PressureML: Modelling Pressure Waves to Generate Large-Scale Water-Usage Insights in Buildings

Tanmaey Gupta
tanmaeygupta99@gmail.com
Microsoft Research
India

Anupam Sobti
anupam.sobti@plaksha.edu.in
Plaksha University
India

Akshay Nambi
akshayn@microsoft.com
Microsoft Research
India

ABSTRACT

Several studies have indicated that delivering insights and feedback on water usage has been effective in curbing water consumption, making it a pivotal component in achieving long-term sustainability objectives. Despite a significant proportion of water consumption originating from large residential and commercial buildings, there is a scarcity of cost-effective and easy-to-integrate solutions that provide water usage insights at a reasonable spatio-temporal granularity in such structures. Furthermore, existing methods for disaggregating water usage necessitate training data and rely on frequent data sampling to capture patterns, both of which pose challenges when scaling up and adapting to new environments. In this work, we aim to solve these challenges through a novel end-to-end approach which records data from pressure sensors and uses time-series classification by DNN models to determine room-wise water consumption in a building. This consumption data is then fed to a novel water disaggregation algorithm which can suggest a set of water-usage events, and has a flexible requirement of training data and sampling granularity. We conduct experiments using our approach and demonstrate its potential as a promising avenue for in-depth exploration, offering valuable insights into water usage on a large scale.

CCS CONCEPTS

• **Computing methodologies → Machine learning approaches**; *Modeling and simulation*; • **Computer systems organization →** *Embedded and cyber-physical systems.*

KEYWORDS

Transfer learning, Time series classification, Pressure hammer waves, Sustainability, Water Disaggregation

ACM Reference Format:
Tanmaey Gupta, Anupam Sobti, and Akshay Nambi. 2023. PressureML: Modelling Pressure Waves to Generate Large-Scale Water-Usage Insights in Buildings. In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23), November 15–16, 2023, Istanbul, Turkey*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3600100.3626636

BuildSys '23, November 15–16, 2023, Istanbul, Turkey

ACM ISBN 979-8-4007-0230-3/23/11…$15.00
https://doi.org/10.1145/3600100.3626636

1 INTRODUCTION

Efficient water use is essential for long-term sustainability goals. To achieve this, it's crucial to provide individuals and organizations with tools to optimize water usage effectively. One such approach is offering water usage insights, which informs decisions about water use and infrastructure design. Studies show that real-time feedback through smart meters and digital solutions can reduce water consumption by up to 30% [15]. Beyond reducing consumption, water usage insights can also lower energy usage in water systems, ensure fair billing, and detect leaks. While a significant percentage of water consumption can be attributed to commercial and residential buildings due to high occupant density, there is a lack of viable solutions which provide these insights on such large-scales. In this work, we present `PressureML`, an end-to-end and scalable approach leveraging DNN to generate real-time water usage feedback in buildings. Delivering water usage insights involves two stages: *1. Recording water usage data at an appropriate level of detail in time and space, and 2. Using disaggregation techniques to identify specific fixture or appliance-level usage.*

Current methods for recording total water usage primarily involve installing water meters in individual houses [4]. These meters come in various types, are installed at the main water inlet, use electromagnetic or ultrasonic pulses to measure water flow, and typically cost between 600 and 1200 dollars. However, when applied to larger residential and commercial settings, these solutions encounter several challenges: *(i) Multiple Inlets:* In such buildings, multiple water inlets are used for different areas, making it challenging to monitor usage from a single source. *(ii) Economical Challenges:* Installing a water meter on every floor or in every unit is not economically feasible. Using fewer meters for multiple areas or units reduces the accuracy of granular usage insights. *(iii) Retrofitting Difficulty:* Retrofitting existing buildings with these meters is not straightforward, often requiring pipe cutting for each unit. *(iv) Maintenance Demands:* These water meters involve complex sensing mechanisms with multiple sensitive components, necessitating high maintenance. HydroSense [10] attempts to calculate volume usage using pressure sensors for each fixture in a house and requires large training data and calibration efforts. Our approach, however, is able to get volume usage for multiple floors simultaneously, and does not require fixture-level training data and calibration.

Once the total water consumption for a specific time period in a house or building area is determined, the next step is to disaggregate this total usage and attribute it to individual fixtures or appliances. Various research efforts have aimed to address this challenge [14, 16, 17]. These studies primarily rely on identifying and encoding distinct features and usage patterns associated with different fixtures or appliances. However, current approaches face three

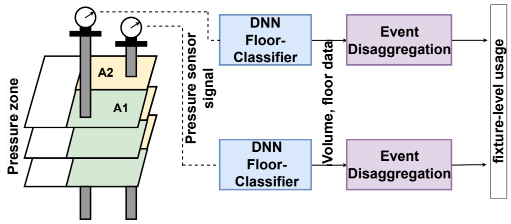

Figure 1: `PressureML` architecture: Pressure sensors are connected to the pipes running through each area (A1, A2) of all houses across floors. Pressure signals are classified into floors, and event disaggregation is performed for the particular floor and area.

significant challenges: *(i) High-Frequency Data Emphasis:* They focus on high-frequency data, which is challenging to collect on a large scale due to power consumption, memory requirements, and data transmission overhead. *(ii) Training Data Requirement:* These approaches depend on the availability of training data, making it difficult to apply them effectively in diverse and previously unseen settings. *(iii) Requirement of additional infrastructure:* These approaches, which suggest installing additional sensing modalities to get usage information, are expensive and not scalable.

`PressureML` aims to address the aforementioned challenges by proposing and evaluating an end-to-end system capable of generating detailed water usage insights on a large scale. Our system is cost-effective and can be easily integrated, making it suitable for various water system configurations in residential and commercial complexes. However, our primary research focus is on high-rise residential apartments, where each house's areas (such as bathrooms and kitchens) on multiple floors are connected via dedicated pipelines supplied by a gravity-based distribution network. For instance, in a house with two bathrooms and a kitchen, there are three vertical main water pipes running through all floors to supply these areas. To tackle this issue, we present a two-stage solution: first, determining the total volume of water used in each area of each house, and then disaggregating this total usage and attributing it to individual fixtures and appliances.

Our contributions focus on a novel approach that shifts from collecting individual readings for all areas within each house to obtaining data for each area across all houses. This approach reduces the need for extensive sensing infrastructure, making it cost-effective, adaptable to multi-inlet setups, and be easily retrofitted at the building's top. It offers flexibility to adjust sensing scope for desired disaggregation granularity and accuracy in large buildings.

The system's core architecture, outlined in Figure 1, will be elaborated in subsequent sections. In the first stage, we install high-resolution pressure sensors at the top of each vertical pipeline within a pressure zone (in contrast to water meters for each house), sampling them at a high frequency. When taps are opened on any floor, they generate unique pressure waves in the pipes, recorded by these sensors. Utilizing deep neural network (DNN) techniques, we model these time-series waves to identify the specific floor(s) of water usage. This stage requires minimal training data due to significant inter-class variability, characteristic patterns, cross-deployment adaptability, and potential use of physics-based simulations. Aggregate water consumption is determined using the time duration of pressure drops. In the second stage, we employ a novel disaggregation algorithm for each area on every floor. This algorithm, unlike traditional methods, does not rely on training data but instead requires general fixture configuration and usage specifications, which are common across deployments and easy to ascertain. It is also adaptable to a range of sampling frequencies. Both stages of our solution, by design, can handle overlapping events, which occur frequently in real-world settings. We assess the initial phase of our solution with a pressure sensor dataset from a test building environment and the second phase with aggregate usage data from a real residential apartment labeled over a one-week period. However, comprehensive evaluations are ongoing and will be discussed in future research, highlighting the need for longer-term data to fully realize the solution's potential.

2 BACKGROUND AND SETUP

In a typical high-rise building, water is supplied through a gravity-based distribution network. An overhead tank (OHT) at the building's top provides water to vertically aligned areas on each floor. Water pressure is kept uniform and within a safe range, typically around 80 psi [1]. However, the upper floors receive lower pressure, while the lower floors have higher pressure. To balance this, the building is divided into pressure zones, each equipped with booster pumps and Pressure Reduction Valves (PRV) to maintain consistent pressure across floors [1].

In this work, we propose to install pressure sensors at points in the distribution network which are directly connected with the end-user equipments like bathroom and kitchen fittings. These points can be the downstream output valves of PRV, booster-pump, or OHT, depending on the network used for that zone. When an end-user valve is opened or closed, the velocity of water changes rapidly, and transient pressure waves called water hammer are generated. These waves get recorded by the installed pressure sensor. The signature and magnitude of the wave depends primarily on the rate of change of velocity of water, distance of the valve to the sensor, and other properties related to the piping structure and fluid in motion [2]. Valve openings create pressure drop, while valve closings increase the pressure. Since all identical areas in a building have a dedicated vertical network, if we use these discriminative factors to identify floor of origin, we can get the aggregate water used through the day in a particular area, as required for the first stage of our solution. Finally, the volume of water flowing upon a pressure drop lasting for time t can be estimated using the equation: $Q = \frac{\Delta P \pi r^4}{8 \mu L} t = \frac{\Delta P}{R_f} t$, where ΔP is the pressure drop and R_f includes length of the pipe and other factors related to infrastructure and fluid used, which are generally constant across the building. Since it is difficult to determine and incorporate all of these factors in calculations, it is more reasonable to empirically determine R_f by recording the volume used at a particular floor and the corresponding pressure drop. We leave its detailed analysis and evaluation to future work. Once the total aggregate volume of each area is calculated as described above, it is fed to a disaggregation algorithm,

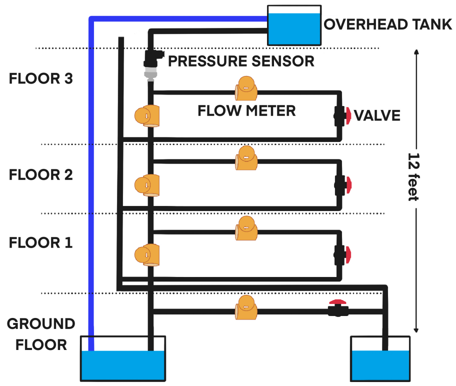

Figure 2: Experimental setup to simulate taps installed on 4 floors and a pressure sensor installed at the top end of the main pipe.

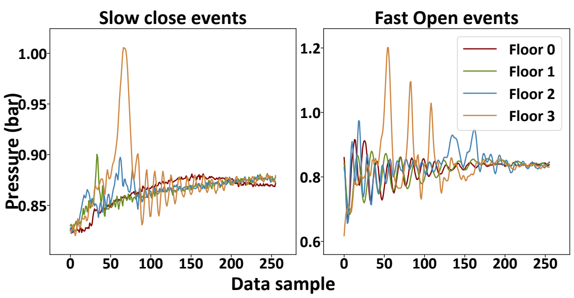

Figure 3: Resampled pressure waveforms recorded by the test setup.

which provides insights related to fixture-level contributions to the total water consumption.

To develop and analyse the first stage of our proposed approach - the floor classification system, we simulated a 4-floor pressure zone of a building using a test setup as shown in figure 2. An OHT was installed and connected to a network of pipes, ultimately ending in user-operated taps installed across 4 floors. A pressure sensor with a sampling frequency of 500Hz was installed at the outlet valve of the OHT which recorded stable pressure under static conditions and transient pressure waves when a tap was manually opened or closed. While collecting data, the tap at each floor is opened and closed at 2 different speeds - fast (0.5-1 sec), and slow (1.5-3 sec). 1 experiment iteration at a floor consists of 4 events: Fast open, Fast close, Slow open and Slow close. Experiments being manual, a small dataset of 4 such iterations per floor was recorded, leading to a total of 32 open samples and 32 close samples across 4 classes (floors). A few recorded waveforms are shown in figure 3.

3 PRESSURE WAVE CLASSIFICATION

As discussed in Section 2, determining the floor of origin for a transient pressure wave is sufficient to calculate aggregate water consumption. These waves, when sensed at a fixed point, exhibit unique signatures and magnitudes due to their reflections and varying travel distances, as evident in Figure 3. We utilize a DNN model to learn these features and identify their floor of origin. The DNN model takes as input a sequence of data points denoting a pressure wave event, and outputs the probability of that wave belonging to each floor. At this stage, we do not distinguish between specific fixtures, which we address in Section 4. To enhance scalability and minimize the need for extensive labeled training data, our methods are designed to operate effectively with smaller datasets, akin to the one described in Section 2. Now, we outline the steps involved in classifying water usage events by their respective floor.

3.1 Automatic event extraction and identification

In order to classify pressure wave events into floors, we first need to extract the transient section from a continuous stream of pressure sensor signal or buffered data. The original pressure signal $p(t)$ is passed through a low-pass filter to highlight the underlying pattern of the signal and eliminate noise. Let this new signal be $p^{'}(t)$. The open/close events are identified by rapid pressure changes and hence we find a forward difference quotient as: $\Delta p^{'}(t) = \frac{p^{'}(t+k)-p^{'}(t)}{k}$. To accommodate different rates of change in pressure due to different valve-operation speeds, multiple such quotients are computed with varying k, whose value is determined in accordance with the sampling frequency of the sensor. In the present case, $k = 50, 100, 500$ are used. Event starting points are marked when the quotient with the largest average value in an event exceeds an empirical threshold (0.01 in our case), while event end points are marked when majority of the quotients are below a threshold (0.005 in our case) for k continuous values. All such events extracted are upsampled/downsampled to have a uniform length of 256 points.

3.2 Data Augmentation

To prevent DNN models used for event classification from overfitting due to limited training data, we employ data augmentation, which has previously shown benefits even in time-series classification (TSC) tasks [13]. A recursive and incremental augmentation approach is used, involving transformations suited for our TSC task. These include *i. Jittering [13]*, which adds Gaussian noise, *ii. Window warp [11]*, which stretches and constricts a window in the time series by a certain factor, *iii. Magnitude warp [11]*, which stretches and constricts the waveform at randomly chosen points across the magnitude dimension, and *iv. DGW-sD [11]* , which uses discriminative shape descriptors for guided warping. Using these augmentations, we increase our dataset size four times to 128 open samples and 128 close samples.

3.3 Classifier model

The primary step of stage 1 of proposed approach is to classify pressure waves into floors of their origin. To achieve this, we use a pretraining-finetuning approach, which, after vision-based tasks, has recently shown performance gains in time-series classification tasks as well [8]. We explore this approach across two axis : model used for training, and the dataset used for pre-training. We pretrain each model on each of the source dataset and then finetune it on our collected pressure dataset, using a 70-30 train-test split. F1 score for floor classification is used to evaluate each pair, and the results are shown in table 1. Models used in this study have shown state-of-the-art results on various time-series classification tasks and vary in terms of inherent layers. For pre-training data, we handpick 4 datasets from the UCR Time series archive [5], which have a large number of labelled samples (>1000), 1 feature dimension, vary in terms of number of classes, sequence length and source of data, and naturally would fit for our downstream task. We use their default train split for the pre-training task. For models with sequence length-dependent blocks like LSTM and Attention, we resample the source length to match with the target length of 256. We also include in this analysis an additional modified dataset (PressureLD, with sequence length =256 and number of classes =4) consisting of pressure sensor signals recorded to detect leaks in a water distribution system [3]. This helps to consider the effect of pre-training on data from the same modality as that of the target task.

Results in table 1 show that the proposed pretraining-finetuning approach is able to classify pressure waves with high F1 scores and can enable classification with less training data for `PressureML`. CNN based models might be a better fit as compared to RNN or Attention based models for sensor data time-series classifications tasks due to less inter-dependence on long and short term context, but a significant presence of patterns in the data. Also, the influence of source dataset used for pre-training is small if chosen reasonably, while model used plays a more crucial role in downstream task performance. The results are promising as the best-performing ResNet model achieves an F1 score of 1 for both open and close events. This calls for a more extensive evaluation study, which we plan to do in future work.

3.4 Aggregate volume calculation

A particular water usage instance is bounded by a open event which marks the start of water flow and a close event, which marks the end of water flow. Based on these floor-wise classified events, a square waveform is computed for each floor with y-axis representing pressure and x-axis representing time. Area under the curve (AUC) is calculated as per flow equation above, with base of the curve being the static pressure level. Entire timeline is then broken into windows of width w, and area falling in that window is denoted as the total water volume consumed in that time window. This approach works for overlapping events as well, wherein greater pressure drop will give higher AUC value for that time-frame.

Since water-usage insights do not require real-time updates, we propose to generate them at the end of each day. Therefore, we record data from the pressure sensor for an entire day before conducting the analysis. This helps in two ways: (1) intensive analysis of pressure signals for classification due to availability of long term data, which would especially be helpful in case of over-lapping signals, and (2) employing usage and statistical heuristics to generate better disaggregation insights, as explained in detail in section 4.

4 DISAGGREGATING EVENTS

This section describes the second stage of our proposed solution, which provides fixture-level water usage insights from aggregate volume usage of a particular area (like bathroom or kitchen), as calculated in section 3.4. To make the solution adaptable to diverse usage patterns, varying sampling frequencies and fittings in different water systems, the proposed water disaggregation algorithm is designed such that it doesn't require labelled training data to learn usage patterns, but rather takes as input only general specifications about the system configuration and equipment used, which is mostly known by domain experts, or can easily be found or measured. This includes the number of fixtures N (different operating modes of each fixture counted separately), their corresponding volume consumed in one usage $v^{(n)}$, and approximate flow-rates $r^{(n)}$. For example, shower uses 4-8 liters/min, while a toilet flush uses 6-10 liters per usage. The proposed algorithm uses combinatorial optimization to find the best subset of fixtures for a given usage instance of continuous non-zero flow. For a usage instance θ with total volume consumed V, we solve:

$$\arg max_{k^{(n)}} \prod_n p(k^{(n)}v|\theta)$$

$$\text{such that } (1-\sigma)V \leq \sum_n k^{(n)}v \leq (1+\sigma)V$$

$$\text{where } k^{(n)} \in \mathbb{Z}, n \in [1, N]$$

where σ is the error margin, and $p(k^{(n)}v|\theta)$ is calculated using general heuristics such as possibility of a fixture's contribution given its flow rate ($r^{(n)}$) and actual water used in each window w of θ, and the approximate number of large events occurring per day. For example, the number of bathing events per day are used to eliminate non-viable combinations. This is possible only when pressure sensor data is not analysed in real time. Due to its inherent approach of combinatorial optimization rather than pattern matching, the proposed method is suitable for overlapping events, where current approaches struggle to perform good. We do preliminary evaluation of the presented water disaggregation algorithm on an aggregate water consumption data of a house in residential building collected at an interval of 5 min, where labelled data of fixture-level usage in a bathroom was available for a duration of 1 week. Overall accuracy of top-3 suggestions of presented method was 78%. Figure 4 shows the temporal water meter data showing total consumption and disaggregation output of our proposed method.

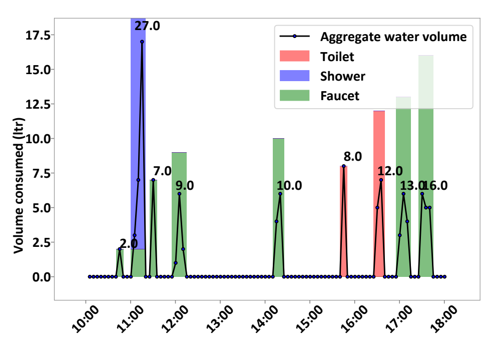

Figure 4: Disaggregated events for total water consumption of a bathroom in a residential apartment.

Table 1: F1 score (Open/Close events) for different models and datasets

Dataset/Model	ResNet[6]	InceptionTime[7]	TST[18]	LSTM-FCN[12]	ConvTran[9]
FordA	1.0/1.0	1.0/0.937	1.0/0.876	1.0/0.968	1.0/0.811
StarLightCurves	1.0/1.0	1.0/1.0	0.934/0.846	1.0/0.968	1.0/0.938
ECG5000	1.0/1.0	0.968/0.937	0.746/0.875	1.0/0.937	1.0/0.968
UWaveGestureLibraryAll	1.0/1.0	0.968/0.937	0.938/0.685	1.0/0.937	0.968/1.0
PressureLD	1.0/1.0	0.968/0.968	0.938/0.841	1.0/0.968	1.0/0.938

5 CONCLUSION AND FUTURE WORK

This paper introduces `PressureML`, an end-to-end method for deriving water usage insights in large-scale buildings, presenting a two-stage solution by first determining aggregate water consumption per floor area by analyzing pressure waves, and then employing an innovative disaggregation algorithm for fixture-level insights. Our approach demonstrates promising potential for accuracy and scalability, as evidenced by our evaluations. Future work will extend these evaluations to confirm the practicality of our approach, and explore important avenues such as handling overlapping events, utilizing physics-based simulations for improved feature learning, testing cross-deployment adaptability, and optimizing the disaggregation algorithm's search space. Our contributions aim to advance the understanding and implementation of water usage insights in real-world scenarios.

REFERENCES

[1] 2022. Designing Domestic Water Systems for High-Rises. https://www.phcppros.com/articles/16415-designing-domestic-water-systems-for-high-rises.

[2] 2022. Water hammer waves. https://web.eng.fiu.edu/arleon/courses/Transient_flows/Lectures/Waterhammer.pdf.

[3] Mohsen Aghashahi, Lina Sela, and M Katherine Banks. 2023. Benchmarking dataset for leak detection and localization in water distribution systems. *Data in Brief* 48 (2023), 109148.

[4] Davide Carboni, Alex Gluhak, Julie A. McCann, and Thomas H. Beach. 2016. Contextualising Water Use in Residential Settings: A Survey of Non-Intrusive Techniques and Approaches. *Sensors* 16, 5 (2016). https://doi.org/10.3390/s16050738

[5] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. 2019. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica* 6, 6 (2019), 1293–1305.

[6] Fawaz et al. 2019. Deep learning for time series classification: a review. *Data mining and knowledge discovery* 33, 4 (2019), 917–963.

[7] Fawaz et al. 2020. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery* 34, 6 (2020), 1936–1962.

[8] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2018. Transfer learning for time series classification. In *2018 IEEE international conference on big data (Big Data)*. IEEE, 1367–1376.

[9] Navid Mohammadi Foumani, Chang Wei Tan, Geoffrey I Webb, and Mahsa Salehi. 2023. Improving Position Encoding of Transformers for Multivariate Time Series Classification. *arXiv preprint arXiv:2305.16642* (2023).

[10] Jon E Froehlich, Eric Larson, Tim Campbell, Conor Haggerty, James Fogarty, and Shwetak N Patel. 2009. HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In *Proceedings of the 11th international conference on Ubiquitous computing*. 235–244.

[11] Brian Kenji Iwana and Seiichi Uchida. 2021. An empirical survey of data augmentation for time series classification with neural networks. *PLOS ONE* 16, 7 (07 2021), 1–32. https://doi.org/10.1371/journal.pone.0254841

[12] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2017. LSTM fully convolutional networks for time series classification. *IEEE access* 6 (2017), 1662–1669.

[13] Bo Liu, Zhenguo Zhang, and Rongyi Cui. 2020. Efficient Time Series Augmentation Methods. In *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. 1004–1009. https://doi.org/10.1109/CISP-BMEI51763.2020.9263602

[14] Laura Pastor-Jabaloyes, FJ Arregui, and R Cobacho. 2018. Water end use disaggregation based on soft computing techniques. *Water* 10, 1 (2018), 46.

[15] Anders L Sønderlund, Joanne R Smith, Christopher J Hutton, Zoran Kapelan, and Dragan Savic. 2016. Effectiveness of smart meter-based consumption feedback in curbing household water use: Knowns and unknowns. *Journal of Water Resources Planning and Management* 142, 12 (2016), 04016060.

[16] Vijay Srinivasan, John Stankovic, and Kamin Whitehouse. 2011. Watersense: Water flow disaggregation using motion sensors. In *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. 19–24.

[17] Bingsheng Wang, Zhiqian Chen, Arnold P Boedihardjo, and Chang-Tien Lu. 2018. Virtual metering: An efficient water disaggregation algorithm via nonintrusive load monitoring. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 4 (2018), 1–30.

[18] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. 2021. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2114–2124.