Monitoring domestic water consumption: a comparative study of model-based and data-driven end-use disaggregation methods

Pavlos V. Pavlou [a,*], Stylianos Filippou [a], Solon Solonos[a], Stelios G. Vrachimis [a], Kleanthis Malialis [a], Demetrios G. Eliades [a], Theocharis Theocarides [a,b] and Marios M. Polycarpou [a,b]

[a] KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus
[b] Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus
*Corresponding author. E-mail: pavlou.v.pavlos@ucy.ac.cy

PVP, 0000-0001-8740-4187; SF, 0000-0001-8123-8309; SGV, 0000-0001-8862-5205; KM, 0000-0003-3432-7434; DGE, 0000-0001-6184-6366; TT, 0000-0001-7222-9152; MMP, 0000-0001-6495-9171

ABSTRACT

Monitoring the water usage of different appliances and informing consumers about it has been shown to have an impact on their behavior toward drinking water conservation. The most practical and cost-effective way to accomplish this is through a non-intrusive approach, that locally analyzes data received from a flow sensor at the main water supply pipe of a household. In this work, we present two different methods addressing the challenges of disaggregating end-use consumption and classifying consumption events. The first method is model-based (MB) and uses a combination of dynamic time wrapping and statistical bounds to analyze four water end-use characteristics. The second, learning-based (LB) method is data-driven and formulates the problem as a time-series classification problem without relying on a priori identification of events. We perform an extensive computational study that includes a comparison between an MB and an LB method, as well as an experimental study to demonstrate the application of the LB method on an edge computing device. Both methods achieve similar F1 scores (LB: 71.73%, MB: 71.04%) with the LB being more precise. The embedded LB method achieves a slightly higher score (72.01%) while enhancing on-site real-time processing, improving security and privacy and enabling cost savings.

Key words: edge computing, multi-label classification, non-intrusive water usage classification, time-series classification

HIGHLIGHTS

- Application and comparison of model-based and data-driven approaches to make predictions based on aggregated water consumption information.
- Classification accuracy per water end-use category depends on the classification approach.
- Real-time implementation of a neural network-based approach on an edge computing device.

1. INTRODUCTION

The constantly increasing gap between water demand and supply has been one of the most important challenges that our world faces (Cosgrove & Loucks 2015). Freshwater availability is depleting quickly due to population growth and climate change putting even more burden on water utilities which must ensure continuous water supply and sufficient pressure during peak times to households (McDonald *et al.* 2014). As a response to these challenges, special emphasis is given to the application of water demand management practices aiming to reduce water use as well as the associated treatment and transportation costs, and the corresponding environmental impact (Inman & Jeffrey 2006).

New investments in data and information technologies enable the collection and transmission of high-resolution data to both utilities and consumers through smart water meters (Di Nardo *et al.* 2021). Smart metering can help improve water systems management by providing insights into water usage patterns (Mazzoni *et al.* 2023) and linking them with socioeconomic characteristics (Steffelbauer *et al.* 2021). Specifically, this information can be used by water utilities to better manage demand during peak hours, thus eliminating the need for further investment to upgrade the existing water infrastructure. In addition,

monitoring how, when, and where household water consumption is used has been shown to have an impact on people's behavior toward water conservation (Cominola *et al.* 2021).

Monitoring household water consumption can be performed with an intrusive approach that considers the direct metering in each water-consuming appliance (e.g., dishwasher, toilet, and shower) or with a non-intrusive approach that considers only the measuring of the total household consumption. Although intrusive monitoring offers more insight into consumer habits, the installation of multiple sensors to monitor each appliance may have a high initial cost and may be inapplicable due to practical considerations. On the other hand, non-intrusive monitoring must be coupled with data analytics and intelligent algorithms to disaggregate the total inflow to different end-uses. Identifying which appliances are in use through non-intrusive water usage classification is challenging since their operation may be overlapping, while specific appliances may operate with intermittent flow making individual consumption events hard to distinguish.

Furthermore, the application of most existing water monitoring methods relies on a combination of sensors, software packages, and cloud-based analysis. These methods can be time consuming and computationally expensive. A possible solution can be given through the application of IoT-edge devices that directly collect and analyze data thus fostering decision-making. In the context of household water monitoring, edge computing devices with machine learning capabilities can be used to identify and analyze water usage patterns, detect devices with excessive water consumption and water usage anomalies, such as leaks, enhance predictive maintenance, and reduce associated costs.

1.1. Literature review

In the following paragraphs, we provide background information on the topic of water usage classification from two different research areas: (1) water-use disaggregation methods developed mainly by researchers in the area of smart water systems, and (2) time-series classification mainly developed by researchers in the area of machine learning.

1.1.1. Water end-use disaggregation methods

Various studies have been proposed to address the challenge of non-intrusive water end-use disaggregation and classification. One of the first disaggregation methods, namely the *Trace Wizard*, was developed based on the decision tree methodology that considers the volume, duration, and flow rate of each end-use event (DeOreo *et al.* 1996; Mayer *et al.* 1999). This method requires a significant amount of data and human intervention for data processing. A similar method was also used for the development of the *Identiflow* software (Kowalski & Marshallsay 2003) while an updated rule-based methodology for automated disaggregation was presented in a more recent study (Mazzoni *et al.* 2021). A different disaggregation tool, namely *Hydrosense*, relies on the use of pressure sensors combined with a Bayesian methodology (Froehlich *et al.* 2009). This approach, however, requires a high initial cost for the deployment of the sensor network. Machine learning and data analytics algorithms were further developed to address the challenge of water end-use disaggregation, with promising results (Pastor-Jabaloyes *et al.* 2018; Meyer *et al.* 2021). In a series of studies, the disaggregation of both single and combined events was addressed by the development of the *Autoflow* model which uses a combination of methods including the Hidden Markov Model (HMM), the dynamic time wrapping (DTW) algorithm, gradient vector filtering, and artificial neural networks (ANN) (Nguyen *et al.* 2013, 2015, 2018). As stressed by the authors, drawbacks that were noted in these studies include the need for a large amount of historical data to train the models, the need for regional data to calibrate and apply the models in areas with different water consumption habits, the need for high-resolution data, and the absence of disaggregation techniques for combined water events.

1.1.2. Time-series classification

Water consumption classification can be formulated as a *time-series classification problem*. Time-series classification is a type of supervised machine learning problem, where time-series data are described by a class label. The difference with other classification problems is that the natural temporal order in the data is significant, and a learning algorithm has to identify and exploit these temporal characteristics. Different types of ANN have been used, e.g., multilayer perception (MLP) models were designed to learn discriminative times series classifiers (Geng & Luo 2019), convolutional neural networks (CNNs) were utilized for the detection of myocardial infractions from electrocardiography data (Strodthoff & Strodthoff 2019), and for risk prediction based on patients' historical medical records (Che *et al.* 2017), and echo state networks for time-series prediction in wireless communication channels were developed by Jaeger & Haas (2004). The RandOm Convolutional KErnel Transform (ROCKET), proposed by Dempster *et al.* (2020), uses a large number of

random convolutional kernels in conjunction with a linear classifier. Moreover, deep learning approaches were deployed to handle multivariate problems; for example, InceptionTime achieves high accuracy through a combination of building on a residual neural network (ResNet) to incorporate Inception modules and ensembling over five multiple random-initial-weight instantiations of the network for greater stability (Ismail Fawaz *et al.* 2020). InceptionTime was trained and tested on a synthetic dataset generated by the authors. Moreover, the TapNet 'few-shot' classification method combines the advantages of both traditional and deep learning approaches and produces a network architecture that can be broken down into three distinct modules (Zhang *et al.* 2020): random dimension permutation, multivariate time-series encoding and attentional prototype learning. This method was tested on real-world multivariate time-series data collected from different applications such as human activity recognition, motion classification, and electrocardiographic/electroen-cephalographic (ECG/EEG) signal classification. Other classification approaches have been employed in addressing issues related to water and wastewater. Hybrid models including wavelet-gene expression programming (WGEP), wavelet-model tree (WMT), and wavelet-evolutionary polynomial regression (WEPR) (Najafzadeh & Zeinolabedini 2018) and ANN including feed-forward back propagation neural network (FFBP-NN) and radial basis function neural network (RBF-NN) (Zeinolabedini & Najafzadeh 2019) were utilized to estimate the daily quantity of sewage sludge. ANN, adaptive neuro-fuzzy interference system (ANFIS) and support vector machine (SVM) were further used in predicting wastewater treatment plant inflow rates (Najafzadeh & Zeinolabedini 2019). Machine learning methods utilizing remote sensing data were further considered to evaluate flood risk using random forest (RF) (Farhadi & Najafzadeh 2021) and river water quality using M5 model tree (MT), multivariate adaptive regression spline (MARS), gene expression programming (GEP), and evolutionary polynomial regression (EPR) (Najafzadeh & Basirian 2023).

1.2. Contributions

The contributions made by this paper are as follows:

1. **Comparison of methodologies**: This paper evaluates and compares two different methodologies for end-use water consumption monitoring using smart meters: the first approach uses an optimization-based event detection to identify individual events and models of events, and the second approach uses a machine learning-based time-series classification method. For the second approach, four different algorithms, namely, SVM, RF, extreme gradient boosting (XGBoost) and MLP, were initially compared.
2. **Real-time implementation on an edge device**: Our second contribution is the implementation of the most efficient machine learning method on an edge computing device that can autonomously analyze and process data on-site, thus demonstrating its real-time applicability for non-intrusive water end-use disaggregation and classification. We transform our model using TensorFlow Lite, and we install it on a microcontroller.
3. **New benchmark dataset**: To objectively compare and assess the drawbacks of each method we introduce a new open benchmark dataset, derived by the STochastic Residential water End-use Model (STREaM) simulator (Cominola *et al.* 2018). The generated dataset is available to be used by the research community in assessing the performance of algorithms and models used for non-intrusive water usage classification. The link to the dataset is provided toward the end of the paper in the Section 'DATA AVAILABILITY STATEMENT'.

2. METHODOLOGY

In this section, we describe the two methods for the disaggregation and classification of water end-uses. The model-based (MB) method relies on detecting the entire usage of an event and the learning-based (LB) method classifies event samples within the dataset without the need of extracting the entire event.

2.1. MB method

The MB method, provided by Pavlou *et al.* (2022), consists of a method that classifies single events and a method that disaggregates combined events into single events (Figure 1).

The two main MB stages of the disaggregation and classification process are as follows:

1. *The offline feature learning stage* analyses the training dataset consisting of labeled data to calculate the statistical bounds of three predefined event features, namely, the event duration, event volume and event flow peak. The algorithm begins by creating event sets from labeled data by extracting observed time-series data with event labels and organizing them into the
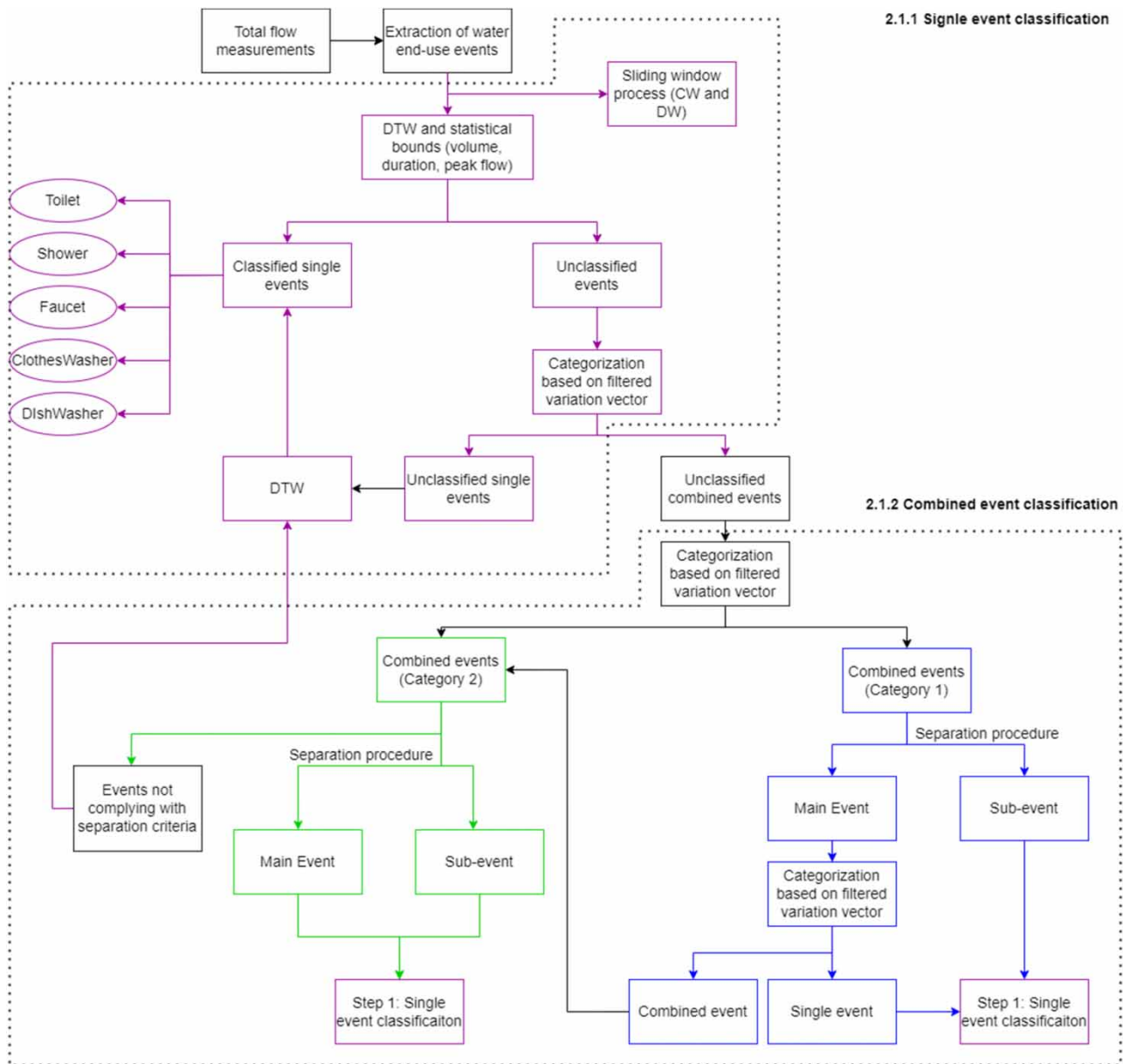
Figure 1 | Water end-use event classification process.

set of events E. Events with the same label l are grouped to form subsets of events $E_l \subset E$. Event signatures (i.e., typical consumption patterns) are not extracted nor analyzed in this stage. The classification process uses the water end-use signatures stored in the data model (see Section 3.1).

2. The *event classification stage* distinguishes individual events in the time series by detecting zero-flow intervals and then processes each event through the single and combined event classification. The classification utilizes the DTW method to analyze each event's consumption pattern and an optimization procedure that uses similarity indices and statistical bounds extracted from *the offline feature learning stage*. The classification of events characterized by intermittent flow, such as dishwashers (DW) and clothes washers (CW), undergoes additional processing. This involves incorporating a time window in the time-series analysis that considers the entire device cycle. A filtered variation technique, that detects flow-rate changes within an event, is used to identify combined events and separate them into single events. Combined events are two or more single events whose occurrence time overlaps.

2.1.1. Single-event classification

A similarity matrix $M_{XY} = W(\tilde{S}_X, \tilde{S}_Y)$ is created, where $\tilde{S}_X$ is the number of investigated events, and $\tilde{S}_Y$ the number of water end-use categories. Both the investigated and labeled events are initially normalized, thus being invariant to scale and offset, as follows:

$$\tilde{S}_t(t) = \frac{S_t(t) - m}{\sigma} \tag{1}$$

$$m = \frac{\sum_{t=1}^{n} S_t}{n} \tag{2}$$

$$\sigma = \sqrt{\frac{\sum_{t=1}^{n} (S_t - m)^2}{n}} \tag{3}$$

where m and σ are the arithmetic mean and standard deviation, respectively.

The highest shape similarity index, W, between the signal of the examined event and each category of labeled water end-use signatures is calculated using the DTW methodology. DTW wraps the time axis to align the data points and calculates the optimal alignment between two time-series according to the following equation:

$$W(i, j) = w(i, j) + \min \{W(i - 1, j), W(i - 1, j - 1), W(i, j - 1)\} \tag{4}$$

where $s = \{s_1, .., s_i, .., s_m\}$ and $t = \{t_1, .., t_j, .., t_n\}$ are the two time series with m and n data points, respectively. Distance metric w is given by $w(i, j) = |s_i - t_j|$ with the possible combinations limited to $(i - 1, j)$, $(i - 1, j - 1)$, $(i, j - 1)$. The accumulated DTW distance $W(m, n)$ is considered the optimal alignment between the two time-series, with the initial condition $W(1, 1) = w(1, 1)$.

The proposed methodology, firstly, detects the operation of intermittent flow devices such as dishwasher (DW) and clothes washer (CW) by applying the DTW approach within a sliding time window with a length equal to the full cycle of operation of the selected appliance. Then it disaggregates the sub-events of these intermittent flow devices within the specified time windows. Following, the DTW algorithm classifies the toilet, shower, and faucet events. A screening procedure takes place to filter out events with volume, duration, and peak flow rate outside the predefined minimum and maximum bounds obtained during the offline learning stage. Events that do not comply with these criteria are marked as unclassified.

A filtering technique is used to categorize unclassified events into either single or combined events. The methodology relies on detecting flow-rate changes that are above a specific threshold. Changes within the flow signal are a strong indication that more than one water-end appliance is active. A vector with the flow-rate changes is calculated as $v_i = f_{i+1} - f_i$, $1 \le i < n$ where $f = (f_1, f_2, .., f_i, .., f_n)$ denotes the event flow-rate points with a duration of n seconds and $v = (v_1, v_2, .., v_i, .., v_{n-1})$ the extracted vector (Figure 2(b)). Single events are distinguished from combined events because they exhibit no fluctuations in the extracted flow-rate variation vector and are classified using only the DTW approach without considering the statistical bounds from the analysis of their features. The remaining unlabeled events are considered combined events.

2.1.2. Combined event classification:

The combined events are categorized into two types: The first type refers to a combined event that includes (a) at least one sub-event that starts during the time another appliance is active and finishes afterward, or (b) at least one sub-event that starts first and finishes during the time of operation of another sub-event. The second type considers sub-events that start and finish within other sub-events.

The identification and separation of the combined events are performed using the flow-rate variation vector described previously. For first-type combined events, the algorithm searches within the starting and ending phases of the variation vector to identify flow-rate rises and flow-rate drops with a similar length that corresponds to the same event. For second-type combined events the algorithm searches within the variation vector to identify the positions where a zero value is followed by a positive value and the positions where a negative value is followed by a zero value (Figure 2(b)). These positions indicate
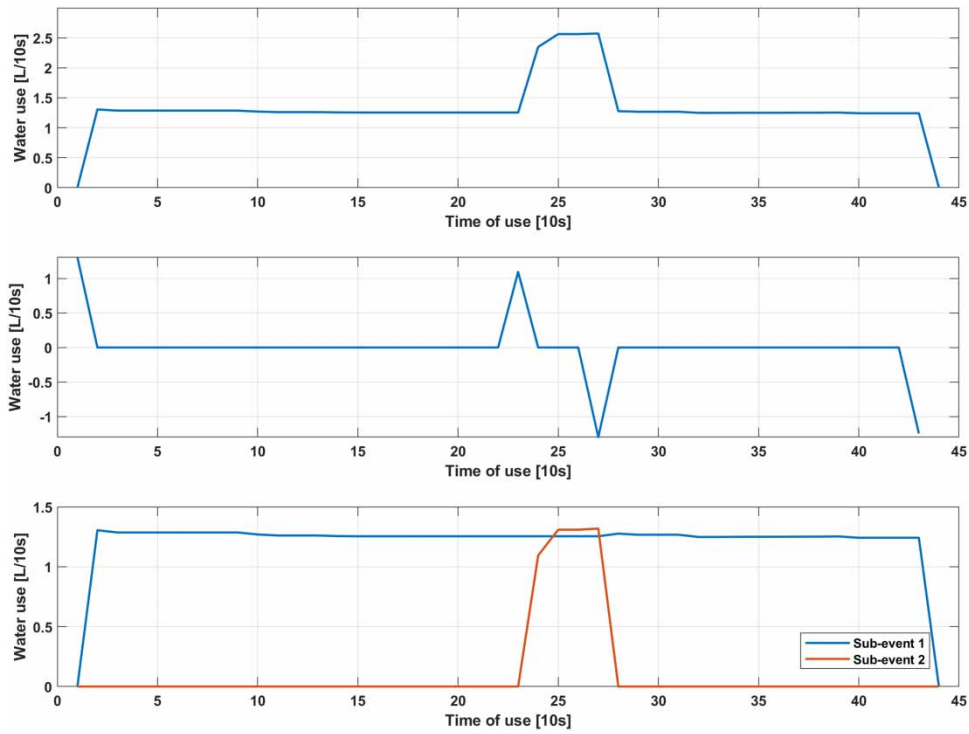
Figure 2 | (a) Combined event as extracted from the dataset, (b) flow-rate variation vector of the combined event, and (c) sub-events extracted from the original event.

the beginning and end of a sub-event within another event. The events extracted through these two separation processes are then labeled using the single-event classification approach. Any events not classified are considered as combined events and thus are processed again through the combined event classification procedure until they are separated into single events. This case refers to a situation where more than two events overlap at the same time.

2.2. LB method

We adopt the learning-based method of Papatheodoulou *et al.* (2022) for our comparative study in this paper. A data generating process provides at each time step t a sequence of instances $S = \{(x^t, y^t)\}_{t=1}^{T}$ from an unknown probability function $p^t(x, y)$, where $T \in [1, \infty)$. The input $x^t = \{x_i^t\}_{i=1}^{d} \in \mathbb{R}^d$ is a d-dimensional vector belonging to input space $X \subset \mathbb{R}^d$. The instances constitute a multivariate time series with d number of time series, and each corresponds to a univariate time series defined as $z_i = \{x_i^t\}_{t=1}^{T} \in \mathbb{R}^T$. The label (i.e., the ground truth) of the classification task is denoted by $y^t \in Y$. When $Y = \{1, \ldots, C\}, C \geq 2$, it is termed multi-class classification, that is, it refers to a task with more than two classes. When $Y \in \{0, 1\}^C$ it is termed multi-label classification, i.e., it assigns to each instance a set of labels. Each digit corresponds to the inclusion (1) or absence (0) of the relevant label.

A classifier f receives a new example x^t at time step t and makes a prediction $\hat{y}^t$ based on a concept $f:X \to Y$ such that $\hat{y}^t = f(x^t)$. We refer to this as time-series classification. To capture the temporal nature of the data, we introduce a memory element, e.g., a sliding window to aid with the prediction task, i.e., $\hat{y}^t = f(x^t, x^{t-1}, \ldots, x^{t-W})$, where W is the window size. Let us define the aggregated sequence $x^{*t} = \{x_1^t \otimes x_2^t \otimes \ldots \otimes x_1^t\}$, where the classifier makes a prediction based solely on the aggregated sequence x^{*t} at time step t. The aggregation operator $\otimes$ depends on the application. We refer to this as aggregated time-series classification.

We formulate the non-intrusive water end-use monitoring problem as an aggregated time-series classification task. For the generation of the dataset, we use a residential water demand simulator (see section 3.1) that synthesizes a sequence of instances S at each time step t based on the water consumption profiles of appliances in U.S. households.

The generated time series consists of five signals, that correspond to the water flow consumption of the toilet (z_1), shower (z_2), faucet (z_3), clothes washer (z_4), and dishwasher (z_5). From the consumption of these five appliances, we compose an aggregated sequence x^{*t}, by summing the water flow consumption from each sequence z_i at each time step t. The aggregated sequence x^{*t} is described by a set of binary labels that correspond to the appliances that were active (1) or inactive (0) at each time step t. To improve the efficacy of the predictions, we use a sliding window approach that allows the classifier to capture information from previous time steps, thus extracting underlying temporal patterns.

2.3. Evaluation methodology

We analyze the results of the classifier using a confusion matrix. The confusion matrix classifies the results into true positives (T_P), true negatives (T_N), false positives (F_P), and false negatives (F_N). Classifiers are typically evaluated using the following accuracy metric given by:

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \tag{5}$$

However, the accuracy metric becomes unsuitable as it is biased toward the majority (normal) class. A widely accepted metric which is less sensitive to imbalance is the F1-score, defined below as the harmonic mean of the model's precision and recall (He & Garcia 2009).

The micro-averaging F1-score or F1-micro is given by:

$$F1 = 2 \cdot \left(\frac{P \cdot R}{P + R} \right) \tag{6}$$

Precision (P) indicates the percentage of true positive indices among the total number of positive indices classified by the model:

$$P = \frac{T_P}{T_P + F_P} \tag{7}$$

and Recall (R) measures the amount of correctly labeled positive cases among the total number of positive cases:

$$R = \frac{T_P}{T_P + F_N} \tag{8}$$

We further evaluate our models using the Cohen's Kappa score and ROC-AUC. Both methods were previously used to evaluate the performance of predictive models (Farhadi & Najafzadeh 2021; Najafzadeh & Basirian 2023). Cohen's Kappa score helps in determining the extent to which the classifiers' decisions agree beyond what could be attributed to random chance (Cohen 1960). It can be calculated by:

$$\text{Cohen's Kappa} = \frac{P_0 - P_e}{1 - P_e} \tag{9}$$

where,

$$P_o = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \tag{10}$$

$$P_e = \left(\frac{T_P + F_P}{T_P + F_P + T_N + F_N} \right) * \left(\frac{T_P + F_N}{T_P + F_P + T_N + F_N} \right) + \left(\frac{T_N + F_P}{T_P + F_P + T_N + F_N} \right) * \left(\frac{T_N + F_N}{T_P + F_P + T_N + F_N} \right) \tag{11}$$

ROC-AUC (Bradley 1997) is a single value that quantifies the area under the receiver operating characteristic curve, which is a graphical plot that illustrates the trade-off between recall and the false positive rate (1-specificity) for various threshold

values. Where specificity (S) is defined by:

$$S = \frac{T_N}{T_N - F_P} \tag{12}$$

3. RESULTS AND DISCUSSION

3.1. Benchmark dataset

This study uses the STochastic Residential water End-use Model (STREaM) (Cominola *et al.* 2018), a modeling software developed to generate synthetic time-series data of a household with a resolution of up to 10 s. STREaM generates a time series of each fixture's water end-use, characterized by its signature, probability distributions of the number of uses per day, single-use durations, water demand contribution, and time of use during the day. STREaM takes into consideration the number of house occupants in the calculation of total household water demand. It was calibrated on a large dataset including observed and disaggregated water end-uses from over 300 single-family households in nine U.S. cities (DeOreo 2011). The following water end-uses were considered in this dataset: toilet, shower, faucet, clothes washer, and dishwasher. The end-uses are further distinguished in standard and high-efficiency appliances, which have different consumption characteristics. The dataset provides the water flow reading at each time step for each fixture and their sum as the total consumption. Figure 3 illustrates the connectivity of a domestic water system. The main flow meter is located at the outflow of a domestic house water tank, thus allowing us to measure water flow for each time step. The system consists of the main water tank, the main water pipe connecting the tank to the manifold, the manifold which distributes the water to the house piping system and the end-use appliances.

The dataset used in this study considers the use of a standard toilet, standard shower, standard faucet, high-efficiency clothes washer, and standard dishwasher in a 2-person household for 180 days (6 months) and it has a resolution of 10 s. Figure 4 depicts the household water consumption for each appliance and the total for a random day. Figure 5 displays the size of each class (i.e., time-series records), to illustrate class imbalance, where no instances indicate that none of the appliances is in use. Class imbalance (He & Garcia 2009) refers to a situation where there is an unequal number of samples in each class, with one or more classes being the majority within the data. There are several implications that affect model performance such as biased predictions in favor of the majority class and poor generalization.

We split the dataset into two main sets: 4.5 months' worth of data are held for training and validation and one set of 1.5 months of data is reserved as a test set. The training and validation subsets correspond to 3- and 1.5-month data, respectively. That equates to a set of 777,600 samples for the training and 388,800 samples for both validation and test sets. The training subset consists of the samples that are given to the model, to identify and learn any underlying patterns of the data. The
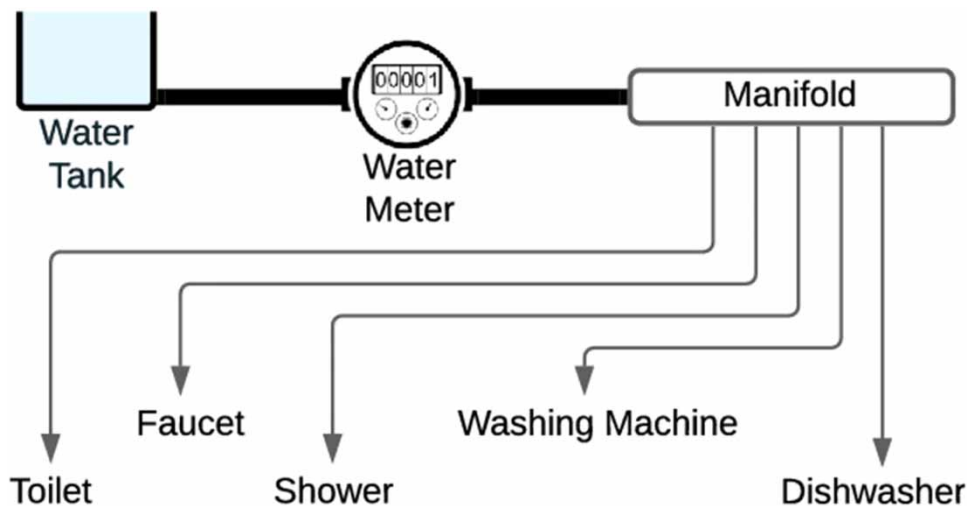

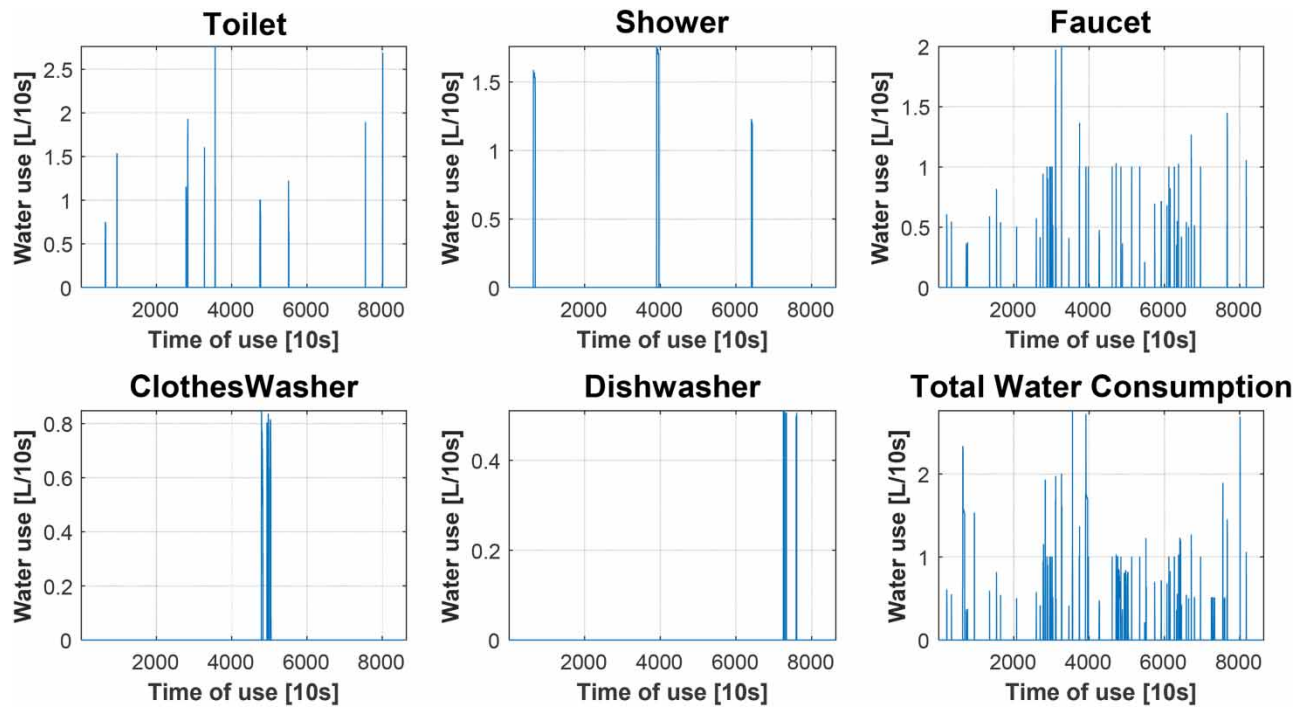**Figure 3** | Piping connectivity in a domestic water system.

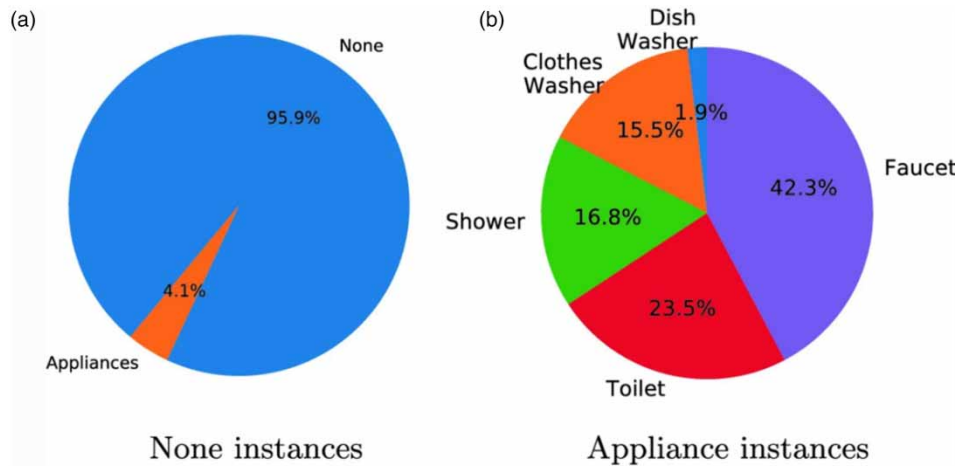**Figure 4** | Daily household water consumption per appliance.

Figure 5 | Dataset's class sizes.

validation subset contains data that are used for evaluation purposes to optimize the model. For the MB method, the training and validation subsets were used as one single training set. The test subset is of unseen samples that are used only to assess the performance of the algorithms, to determine how well the algorithms can generalize on unseen data. The link to the dataset is provided toward the end of the paper in the Section 'DATA AVAILABILITY STATEMENT'.

3.2. Comparative study

3.2.1. MB method

As previously described, the MB method relies on detecting the entire usage time when a water appliance is active and examining its usage features. The three end-use characteristics of the events included in the training set, namely duration, volume,

and peak flow, are analyzed during the *offline feature learning stage* and the corresponding 95% confidence intervals of each water end-use category were calculated (Table 1). The 95% range of confidence intervals is chosen as the optimal range since it resulted in the highest classification accuracy with the optimum number of false positive and true positive predictions simultaneously. For the intermittent flow devices, the DW and the CW, the statistical analysis considers the event characteristics of every single event that is included in the full cycle of operation.

According to the analysis, the toilet, faucet, and DW events have similar event characteristics, specifically for consumption duration and peak flow. Similarly, the calculated event volume bounds are identical as well, although the DW and faucets can generate events with less volume than toilets and faucets. The shower events exhibit longer durations and larger consumption volumes than the rest of the events which can play a significant role during the identification process. The extracted confidence intervals for the CW category indicate similarity with toilet, faucet, and DW events, although events with higher consumption duration, volume and flow rate can occur. In general, the end-use characteristics for most of the water appliances are not very distinctive; this may be due to the data resolution of 10 s. Datasets with higher resolution can provide more information regarding water end-use characteristics and have been more useful for disaggregation purposes (Pavlou *et al.* 2022). This highlights the importance of using the DTW algorithm which enhances the classification methodology through signal analysis and pattern recognition.

3.2.2. Learning-based method

The following four algorithms are considered for the LB method: SVM, RF, extreme gradient boosting (XGBoost), and MLP. SVM is a machine learning algorithm that separates data points of multiple classes in a high-dimensional feature space by finding the optimal hyperplane (Cortes & Vapnik 1995). RF is a tree-based, ensemble learning algorithm, i.e., it depends on multiple tree-based learners which make individual predictions that are then averaged together (Bishop & Nasrabadi 2006). Typically, the more trees it has, the more robust the model is as its performance does not rely on a single tree. XGBoost is a machine-learning technique that produces a prediction model in the form of an ensemble of weak prediction models, which are typically tree-based (Chen & Guestrin 2016). This technique builds a model in a stage-wise fashion and combines weak learners into a single strong learner. As each weak learner is added, a new model is fitted to provide a more accurate estimation. The XGBoost classifier is a tree-based ensemble machine learning algorithm with Gradient Boosting as its main component. Moreover, XGBoost can handle missing values on its own and it is very effective and efficient in terms of performance as well as training time even on large datasets. MLP is a feed-forward neural network that consists of an input and an output layer and can have multiple hidden layers (Bishop & Nasrabadi 2006). MLP uses the backpropagation algorithm for training which computes the gradient of the loss function with respect to the weights of the neural network.

We have tuned all the algorithms, using the classifier chain (CC) method and a window size equal to 120 time steps. CC is capable of exploiting correlations among target variables. In a multi-label classification setting with N-classes, N-binary classifiers are assigned a number that corresponds to their order in the classifier's chain. The training process follows the order of the models in the chain, where each binary classifier is fit on the available training data with the addition of the actual target labels of the classes whose models were assigned a lower order in the chain. Table 2 presents the performance of each classifier, with the XGBoost and MLP achieving a higher and very similar performance compared to SVM and RF. Specifically, XGBoost and MLP achieved an *F*-score of 71.98 and 71.73%, respectively, whereas RF achieved 55.75% and SVM achieved 65.29%. We opted to proceed using the MLP model since the performance is close to the XGBoost model and it is more applicable for integration on a microprocessor.

Then, we investigate three different sliding window sizes on the multilayer perceptron (MLP) model that capture the previous 60, 120, and 240 time steps that correspond to 10-, 20-, and 40-min intervals respectively. Table 3 shows the MLP

Table 1 | 95% confidence intervals obtained for the water end-use features: volume, duration, peak flow

	Toilet	Shower	Faucet	CW	DW
Duration (s)	10–130	70–890	10–100	10–330	10–140
Volume (L)	4–13	9–117	1–7	0.44–28.26	0.09–7.60
Peak flow (L/10 s)	0.4–3.23	0.59–1.95	0.23–1.46	0.50–2.35	0.39–1.60

Table 2 | Comparison of different classifiers

	MLP	XGBoost	RF	SVM
Accuracy (%)	98.89	98.78	96.51	98.34
F1-Micro (%)	71.73	71.98	55.75	65.29

Table 3 | Performance of MLP (CC) with different window sizes

	Window 60	Window 120	Window 240
Accuracy (%)	98.65	98.89	98.7
F1-Micro (%)	70.34	71.73	69.9

performance utilizing different window sizes. The best performance with an F1-score of 71.73% is obtained with a window size of 120 time steps.

3.2.3. Comparison of two methods

The results obtained from the two disaggregation and classification methods are presented in Table 4. The results were extracted by analyzing the T_P, T_N, F_P, and F_N samples of each water end-use fixture. The results from the MB method were initially calculated in terms of the number of events and then converted into the number of samples per event to be comparable with the LB method. A sample corresponds to a single data point. An event includes all the consecutive data points that correspond to the operation of a device. Confusion matrices were further created per device per approach and presented in Figures 6 and 7. For the LB method, the confusion matrices were obtained using the MLP (CC) with a window size of 120 time steps. Figure 8 demonstrates the confusion matrices of the precision of each method including the misclassification percentage per device. The confusion matrices are computed in a class-wise fashion where the multi-class data are treated as if they were binarized under a one-versus-rest transformation. The y-axis includes the actual class labels and the x-axis shows the algorithm's predictions.

   Overall, both methods reach approximately the same total F1 classification score (LB: 71.73% and MB: 71.04%) and Cohen's Kappa score (LB: 71.62% and MB: 70.82%). The MB method was able to identify more event samples than the LB method (LB: 66.83% and MB: 70.21%) demonstrating, however, a lower precision (LB: 77.61% and MB: 71.88%). More specifically, each method demonstrates a different level of efficiency in detecting each water end-use according to the presented results. The MB method is slightly better at classifying shower and CW events while the LB method can detect faucet events more accurately. Both methods exhibit approximately the same level of effectiveness in identifying toilet events and a low score for DW classification. Moreover, we compare the two methods by calculating the ROC-AUC

Table 4 | Classification results obtained using the Recall, Precision, F1-score and Cohen's Kappa metrics for the model-based and learning-based methods

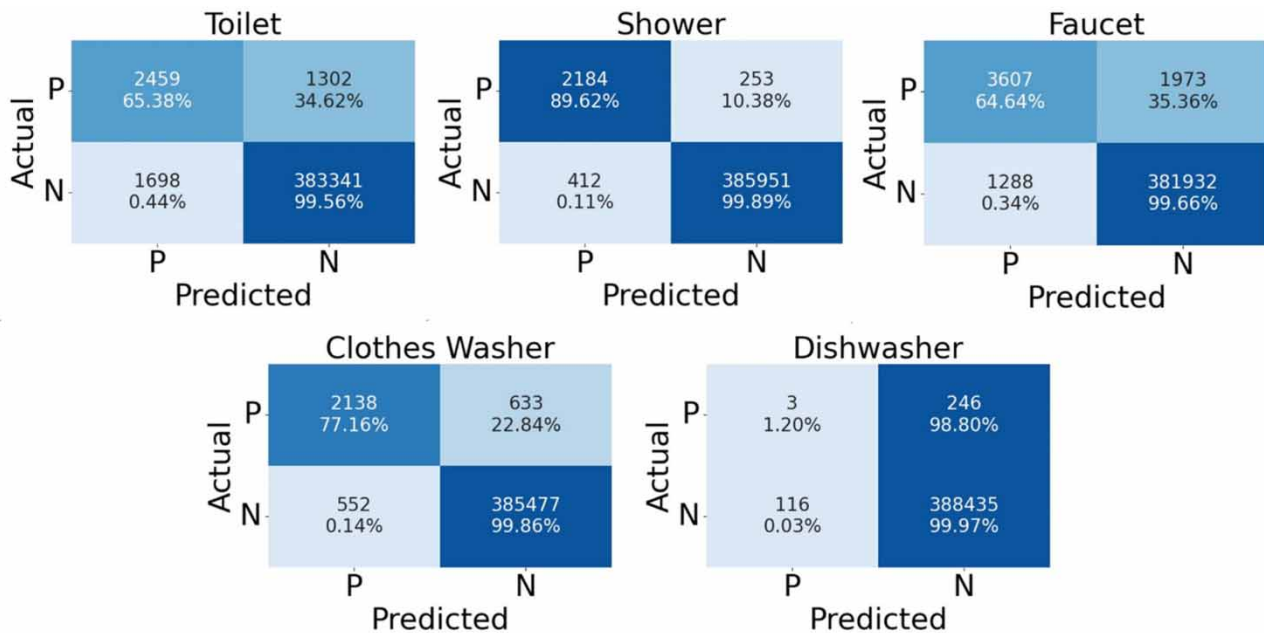|	Learning-based (LB)				Model-based (MB)			
	Recall (%)	Precision (%)	F1-score (%)	Cohen's Kappa (%)	Recall (%)	Precision (%)	F1-score (%)	Cohen's Kappa (%)
Toilet	53.52	71.92	61.37	61.05	65.38	59.15	62.11	61.72
Shower	70.17	90.02	78.86	78.83	89.62	84.13	86.79	86.70
Faucet	79.78	73.51	76.52	76.16	64.64	73.69	68.87	68.45
CW	64.46	86.71	73.95	71.76	77.16	79.48	78.30	78.15
DW	4.82	41.37	8.63	8.62	1.20	2.52	1.63	1.59
Total	66.83	77.61	71.73	71.62	70.21	71.88	71.04	70.82

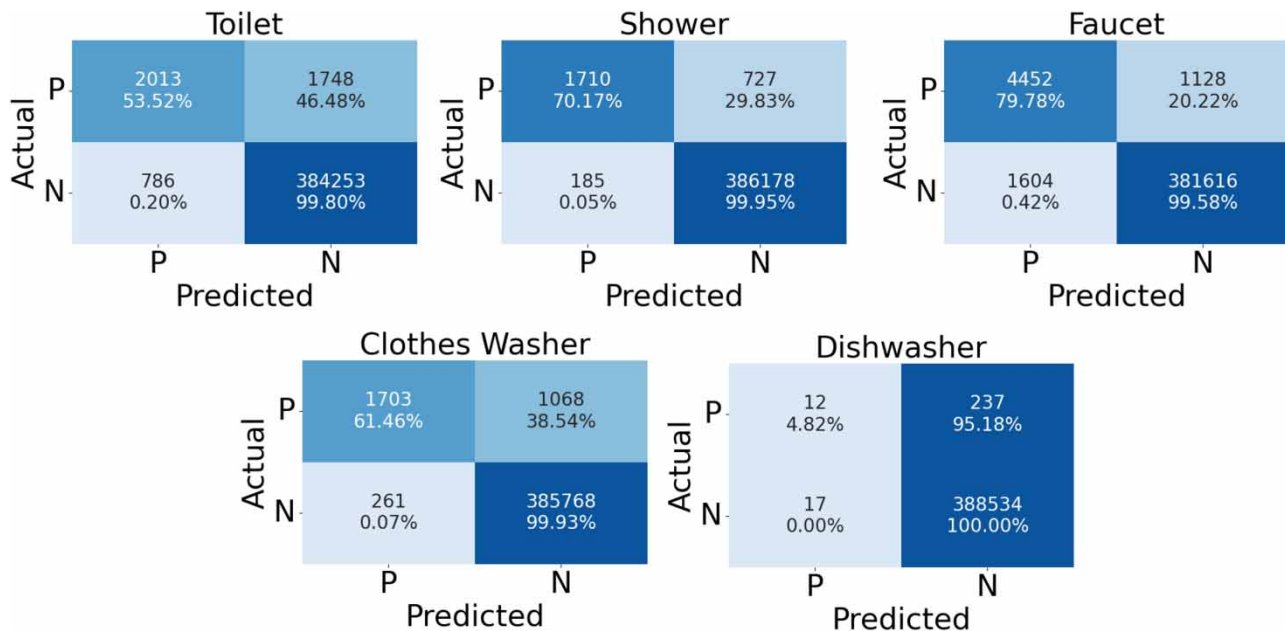**Figure 6** | Confusion matrices of the model-based method.

Figure 7 | Confusion matrices of the learning-based method.

score, for which the MB method achieved 84.93% and the LB method 83.36%. A more detailed analysis of the results per water appliance is presented below.

Toilet: Both models reach a similar F1-score regarding toilet classification (LB: 61.37% and MB: 62.11%), with the LB method showing a higher precision and the MB method being able to identify more samples labeled as 'Toilet'. As demonstrated in Figure 7, the LB model has difficulty in identifying the 'Toilet' class as it manages to correctly identify when the toilet was used approximately half of the time (53.52%). The MB model is slightly better since it manages to identify
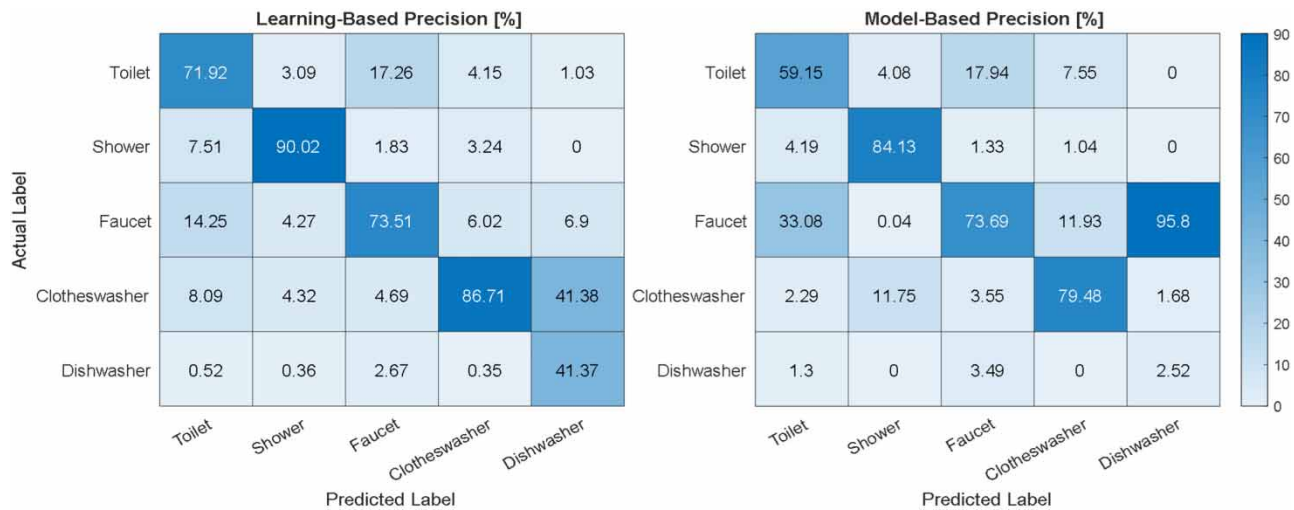
Figure 8 | Confusion matrix of the precision of the learning-based (left) and model-based (right) proposed methods in water end-use classification.

65.83% of the toilet samples (Figure 6). Toilet events have a fixed mechanical operation/signature which can be detected using the DTW algorithm, thus explaining the higher detection score for the MB model. In both cases, the models misclassify the 'Toilet' class, mainly, with the faucet class (Figure 8).

Shower: Regarding the 'Shower' class, the LB model identifies correctly almost 70.17%, and the MB model approximately 89.62% of the samples that are labeled as a shower. Also, both models have few false positives as it distinguishes the rest of the samples with ease. In that case, the use of statistical bounds extracted through the offline learning stage enhances the MB method to identify more shower events due to their distinctive consumption volume and duration. On the other hand, the application of the usage characteristics in the classification process led to lower MB precision (84.13%) compared to the LB method (90.02%). CW events with larger consumption duration and volume were misclassified as shower events (Figure 8).

Faucet: For the 'Faucet' category, we observe that the classifier of the LB model learned to identify cases with higher accuracy (76.52%) than the MB method (68.87%). In terms of precision, both methods are considered equal, reaching a score of approximately 74%. The misclassified faucet events were mostly confused as toilet events. These two water appliances constitute most of the household water end-uses which can explain their frequent misclassification. In addition, as depicted in Figure 8, many missed faucet events were identified by the algorithm as DW events.

Clothes washer: According to the confusion matrix for the 'Clothes washer' class, the LB and MB methods correctly identify 61.46 and 77.16% of the cases where the CW was in operation, respectively. In addition, only a few misclassifications of 'Clothes washer' events are noted. The higher MB score indicates the usefulness of applying a sliding window to detect the full operation cycle of intermittent flow devices.

Dishwasher: On the other hand, both models exhibit difficulty in correctly classifying any of the dishwasher samples. Specifically, the MB and LB algorithms manage to correctly identify only 3 and 12 samples out of the total 249 samples when the DW was in use. The low performance is attributed partly to the fact that the DW constitutes the minority class with just 1.9% relative to the other class as shown in Figure 5, and partly to the fact that the DW cycle exhibits intermittent behavior, thus making it harder for the model to distinguish between the DW and other devices. Further, although the DW generally exhibits lower consumption volume, duration, and flow rate than the other events, extracting distinctive statistical bounds was impossible due to the low data resolution.

The two methods present differences regarding their complexity, computational time and hardware implementation. The LB method requires a large amount of data to achieve high performance. With the introduction of the window size, the input to the model becomes even larger compared to other approaches thus becoming more computationally expensive. The same applies during the training phase of the model. On the other hand, this data-driven method with the introduction of more data in the training phase leads to higher model accuracy. Furthermore, the development of the LB method based on an MLP

model gives us the capability to transfer the model to a microprocessor. Regarding the MB method, the algorithm does not need a large amount of labeled data for the training phase thus making it faster compared to the LB method. Historical data are essentially needed for the acquisition of event consumption signatures and the calculation of the statistical bounds for each event feature. On the other hand, the MB algorithm uses software packages of which the computational and memory needs may prohibit its implementation on edge intelligence devices. A more suitable implementation strategy would be the processing of data in an online platform, which is associated with higher communication costs.

3.3. Experimental demonstration of the learning-based method for edge-device predictions

To demonstrate experimentally the applicability of the method, the LB model was further implemented on a microcontroller aiming for real-time water-use event classification of four appliances, namely the toilet, shower, faucet, and clothes washer by using the total flow measurement as input. The DW category was not included in this case study due to its low classification score as presented in Section 3.2.3.

Microcontroller. The microcontroller 'BALoRa V0.1' which was developed in the KIOS Research and Innovation Center of Excellence was used for the deployment of the machine learning model. The proposed microcontroller integrates the ESP32-PICO-D4 which is a compact version of the dual-core processor ESP32 microcontroller (Espressif 2022). It includes an integrated Serial Peripheral Interface (SPI) flash memory of 4 MB and 520 KB of SRAM and was programed using PlatformIO. This module was chosen due to its robust performance, ultra-low power consumption, and its sufficient resources for running the program and processing data efficiently.

Implementation details. The implementation process included initially the transformation of the proposed MLP model from a 'TensorFlow' type to 'TensorFlow Lite' type to deploy the MLP classifier on the microcontroller (Abadi et al. 2016). An optimized FlatBuffer format (Flatbuffers 2014) was utilized which allowed us to transform and optimize the classifier to run on very low power on a microcontroller. The machine learning model was adjusted to comply with the microcontroller's low RAM and storage memory. This was achieved by converting the TensorFlow type, with an initial size of 429 kB, into the TensorFlow Lite, thus reducing the model's size to 83 kB. Next, we used the TF Lite Microcontrollers libraries by including the header files in PlatformIO to run the model on the microcontroller. The headers provide the necessary information used by the interpreter to run the model such as the debug information, the code to load and run the model, the schema for the TF Lite, the model file format, and the versioning information for the schema. Further, the main script was programed to interact with the TF Lite model thus being able to provide input and receive output signals. The source code was then utilized for the development of the model on the microcontroller using the PlatformIO. The microcontroller was connected to a PC via a USB in a serial connection and the simulated water flow measurements were sent to the microcontroller through a Matlab application using an Excel file (Figure 9). Each time a new variable was sent to the microcontroller, the main function invoked the neural network and outputted a prediction of an appliance being on or off.

Results. The tinyML model was deployed on the microprocessor with a window size of 60 time steps and the classification results are presented in Table 5 and in Figure 10. The specific window size was selected due to the microcontrollers' low RAM and storage memory. Moreover, due to the microcontrollers' low RAM and storage memory classification input data were processed into 2-bit floats from a 32-bit float. Even though the classification method is the same, the tinyML model achieved a slightly higher F1-score (72.01%) and Cohen's Kappa score (71.82%) compared to the LB method (71.73 and 71.62%) due to the numeric bit modification of the input data. Specifically, higher F1 scores were recorded for the shower (83.37%), faucet (77.32%), and DW (21.64%) categories and a lower F1-score for the toilet (60.79%) and CW (67.33%) categories.

4. CONCLUSIONS AND FUTURE WORK

Non-intrusive water usage disaggregation and classification is a challenging task that includes the identification of both single and combined events with overlapping use while specific appliances may operate with intermittent flow, making individual consumption events hard to distinguish.

In this work, we presented and compared two different methods to overcome this challenge, an MB method and a learning-based method. The objective was to identify which method is more suitable while highlighting the advantages and disadvantages of each method. The MB algorithm uses a combination of methods including DTW, statistical bounds, variation vectors, and sliding time windows to analyze the consumption pattern, the volume, the duration, and the flow rate of an event in order to correctly classify it. For the LB method, four different algorithms, namely, SVM, RF, XGBoost, and MLP, were initially
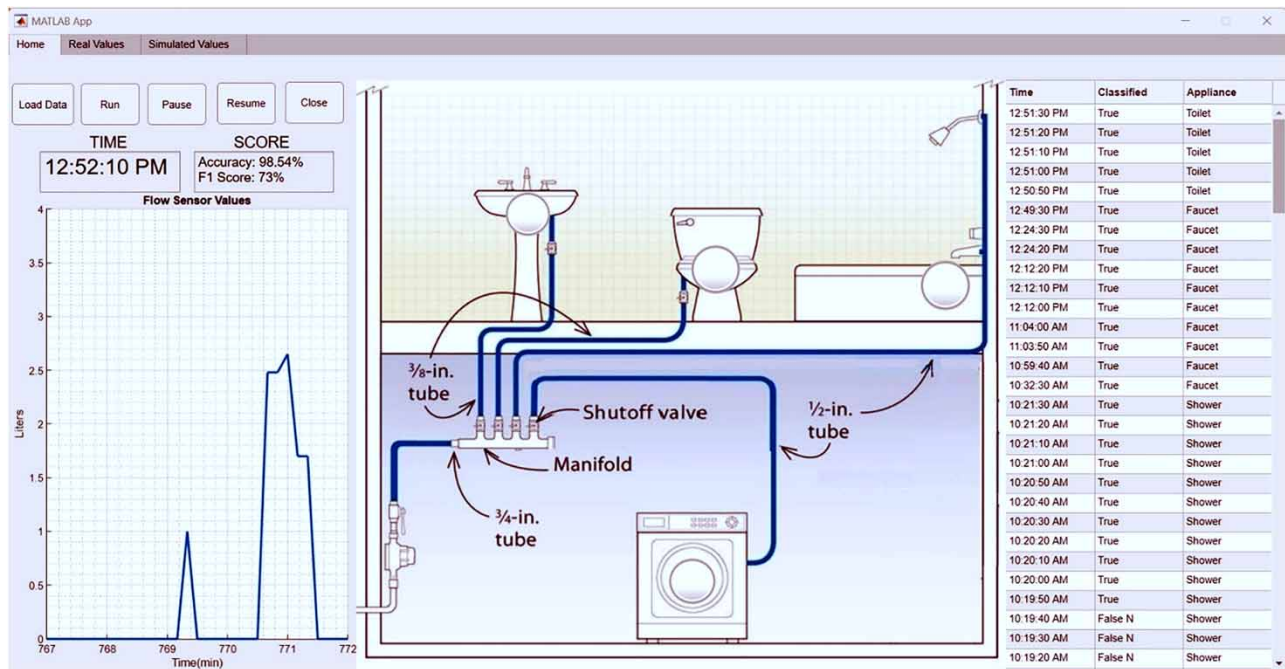
Figure 9 | The graphical user interface of the tinyML model for real-time visualization of the data and the predictions.

Table 5 | Performance of MLP model on the microprocessor

| | Edge-device application (Window 60) | | | |
	Recall (%)	Precision (%)	F1-score (%)	Cohen's Kappa (%)
Toilet	54.53	68.69	60.79	60.46
Shower	76.77	91.22	83.37	83.28
Faucet	79.53	75.23	77.32	76.98
CW	54.24	88.77	67.33	67.16
DW	13.25	58.93	21.64	21.26
Total	66.87	78.01	72.01	71.82

compared. The most efficient LB algorithm was found to be the MLP that utilizes a windowing feature to capture the temporal aspects of the data. The chosen LB method was then compared with the MB method utilizing three evaluation metrics, namely, the F1-score, the Cohen's Kappa, and the ROC-AUC. Both methods demonstrate approximately the same level of effectiveness, with the MB method showing a higher accuracy in classifying shower and CW events, and the LB method being better in detecting faucet events. Identified difficulties are the class imbalance and the noisy information as a result of the time-series aggregation. Further, data with higher resolution (e.g., at 1 s) would be more suitable for the application of these techniques. The LB method demands extensive data, contributing to computational complexity, while the MB method that is faster may face implementation challenges on edge devices due to computational and memory requirements.

An important contribution of this work has been the development of an experimental demonstration, where the LB method is implemented into an embedded hardware solution that can autonomously analyze and process data on-site and in real-time, thereby improving data security and privacy. Additionally, the use of an optimized tinyML algorithm allows us to utilize low-powered devices with a smaller footprint leading to reduced power consumption, memory usage, and eventually cost savings.
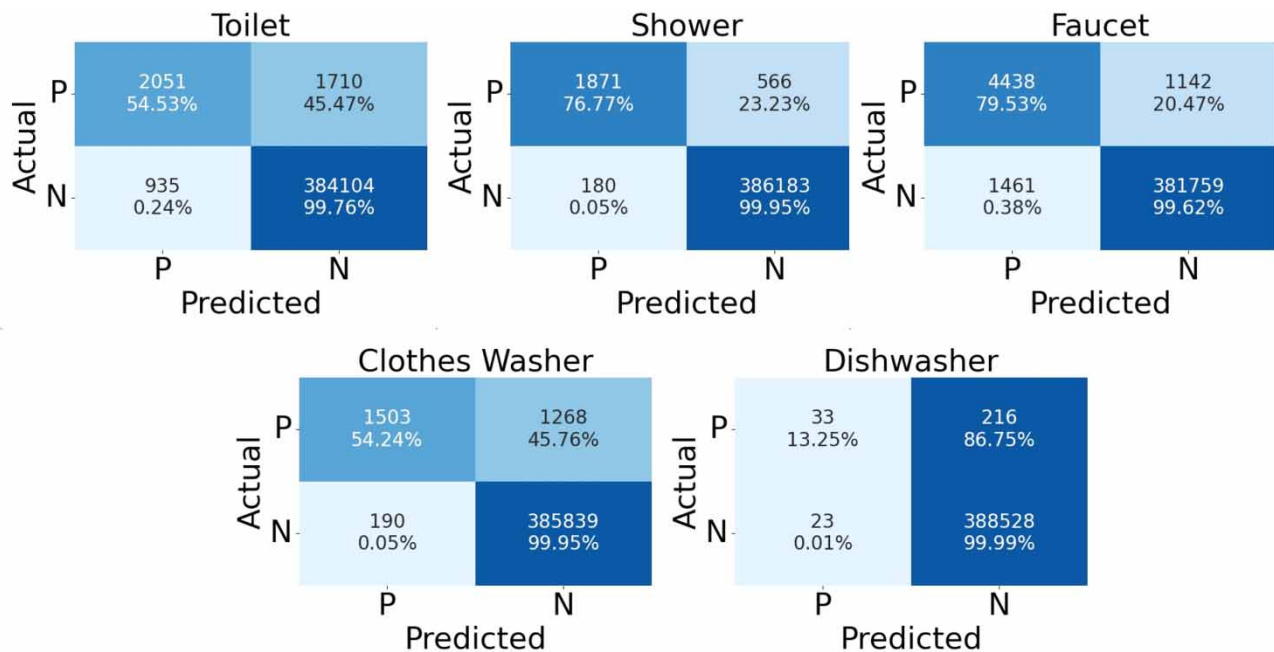
Figure 10 | Confusion matrix of the tinyML model.

Future work will attempt to improve the accuracy of the two methodologies and the corresponding LB hardware solution. For the MB, the separation process of combined events can be investigated in more depth considering all possible combinations while the LB model can be further improved to better capture longer temporal correlations using deep neural models, such as Long Short-Term Memories (LSTM) and CNNs. Lastly, the applicability of these methods is further suggested to be tested in datasets including new water appliances and considering the presence of leakages.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The generated dataset is available in the following repository: https://github.com/KIOS-Research/Water-Usage-Dataset.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M. & Zheng, X. 2016 Tensorflow: large-scale machine learning on heterogeneous distributed systems. *The Library*. http://arxiv.org/abs/1603.04467.

Bishop, C. M. & Nasrabadi, N. M. 2006 *Pattern Recognition and Machine Learning*. Springer-Verlag, Berlin, Heidelberg.

Bradley, A. P. 1997 The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **30** (7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2.

Che, Z., Cheng, Y., Zhai, S., Sun, Z. & Liu, Y. 2017 Boosting deep learning risk prediction with generative adversarial networks for electronic health records. In: *2017 IEEE International Conference on Data Mining (ICDM)*. https://doi.org/10.1109/ICDM.2017.93.

Chen, T. & Guestrin, C. 2016 XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785.

Cohen, J. 1960 A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20** (1), 37–46. https://doi.org/10.1177/001316446002000104.

Cominola, A., Giuliani, M., Castelletti, A., Rosenberg, D. E. & Abdallah, A. M. 2018 Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management. *Environmental Modelling & Software* **102**, 199–212. https://doi.org/10.1016/j.envsoft.2017.11.022.

Cominola, A., Giuliani, M., Castelletti, A., Fraternali, P., Gonzalez, S. L. H., Herrero, J. C. G., Novak, J. & Rizzoli, A. E. 2021 Long-term water conservation is fostered by smart meter-based feedback and digital user engagement. *Npj Clean Water* **4** (1), 29. https://doi.org/10.1038/s41545-021-00119-0.

Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Machine Learning* **20**, 273–297. https://doi.org/10.1023/A:1022627411411.

Cosgrove, W. J. & Loucks, D. P. 2015 Water management: Current and future challenges and research directions. *Water Resources Research* **51** (6), 4823–4839. https://doi.org/10.1002/2014WR016869.

Dempster, A., Petitjean, F. & Webb, G. I. 2020 ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery* **34** (5), 1454–1495. https://doi.org/10.1007/s10618-020-00701-z.

DeOreo, W. B. 2011 Analysis of water use in new single family homes. *By Aquacraft. For Salt Lake City Corporation and US EPA*.

DeOreo, W. B., Heaney, J. P. & Mayer, P. W. 1996 Flow trace analysis to access water use. *Journal – American Water Works Association* **88** (1), 79–90. https://doi.org/10.1002/j.1551-8833.1996.tb06487.x.

Di Nardo, A., Boccelli, D. L., Herrera, M., Creaco, E., Cominola, A., Sitzenfrei, R. & Taormina, R. 2021 Smart urban water networks: Solutions, trends and challenges. *Water* **13** (4), 501. https://doi.org/10.3390/w13040501.

Espressif 2022 *ESP32-PICO-D4 Datasheet*. Available from: http://espressif.com/sites/default/files/documentation/esp32-pico-d4_datasheet_en.pdf (accessed 3 April 2023).

Farhadi, H. & Najafzadeh, M. 2021 Flood risk mapping by remote sensing data and random forest technique. *Water* **13** (21), 3115. https://doi.org/10.3390/w13213115.

Flatbuffers 2014 *Flatbuffers*. Available from: https://google.github.io/flatbuffers/ (accessed 3 April 2023).

Froehlich, J. E., Larson, E., Campbell, T., Haggerty, C., Fogarty, J. & Patel, S. N. 2009 HydroSense. In *Proceedings of the 11th International Conference on Ubiquitous Computing*. https://doi.org/10.1145/1620545.1620581.

Geng, Y. & Luo, X. 2019 Cost-sensitive convolutional neural networks for imbalanced time series classification. *Intelligent Data Analysis* **23** (2), 357–370. https://doi.org/10.3233/IDA-183831.

He, H. & Garcia, E. A. 2009 Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* **21** (9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239.

Inman, D. & Jeffrey, P. 2006 A review of residential water conservation tool performance and influences on implementation effectiveness. *Urban Water Journal* **3** (3), 127–143. https://doi.org/10.1080/15730620600961288.

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A. & Petitjean, F. 2020 Inceptiontime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery* **34** (6), 1936–1962. https://doi.org/10.1007/s10618-020-00710-y.

Jaeger, H. & Haas, H. 2004 Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304** (5667), 78–80. https://doi.org/10.1126/science.1091277.

Kowalski, M. & Marshallsay, D. 2003 A system for improved assessment of domestic water use components. In: *Proceedings of 2nd International Water Association Conference on Efficient Use and Management of Urban Water Supply*, London, UK, International Water Association.

Mayer, P. W., Deoreo, W. B., Opitz, E. M., Kiefer, J. C., Davis, W. Y., Dziegielewski, B. & Nelson, J. O. 1999 *Residential End Uses of Water*. American Water Works Association Research Foundation, Denver, CO, USA.

Mazzoni, F., Alvisi, S., Franchini, M., Ferraris, M. & Kapelan, Z. 2021 Automated household water end-use disaggregation through rule-based methodology. *Journal of Water Resources Planning and Management* **147** (6), 04021024. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001379.

Mazzoni, F., Alvisi, S., Blokker, M., Buchberger, S. G., Castelletti, A., Cominola, A., Gross, M.-P., Jacobs, H. E., Mayer, P., Steffelbauer, D. B., Stewart, R. A., Stillwell, A. S., Tzatchkov, V., Yamanaka, V.-H. A. & Franchini, M. 2023 Investigating the characteristics of residential end uses of water: A worldwide review. *Water Research* **230**, 119500. https://doi.org/10.1016/j.watres.2022.119500.

McDonald, R. I., Weber, K., Padowski, J., Flörke, M., Schneider, C., Green, P. A., Gleeson, T., Eckman, S., Lehner, B., Balk, D., Boucher, T., Grill, G. & Montgomery, M. 2014 Water on an urban planet: Urbanization and the reach of urban water infrastructure. *Global Environmental Change* **27**, 96–105. https://doi.org/10.1016/j.gloenvcha.2014.04.022.

Meyer, B. E., Nguyen, K., Beal, C. D., Jacobs, H. E. & Buchberger, S. G. 2021 Classifying household water use events into indoor and outdoor use: improving the benefits of basic smart meter data sets. *Journal of Water Resources Planning and Management* **147** (12), 04021079. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001471.

Najafzadeh, M. & Basirian, S. 2023 Evaluation of river water quality index using remote sensing and artificial intelligence models. *Remote Sensing* **15** (9), 2359. https://doi.org/10.3390/rs15092359.

Najafzadeh, M. & Zeinolabedini, M. 2018 Derivation of optimal equations for prediction of sewage sludge quantity using wavelet conjunction models: An environmental assessment. *Environmental Science and Pollution Research* **25** (23), 22931–22943. https://doi.org/10.1007/s11356-018-1975-5.

Najafzadeh, M. & Zeinolabedini, M. 2019 Prognostication of waste water treatment plant performance using efficient soft computing models: An environmental evaluation. *Measurement: Journal of the International Measurement Confederation* **138**, 690–701. https://doi.org/10.1016/j.measurement.2019.02.014.

Nguyen, K. A., Stewart, R. A. & Zhang, H. 2013 An intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Environmental Modelling & Software* **47**, 108–127. https://doi.org/10.1016/j.envsoft.2013.05.002.

Nguyen, K. A., Stewart, R. A., Zhang, H. & Jones, C. 2015 Intelligent autonomous system for residential water end use classification: Autoflow. *Applied Soft Computing* **31**, 118–131. https://doi.org/10.1016/j.asoc.2015.03.007.

Nguyen, K. A., Stewart, R. A., Zhang, H. & Sahin, O. 2018 An adaptive model for the autonomous monitoring and management of water end use. *Smart Water* **3** (1), 5. https://doi.org/10.1186/s40713-018-0012-7.

Papatheodoulou, D., Pavlou, P., Vrachimis, S. G., Malialis, K., Eliades, D. G., Theocharides, T., 2022 A multi-label time series classification approach for non-intrusive water end-use monitoring. In: *Artificial Intelligence Applications and Innovations* (Maglogiannis, I., Iliadis, L., Macintyre, J. & Cortez, P. eds.). Springer International Publishing, pp. 53–66. https://doi.org/10.1007/978-3-031-08337-2_5.

Pastor-Jabaloyes, L., Arregui, F. & Cobacho, R. 2018 Water end use disaggregation based on soft computing techniques. *Water* **10** (1), 46. https://doi.org/10.3390/w10010046.

Pavlou, P., Vrachimis, S. G., Eliades, D. G. & Polycarpou, M. M. 2022 Non-intrusive water usage classification considering limited training data. In: *2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry*. https://doi.org/10.4995/WDSA-CCWI2022.2022.14829.

Steffelbauer, D. B., Blokker, E. J. M., Buchberger, S. G., Knobbe, A. & Abraham, E. 2021 Dynamic time warping clustering to discover socioeconomic characteristics in smart water meter data. *Journal of Water Resources Planning and Management* **147**, 6. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001360.

Strodthoff, N. & Strodthoff, C. 2019 Detecting and interpreting myocardial infarction using fully convolutional neural networks. *Physiological Measurement* **40** (1), 015001. https://doi.org/10.1088/1361-6579/aaf34d.

Zeinolabedini, M. & Najafzadeh, M. 2019 Comparative study of different wavelet-based neural network models to predict sewage sludge quantity in wastewater treatment plant. *Environmental Monitoring and Assessment* **191** (3), 163. https://doi.org/10.1007/s10661-019-7196-7.

Zhang, X., Gao, Y., Lin, J. & Lu, C.-T. 2020 Tapnet: Multivariate time series classification with attentional prototypical network. *Proceedings of the AAAI Conference on Artificial Intelligence* **34** (04), 6845–6852. https://doi.org/10.1609/aaai.v34i04.6165.