Brick-by-Brick: Cyber-Physical Building Data Classification Challenge

Arian Prabowo
arian.prabowo@unsw.edu.au
University of New South Wales (UNSW)
Sydney, NSW, Australia

Xiachong Lin
dawn.lin@student.unsw.edu.au
UNSW
Sydney, NSW, Australia

Imran Razzak
imran.razzak@unsw.edu.au
UNSW
Sydney, NSW, Australia

Hao Xue
hao.xue1@unsw.edu.au
UNSW
Sydney, NSW, Australia

Matthew Amos
matt.amos@csiro.au
Energy, CSIRO
Newcastle, NSW, Australia

Stephen D. White
stephen.d.white@csiro.au
Energy, CSIRO
Newcastle, NSW, Australia

Flora D. Salim
flora.salim@unsw.edu.au
UNSW
Sydney, NSW, Australia

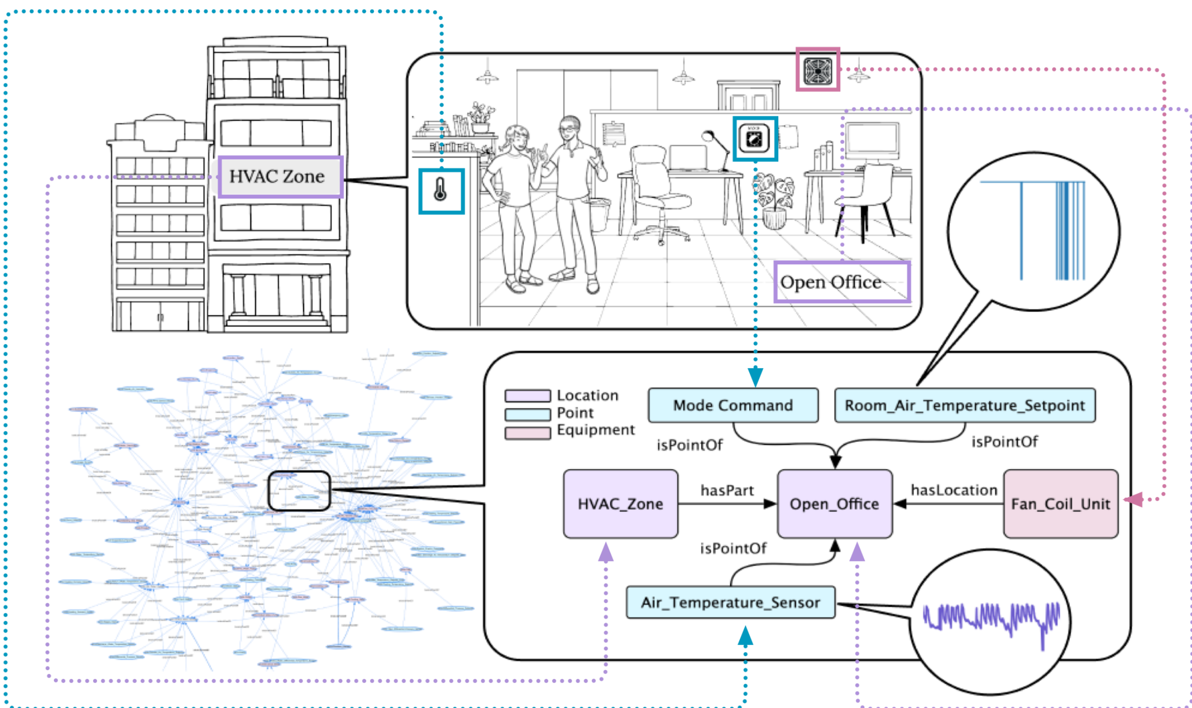

Figure 1: Scaling up the web integration of buildings is important to improve human wellbeing, energy efficiency, and tackling climate change. A significant barrier is in automating the semantic labeling of data streams from IoT devices in buildings. In the Brick-by-brick challenge (www.aicrowd.com/challenges/brick-by-brick-2024), the participants classify the timeseries from IoT devices, according to a standardised semantic schema. Figure is from [11]

Abstract

Buildings account to nearly one-third of global energy consumption and carbon emissions, making their optimization essential in combating climate change. Cyber-Physical Buildings, enabled by the integration of Internet-of-Things (IoT) devices and advanced data analytics tools like AI, offer a smart and effective approach to energy management. A key challenge, however, lies in automating the semantic labeling of IoT devices to ensure machine-interpretable data. The "Brick-by-Brick: Cyber-Physical Building Data Classification Challenge" aims to tackle this challenge by classifying time-series data from IoT devices within buildings. Participants will engage with a dataset consisting of over 10,000 time-series streams collected over three years across three buildings, representing 91 unique semantic classes. Both the dataset and baselines are established in a published paper. With a total prize pool of 20,000 AUD, the competition is ready to launch in December 2024 and run through February 2025, hosted by AIcrowd. This challenge invites researchers, practitioners, and technologists to drive AI-enabled solutions for advancing the next generation of environmentally sustainable cyber-physical buildings. Additional details on the dataset, benchmark, and code can be found in the official repository (https://github.com/cruiseresearchgroup/DIEF_BTS). The challenge was published on AIcrowd www.aicrowd.com/challenges/brick-by-brick-2024.

CCS Concepts

• Information systems → Sensor networks; **Data stream mining**; **Data analytics**; **• Computing methodologies → Ontology engineering**; **Machine learning**.

Keywords

building, timeseries, ontology, machine learning, classification

ACM Reference Format:

Arian Prabowo, Xiachong Lin, Imran Razzak, Hao Xue, Matthew Amos, Stephen D. White, and Flora D. Salim. 2025. Brick-by-Brick: Cyber-Physical Building Data Classification Challenge. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3701716.3718483

1 Introduction

The building sector is a predominant energy consumer, using around 40% of the world's energy. It also responsible for 30% of global CO2 emissions according to the International Energy Agency (IEA) [4, 13]. Therefore, a significant improvement in building energy efficiency is necessary to reduce energy consumption and achieve net-zero emissions. Incorporating digital technologies such as data-driven analytics applications with Building Management Systems (BMS) can help to achieve this goal by providing advanced tools and models for monitoring energy usage, detecting wasteful and inefficient operations, demand management [2]. The utilisation of such analytics applications requires classification (or mapping) numerous heterogeneous points (e.g., sensor, control command, setpoint, alarm, etc.) according to a uniform metadata schema (like Brick ontology) for buildings. Unfortunately, most of the existing buildings uses different metadata conventions/schema designed to satisfy their unique needs.

This restricts widespread utilisation of energy analytics applications among millions of existing and new buildings, and makes them tightly coupled to the specific buildings for which they are developed for. The current practice to classifying (or mapping) the points in buildings by following a standard schema is a manual process which is time and labour intensive. It also requires significant effort from highly specialised domain experts, however, yet ultimately, is an exercise that is susceptible to errors.

Table 1: Data partition statistics. "LB" is the public leader board

	BTS_A	**BTS_B**	**BTS_C**
n_StreamID	8,349	851	5,347
Timeseries Size (zip)	8.48 GB	1.31 GB	8.98 GB
Train Date Start	01-Jan-21	01-Jan-21	01-Jan-21
Train Date End	31-Mar-21	24-Dec-21	01-Jan-21
Test LB Date Start	01-Apr-21	01-Jan-22	01-Jan-21
Test LB Date End	24-Dec-22	24-Dec-22	24-Dec-21
Test Secret Date Start	01-Jan-23	01-Jan-23	01-Jan-22
Test Secret Date End	24-Dec-23	24-Dec-23	24-Dec-23
	Train	**Test LB**	**Test Secret**
n_StreamID	8,653	13,500	13,570
n_chunk	31,839	165,527	163,912
Timeseries Size (zip)	1.29 GB	11.54 GB	

To streamline this mapping process and facilitate the widespread portability of analytics applications, one approach is to automatically classify the points based their time-series data recorded over time, according to the Brick ontology (https://brickschema.org). These time series encompass a variety of data types, including sensors, equipment operational status, control signal, electrical metering, etc. Automating this classification process would significantly reduce the time and resources required to integrate data-driven tech solutions into buildings, thereby democratizing access to innovative building management solutions and advancing the goal of sustainable and efficient building operations [1, 3, 5–7, 12].

In this challenge, the task is to classify heterogeneous building points by using their time series data (or features from such data) as input, according to the Brick ontology. Check the official repository for the dataset and codes.

2 Dataset and Baselines

The time series data comes from 3 anonymised buildings in Australia. We combine all the time series data from the 3 buildings into one, before partitioning them into training data, leader board test

ACM ISBN 979-8-4007-1331-6/2025/04
https://doi.org/10.1145/3701716.3718483

data, and secret competition test data. The partitioning is purposedly non-proportional. The idea being, we are testing the generalisation capabilities of the algorithms to buildings with various degree of representation in the training data. Approximately, 20%, 45%, and 35% of the time series are assigned to the private, public, and secret sets, respectively. However, note that the length of each time series can vary. More details about the partition statistic can be found in Table 1. Finally, we split each timeseries data into shorter duration data sets of different length. Similarly, we split it to evaluate the algorithms performance with various observation windows.

More information about the dataset, including how to access it, as well as the code to reproduce the baselines, are available from the official repository. Each time series consists of a chronologically ordered timestamp and value pair from a point in a building, such as sensor (e.g. degrees Celsius readings from a temperature sensor) or setpoint (e.g. temperature setpoint related to a Variable Air Volume box), or alarm (e.g., High CO2 Alarm), etc. There is no additional information such as units. The participants can expect the timeseries to have irregular sampling rate. The timestamp is in relative time, not absolute time. The timeseries data has not been cleaned, and typical issues typically found with raw signal data can be expected. The timeseries durations range from two to eight weeks.

Table 1 describes how the data is partitioned. Note that the training data contains zero data from BTS_C. This is on purpose to evaluate the capabilities of the proposed algorithms on a completely new buildings. Roughly speaking the number of stream ID in the training partition is BTS_A + BTS_B, while for the testing dataset, are the stream ID from all three. However, they are not the same, as some StreamID are simply labelled as points, or only contains zero values, and thus are removed. The participants will be provided the timeseries data for the entire test sets and will submit predictions on the entire test sets as well. Approximately half of the submissions will be evaluated for the leader board. We then split each timeseries data into shorter duration data called chucks. The challenge is to classify each chunk.

This dataset and the relevant baselines are described in details in the official release paper [11]. More information can also be obtained via the official repository. The dataset will be released under CC BY 4.0. Its prior use in various research studies [8–10] indicates its quality and suitability for use in a competition.

3 Problem Statement and Evaluation metrics

This is a multilabel timeseries classification task. Figure 2 provide visualisation of the task. The labels are 94 different Point sub-classes in a modified Brick schema, that are found in the buildings in the dataset. This is based on Brick ontology version 1.2.1 (https://brick.andrew.cmu.edu/ontology/1.2/classes/Point/). We modified the existing Brick schema by combining some of labels together. For example, various different Alarms are combined into a single class. Each timeseries can have multiple positive (the true label and all their parent class), zero (all their subclasses), and negative labels (all others). The idea is not penalising the participants for making a prediction that is more specific than the ground truth. Once this is established, typical classification metrics such as accuracy or F1 can be derived. We are using F1 as the evaluation metric.

4 Baselines

Table 2: Baseline results in Public Leaderboard Partition. The main metric is the (macro) F1-score

Method	Accuracy	**F1-score**	mAP
Zero	0.977	0.000	0.000
Random Uniform	0.500	0.036	0.023
Random Proportional	0.956	0.010	0.023
Mode	0.966	0.000	0.000
One	0.023	0.040	0.023
Transformer default	0.960	0.020	0.023
Transformer [11]	0.969	0.015	0.023
Transformer hp-tuned	0.879	0.031	0.023

We performed further exploration into this task and dataset by conducting a series of experiments to established multiple baselines. Previously [11], we have performed a series of benchmarking experiment on the same task with a slightly different setup and metrics. For brevity, we do not elaborate various experimental details here as they can be found in the previous paper [11].

For this competition, we repeated the best performing model (transformer) with the new setup, with three different hyperparameters: default, the one used in the previous benchmark, and tuned for the new setup. Hyperparameter tuning provided slight improvements as expected. The result is shown in Table 2. The best performing model is the naive model **One**, where it simply returns true for every label. This indicates that this is an unsolved problem with significant potential for new discoveries.

5 Competition Format, Implementation Details, and Timeline

Participants will utilise the AIcrowd (https://www.aicrowd.com/) platform to submit CSV file containing their predictions. Allowed submission frequency/timing will be 10 times a day. The performance on the leader board test set will be published on the leader board. The winner will be determined by the combined score of their final submission across both leader board and secret set. Q&A will be done using the official forum on the AICrowd platform. We divided the competition to 2 rounds, to allow possible revisions if such need may arise.

The standard competition rules from AICrowd is used to ensure fairness. This includes requirements for documentation of solutions and code validations. Two major additional terms are:

(1) Participants can only train the model using the provided dataset. No other external datasets are allowed.
(2) The organiser will provide a minimum performance requirement to be eligible for the prizes.

Timeline:

- Monday, 9 December 2024: Launch Round 1
- Wednesday, 9 January 2025: Launch Round 2
- Monday, 3 February 2025: Completion of Competition

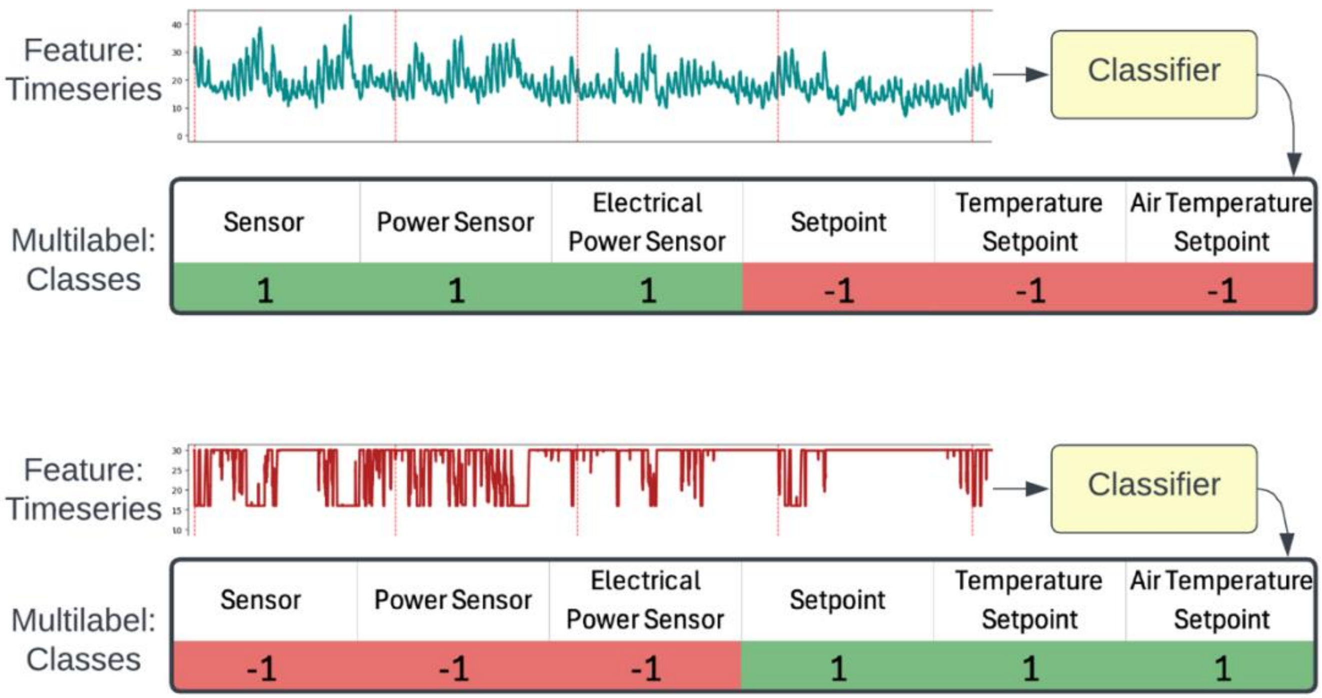

Figure 2: Visualisation of the multi-label timeseries classification task.

- Friday, 7 February 2025: Winner Announced
- 28 April - 2 May 2025: The Web Conf. workshop

6 Organising team

The members of the competition organising team is gender balanced and diverse in terms of their career stages and cultural background. We are committed to encourage broad participation to the competition through the AICrowd platform.

This challenge involves key contributors and collaborators from various institutions. The primary contact for the platform is AICrowd. The primary contacts for the challenge are Arian Prabowo (UNSW) and Flora D. Salim (UNSW), with Xiachong (Dawn) Lin (UNSW, CSIRO) as the secondary contact. The primary contact for the dataset is Matthew Amos (CSIRO). Other coordinators include Imran Razzak (UNSW), Hao Xue (UNSW), and Stephen D. White (CSIRO). Emily W. Yap (UOW) is a collaborator. Members of the competition steering committee include Frank Zeicher (IoT Alliance Australia), Tracey Colley (RACE2030), Clayton Miller (NUS), Clayton McDowell (UOW), Igor Sartori (SINTEF), and Gabe Fierro (NREL). Participating organizations are the Energy business unit of Commonwealth Scientific and Industrial Research Organisation (CSIRO Energy), the University of New South Wales (UNSW), and RACE for 2030.

7 Awards

The total prize pool is $20,000 AUD, to be distributed to 5 winners, including the in-person presentation grant to be paid to winning teams who present their solutions in person in the WWW'25 conference. This remuneration is to be paid after the presentation.

Gold 1 winner of $5,000 prize money and $2,500 grant.
Silver 1 winner of $3,000 prize money and $2,000 grant.
Bronze 3 winners of $1,000 prize money and $1,500 grant.

The winners are strongly encouraged to submit their report to the WWW workshop conference proceeding, as well as invitation to co-author the competition paper. Additional certificate for honourable mentions might be available.

8 Relevance to the Web Conference

This competition focuses on accelerating the integration of web-of-things data from buildings into the Web, addressing a critical need for machine-readable, interoperable data. It bridges the cyber-physical web data gap by integrating physical objects within buildings into the Web through the construction of a semantic graph, leveraging a standardized ontology in the building domain, brought together with a large-scale web-of-things (WoT) data for the very first time. By automating the semantic labeling of WoT device data, it bridges the gap between buildings and Web-based frameworks, enabling smarter, more sustainable management systems. Using AI and machine learning, the competition brings us one step closer to tackling pressing sustainability issues, such as reducing energy consumption and carbon emissions. This alignment between AI-driven solutions and Web technologies highlights the potential for transformative advancements in sustainable urban development, making this competition a valuable contribution to The Web Conference community.

9 Conclusion

The "Brick by Brick: Automating Building Data Classification Challenge" represents a significant step forward in addressing the global challenge of building sustainability through the integration of Web technologies, AI, and data science. By automating the semantic labeling of IoT device data, this competition enables smarter, more efficient building management systems, reducing energy consumption and carbon emissions on a global scale.

Through a diverse and dataset, well-established baselines, and rigorous evaluation metrics, the challenge provides a platform for researchers and practitioners to push the boundaries of AI-driven solutions in environmental sustainability. With a substantial prize pool and opportunities for real-world impact, the competition bridges innovation and practical application, encouraging collaboration and advancement in the domain of smart buildings.

Acknowledgments

This research is supported by the NSW Government through the CSIRO's NSW Digital Infrastructure Energy Flexibility (DIEF) project, funded under the Net Zero Plan Stage 1: 2020-2030.

This project is also funded by the Reliable Affordable Clean Energy for 2030 (RACE for 2030) Cooperative Research Centre.

This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS enabled capability supported by the Australian Government.

References

[1] Mahathir Almashor, Mashud Rana, John McCulloch, Ashfaqur Rahman, and Subbu Sethuvenkatraman. 2023. What's The Point: AutoEncoding Building Point Names. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 256–260.

[2] David Blum, José Candanedo, Zhelun Chen, Gabe Fierro, Virginia Gori, Hicham Johra, Henrik Madsen, Anna Marszal-Pomianowska, Zheng O'Neill, Ojas Pradhan, Dimitrios Rovas, Francesco Sacco, Sofia Stensson, Christian A Thilker, Charalampos Vallianos, Jin Wen, and Stephen White. 2023. *Data-Driven Smart Buildings: State-of-the-Art Review*. CSIRO, Australia. https://doi.org/10.4225/08/6219e931e104b

[3] Jingkun Gao, Joern Ploennigs, and Mario Berges. 2015. A data-driven meta-data inference framework for building automation systems. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. 23–32.

[4] M. González-Torres, L. Pérez-Lombard, Juan F. Coronel, Ismael R. Maestre, and Da Yan. 2022. A review on buildings energy information: Trends, end-uses, fuels and drivers. *Energy Reports* 8 (2022), 626–637. https://doi.org/10.1016/j.egyr.2021.11.280

[5] Dezhi Hong, Hongning Wang, Jorge Ortiz, and Kamin Whitehouse. 2015. The building adapter: Towards quickly applying building analytics at scale. In *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments*. 123–132.

[6] Jason Koh, Bharathan Balaji, Dhiman Sengupta, Julian McAuley, Rajesh Gupta, and Yuvraj Agarwal. 2018. Scrabble: transferrable semi-automated semantic metadata normalization using intermediate representation. In *Proceedings of the 5th Conference on Systems for Built Environments*. 11–20.

[7] Jason Koh, Dezhi Hong, Rajesh Gupta, Kamin Whitehouse, Hongning Wang, and Yuvraj Agarwal. 2018. Plaster: An integration, benchmark, and development framework for metadata normalization methods. In *Proceedings of the 5th Conference on Systems for Built Environments*. 1–10.

[8] Xiachong Lin, Arian Prabowo, Imran Razzak, Hao Xue, Matthew Amos, Sam Behrens, and Flora D Salim. 2024. BiTSA: Leveraging Time Series Foundation Model for Building Energy Analytics. *proceedings of IEEE International Conference on Data Mining* (2024).

[9] Xiachong Lin, Arian Prabowo, Imran Razzak, Hao Xue, Matthew Amos, Sam Behrens, and Flora D Salim. 2024. Exploring Capabilities of Time Series Foundation Models in Building Analytics. *arXiv preprint arXiv:2411.08888* (2024).

[10] Xiachong Lin, Arian Prabowo, Imran Razzak, Hao Xue, Matthew Amos, Sam Behrens, and Flora D Salim. 2025. A Gap in Time: The Challenge of Processing Heterogeneous IoT Data in Digitalized Buildings. *IEEE Pervasive Computing Magazine* (2025).

[11] Arian Prabowo, Xiachong Lin, Imran Razzak, Hao Xue, Emily W Yap, Matthew Amos, and Flora D Salim. 2024. BTS: Building Timeseries Dataset: Empowering Large-Scale Building Analytics. *Thirty-Eighth Annual Conference on Neural Information Processing Systems* (2024).

[12] Mashud Rana, Ashfaqur Rahman, Mahathir Almashor, John McCulloch, and Subbu Sethuvenkatraman. 2023. Automatic Classification of Sensors in Buildings: Learning from Time Series Data. In *Australasian Joint Conference on Artificial Intelligence*. Springer, 367–378.

[13] Flora D Salim, Bing Dong, Mohamed Ouf, Qi Wang, Ilaria Pigliautile, Xuyuan Kang, Tianzhen Hong, Wenbo Wu, Yapan Liu, Shakila Khan Rumi, et al. 2020. Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey. *Building and Environment* 183 (2020), 106964.