphade, S. K. Mamidipalli, and J. Hutchinson, "Heat pump detection from coarse grained smart meter data with positive and unlabeled learning," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1330–1338.

[15] K. Hopf, M. Sodenkamp, and T. Staake, "Enhancing energy efficiency in the residential sector with smart meter data analytics," *Electronic Markets*, vol. 28, no. 4, pp. 453–473, 2018.

[16] X. She, L. Cong, B. Nie, G. Leng, H. Peng, Y. Chen, X. Zhang, T. Wen, H. Yang, and Y. Luo, "Energy-efficient and-economic technologies for air conditioning with vapor compression refrigeration: A comprehensive review," *Applied Energy*, vol. 232, pp. 157–186, 2018.

[17] A. Weigert, K. Hopf, S. A. Günther, and T. Staake, "Heat pump inspections result in large energy savings when a pre-selection of households is performed: A promising use case of smart meter data," *Energy Policy*, vol. forthcoming, 2022.

[18] M. Mahmoudi, M. Dehghan, H. Haghgou, and M. Keyanpour-Rad, "Techno-economic performance of photovoltaic-powered air-conditioning heat pumps with variable-speed and fixed-speed compression systems," *Sustainable Energy Technologies and Assessments*, vol. 45, p. 101113, 2021.

[19] G. Bagarella, R. Lazzarin, and M. Noro, "Sizing strategy of on–off and modulating heat pump systems based on annual energy analysis," *International journal of refrigeration*, vol. 65, pp. 183–193, 2016.

[20] M. Szreder and M. Miara, "Effect of heat capacity modulation of heat pump to meet variable hot water demand," *Applied thermal engineering*, vol. 165, p. 114591, 2020.

[21] C. Aprea, R. Mastrullo, and C. Renno, "Experimental analysis of the scroll compressor performances varying its speed," *Applied thermal engineering*, vol. 26, no. 10, pp. 983–992, 2006.

[22] S. Shao, W. Shi, X. Li, and H. Chen, "Performance representation of variable-speed compressor for inverter air conditioners based on experimental data," *International journal of refrigeration*, vol. 27, no. 8, pp. 805–815, 2004.

[23] L. Zhao, L. Zhao, Q. Zhang, and G. Ding, "Theoretical and basic experimental analysis on load adjustment of geothermal heat pump systems," *Energy Conversion and Management*, vol. 44, no. 1, pp. 1–9, 2003.

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

[25] C. Betzold, S. Bordin, A. Dentel, and G. Harhausen, "Control strategies for modulating heat pumps in a plus energy building," in *Heat Pump Conference*, 2021.

[26] T. Péan, R. Costa-Castelló, E. Fuentes, and J. Salom, "Experimental testing of variable speed heat pump control strategies for enhancing energy flexibility in buildings," *IEEE access*, vol. 7, pp. 37 071–37 087, 2019.

[27] T. Q. Péan, J. Salom, and R. Costa-Castelló, "Review of control strategies for improving the energy flexibility provided by heat pump systems in buildings," *Journal of Process Control*, vol. 74, pp. 35–49, 2019.

[28] L. Langer and T. Volling, "An optimal home energy management system for modulating heat pumps and photovoltaic systems," *Applied Energy*, vol. 278, p. 115661, 2020.

[29] D. Fischer, F. Rautenberg, T. Wirtz, B. Wille-Haussmann, and H. Madani, "Smart meter enabled control for variable speed heat pumps to increase pv self-consumption," in *Proceedings of the 24th IIR International Congress of Refrigeration, Yokohama, Japan*, 2015, pp. 16–22.

[30] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2018.

[31] S.-I. Yang, C. Shen *et al.*, "A review of electric load classification in smart grid environment," *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 103–110, 2013.

[32] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68–80, 2012.

[33] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology*, vol. 20, no. 2, pp. 117–129, 2015.

[34] Y. Cui, R. Yan, R. Sharma, T. Saha, and N. Horrocks, "Realizing multifractality of smart meter data for household characteristic prediction," *International Journal of Electrical Power & Energy Systems*, vol. 139, p. 108003, 2022.

[35] G. Sun, Y. Cong, D. Hou, H. Fan, X. Xu, and H. Yu, "Joint household characteristic prediction via smart meter data," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1834–1844, 2017.

[36] C. Beckel, L. Sadamori, T. Staake, and S. Santini, "Revealing household characteristics from smart meter data," *Energy*, vol. 78, pp. 397–410, 2014.

[37] S. Yan, K. Li, F. Wang, X. Ge, X. Lu, Z. Mi, H. Chen, and S. Chang, "Time–frequency feature combination based household characteristic identification approach using smart meter data," *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 2251–2262, 2020.

[38] M. A. Devlin and B. P. Hayes, "Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data," *IEEE transactions on consumer electronics*, vol. 65, no. 3, pp. 339–348, 2019.

[39] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, "A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem," *arXiv preprint arXiv:1703.00785*, 2017.

[40] J. Revuelta Herrero, Á. Lozano Murciego, A. López Barriuso, D. Hernández de la Iglesia, G. Villarrubia González, J. M. Corchado Rodríguez, and R. Carreira, "Non intrusive load monitoring (NILM): A state of the art," in *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, 2017, pp. 125–138.

[41] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.

[42] M. Zhuang, M. Shahidehpour, and Z. Li, "An overview of non-intrusive load monitoring: Approaches, business applications, and challenges," in *2018 international conference on power system technology (POWERCON)*. IEEE, 2018, pp. 4291–4299.

[43] F. Culière, L. Leduc, and A. Belikov, "Bayesian model of electrical heating disaggregation," in *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, 2020, pp. 25–29.

[44] K. Kouzelis, Z. H. Tan, B. Bak-Jensen, J. R. Pillai, and E. Ritchie, "Estimation of residential heat pump consumption for flexibility market applications," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1852–1864, 2015.

[45] C. Deb, M. Frei, J. Hofer, and A. Schlueter, "Automated load disaggregation for residences with electrical resistance heating," *Energy and Buildings*, vol. 182, pp. 61–74, 2019.

[46] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psychological bulletin*, vol. 76, no. 5, p. 378, 1971.

[47] A. J. Viera, J. M. Garrett *et al.*, "Understanding interobserver agreement: the kappa statistic," *Fam med*, vol. 37, no. 5, pp. 360–363, 2005.

[48] C. G. Northcutt, A. Athalye, and J. Mueller, "Pervasive label errors in test sets destabilize machine learning benchmarks," *arXiv preprint arXiv:2103.14749*, 2021.

[49] B. Najafi, M. Depalo, F. Rinaldi, and R. Arghandeh, "Building characterization through smart meter data analytics: Determination of the most influential temporal and importance-in-prediction based features," *Energy and Buildings*, vol. 234, p. 110671, 2021.

[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.