Anomaly detection on household appliances based on variational autoencoders

# Anomaly detection on household appliances based on variational autoencoders

Marco Castangia[a,*], Riccardo Sappa[b], Awet Abraha Girmay[b], Christian Camarda[b], Enrico Macii[c], Edoardo Patti[a]

[a] *Department of Control and Computer Engineering, Politecnico di Torino, 10129 Torino, Italy, email: {name.surname}@polito.it*
[b] *Midiori, Italy, email: {name.surname}@midorisrl.eu*
[c] *Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, 10129 Torino, Italy, email: {name.surname}@polito.it*

**Abstract**

Electrical anomalies in residential buildings represent a serious problem that can unpredictably change the power profiles of end-users, causing a sub-optimal energy distribution. In addition, electrical faults can cause unnoticed energy wastages and higher energy bills, or even severe damages for properties and people in the most critical cases. In this paper, we introduce a novel anomaly detection method for detecting electrical faults in household appliances based on the analysis of their power signatures with unsupervised deep learning techniques. For this purpose, we trained a variational autoencoder to reconstruct the power signatures of three commonly used devices: the dishwasher, the washing machine and the dryer. For each use case, we injected several randomly generated anomalies that simulate to our best realistic electrical faults in these devices. To demonstrate the effectiveness of our method, we compared the accuracy of the variational autoencoder with the classification performance of a one-class support vector machine (OC-SVM) trained with two manual features: the energy consumption and duration of the appliance's operations. The variational autoencoder showed higher classification accuracy with respect to the OC-SVM, reporting an F1-score greater than 90% in all the use cases. Most importantly, the results demonstrate that deep anomaly detection methods outperform traditional algorithms based on handcrafted features, allowing to better characterize the set of normal cycles and produce more precise alerts for the monitored devices.

*Keywords:* smart grids, household appliance, anomaly detection, unsupervised learning,

*marco.castangia@polito.it

variational autoencoder, deep learning.

*2010 MSC:* 00-01, 99-00

## 1. Introduction

The accelerating depletion of fossil fuels and the impending rise of worldwide temperatures require an urgent intervention from governments to find more sustainable energy sources (COP26, 2021). As for now, renewable energy sources already provide for a quarter of the world's electricity supply and shows a very promising upward trend in several countries (Ritchie & Roser, 2020). Despite that fact, renewable sources such as solar and wind still pose important challenges for their full integration in our energy mix, mainly because their intermittent nature makes them less reliable (Castangia et al., 2021a).

The development of Smart Grids enables new opportunities for grid operators, that can now leverage the capabilities of Information and Communication Technologies (ICT) to close the gap between energy providers and consumers. In particular, consumers can be involved in demand-response (DR) programs in which they commit to change their energy demand in response to variations in the available power supply from utilities (Siano, 2014). With the higher penetration of renewable sources into the electricity grids, DR programs are mainly used to match the energy demand of end-users with the unstable power supply of renewables. To maintain the balance between power supply and demand, grid operators need accurate estimates on both sides of the grid (Mocanu et al., 2016).

The presence of electrical anomalies in the power consumption of residential buildings can seriously hinder the optimal scheduling of energy distribution. In fact, unpredictable electrical loads can easily invalidate the estimates of grid operators about the future end-users' consumption, causing a sub-optimal energy distribution. For this reason, utilities are interested in promptly identifying electrical anomalies and get rid of them to improve their models of the consumers' power profiles. In addition to that, there is an obvious energy wastage for the end-user, that ends up paying higher costs in the energy bill. In the most critical cases, electrical anomalies can even evolve in more severe damages, with a serious risk for both properties and people.

The only way to truly fix an electrical fault is to find the responsible device, which can be repaired or replaced by the end-user afterwards. It has been largely demonstrated that anomaly detection performed at the aggregate-level, i.e. at the main meter of the house, do not provide enough details to find the faulty device (Himeur et al., 2021). Conversely, power consumption collected at the appliance-level can provide the right level of details to evaluate the health status of the appliance. The development of Internet of Things (IoT) allowed the creation of more affordable monitoring devices that can be deployed at scale in our houses. For example, monitoring devices such as Smart Plugs can be used to monitor the power consumption of the major household's appliances, including fridges, dishwashers, washing machines and air conditioners (Alsalemi et al., 2020).

In this study, we introduce a new approach to find electrical anomalies in the power consumption of household appliances through the analysis of their power signatures with deep learning techniques. In particular, we exploit a public data set of power measurements collected directly from the monitored device at regular intervals of 1 second. To demonstrate the effectiveness of our anomaly detection framework, we analyzed the power signatures of three common household appliances: the dishwasher, the washing machine and the clothes' dryer. Similarly to other anomaly detection methods, we want to model the normal behaviours in the data set and recognize irregular instances at test time. However, this task can be particularly challenging in the presence of household appliances with more complex power signatures and multiple modes of operation. In those cases, traditional anomaly detection algorithms risk to overfit the normal behaviours of the appliance if their hyper-parameters are not properly chosen a priori, such as the expected future percentage of anomalies. In addition, the extraction of handcrafted features is a tedious task that can easily lead to sub-optimal features that do not fully characterize the raw data set. To cope with these problems, we adopted deep learning techniques that are well-known for their capabilities of automatically learning complex features and can model very convoluted input sequences (Oord et al., 2016). More specifically, we implemented a variational autoencoder to model the normal operation mode of the monitored appliance. Then, at test time we compare the reconstruction error of the autoencoder with an anomaly threshold previously estimated from the training errors, looking for significant deviations. Whenever the reconstruction error exceeds our threshold, we found an anomalous operation cycle.

With respect to our previous work (Castangia et al., 2021b), we significantly extended our investigation on anomalous household appliances in the following ways:

- The set of analyzed devices has been expanded from low-power equipments (e.g. fridge, stereos, TVs, baseline) to more energy-intensive appliances such as the washing machine, the dishwasher and the clothes' dryers.
- The study of the device's power consumption has been refined, passing from the hourly analysis of energy consumption to the study of instantaneous power measurements, with a tremendous increase in the amount of details.

The rest of this paper is organized as follows: Section 2 provides an overview of the previous works on the task of anomaly detection at the appliance-level. Section 3 gives a description of the data set. Section 4 describes the different stages of our anomaly detection framework. Section 5 presents the adopted evaluation procedure, followed by an in depth discussion of the classification results for the different devices analyzed. Finally, Section 7 highlights the most significant achievements of this work, providing some directions for potential improvements.

## 2. Related works

Many researchers studied anomalies in household appliances from a behavioural point of view, exploiting the presence of patterns in the daily habits of end-users. Mao et al. in (Mao et al., 2019) present an anomaly detection framework based on frequent pattern mining aimed at identifying anomalies in the daily usage of various home appliances. The proposed framework monitors the number of daily operations and consider as anomalous those days that significantly deviate from the expected number of appliance's usages. To distinguish between normal and anomalous days, the authors altered the number of operations in normal days and trained a classifier to recognize those anomalous instances. Patricio et al. in (Patricio et al., 2019) analyzed the operations of many devices during the day with the goal of finding anomalous behaviours in elderly people. The authors firstly extracted a lower data representation of the hourly power consumption of the target device with an autoencoder neural network. Then, they trained a random forest classifier to identify anomalous behaviours, that were previously generated by changing the daily patterns of the devices during the day. Similarly, Gonzalez et al. in (Gonzalez et al., 2021) proposed a method to identify anomalous behaviours in elderly people based on deep autoencoders. In particular, the authors compared the performance of autoencoders and variational autoencoders in extracting relevant representations of the usage patterns. The extracted representation are given as input to a random forest classifier for predicting abnormal behaviours, with the help of synthetically generated anomalous patterns. In the end, the solution with the variational autoencoder seems to be the most effective for all the appliances analyzed.

Another interesting field of research investigates the detection of faulty appliances through the analysis of their power consumption. Himeur et al. in (Himeur et al., 2020) extracted a set of micro-moments classes by analyzing each device in terms of its power range, maximum operation time and standby consumption. To find anomalous power consumption, the authors trained a deep neural network to recognize the extracted micro-moments classes, including the anomalous ones. Hosseini et al. in (Hosseini et al., 2020) introduced a semi-supervised framework for detecting electrical anomalies in the power consumption of a fridge. The authors characterized the fridge's power profile in terms of its daily energy consumption. The normal values are extracted by monitoring the fridge's power during a first calibration period. Their algorithm recognizes an anomalous day whenever the daily consumption exceeds the mean value by more than three standard deviations. Benninger et al. in (Benninger et al., 2020) presented a completely unsupervised anomaly detection framework that works in real-time with high frequency (1Hz) current measurements. The proposed framework firstly applies a k-means clustering to divide the samples into the different operation modes of the device. Then, a one-class support vector machine (OC-SVM) is trained for each operation mode, assuming a 20% of samples belonging to the anomalous class in normal conditions. In the test phase, samples that deviates more than 20% from their assigned operation mode are considered anomalous for that device. Rashid et al in (Rashid et al., 2019) tried to find anomalies in household appliances by extracting their power consumption with non-intrusive load monitoring (NILM) techniques. To find the anomalies, they applied some thresholds to the energy consumption and duration of the device's activation based on their expected normal power profile. The authors concluded that current state-of-the-art NILM techniques are not sufficiently accurate to perform anomaly detection, which is otherwise very effective when working with sub-metered data. In our previous work (Castangia et al., 2021b), we analyzed the power consumption of three different sources of power absorption, namely the baseline, the fridge and some electronic devices (TVs, stereos), with the aim of finding both electrical and behavioural anomalies. In detail, we analyzed the selected devices in terms of their daily/hourly energy consumption, duration and time of the day. In this paper, which extends our previous work, we want to elaborate further on the study of anomalous appliances by extending the set of analyzed devices to more energy-intensive equipments.

In particular, instead of studying the hourly energy consumption of the monitored devices, we want to investigate the characteristics of their instantaneous power measurements, which provide a greater level of detail. As a matter fact, the power signature of the monitored device reflects the functioning of its internal electrical components. Consequently, we can recognize electrical faults in the appliance by analyzing the temporal features of its power consumption.

The proposed solution based on the use of deep learning techniques is specifically aimed at solving the limitations of previous literature works. First of all, we avoided using a supervised learning approach because it is unfeasible to know in advance all the possible anomalous behaviours for a certain appliance. Indeed, a better method consists of using an unsupervised approach, considering any deviation from the expected behaviour as a potential anomaly in the monitored device. In the same spirit, we avoided models that assume a predetermined percentage of anomalous instances in the data set, such as the OC-SVM proposed in (Benninger et al., 2020). Conversely, we preferred to use a more adaptive anomaly threshold, which is determined by the specific characteristics of the training data at hand. Most importantly, we escaped the use of manual features by employing a deep learning approach, which automatically derives useful features as part of its training process (Chalapathy & Chawla, 2019). In fact, handcrafted features risk to provide only a limited description of the appliance's behaviour and may be unsuitable to characterize more complex devices presenting multiple modes of operations. For the sake of clarity, Table 1 provides an overview of the main differences between our methodology and those proposed in the literature. To demonstrate the advantage of deep learning with respect to previous solutions employing manual features, we compared our anomaly detection method with a one-class support vector machine trained with two features: the energy consumption and the duration of the appliance's power signatures.

Table 1: Differences between our methodology and the related works focusing on anomalous appliance detection.

| **Reference** | **Learning method** | **Anomaly threshold** | **Feature extraction** |
|---|---|---|---|
| (Himeur et al., 2020) | supervised | adaptive | manual |
| (Hosseini et al., 2020) | unsupervised | adaptive | manual |
| (Benninger et al., 2020) | unsupervised | hard-coded | manual |
| (Rashid et al., 2019) | unsupervised | adaptive | manual |
| Proposed solution | unsupervised | adaptive | automatic |

## 3. Dataset

To run our experiments we used the Electricity Consumption and Occupancy (ECO) data set (Beckel et al., 2014) collected from 6 households in Switzerland between June 2012 and January 2013. This data set is particularly suitable for our purposes because it provides power measurements at the appliance-level with a very high resolution (1Hz). Besides, the long monitoring period (8 months) guarantees a considerable number of appliance's operations and a good diversity of appliance's programs, which is important to test the robustness of our approach. As a matter of fact, we decided to focus only on household appliances that present a repeatable power signature, which is determined mainly by the specific appliance's program selected by the end-user. The most notable appliances that fall into this category are the washing machine, the dishwasher and the dryer. The reason behind our choice is that appliance's programs present a certain regularity over time, that can be modeled and exploited to detect major anomalies in the electrical components of the device. For this category of appliances, simple statistics such as the total energy consumption and duration of the operation do not totally describe the operational behaviour of the electrical components, thus preventing the identification of more subtle anomalies. In fact, the possibility of selecting multiple programs in the same appliance may determine the presence of two different operation modes with almost the same energy consumption and duration. In these cases, the operations are discernible only by considering the temporal correlations of the different phases that constitute their specific power signatures. Table 2 reports the initial number of operation cycles that we extracted from the raw power measurements of the washing machine, the dishwasher and the dryer in the ECO data set. The segmentation procedure used to extract the operation cycles is described in the next Section 4.

Table 2: Data set.

| Appliance | Building | Num. of cycles | Monitoring period |
|---|---|---|---|
| Dishwasher | 2 | 62 | 2012/06/01 - 2013/01/31 |
| Dryer | 1 | 78 | 2012/06/01 - 2013/01/23 |
| Washing machine | 1 | 274 | 2012/06/01 - 2013/01/23 |

## 4. Methodology

Figure 1 shows the complete pipeline of our anomaly detection framework. The pipeline can be divided in two main steps: the training phase and the test phase. Both phases firstly involve the separation of the individual operation cycles and their preparation to be processed by the variational autoencoder. The purpose of the training procedure is to properly model the power profiles of the normal operation cycles and determine the best anomaly threshold to separate the normal cycles from the anomalous ones. In the test phase, we apply the variational autoencoder to reconstruct the power signatures observed at inference time. To recognize anomalous instances, we compare the reconstruction errors with the anomaly threshold determined at the end of the training phase: we consider anomalous all those instances whose reconstruction error exceeds our anomaly threshold. In the following, we provide a thorough explanation of the different components of our pipeline. In addition, we give a brief description of the procedures that we used to generate the synthetic anomalies to test our framework.

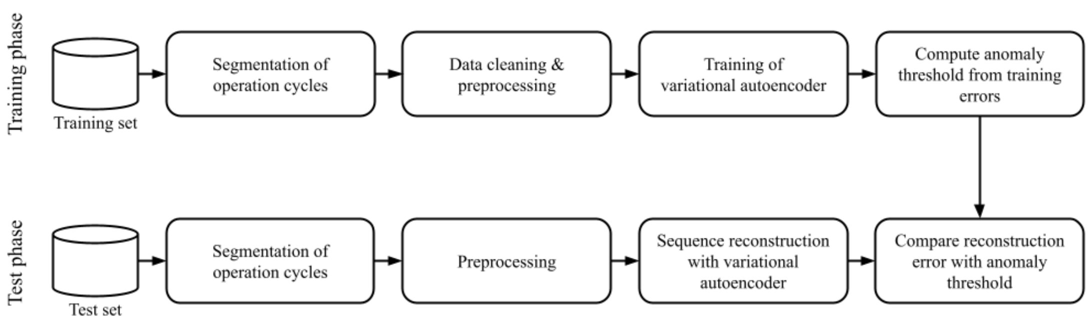



Figure 1: Pipeline of our proposed anomaly detection framework.

### *4.1. Segmentation of operation cycles*

The segmentation of the individual appliance's operations is the very first stage of our data pipeline. Indeed, we need to analyze each appliance's activation separately in order to model the typical power signature of the monitored device. Massidda et al. in (Massidda et al., 2020) devised a simple segmentation procedure that processes the power measurements from the sub-meter in linear time, and gives the start and stop times of each appliance's activation. The algorithm requires three input parameters that the authors called *min power*, *min off* and *min on*. The first argument (*min power*) specifies the minimum power threshold to consider the appliance in its operational state, i.e. active. The second argument (*min off*) is the minimum elapsed time with the power consumption below the *min power* threshold to assume that the appliance's operation has been completed. Finally, the third argument (*min on*) specifies the minimum elapsed time with the power consumption above the *min power* threshold to consider the appliance's state as active. Table 3 reports the arguments of the segmentation procedure for the three household appliances analyzed in this work. The best values for these parameters were found in an empirical way by studying the duration and the minimum power consumption of each device. The same parameters can work also with other data sets since they are based on universal features that characterize the power signatures of the monitored devices. More advanced segmentation procedures based on deep neural networks can be implemented, but they are out of the scope of this work and do not provide significant benefits with respect to our method.

Table 3: Selected values for the main parameters of the segmentation procedure.

| | *min power* [W] | *min on* [s] | *min off* [s] |
|---|---|---|---|
| Dishwasher | 10 | 120 | 180 |
| Dryer | 10 | 120 | 150 |
| Washing machine | 10 | 120 | 180 |

### *4.2. Data cleaning*

Before feeding the data into our variational autoencoder, we must make sure that the training set contains only normal operation cycles. Unfortunately, the appliance's operations extracted by the segmentation procedure may contain spurious cycles that do not represent authentic power signatures. An incorrect segmentation can have three possible causes: i) an error in the sub-meter measurements, ii) a voluntary interruption of the operation by the end-user, and iii) a failure of the segmentation algorithm in separating very close consecutive cycles. In either case, the result is a power signature that is uncommon and therefore not representative of the normal operational cycle of the target device. To detect spurious operation cycles we exploited the Isolation Forest algorithm (Liu et al., 2008), fitted with the duration and the energy consumption of the individual cycles. In this way, every appliance's activation with an unusual duration or an unexpected energy consumption has been removed from the data set.

### 4.3. Preprocessing

The variational autoencoder uses as input the sub-metered power consumption of the monitored appliance measured with a sampling frequency of 1 Hz. In particular, the model has been designed to handle sequences of instantaneous power measurements whose maximum length has been set to 8000 samples, which correspond roughly to 2 hours of activity. For this reason, all input sequences have been capped to 8000 samples to ensure consistency in the training set. Shorter sequences have been zero padded up to the right number of samples. To improve the training procedure, we normalized the input sequences with the following formula:

$$x_{scaled} = \frac{x - \mu_x}{\sigma_x} \tag{1}$$

where $x$ is the power signature, $\mu_x$ is the mean power in the training set and $\sigma_x$ is the standard deviation. The same values are used at test time to prepare the input sequences for the analysis.

### 4.4. Variational autoencoder

Autoencoders are neural networks that are trained to minimize the reconstruction error between the inputs and the outputs in the presence of some constraint (Hinton & Salakhutdinov, 2006). The most commonly used constraint consists on reducing the dimensionality of the hidden layers, such that the neural network retains only the most relevant features necessary to reconstruct the input signal. The variational autoencoder was firstly introduced in (Kingma & Welling, 2013) as an effective way for performing Bayesian inference in the presence of continuous latent variables with intractable posterior distributions. Differently from classical autoencoders, that always have a deterministic output at inference time, variational autoencoders present a probabilistic component that partially depends on randomness, even at inference time. More specifically, the variables composing the latent representation are sampled from a multivariate random distribution, whose mean and standard deviation are determined by the encoder network. In addition to the typical reconstruction error, variational autoencoders are also trained to minimize the Kullback–Leibler divergence between the distribution of the extracted latent representations and a Gaussian distribution. This second loss acts as a regularization factor and pushes the network to learn better data representations. The latent loss is described in Equation 2, where $Z_{dim}$ is the number of latent variables of our model, and $y_i$ and $\mu$ are the outputs of the encoder network. Notice that the encoder predicts $y_i = log(\sigma^2)$ rather than generating directly $\sigma$ because it leads to better convergence during training.

$$\mathcal{L} = -\frac{1}{2}\sum_{i=1}^{Z_{dim}} 1 + y_i - e^{y_i} - \mu_i^2 \tag{2}$$

The full architecture of the variational autoencoder used in this work is described in Figure 2. The encoder network is composed by a stack of six convolutional layers that drastically reduce the dimension of the input sequences. Then, the encoder terminates with two dense layers that generates the mean and the variance of the latent variables, respectively. The parameters are combined with the output of a normal distribution with zero mean and unit variance to produce the final latent representation. Finally, the latent variables are passed to a decoder network with a stack of six transposed convolutional layers, that brings everything to the original sequence length.

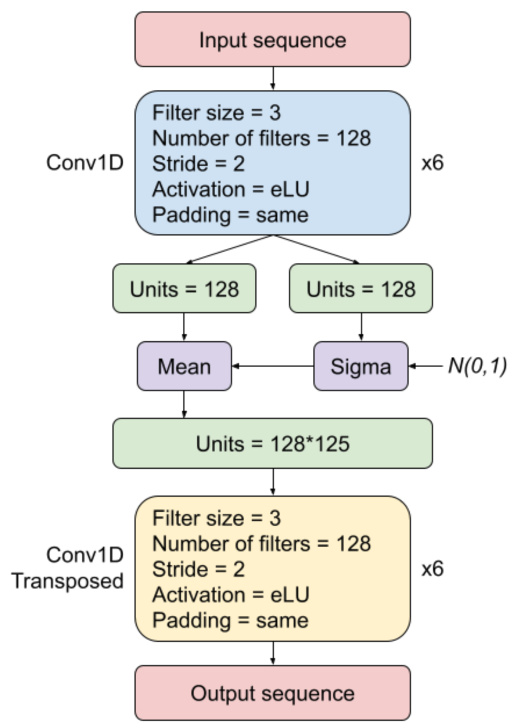



Figure 2: Architecture of the variational autoencoder.

The hyperparameters of the training procedure are reported in Table 4. We used the adaptive moment estimation (Adam) algorithm with default arguments to optimize the network's parameters. The training can run for at most 500 epochs. We kept 20% of the training set for validating the network's performance at the end of each epoch, stopping the training if the loss does not show improvements for more than 25 epochs (early stopping criteria). A default batch size of 32 has been used. The network is implemented in Python with the help of the Keras library (Chollet et al., 2015)

Table 4: Training hyperparameters of the variational autoencoder.

| Hyperparameter | Value |
|---|---|
| Optimizer | Adam |
| Learning rate | 0.001 |
| Epochs | 500 |
| Batch size | 32 |
| Stopping criteria | early stopping with patience equal to 25 |

### *4.5. Synthetic anomalies in the washing machine*

The power signature of the washing machine is generally characterized by two main features: the water heating and the spin cycles. The spin cycles represent only a minor contribution to the total energy consumption of the washing machine. Indeed, it has been observed that more than 90% of the energy consumption in a single operation is due to the water heating stage (Issi & Kaplan, 2018). For this reason, we modeled electrical anomalies in the washing machine as unexpected water heating stages randomly inserted in the operation. A detailed description of the procedure for generating washing machine's anomalies is reported in Algorithm 1. Figure 3 shows an artificial anomaly randomly generated for the washing machine, in which the duration of the heating stage has been significantly altered, resulting in an increase of the energy consumption. Indeed, detecting this kind of anomalies is very important for assessing the energy efficiency of the device, that can unexpectedly consume more power than what is actually required by the selected program.

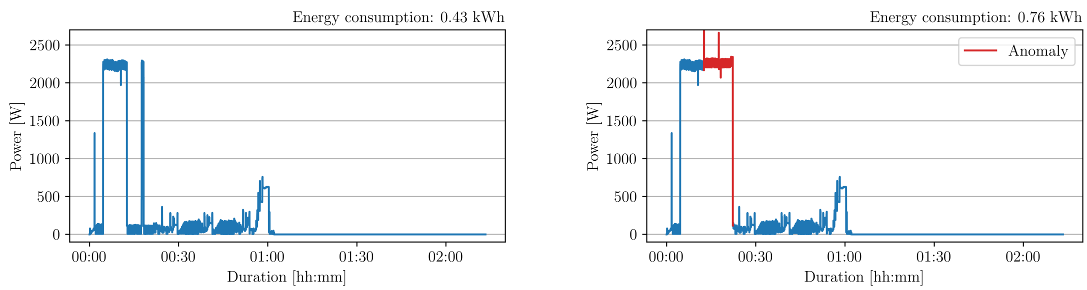
**Algorithm 1** Steps for generating anomalous washing machine's operations.

1: Use the end of the first heating stage as the *starting point* of the anomaly;

2: If there are no heating stages (cold wash), then randomly select a *starting point* between 2 and 20 minutes from the beginning of the operation;

3: Pick up the *end point* of the anomaly by randomly selecting a duration time between 5 and 20 minutes;

4: Increase the power consumption from the *starting point* to the *end point* by 2200 W, except for the samples that are already in a heating stage, i.e. where the power is greater than 1500 W.




(a) Normal washing machine operation

(b) Anomalous washing machine operation

Figure 3: Examples of a normal washing machine operation and the corresponding anomalous cycle generated.

### *4.6. Synthetic anomalies in the dishwasher*

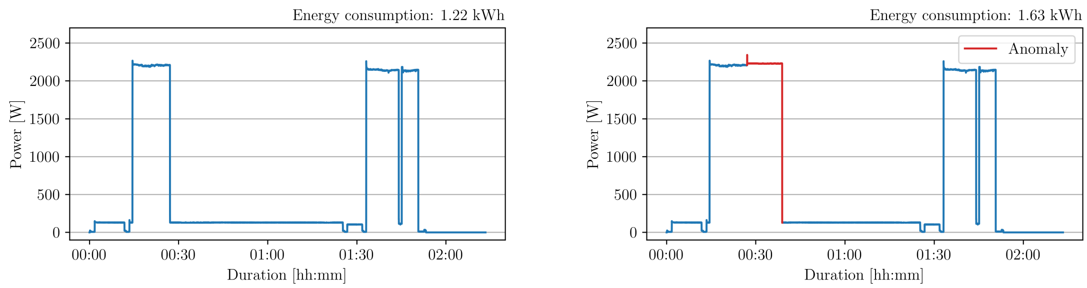
The dishwasher presents a power signature characterized by the alternation of multiple heating stages with different duration. Each dishwasher's program has its own specific pattern, that makes the task of anomaly detection more challenging. Electrical anomalies in the dishwasher can be represented by a corruption in the duration of the different heating stages constituting the single operation cycle. Therefore, we decided to randomly change the duration of the heating stages in order to simulate potential electrical anomalies in the dishwasher. The different steps for generating the dishwasher's anomalies are described in Algorithm 2. Figure 4 depicts an anomaly of this kind, in which the duration of the first heating cycle has been incremented to simulate a general electrical anomaly in the dishwasher. Similarly to the case of the washing machine, this kind of anomalies are useful for detecting inefficiency in the appliance's operations, which may cause a systematic increase in the overall energy consumption of the device.

**Algorithm 2** Steps for generating anomalous dishwasher's operations.

1: Use the end of a randomly selected heating stage as the *starting point* of the anomaly;

2: Pick up the *end point* of the anomaly by randomly selecting a duration time between 5 and 20 minutes;

3: Increase the power consumption from the *starting point* to the *end point* by 2100 W, except for the samples that are already in a heating stage, i.e. where the power is greater than 1500 W.




(a) Normal dishwasher operation

(b) Anomalous dishwasher operation

Figure 4: Examples of a normal dishwasher operation and the corresponding anomalous cycle generated.

### *4.7. Synthetic anomalies in the dryer*

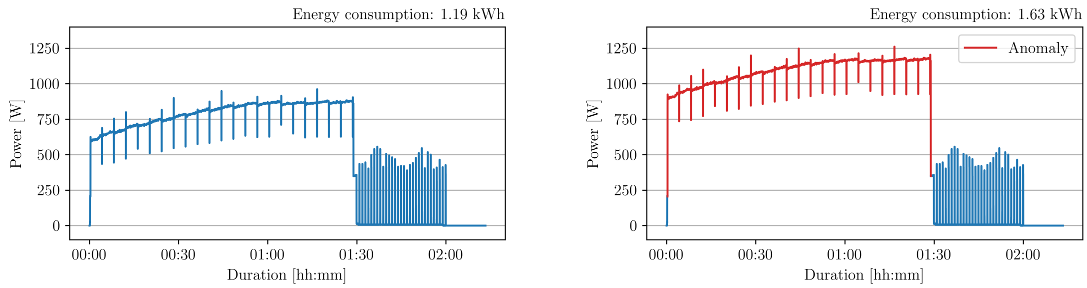
The dryer shows a simple power signature with a single long heating stage that interests a major part of its operation cycle. Indeed, we do not have a complex pattern to model since there is no alternation of heating stages in the dryer's case. However, we can still replicate a realistic electrical anomaly by increasing the overall power consumption of the principal heating phase. The pseudo code for generating the dryer's anomalies is reported in Algorithm 3. Figure 5 shows an anomaly in the power signature of the dryer in which the power consumption of the heating stage has been increased by 300 W. Similarly to the other use cases, an increase in the energy consumption is an important concern for the efficiency of the household appliance.

**Algorithm 3** Steps for generating anomalous dryer's operations.

1: Use the start of the main heating stage as the *starting point* of the anomaly;

2: Use the end of the main heating stage as the *end point* of the anomaly;

3: Increase the power consumption from the *starting point* to the *end point* by 300 W.




(a) Normal dryer operation

(b) Anomalous dryer operation

Figure 5: Examples of a normal dryer operation and the corresponding anomalous cycle generated.

## 5. Results

In the following, we present the results of our deep anomaly detection method applied to the different appliances considered in this study. Firstly, we introduce the validation procedure and the classification metrics that we used to evaluate our methodology. Then, we describe the computation of the anomaly threshold that we used to distinguish between normal and anomalous instances. To conclude, we compare the classification scores of the variational autoencoder with the results obtained by a classical anomaly detection algorithm trained on some handcrafted features, providing some motivations for the superiority of our method.

### *5.1. K-fold cross validation*

In general, it is desirable to have both a large training set and a wide validation set to test the performance of our models in a vast range of different scenarios. In this study, we are dealing with a limited amount of data points, which collides with the previous expectations. However, we can still perform an extensive evaluation by employing a *K-fold cross validation*. In detail, we run a 5-fold cross validation for each appliance, training the variarional autoencoders on the first 4 folds and testing on the remaining one. After each iteration, the data set is rotated by one fold to generate a new partition between training and test sets. In this way, the model is evaluated on unseen data points at every iteration, which guarantees an exhaustive evaluation of our methodology (Bengio & Grandvalet, 2004).

The synthetic anomalies described in Section 4 were injected in the test set at the beginning of each iteration of the K-fold cross validation. In the real-world, anomalies usually constitute only a minor percentage of the data set, since they are uncommon data points by definition. In our case, we decided to keep an equal percentage of normal and anomalous instances in order to evaluate the performance of our approach in a variety of different scenarios. Furthermore, we are not limited by the number of anomalies because they were randomly generated at each iteration. Thus, each test set created during the K-fold cross validation contains an equal amount of normal and anomalous operation cycles. Please, notice that the anomalous instances in the test set were generated by altering the normal operation cycles of the same test set.

### *5.2. Evaluation metrics*

The task of anomaly detection can be easily seen as a binary classification problem, where the negative class represents the normal data points and the positive class corresponds to the anomalous instances. Therefore, we adopted the typical classification metrics listed below:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{6}$$

where true positives (TP) indicate truly anomalous cycles classified as anomalous, true negatives (TN) are normal cycles correctly predicted as normal, false positives (FP) denote normal instances wrongly predicted as anomalies and false negatives (FN) are anomalous instances erroneously predicted as normal. In brief, the accuracy just reports how many times our predictions are actually correct. The precision gives us the ratio of true positives over the total number of positive predictions, whereas the recall gives the ratio of true positives over the total number of positive instances. To summarize precision and recall, we can compute their harmonic mean which is expressed by the F1-score.

### *5.3. Anomaly threshold*

The reconstruction error can be used to distinguish between normal and anomalous instances. As a matter of fact, normal operation cycles are reconstructed quite accurately by the autoencoder, since they have been already seen in the training set. Conversely, anomalous operation cycles present a high reconstruction error because they are completely unknown to the autoencoder and do not respect the training data distribution. In this work, we defined the *reconstruction error* as the mean absolute error between the input sequence $y$ and the reconstructed sequence $\hat{y}$, both with length equal to $n$.

$$reconstruction\ error = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

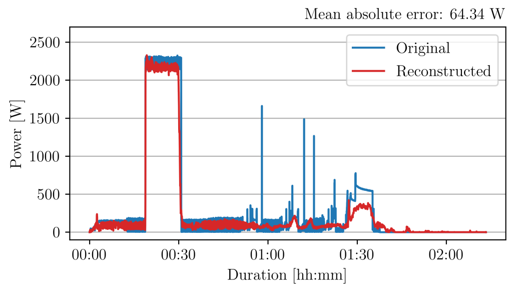
Once the reconstruction error has been defined, we must select an anomaly threshold to optimally separate normal and anomalous instances at inference time. In a first attempt, we tried to use the maximum reconstruction error achieved in the training set as our threshold. However, we early found that the maximum error on the training set is susceptible to the presence of outliers in the training data. Therefore, we finally decided to use an anomaly threshold equal to three standard deviations above the mean reconstruction error obtained on the training set. Indeed, according to the empirical rule of statistics, we can assume that 99.7% of data lies within three standard deviations of the mean. In this way, the threshold is more robust and is not contaminated by potential outliers in the training set.

### *5.4. Analysis of the classification performance*

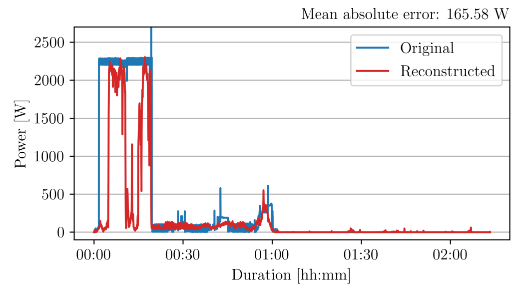
To test the effectiveness of the variational autoencoder, we compared our model with a one-class support vector machine (OC-SVM) (Yin et al., 2014) trained with two features, i.e. the total energy consumption and the duration of the operation cycles. The OC-SVM has been favoured by optimizing both the *contamination factor* and the *gamma* parameter in order to maximize the F1-score obtained on the test set. Notice that in a real-world scenario those parameters must be selected aprioristically, depending on the specific characteristics of the data set and the expected percentage of anomalous instances. Table 5 reports the classification results of the two models for the three appliances considered in this study. In particular, we reported the mean scores obtained across the five iterations of the K-fold cross validation. In all the use cases, the variational autoencoder overcomes the classification performance of the OC-SVM algorithm, both in terms of accuracy and F1-score. In particular, the accuracy of the variational autoencoder is greater than 90% for all the appliances analyzed, outperforming the results of the OC-SVM by more than 10% for the dryer and by more than 4% for both the washing machine and the dishwasher. The same improvements can be noticed on the F1-score, which exceeds the 90% threshold in all the use cases. Figure 6 shows the difference in the reconstruction errors when the variational autoencoder is fed with a normal cycle or an anomalous one. In the latter case, the reconstruction error is way higher, since the variational autoencoder is not able to properly reconstruct unknown operation cycles that deviate from normal power signatures. The proposed approach also proved particularly efficient in terms of computational costs, requiring less than 5 ms to reconstruct a single power signature and compare the reconstruction error with the reference anomaly threshold. The computational times demonstrate that our anomaly detection framework can be also deployed on edge computing devices with minor processing capabilities, including smart plugs and smart appliances.

Table 5: Classification results for the different appliances.

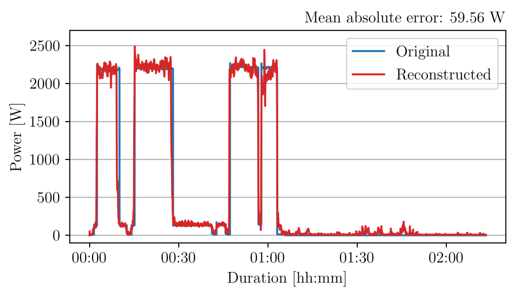
| **Appliance** | **Model** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
|---|---|---|---|---|---|
| Washing machine | OC-SVM | 0.88 | 0.83 | **0.95** | 0.89 |
| | VAE | **0.92** | **0.97** | 0.86 | **0.91** |
| Dishwasher | OC-SVM | 0.91 | 0.91 | 0.93 | 0.91 |
| | VAE | **0.95** | **0.95** | **0.95** | **0.95** |
| Dryer | OC-SVM | 0.84 | 0.81 | 0.90 | 0.85 |
| | VAE | **0.98** | **0.97** | **0.98** | **0.98** |



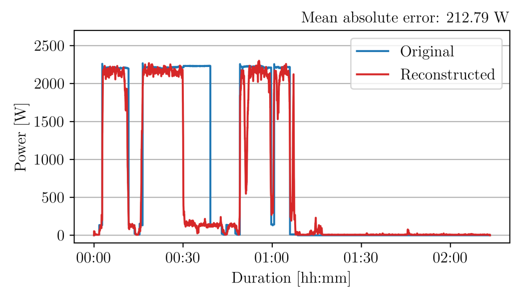

(a) Normal washing machine operation



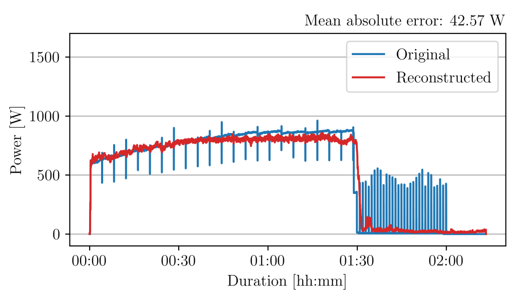

(b) Anomalous washing machine operation



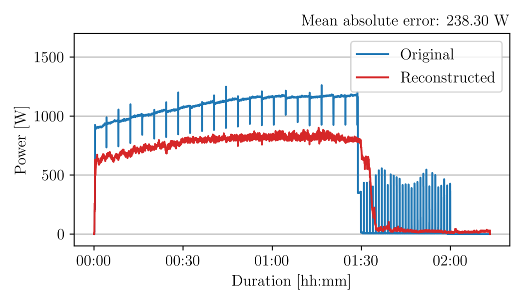

(c) Normal dishwasher operation




(d) Anomalous dishwasher operation




(e) Normal dryer operation



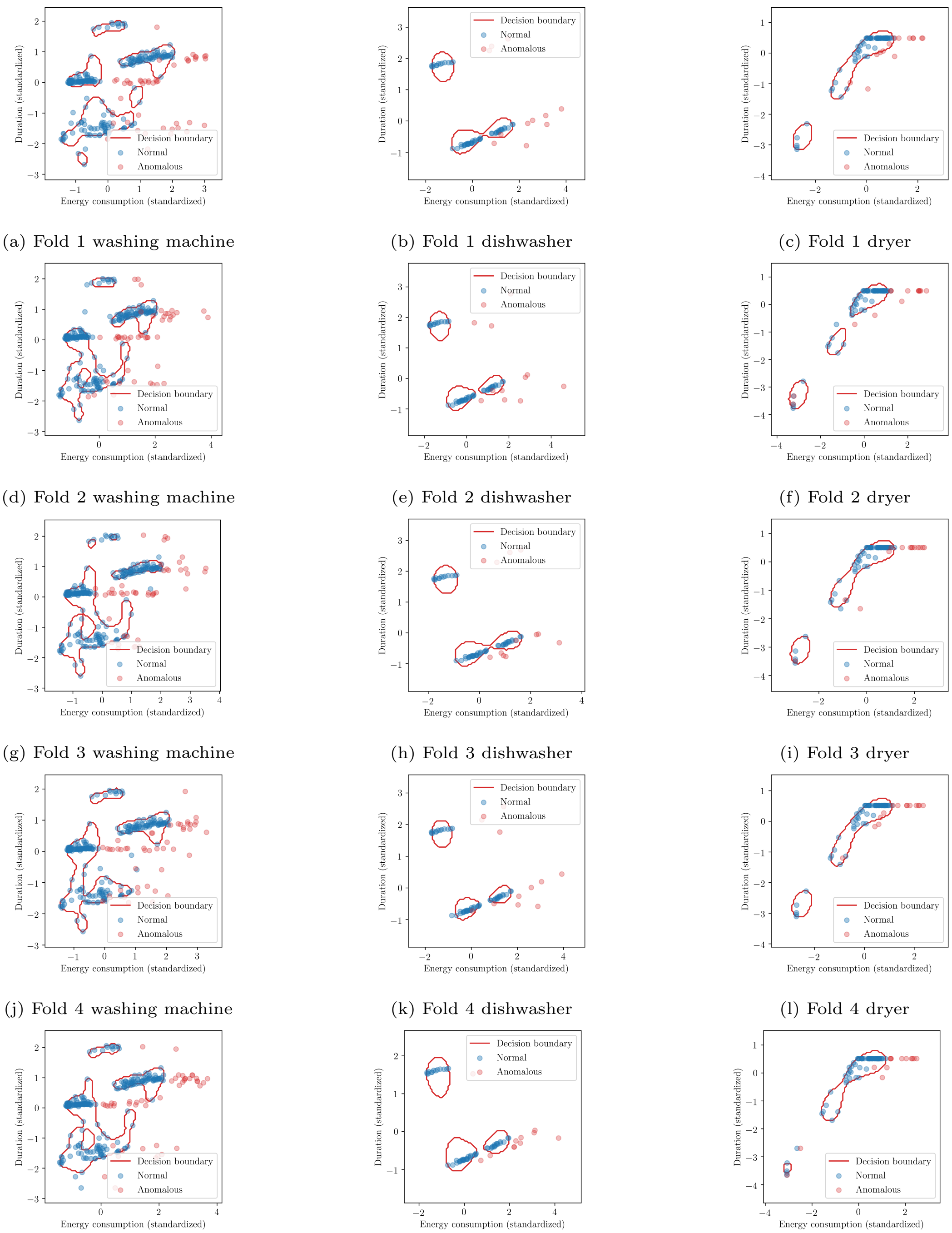

(f) Anomalous dryer operation

Figure 6: Difference in the signal reconstruction between normal (left column) and anomalous (right column) operation cycles in the washing machine, the dishwasher and the dryer, respectively.

The advantage of our anomaly detection method with respect to the OC-SVM algorithm lies in the greater generalization capabilities of the variational autoencoder, which seems to better cope with the diversity of the different appliance's programs. With reference to the results of Table 5, the OC-SVM showed lower precision scores with respect to the variational autoencoder, revealing a significant number of false positives in all the use cases. Therefore, we can deduce that the OC-SVM is more prone to overfitting the distribution of the training samples in comparison to the deep autoencoders, causing a higher number of false positives at test time. Figure 7 shows the optimal decision boundaries described by the OC-SVM once trained on the set of normal instances. Each column reports the decision boundaries obtained on each fold of the cross-validation procedure for a single appliance. The plots reported in Figure 7 clearly show the complexity of the operation cycles when represented in terms of their total energy consumption (horizontal axis) and duration (vertical axis). Indeed, the presence of multiple appliance's programs with different cardinalities makes it difficult for the OC-SVM algorithm to find an optimal decision boundary to separate normal instances from the anomalous ones. Furthermore, the decision boundaries vary based on the specific input parameters of the OC-SVM algorithm and must be optimized depending on the specific characteristics of the training set. However, in the real-world those parameters cannot be optimized and must be selected a priori based on precise assumptions on the number of anomalies and the data distribution.

Figure 8 shows the distributions of the reconstruction errors in the five folds of the cross validation procedure for the washing machine (first column), the dishwasher (second column) and the dryer (third column). Each Figure reports the reconstruction errors of both normal (histograms in blue) and anomalous instances (histograms in red), including the anomaly threshold determined from the distribution of the training errors. Notice that the position of the threshold changes across the different folds, showing its adaptability on the specific distribution of the training errors. Most importantly, the histograms reported in Figure 8 demonstrate that the reconstruction error is an excellent criterion to separate normal instances from the anomalous ones, independently of the nature of the original input data. Indeed, the variational autoencoder makes sure that normal instances obtain a low reconstruction error, whereas the errors of abnormal instances exceed the anomaly threshold determined at the end of the training phase.

To conclude, the variational autoencoder achieves a greater generalization performance with respect to the OC-SVM thanks to an improved data representation that facilitates the separation between normal and anomalous instances. In addition, the variational autoencoder is more indicated for a real-world application because it does not require any aprioristic assumption on its parameters, which are instead dynamically determined based on the specific characteristics of the data set at hand.




(a) Fold 1 washing machine (b) Fold 1 dishwasher (c) Fold 1 dryer

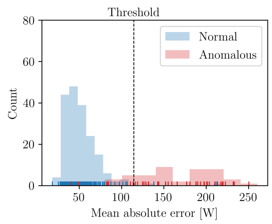
(d) Fold 2 washing machine (e) Fold 2 dishwasher (f) Fold 2 dryer

(g) Fold 3 washing machine (h) Fold 3 dishwasher (i) Fold 3 dryer

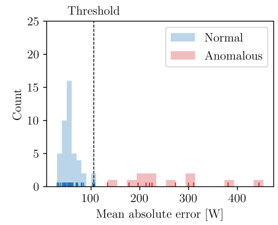
(j) Fold 4 washing machine (k) Fold 4 dishwasher (l) Fold 4 dryer

(m) Fold 5 washing machine (n) Fold 5 dishwasher (o) Fold 5 dryer

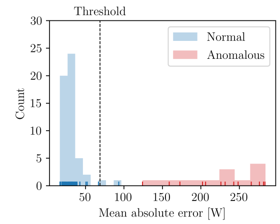
Figure 7: Decision boundaries of the OC-SVM in the five folds of the cross validation procedure for the washing machine (first column), the dishwasher (second column) and the dryer (third column).



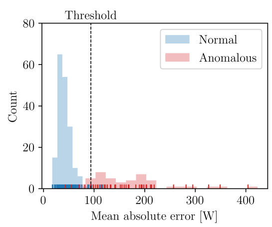

(a) Fold 1 washing machine



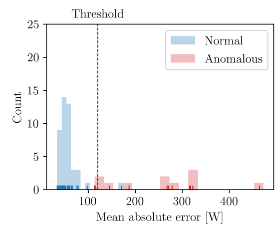

(b) Fold 1 dishwasher



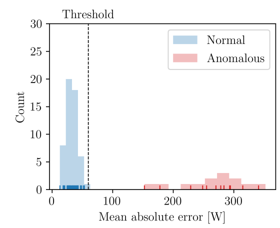

(c) Fold 1 dryer



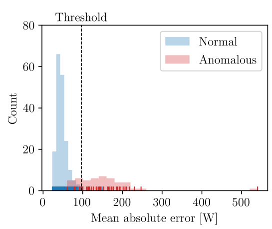

(d) Fold 2 washing machine



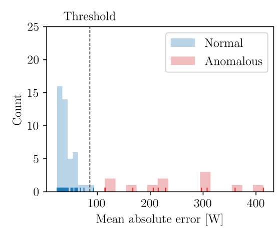

(e) Fold 2 dishwasher



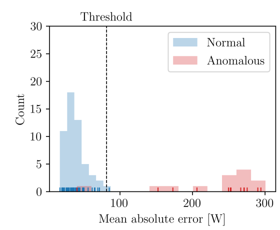

(f) Fold 2 dryer



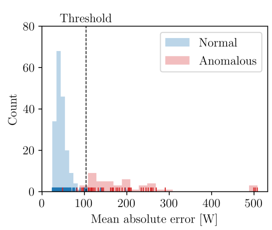

(g) Fold 3 washing machine



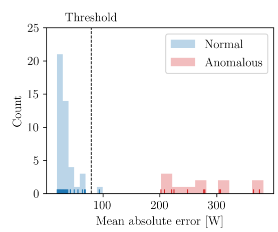

(h) Fold 3 dishwasher



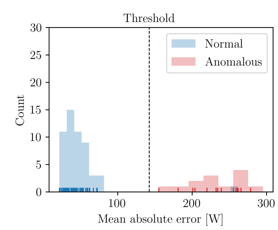

(i) Fold 3 dryer



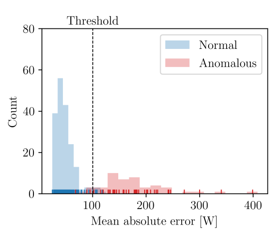

(j) Fold 4 washing machine



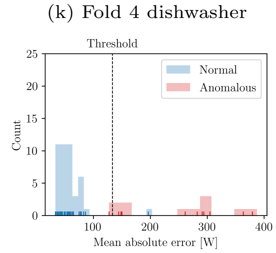

(k) Fold 4 dishwasher



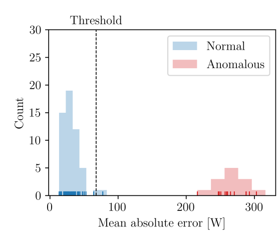

(l) Fold 4 dryer




(m) Fold 5 washing machine




(n) Fold 5 dishwasher




(o) Fold 5 dryer

Figure 8: Distribution of the reconstruction errors in the five folds of the cross validation procedure for the washing machine (first column), the dishwasher (second column) and the dryer (third column).

## 6. Limitations

Similarly to other anomaly detection methods, the proposed approach can suffer from a low representation of the complete set of operational cycles for the monitored device. Indeed, we always assume that the set of normal instances used to train our variational autoencoder already covers the vast majority of operation modes selected by the end-user. If this condition is not respected, then we can encounter a false positive whenever the user selects an appliance's program that was not present in the original training set. The only way to overcome this limitation consists of collecting a sufficiently large training set before deploying our variational autoencoder. In general, a long monitoring period guarantees a good coverage of the full set of operation cycles for a specific appliance, due to the fact that end-users usually present habits in the selection of their programs. It is worth mentioning that the training period for the appliances analyzed in this work is less than six months and test data rarely deviates from the distribution of the training data. This fact demonstrates that with a sufficiently long monitoring period the occurrence of unseen operation cycles represents an extremely uncommon event. Nevertheless, we must also consider that our method can be implemented directly by manufacturing companies to monitor the health status of their appliances and offer maintenance services to their customers. In this second scenario, we can assume that manufacturers already know the complete set of appliance's programs to train the variational autoencoder, thus completely avoiding the occurrence of unseen operation cycles at inference time. To summarize the limitations, we can state that anomaly detection methods generally require a sufficiently long monitoring period to learn the characteristics of normal instances, and the same is true for deep learning-based anomaly detection methods such as the one presented in this work.

## 7. Conclusion

In this study, we presented a new framework for detecting electrical anomalies in the power consumption of household appliances based on the analysis of their power signatures with unsupervised deep learning techniques. The methodology exploits the capabilities of variational autoencoders in order to overcome the limitations of previous methods based on manual features. In fact, deep learning allows us to automatically select the most relevant features to describe the power signatures, thus avoiding the cumbersome task of extracting handcrafted features. In addition, variational autoencoders work in a completely unsupervised fashion, without the need for synthetic anomalous instances that can lead to high generalization errors.

The framework has been tested on three different household appliances with quite dissimilar power signatures, which are the dishwasher, the washing machine and the clothes' dryer. The results show very good classification performance on all the three use cases, reporting an F1-score greater than 90% on average. The proposed method outperformed the one-class support vector machine trained with classical features such as the total energy consumption and the duration of the operation cycles, showing that deep learning can better characterize the set of normal instances with respect to traditional anomaly detection methods.

The detection of anomalous appliances can significantly improve the estimation of the end users' power profiles, allowing grid operators to perform a more accurate scheduling of the energy distribution. Furthermore, the end-users can easily save energy by repairing or substituting faulty appliances that cause an increment in their overall energy demand. Finally, manufacturing companies can harness smart appliances with anomaly detection capabilities in order to automatically notify anomalies to maintainers and trigger their interventions when necessary.

As a future work, we want to apply the proposed method to a set of realistic electrical anomalies observed in true faulty appliances. Alternatively, we will also extend this work to analyze behavioural anomalies affecting the usage patterns of the end-users.

## Acknowledgement

This research work has been funded by the European Commission through the EU H2020 research and innovation programme, ECSEL Joint Undertaking, and National Funding Authorities from 18 countries under the Arrowhead Tools research project with Grant Agreement no. 826452.

## References

Alsalemi, A., Himeur, Y., Bensaali, F., & Amira, A. (2020). Appliance-level monitoring with micro-moment smart plugs. *arXiv preprint arXiv:2012.05787*, .

Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., & Santini, S. (2014). The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys 2014). Memphis, TN, USA* (pp. 80–89). ACM.

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, *5*, 1089–1105.

Benninger, M., Hofmann, M., & Liebschner, M. (2020). Transferable and self-learning online monitoring system for electrical household appliances. *Procedia Computer Science*, *176*, 148–155.

Castangia, M., Aliberti, A., Bottaccioli, L., Macii, E., & Patti, E. (2021a). A compound of feature selection techniques to improve solar radiation forecasting. *Expert Systems with Applications*, *178*, 114979.

Castangia, M., Sappa, R., Girmay, A. A., Camarda, C., Macii, E., & Patti, E. (2021b). Detection of anomalies in household appliances from disaggregated load consumption. In *2021 International Conference on Smart Energy Systems and Technologies (SEST)* (pp. 1–6). IEEE.

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, .

Chollet, F. et al. (2015). Keras. `https://keras.io`.

COP26 (2021). Un climate change conference uk 2021. `https://ukcop26.org/`. Accessed: 2021-12-01.

Gonzalez, D., Patricio, M. A., Berlanga, A., & Molina, J. M. (2021). Variational autoencoders for anomaly detection in the behaviour of the elderly using electricity consumption data. *Expert Systems*, (p. e12744).

Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2020). A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. *Cognitive Computation*, *12*, 1381–1401.

Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., & Amira, A. (2021). Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*, *287*, 116601.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*, 504–507.

Hosseini, S. S., Agbossou, K., Kelouwani, S., Cardenas, A., & Henao, N. (2020). A practical approach to residential appliances on-line anomaly detection: A case study of standard and smart refrigerators. *IEEE Access*, *8*, 57905–57922.

Issi, F., & Kaplan, O. (2018). The determination of load profiles and power consumptions of home appliances. *Energies*, *11*, 607.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, .

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining* (pp. 413–422). IEEE.

Mao, J., Zhu, S., Bian, J., Lin, Q., & Liu, J. (2019). Anomalous power-usage behavior detection from smart home wireless communications. *Journal of Communications and Information Networks*, *4*, 13–23.

Massidda, L., Marrocu, M., & Manca, S. (2020). Non-intrusive load disaggregation by convolutional neural network and multilabel classification. *Applied Sciences*, *10*, 1454.

Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, *6*, 91–99.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, .

Patricio, M. A., González, D., Molina, J. M., & Berlanga, A. (2019). Analysis of the consumption of household appliances for the detection of anomalies in the behaviour of older people. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 60–68). Springer.

Rashid, H., Singh, P., Stankovic, V., & Stankovic, L. (2019). Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour? *Applied energy*, *238*, 796–805.

Ritchie, H., & Roser, M. (2020). Energy. *Our World in Data*, . Https://ourworldindata.org/energy.

Siano, P. (2014). Demand response and smart grids—a survey. *Renewable and sustainable energy reviews*, *30*, 461–478.

Yin, S., Zhu, X., & Jing, C. (2014). Fault detection based on a robust one class support vector machine. *Neurocomputing*, *145*, 263–268.