ORIGINAL RESEARCH Open access

Integrating Context-Aware Transformers and Reinforcement Learning for Intelligent District Heating Control

Fatima El Idrissi[1*], Samira Bennani[1]

Abstract

With district heating systems progressing toward climate-neutral targets, managing their operations amid varied building features and shifting weather patterns poses a persistent difficulty. Conventional control techniques generally lack sufficient flexibility to handle individual building behaviors and to ensure immediate compliance with critical safety requirements. This paper introduces a unified machine learning system that fuses an adaptive context-aware transformer with deep reinforcement learning to resolve these issues. The developed method positions the adaptive context-aware transformer as a predictive simulation environment inside a Deep Q-Network (DQN) setup, facilitating intelligent, tailored regulation of district heating operations. Drawing on actual data collected from 148 residential buildings located in Sweden and Finland, the system applies contextual embeddings along with time-based inputs to accurately project indoor temperature paths, reaching root mean square error levels of 0.18–0.24 °C in Swedish structures and 0.26–0.32 °C in Finnish ones. The DQN controller applies these projections to fine-tune heating supply, all while respecting safety boundaries and protecting resident comfort. Testing outcomes indicate major reductions in energy use, particularly with mid-rise buildings attaining as much as 14.85% lower consumption, alongside top seasonal cuts above 20% in spring months. This combined strategy highlights strong opportunities for enhanced energy efficiency and stable indoor environment regulation within modern district heating infrastructures.

Keywords Residential buildings, Adaptive context-aware transformer, District heating, Energy optimization, Reinforcement learning (RL)

*Correspondence:
Fatima El Idrissi
fatima.idrissi@gmail.com
[1] Department of AI Applications in Smart Grids, Faculty of Engineering, Mohammed V University, Rabat, Morocco

Introduction

District heating is an essential part of eco-friendly city energy infrastructure, providing heat to buildings from central sources via well-insulated pipe networks. In Sweden, this system underpins residential heating, covering about 93% of apartment complexes. It holds nearly 62% of the total heating market in residential and service areas [1]. In 2023, Swedish district heating supplied 49.8 TWh of thermal energy, with residential buildings accounting for roughly 31.2 TWh — enough to warm more than 2 million households.

Even with its widespread use, effective regulation of district heating remains a notable challenge. Standard control methods, which mostly rely on outdoor temperature adjustment curves and simple feedback loops, often cannot adapt to the diverse thermal profiles of separate buildings or evolving occupant habits. These weaknesses are exacerbated by Sweden's harsh, variable climate, as highlighted in Kensby [2]'s examination of thermal energy storage in homes and Mangold *et al*. [3]'s extensive research on energy performance variability across the Swedish building inventory.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Successfully tackling these issues requires context-aware regulation that embeds building-specific details — including location, construction traits, number of floors, occupant density, and usage category — directly into decision-making. Here, context-awareness means the system's ability to understand and respond to each building's unique thermal behavior, leading to better indoor temperature forecasts and more personalized adjustments. Absent this level of adaptation, classic models tend toward oversimplification, leading to wasted energy, poor comfort, and unnecessary consumption.

In addition to reducing energy demand and maintaining comfort, control strategies must strictly adhere to operational safety rules. Supply pipe temperatures need to stay within secure limits to prevent damage, equipment wear, or discomfort for users. Consequently, performing optimization while meeting these strict safety conditions adds a significant layer of complexity.

Advances in artificial intelligence, such as Wang *et al*. [4]'s thorough comparison of shallow and deep learning for predicting building thermal loads and Li *et al*. [5]'s creative transformer architecture for forecasting cooling demand, provide valuable tools for managing these challenges.

Still, current solutions have three important shortcomings: (1) limited use of building-specific context in prediction models [5]; (2) weak handling of interactions between indoor, outdoor, and supply temperatures [6]; and (3) few reinforcement learning methods capable of minimizing energy use without compromising comfort or safety [7]. To fill these gaps, this research proposes an advanced integrated system that improves upon existing work through:

• Upgrading our earlier Adaptive Context-Aware Transformer [8] to act as a realistic environment simulator for reinforcement learning, delivering precise and responsive indoor temperature estimates for different buildings under changing control inputs.

• Deploying a Deep Q-network (DQN) agent that treats the adaptive context-aware transformer as its simulation environment to enable responsive heating adjustments under safety limits.

• Furthermore, testing coordinated control across multiple buildings that generalizes well to varied building categories while protecting each structure's individual results.

Related Work

The integration of artificial intelligence into energy infrastructure is a key domain for confronting international climate issues and advancing energy performance objectives. District heating networks, which meet about 12% of global space heating demand, hold considerable promise for intelligent optimization via state-of-the-art artificial intelligence methods. Employing machine learning in energy infrastructure has become a key topic uniting energy policy specialists and artificial intelligence researchers. Artificial intelligence solutions applied to energy infrastructure have unlocked revolutionary capabilities across numerous sectors, from demand estimation and grid balancing to building energy management. Deep learning frameworks have excelled at detecting complex temporal and spatial trends in energy usage across multiple contexts. Utilizing reinforcement learning within energy regulation systems signals a transition from passive responses to anticipatory energy governance, empowering systems to acquire superior tactics through engagement with an evolving environment.

District heating regulation technologies have advanced substantially over recent decades, shifting from simple temperature control via control curves to complex AI-enabled platforms. Werner [9]'s extensive assessment of district heating growth highlighted this progression. In contrast, Lund *et al*. [10] investigated the integration of renewable energy sources and waste heat utilization systems, revealing how innovations mirror rising expectations for efficiency and long-term viability. Gustafsson *et al*. [11] improved district heating substation operations using revised control methods, while Carli *et al*. [12] designed Internet-of-Things systems for model-predictive regulation of heating, ventilation, and air

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

conditioning systems in intelligent structures. Serale *et al*. [13] analyzed model-predictive regulation to improve energy performance in building heating, ventilation, and air conditioning across diverse climatic scenarios and building types, and Wang *et al*. [4] assessed building thermal load estimation using machine learning methods. Ahmad *et al*. [14] formulated short-term energy estimation methods for district-wide load supervision, while Wang and Hong [15] evaluated the prospects and challenges of reinforcement learning for building oversight. Weinberg *et al*. [16] supplied an extensive survey of reinforcement learning deployments in building energy platforms from a computational viewpoint.

Traditional HVAC/central heating control methods

District heating regulation development can be followed through three primary technology phases, each delivering important advances in performance and regulation capacity:

First-generation systems (1970–1990) chiefly depended on outdoor temperature compensation curves to modify supply temperature, according to records from Gustafsson *et al*. [11]. Werner [9]'s robust evaluation of district heating infrastructures, although dependable, delivered solely elementary capabilities and lacked adaptability to unique building traits or resident usage trends. Worldwide investigations in that timeframe primarily aimed at raising system dependability and foundational efficiency.

Second-generation systems (1990–2010) incorporated Model Predictive Control (MPC), integrating weather projections and building thermal simulations. Investigations by Carli *et al*. [12] revealed that internet-of-things-supported MPC deployments could yield 15%–25% reductions in energy consumption relative to first-generation techniques. Nevertheless, these benefits came with increased computational loads and challenges in maintaining precise building simulations. Equivalent conclusions emerged from Serale *et al*. [13]'s global analyses, which showed that MPC deployments achieved reliable energy reductions across diverse climate conditions.

Today's third generation of regulation systems has incorporated sophisticated elements such as thermal storage applications outlined by Kensby *et al.* [17] and flexible setpoint calibration. Mangold *et al*. [3] examined energy performance statistics from various building inventories, illustrating how disparities in thermal attributes strongly influence regulatory outcomes. This phase has further incorporated renewable energy integration and waste heat recovery mechanisms reviewed by Lund *et al*. [10], demanding even more refined regulation strategies.

Predictive models for energy and sensor data

Machine learning's emergence has revolutionized forecasting techniques in building energy platforms. Initial statistical procedures have been superseded by deep learning strategies, notably for processing complex temporal sequences typical of district heating operations. Wang *et al*. [4] investigated the estimation of building thermal loads using shallow and deep machine learning methods. Worldwide inquiries have indicated ongoing improvements in estimation precision with these contemporary strategies across varied climate regions and building types, as evidenced by Ahmad *et al*. [14] in their district-scale energy estimation analysis.

Transformer designs have shown notable potential here, thanks to their self-attention mechanisms, which are particularly well-suited for time-series estimation in building energy applications. Ni *et al*. [18] highlighted the benefits of deep learning techniques for multi-horizon building energy projections. In contrast, Tian *et al*. [19] reported positive outcomes with data-supported parallel estimation procedures for building energy utilization. While conventional techniques primarily emphasized temporal sequences, recent advances have underscored the need to integrate contextual elements. In our earlier study, we proposed context-aware frameworks for indoor temperature estimation and demonstrated their benefits across different building types [20]. Expanding on this base, we constructed an extensive adaptive context-embedding configuration that further improved estimation precision by directly weaving building-specific attributes into the transformer design [8]. This configuration improvement reaffirmed our initial conclusions and displayed systematic ways to embed contextual information into estimation frameworks. Tian *et al*. [19] and Wang *et al*. [4] demonstrated that combining

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

building context with temporal components has become vital for reliable estimation, especially in settings with diverse building attributes and fluctuating atmospheric conditions [4, 19].

Newer initiatives have addressed specific challenges in district heating operations. Wang *et al.* [4] formulated a combined technique that combined CNN and LSTM frameworks, achieving a 17% boost in estimation precision over standard procedures. This design evolution has proven especially useful for grasping both spatial and temporal linkages in building thermal responses, an observation reinforced by Wei *et al.* [21] in several worldwide applications.

Reinforcement learning-based control

Reinforcement learning delivers a fundamental transformation in building regulation platforms, enabling adaptive and optimization features that surpass those of conventional techniques. Al Sayed *et al.* [22] provided a thorough examination of deep reinforcement learning deployments in HVAC oversight, noting the advantages and challenges in real-world applications across different regions and climates.

Recent investigations have prioritized multi-building regulation tactics. Tian *et al.* [19] illustrated the advantages of DQN-based methods for navigating system complexity while preserving each building's outcomes. This extends the foundational work of Nagy *et al*. [23] and Wang and Hong [15], who verified the viability of deep reinforcement learning for optimal space heating management. Global inquiries have also corroborated these outcomes across various building types and climate zones.

Safety, along with dependability in reinforcement learning-powered controllers, has emerged as a prominent research domain. Deng *et al*. [24] proposed a secure reinforcement learning framework for district heating operations that addresses the essential requirements for constraint satisfaction and stability assurance. This effort aligns with the global efforts of Weinberg *et al*. [16] to create resilient regulatory platforms capable of sustaining outcomes under variable circumstances.

Combining forecasting models with reinforcement learning has led to marked improvements in estimation precision. Mocanu *et al*. [25] demonstrated effective online building energy optimization via deep reinforcement learning, while Dalamagkidis *et al.* [26] achieved a robust balance between thermal comfort and energy savings. These tactics have been confirmed by Langer and Volling [27] across multiple international environments, yielding steady performance gains.

Research gaps

Even with recent developments, three important shortcomings still exist in current solutions, reducing their usefulness in real applications. The first issue is the limited inclusion of detailed building-related context in prediction models, leading to overly basic control methods. Although Wang *et al*. [4] and Ahmad *et al*. [14] reported better results using machine learning, these techniques treat buildings as uniform structures and do not adequately account for details such as structural features, location, or usage type. The second shortcoming involves weak modeling of the complicated links between indoor, outdoor, and supply temperatures, which limits how well control systems perform. Most existing techniques treat temperature factors separately rather than understanding their evolving relationships. Although Tian *et al.* [19] and Ni *et al*. [18] addressed some aspects using improved forecasting tools, the complete modeling of thermal interactions remains underdeveloped. The third limitation is the lack of reinforcement learning approaches that can simultaneously optimize energy use while guaranteeing comfort and system safety. While Al Sayed *et al*. [22] examined various uses and Mocanu *et al.* [25] demonstrated optimization capabilities, most current solutions focus on a single goal and do not balance efficiency, comfort, and safety. These shortcomings led us to develop a combined method that integrates adaptive context-aware transformer models and deep reinforcement learning to address building-specific needs and jointly optimize multiple objectives.

Data Preparation

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Raw data

The dataset for this research is derived from actual sensor readings collected in buildings in Sweden and Finland. It covers buildings located in three Swedish regions (Norrland, Svealand, and Götaland) plus two Finnish regions (Norrland and Österland). EcoGuard AB supplied the information, covering time periods of 3 to 10 years with readings every 15 minutes.

Descriptions of the buildings from both nations appear in **Table 1**. Swedish buildings are entirely residential and spread across three regions: Norrland, with the cities of Skellefteå and Luleå; Svealand, with the cities of Stockholm, Västeras, Karlstad, and Skövde; and Götaland, with the cities of Trollhättan, Gothenburg, and Malmö. Finnish buildings offer more variety, including commercial and care facilities located in two regions: Norrland, with the cities of Oulu and Tyrnävä; and Österland, with the cities of Vantaa, Helsinki, Espoo, Vaasa, Järvenpää, Kronoby, Raisio, Pori, Vihti, and Mikkeli.

Table 1. Information on Buildings data from Sweden and Finland.

Region	City (Location)	Building type	Number of buildings
Norrland	Skellefteå	Residential building	6
	Luleå	Residential building	21
	Oulu	Commercial building	1
	Tyrnävä	Commercial building	1
Svealand	Stockholm	Residential building	16
	Västerås	Residential building	6
	Karlstad	Residential building	27
	Skövde	Residential building	7
Götaland	Trollhättan	Residential building	6
	Gothenburg	Residential building	19
	Malmö	Residential building	12
Österland	Vantaa	Residential + Care + Commercial buildings	8
	Helsinki	Residential + Commercial buildings	3
	Espoo	Commercial building	7
	Vaasa	Residential building	1
	Järvenpää	Residential building	1
	Kronoby	Residential building	1
	Raisio	Residential building	1
	Pori	Residential building	1
	Vihti	Care building	1
	Mikkeli	Commercial + Care buildings	2
Total number of buildings			148

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Three essential sensor categories were installed in every building: indoor temperature sensors in each apartment or home, pipe temperature sensors for district heating input to the building, and outdoor temperature sensors. To obtain a single typical indoor temperature value per building, all indoor readings from measured areas were averaged. As a result, the complete dataset includes three main variables per building: the averaged indoor temperature (T), the outdoor temperature (OT), and the pipe temperature (PT).

For the reinforcement learning part, five buildings located in Stockholm, Sweden, were chosen. These had reliable data covering three full years (2020–2023) and shared similar operating conditions, providing a strong base for testing RL control approaches.

Context-aware data preparation

Alongside the original measurements, we added metadata describing the buildings using various contextual characteristics related to their location and physical makeup. All buildings were sorted by height profile into three groups: Low Rise (LR) for structures with 1–3 floors, Mid Rise (MR) for 4–6 floors, and High Rise (HR) for those with 7 or more floors.

To better show building size in the horizontal direction, the Apartment Density Index (ADI) was created. This measure indicates the spread across the ground level and works together with the vertical grouping. In the ADI calculation, “spaces” refer to apartments in residential buildings and to functional zones (offices, patient rooms, service areas) in other building types. The ADI is calculated as follows:

$$\mathrm{ADI} = \frac{S}{F} \quad (1)$$

Where S represents the number of spaces in the building, and F represents the number of floors.

Full metadata also records the precise floor numbers for each building. **Figure 1a** shows the geographic distribution of buildings across the two countries, with city positions marked. **Figure 1b** displays sample buildings from the different rise groups, while **Table 2** provides metadata examples for typical buildings from various sites and categories.

Table 2. Representation of Meta-data for a subset of Buildings.

Land	Building type	City	Rise	ADI	Floors
Norrland	Residential	Skellefteä	Mid Rise	4	4
Norrland	Commercial	Oulu	Mid Rise	4.50	5
Svealand	Residential	Västerås	High Rise	15	7
Svealand	Residential	Stockholm	High Rise	7.66	8
Svealand	Residential	Karlstad	Low Rise	7	3
Götaland	Residential	Skövde	Mid Rise	8	4
Götaland	Residential	Gothenburg	Mid Rise	18	6
Österland	Care building	Vihti	Low Rise	7	3
Österland	Commercial	Mikkeli	Mid Rise	4	4

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

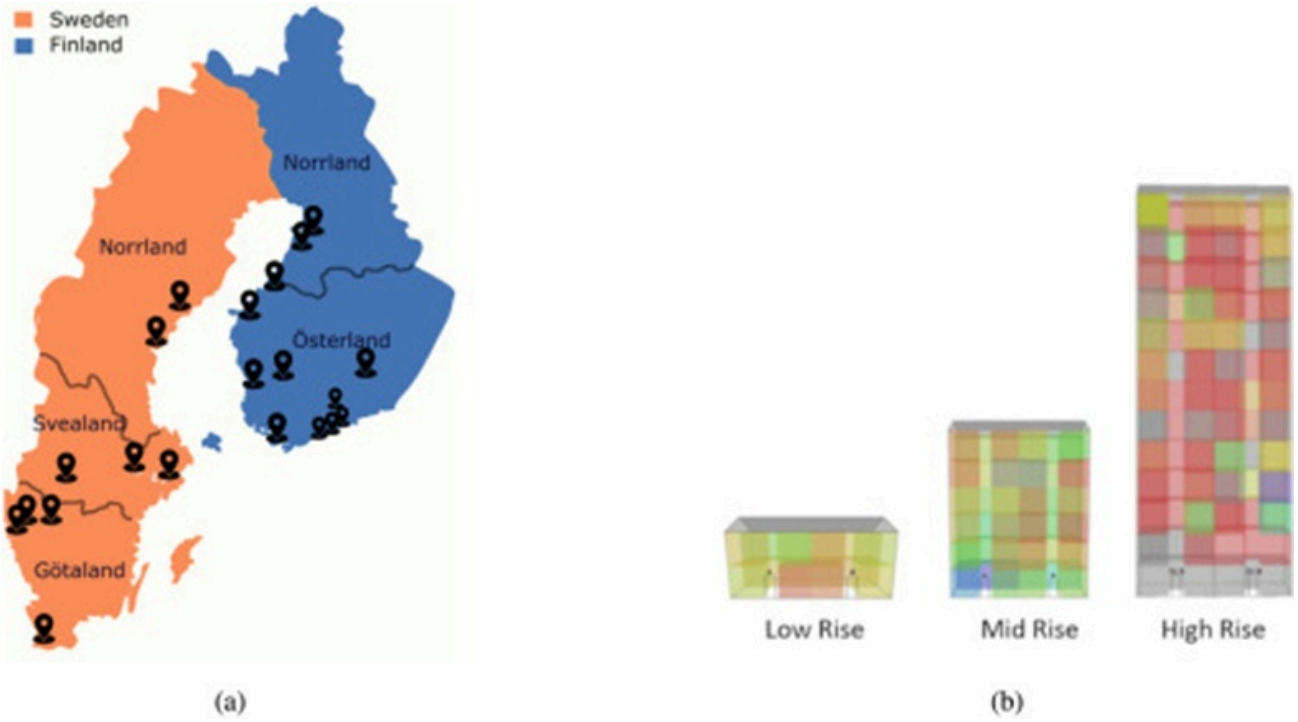


Figure 1. Visualization of Geographical and building context. (a) Illustration of lands in Sweden (orange)- Norrland, Svealand, Götaland, and Finland (blue)- Norrland, Österland, and pinpoint locations of cities. (b) Illustration of residential buildings with the context of Rise.

Data pre-processing

Multivariate time-series data were collected at 15-minute intervals. This data was converted to hourly intervals to meet the operational needs of the district heating control system. The following steps were carried out for each building:

• Normalization of temperature values to the range [−1, 1] to create uniform scaling among different measurements and buildings.

• Temporal features (TF) were pulled from the timestamps to show repeating cycles in heating requirements. These features cover Year, Month, Week of the Year, Day of the Month, Day of the Week, and Hour of the Day. This helps the model detect both long-term seasonal shifts and everyday usage habits.

The full dataset was organized to match actual use conditions, using 70% for the training set, 30% for the validation set, and the newest 30% saved for the test set. This division enabled robust model development and proper testing. For the environment prediction task using the Adaptive Context-aware transformer model, the averaged indoor temperature was the main target, while outdoor and pipe temperatures were the input variables. Temporal features were turned into a positional encoding matrix. The same preparation process was followed for reinforcement learning data to keep everything consistent between prediction and control activities, enabling a combined assessment of both.

Proposed Approach

Finding the best way to operate district heating systems is a key challenge when aiming for high energy efficiency, good occupant comfort, and stable control simultaneously. Standard control techniques rely on reactive methods or simple

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

outdoor-temperature-based curves that do not account for each building's specific thermal features. This uniform approach often results in wasted energy, inconsistent indoor conditions, and reduced comfort. A better solution requires a smart system that learns each building's thermal patterns, predicts upcoming changes, and applies precise adjustments to maintain high comfort while cutting energy use. To solve these problems, we built a combined system that combines deep learning prediction capabilities with reinforcement learning decision-making.

The main new element here is the use of adaptive context-aware transformers as the environment model within the DQN system. This is very different from standard thermal models for buildings in district heating. Normal district heating relies on reactive curves tied to current outdoor temperatures, without looking ahead. Our method implements a predictive control approach in which the transformer environment uses weather forecasts and unique building details to estimate 6-hour temperature paths. This lets the DQN agent choose forward-looking actions for district heating supply. The setup gives two major improvements over usual models: (1) forward-looking control thanks to weather forecast use, and (2) custom building adaptation using contextual embeddings instead of standard thermal values.

The suggested framework offers an integrated system for improving district heating in buildings, shown in **Figure 2**. It connects an adaptive context-aware transformer model to reinforcement learning, forming a closed control loop with four central parts: the Environment (adaptive context-aware transformer), the Reinforcement Learning Agent (Deep Q-Network), the Action Space (pipe temperature changes), and the Reward (based on present conditions and forecasts). The Environment handles building-specific information and temperature inputs, while the RL Agent improves control rules through repeated interaction. Later sections explain the Adaptive Context-aware Transformer architecture, its connection to the DQN agent, the state and action definitions, the multi-objective reward design, and how all these components combine to support district heating management.

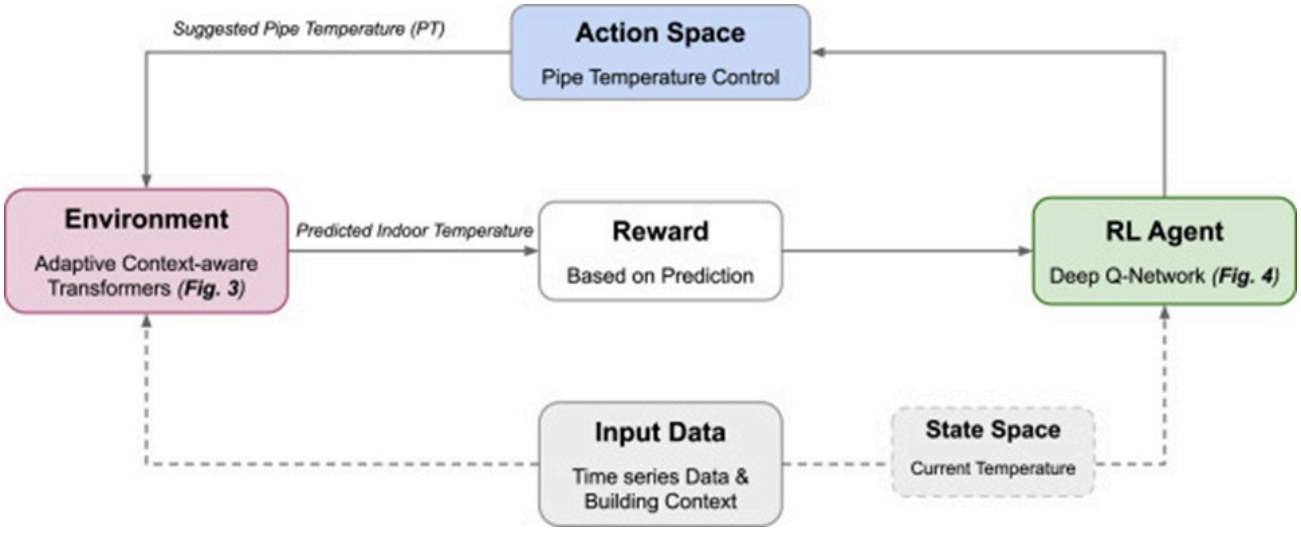


Figure 2. Integrated machine learning framework architecture showing the interaction between adaptive context-aware transformer environment and Deep Q-Network agent for district heating control, including data flow paths from time series sensor data and building context through reward calculation mechanisms to pipe temperature control feedback loops.

Adaptive context-aware transformer as environment

Our previous model [8] serves as the basis for this work, and we extend its adaptive context-aware transformer design to serve as an environment simulator for reinforcement learning. The fundamental architecture remains akin to our earlier contribution. Yet, two significant extensions are introduced: (1) the geographical domain is widened to include Finland together with Sweden, and (2) building type is incorporated as an additional contextual variable.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

The original architecture was built on the standard transformer model [28] and tailored embedding representations to building-specific attributes for indoor temperature forecasting. For the present study, we augment this model to encode a broader set of contextual metadata describing residential buildings, alongside temporal features extracted from multivariate input time series. The contextual metadata now covers Land (the traditional historical regions spanning both Sweden and Finland), City, Rise (building height), ADI (Apartment Density Index), Floors, and Building type. In this setting, 'Lands' designates the traditional historical regions of Sweden and Finland, as visualized in **Figure 1a**. The modified adaptive context-aware transformer architecture, employed as an environment simulator, is shown in **Figure 3.**

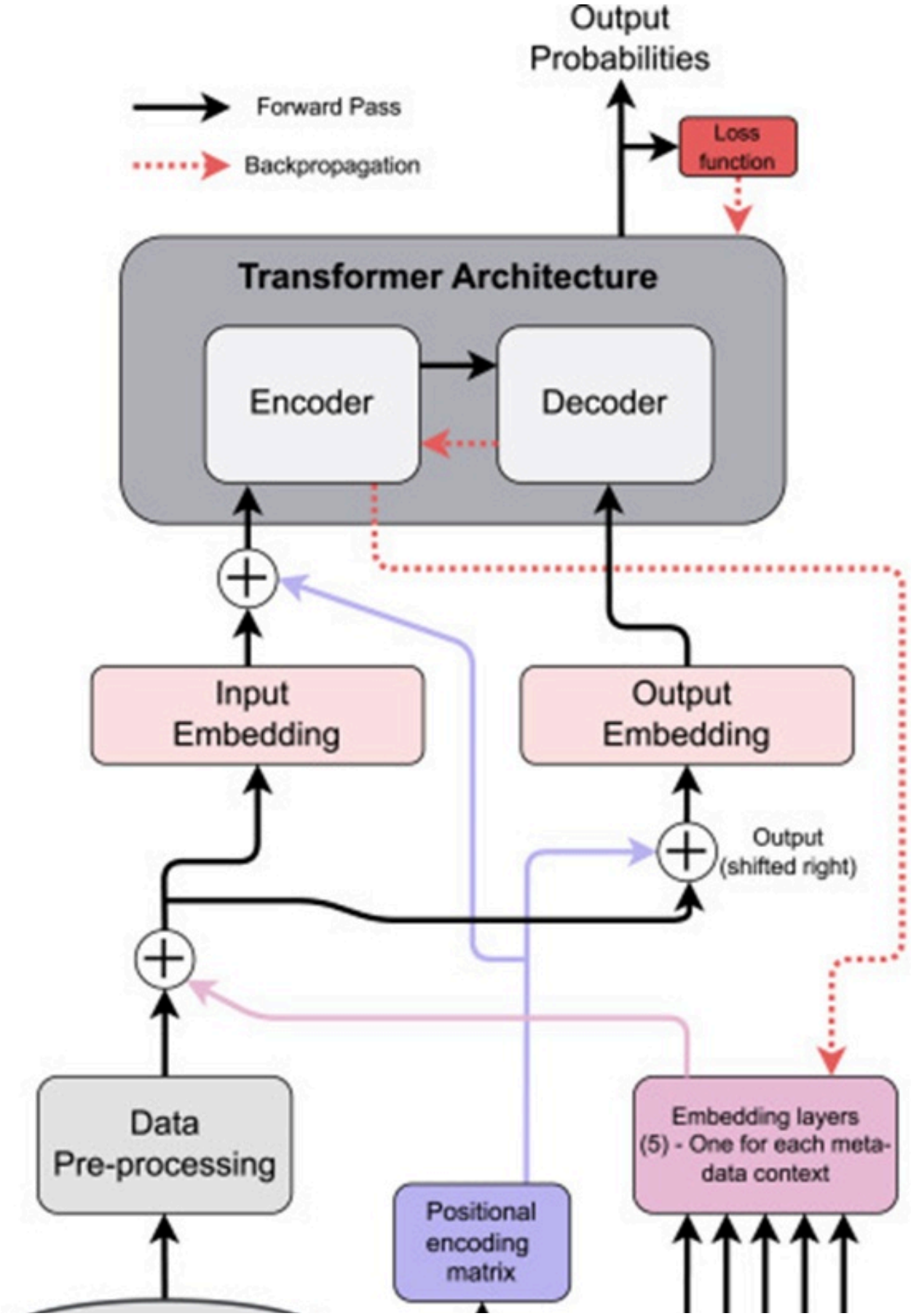
Normalisation is then applied to the temporal features to achieve uniformity. The normalisation for any given temporal feature is obtained through the following computation:

$$TF_{i,norm} = \frac{TF_i - min(TF_i)}{max(TF_i) - min(TF_i)} \quad (2)$$

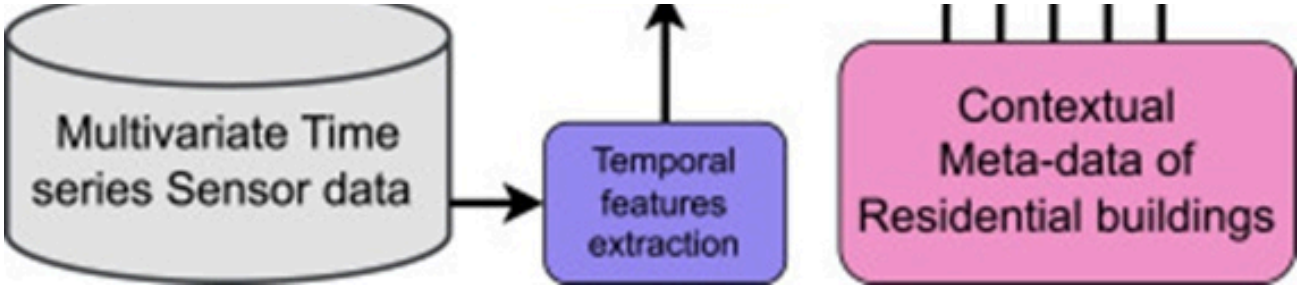
Where indicates the raw temporal feature value, while min() and max() correspond to the minimum and maximum values recorded for that feature within the entire dataset. Once these normalised features are aggregated, they form a temporal positional encoding matrix whose rows capture the timestamped time series data.

The primary contribution of this investigation is the deployment of the expanded adaptive context-aware transformer model as an environment simulator embedded within a reinforcement learning framework. A predictive control loop is thereby established, wherein the model yields accurate indoor temperature predictions as a function of pipe temperature adjustments. The system processes 48 hours of historical data to predict indoor temperature over the next 6 hours, allowing the reinforcement learning agent to evaluate candidate control actions based on their anticipated outcomes. Such a forward-looking mechanism equips the agent to account for both immediate effects and near-term future repercussions of its decisions, giving rise to a proactive control strategy that foresees building thermal responses rather than merely reacting to the present situation.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press



© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press



$$\mathbf{e}_i = E_i(A_i), \quad (3)$$

The aggregated feature vector F for an instance is subsequently assembled by concatenating these embedding vectors:

$$\mathbf{F} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n], \quad (4)$$

where n signifies the total number of features, this vector F, constituting an embedded representation, is supplied to the transformer model together with the pre-processed input time series, thereby permitting the model to draw insights from a rich, high-dimensional feature space.

Fusion of the contextual embeddings with the temporal features is accomplished by concatenation and projection at the transformer's input. Each item of contextual building data (land, city, rise, ADI, floors, building type) is mapped to a 10-dimensional embedding vector (using the optimal embedding size factor). Concatenating these six 10-dimensional embeddings produces a 60-dimensional contextual vector F **(Eq. 4)**. The vector is then projected into the transformer's hidden space (64 dimensions) and summed over all time steps in the 48-hour temporal sequence. As a result, each timestep conveys both temporal cues (time patterns) and building-specific cues (building characteristics), enabling the model to learn how disparate building types react over time.

As in our prior work [8], embedding layers map variables to a high-dimensional vector space of dimension D, matching the transformer's hidden layers. Through gradient descent, these embeddings are iteratively tuned to capture underlying building contextual patterns.

Reinforcement learning (DQN)

Reinforcement Learning (RL) constitutes a machine learning paradigm in which an agent acquires the ability to make sequential choices by interacting with an environment [29]. In this study, we adopt the Deep Q-Network (DQN), a trailblazing deep reinforcement learning algorithm proposed by Mnih *et al*. [30] that combines Q-learning with deep neural networks to handle high-dimensional state spaces. DQN overcomes conventional Q-learning's constraints by employing a neural network to approximate the Q-function, thus managing the complex state–action mappings demanded by building temperature control.

State space

Designing the state space is pivotal for encapsulating the temporal dynamics of building thermal behaviour while upholding computational tractability. At each timestep t, the state is portrayed as a 9-dimensional vector that covers both present conditions and future projections:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$s_t = [T_t^{in}, T_t^{out}, T_t^{pipe}, T_{(t+1:t+6)}^{out}] \quad (5)$$

Where:

- T_t^{in} corresponds to the current indoor temperature (Average Value)
- T_t^{out} denotes the current outdoor temperature
- T_t^{pipe} represents the current pipe temperature
- $T_{(t+1:t+6)}^{out}$ = [T_{t+1}^{out}, ..., T_{t+6}^{out}] captures the 6-hour outdoor temperature forecast

The complete state space S is formally specified as:

$$S = s_t \in \mathbb{R}^9 : T_t^{in}, T_t^{out}, T_t^{pipe} \in [20°C, 70°C] \quad (6)$$

Such a forward-looking state formulation enables the agent to devise anticipatory control strategies. This capability proves especially vital for buildings with considerable thermal inertia, where purely reactive control might fall short. The incorporation of future outdoor temperature predictions permits the agent to foresee and proactively address impending weather variations, while the heating limits reflect the system's physically meaningful constraints.

Action space

The action space reconciles control granularity with system stability by combining discrete steps with scaling factors. At every timestep t, the agent selects a base action drawn from the discrete action space A = {-4.0, -2.0, 0, +2.0, +4.0}, which denotes pipe temperature adjustments in °C. The choice of discrete action increments ({−4.0, −2.0, 0, + 2.0, + 4.0}) was intended to strike a balance between system responsiveness and operational stability. Smaller increments (±1.0 °C) provided insufficient control authority in buildings with elevated thermal inertia, necessitating multiple consecutive actions to realise desired temperature shifts and resulting in slower responses. Increments of larger magnitude (±6.0 °C or greater) induced excessive temperature swings and possible comfort infringements owing to system overshoot, especially within buildings of low thermal mass. The scaling mechanism (7) further hones control granularity in accordance with temperature error, effectively yielding 15 discrete control levels that range from ±0.5 °C to ±6.0 °C. This scheme achieves an optimal compromise between control precision and system stability: finely tuned control near setpoint temperatures and vigorous correction whenever significant deviations occur. The final action is derived by multiplying the base action by a scaling factor ζ:

$$a_{final} = a_t \cdot \zeta, \text{ where } \zeta \in \{0.5, 1.0, 1.5\} \quad (7)$$

The scaling factor ζ is established adaptively based on the temperature error $e_t = |T_{current} + T_{desired}|$:

$$\zeta = \begin{cases} 0.5, & \text{if } e_t \leq 0.5°C \\ 1.0, & \text{if } 0.5°C < e_t \leq 2.0°C \\ 1.5, & \text{if } e_t > 2.0°C \end{cases} \quad (8)$$

Consequently, 15 possible actions emerge through the pairing of five base actions and three scaling factors. As an illustration, a base action of + 2.0 °C may translate into actual adjustments of + 1.0 °C (ζ = 0.5), + 2.0 °C (ζ = 1.0), or +

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

3.0 °C (ζ = 1.5). This scaling design enables more precise control responses: modest adjustments near the target temperature and substantial adjustments when marked corrections are necessary.

To maintain safe operation and prevent thermal shock to the heating system, adaptive safety bounds (20–70 °C) are introduced that dynamically limit the maximum permissible temperature adjustments based on current system conditions, building thermal state, and operational constraints. The 20–70 °C pipe temperature limits delineate the operational envelope of district heating systems, with 20 °C as the lowest temperature the system can deliver and 70 °C as the maximum safe operating temperature to avert thermal stress and equipment damage. Domain experts supplied these safety constraints. Constraints are imposed at the agent level during action selection via a clipping mechanism. Once the DQN agent picks an action, the resulting pipe temperature is computed through:

$$T_{t+1}^{pipe} = \text{clip}(T_t^{pipe} + a_t, 20, 70) \quad (9)$$

This guarantees that every control action resides within safe operational boundaries irrespective of the agent's policy. This arrangement hinders the agent from acquiring potentially hazardous control behaviours while preserving the latitude to examine the entire safe operating spectrum during training.

Reward function

The reward function is engineered to sustain ideal indoor comfort by maintaining building temperatures at a 21 °C target setpoint while balancing multiple control objectives and system stability. The 21 °C indoor temperature target was adopted in accordance with standard comfort specifications for residential buildings. We construct a reward function that incorporates temperature regulation, responsiveness to outdoor conditions, and stability of control actions:

$$R = \alpha R_{temp} + \beta R_{weather} + \gamma R_{stability} + \delta R_{safety} \quad (10)$$

For the temperature maintenance component (R_temp), an exponential decay formulation is employed to create strong incentives for remaining near the 21 °C setpoint, with progressively heavier penalties applied as deviations grow larger:

$$R_{temp} = \begin{cases} 10 \cdot e^{-|T_{pred}-21|}, \text{if } -1.0 \le |T_{pred} - 21| \le 1.0 \\ 10 \cdot e^{-|T_{pred}-21|} - 3|T_{pred} - 21|, \text{otherwise} \end{cases} \quad (11)$$

Where T_pred denotes the predicted indoor temperature, this structure assigns high rewards for keeping temperatures within ±1.0 °C of 21 °C, while imposing steeper penalties for larger departures to ensure occupant comfort.

The weather responsiveness term (R_weather) pushes the agent to modulate pipe temperature in step with outdoor temperature variations:

$$R_{weather} = -|\Delta PT - \lambda \cdot \Delta OT| \quad (12)$$

ΔPT and ΔOT capture the respective changes in pipe and outdoor temperatures, and λ serves as a scaling coefficient that encodes the strength of the correlation between control responses and outdoor temperature shifts.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

The stability component (R_stability) is essential for suppressing sudden temperature oscillations that could compromise both occupant comfort and system efficiency:

$$R_{stability} = -\sum_{t=1}^{n-1}(T_{t+1} - T_t)^2 \quad (13)$$

By levying a quadratic penalty on successive temperature differences, abrupt indoor temperature shifts are discouraged. Summation over the prior n-1 timesteps steers the agent toward smooth temperature progressions rather than sharp adjustments. This consideration carries particular weight in building heating applications, where rapid temperature swings tend to elevate energy consumption and degrade occupant comfort.

Finally, R_safety delivers a binary reward (+ 1.0) whenever pipe temperature stays within the safe operating envelope of 20–70 °C:

$$R_{safety} = \begin{cases} 1.0, & \text{if } 20°C \leq T_{pipe} \leq 70°C \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

The coefficients α = 1.0, β = 0.5, γ = 0.3, and δ = 0.2 were arrived at through hyper-parameter tuning conducted over the training period. Multiple coefficient combinations were trialled and judged by the agent's capacity to preserve temperature stability and comfort. The selected values reflect their success in reconciling the temperature maintenance priority with both weather responsiveness and system stability. To foster stable training dynamics, the aggregate reward is confined within [−10, 10]. This reward formulation effectively teaches the agent to uphold the 21 °C indoor temperature target while adapting suitably to weather changes and system limitations.

DQN architecture

The DQN architecture, visualised in **Figure 4**, realises a complete learning framework for building temperature regulation. It consists of several mutually interacting elements that collaborate to acquire and execute optimal control policies. Central to the system are two neural networks — the DQN Model and the Target Model — both sharing the same configuration: an input layer of 9 neurons, two hidden layers of 256 neurons each, with ReLU activations, and an output layer of 5 neurons matching the discrete action space.

The learning sequence initiates with the Experience Replay Buffer, which accumulates tuples comprising state transitions, actions, and rewards. These stored experiences are sampled during training to sever temporal correlations and bolster learning stability. The Q-Value Calculator subsequently operates on these samples to derive the temporal difference error, drawing upon both the current DQN Model and the Target Model to generate consistent Q-value estimates.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

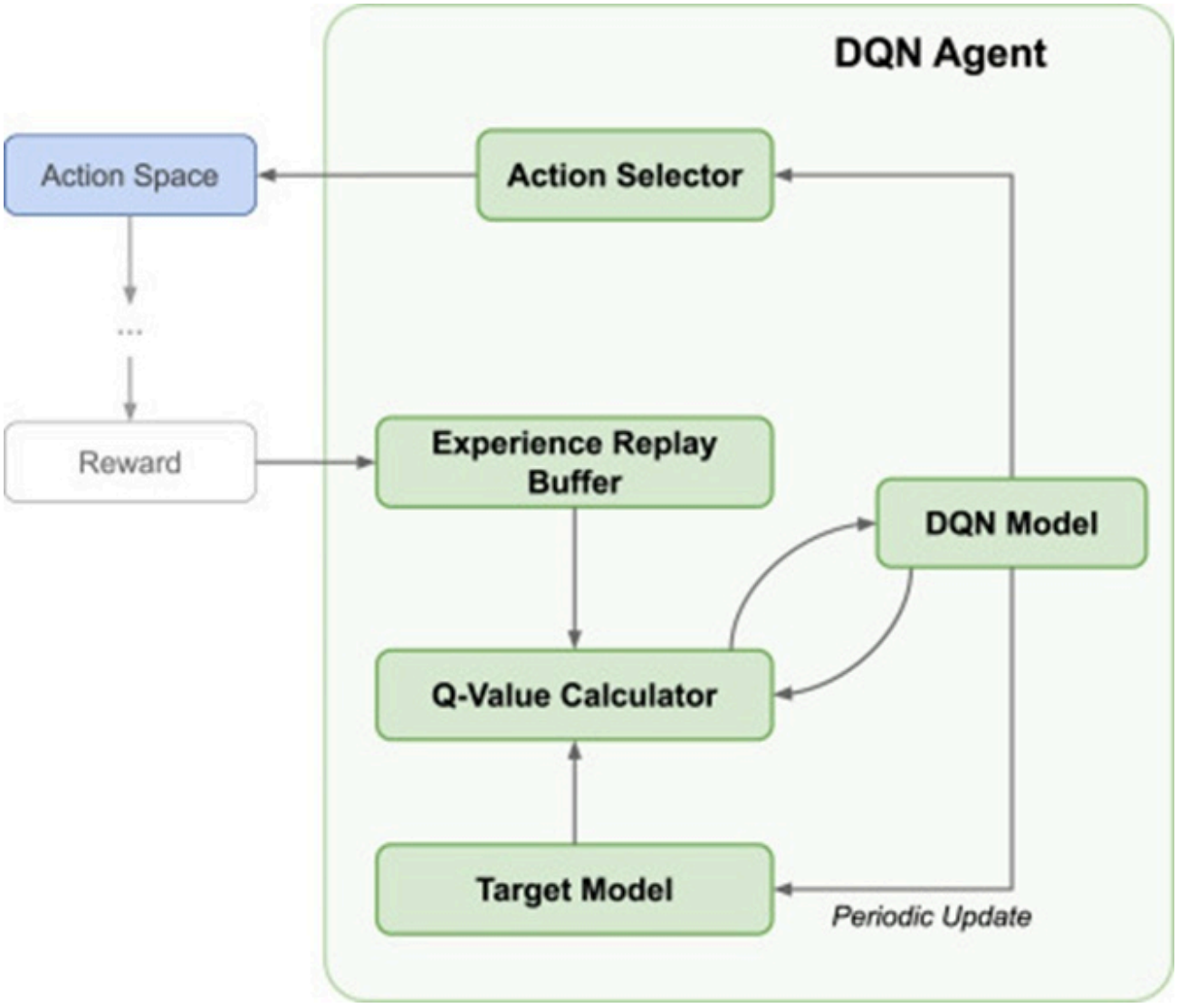


Figure 4. Deep Q-Network (DQN) architecture showing the integration of experience replay, Q-value calculation, and model components.

The Target Model is refreshed from the DQN Model every 5 episodes (denoted as “Periodic Update” in **Figure 4**), providing a stable anchor for Q-value calculation. This dual Q-learning scheme notably reduces overestimation bias throughout the learning process. The Action Selector deploys an ϵ-greedy approach, toggling between exploratory random actions and exploitation of the learned policy as guided by the DQN Model’s output.

Training proceeds via the Adam optimiser configured with a learning rate of 0.001, and gradient clipping is enforced to maintain update stability. The Reward Calculator scores the agent’s actions against multiple criteria and routes this evaluative signal back to the Experience Replay Buffer, closing the learning feedback loop. The entire system functions within the predefined Action Space comprising five discrete temperature adjustment actions.

The coordinated operation of these components yields a learning system capable of adapting to changing environmental conditions while maintaining stable, efficient control policies for pipe temperature management.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Integration of adaptive context-aware transformer environment and RL agent

The essence of our integration strategy rests on forging a bridge between prediction and control. By uniting the adaptive context-aware transformer's proficiency in modelling individual building thermal dynamics with the DQN's strength in learning optimal control strategies, we have produced a system that simultaneously tailors itself to specific building characteristics while generating control policies that generalise successfully across heterogeneous building stocks. This fusion yields a holistic control framework that harnesses both predictive modelling and reinforcement learning to build a temperature management system. The transformer model operates as a differentiable environment simulator, supplying the DQN agent with precise forecasts while factoring in building-specific attributes via contextual embeddings.

At each timestep t, the system operates within the state space outlined in **Eq. 5**, where states are enhanced by building-specific contextual embeddings F_building from **Eq. 4** that encapsulate metadata such as building type and geographical location. The integration unfolds as a structured interaction cycle: the transformer initially ingests a 48-timestep historical window to ascertain the present state and project future temperature behaviour:

$$T_{t+1:t+6}^{pred} = f_{\text{transformer}}(s_{t-47:t}, F_{\text{building}}) \quad (15)$$

Thereafter, the DQN agent selects an action a_t from the discrete action set A = {-4.0, -2.0, 0, +2.0, +4.0}, corresponding to pipe temperature adjustments measured in °C. The selected action gets bounded by adaptive safety limits to retain operation within the permissible state space:

$$T_{t+1}^{pipe} = \text{clip}(T_t^{pipe} + a_t, 20, 70) \quad (16)$$

The transformer environment applies this action and projects the ensuing indoor temperature path. The system then derives the reward using **Eq. 10**, measuring the action's effectiveness across several dimensions: temperature maintenance, weather responsiveness, stability, and satisfaction of safety constraints.

To support learning across a portfolio of buildings while preserving building-specific adaptability, the system adopts a shared replay buffer $\mathcal{B}$ that holds experiences structured as tuples:

$$(s_t, a_t, r_t, s_{t+1}, F_{\text{building}}) \in \mathcal{B} \quad (17)$$

Throughout training, experiences are drawn from this buffer to update the DQN agent's policy. Embedding building representations F_building within the state allows the agent to develop building-tailored control strategies while extrapolating across buildings of comparable types.

This unified architecture enables dependable temperature regulation across diverse building types while accommodating individual building idiosyncrasies and variable weather regimes. The marriage of the transformer's forecasting capacity with the DQN's learning ability results in a control system that maintains comfortable indoor conditions, adheres to system constraints, and enhances energy utilisation.

Evaluation metrics

Evaluation metrics for adaptive context-aware transformer (environment)

To assess the predictive fidelity of the adaptive context-aware transformer, we utilised Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2) as benchmarks for gauging model prediction

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

accuracy. Both RMSE and MAE are standard metrics in the literature for quantifying the divergence between predicted outputs and ground-truth values. RMSE factors in error magnitude and assigns disproportionately heavier penalties to larger errors, whereas MAE treats all error magnitudes uniformly. The R^2 The metric quantifies the share of variance in the dependent variable (actual values) attributable to the independent variable (predicted values). An R^2 reading of 1 denotes a flawless data fit, whereas a score of 0 implies the model accounts for none of the data's variability. Collectively, these metrics provide valuable diagnostic insight into time-series prediction model performance, aiding model selection and iterative improvement.

The evaluation metrics are derived as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (18)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (19)$$

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}\right] \quad (20)$$

Where $\widehat{y_i}$ denotes the predicted value, y_i is the corresponding actual value, $\overline{y}$ stands for the output mean, and N represents the total count of samples in the test dataset.

Evaluation of RL DQN-agent

The DQN-agent-based control system's performance is appraised through two primary metrics:

1. Control performance: This metric quantifies how effectively the agent sustains the target indoor temperature across diverse environmental and operational settings. Control performance is corroborated by juxtaposing the indoor temperature predictions from the adaptive context-aware transformer model with real indoor measurements. A stable indoor temperature of roughly 21 °C, maintained irrespective of outdoor variability, signals robust control capability.

2. Energy efficiency: The system's energy draw is computed from the thermal differential between the pipe supply temperature (T_pipe) and the indoor temperature (T_indoor). Energy consumption (E) is formulated as:

$$E = c \cdot f \cdot \Delta T \cdot t \quad (21)$$

Where:

• c denotes the specific heat capacity of water (4.186 kJ/kg K),

• f is the assumed water flow rate (2.0 litres per second),

• $\Delta T = T_{\text{pipe}} - T_{\text{indoor}}$,

• t represents the time interval (in hours).

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Energy quantities are expressed in kilowatt-hours (kWh), using the conversion factor 1 kWh = 3.6 × 10^3 kJ. For energy efficiency assessment, consumption under DQN control is benchmarked against the baseline (original system). Energy savings (S) are then calculated as:

$$S = \frac{En_{\text{original}} - En_{\text{predicted}}}{En_{\text{original}}} \times 100\% \quad (22)$$

Where $En_{original}$ and $En_{predicted}$ correspond to the energy consumed by the original and DQN-controlled systems, respectively. This metric yields a quantitative indicator of the DQN agent's effectiveness in lowering energy usage while preserving control performance.

Experimental Setup

Baseline traditional control system

The reference control infrastructure consists of the outdoor temperature compensation curves that district heating networks presently run in production. Within this well-established industrial methodology, a linear relationship is posited between outdoor temperature (OT) and pipe supply temperature (PT), with the underlying logic that a drop in outdoor conditions warrants an increase in supply temperature. For each building, a dedicated linear compensation curve has been derived by fitting against its historical operational logs. The generic linear specification takes the form $PT = a \times OT + b$, within which a captures the temperature slope coefficient — effectively governing how sharply the system reacts to swings in outdoor temperature — whereas b encodes the base temperature offset (°C), which fixes the floor for operating temperature.

$$\text{Building 24585:} \quad PT = -0.618 \times OT + 34.9 \quad (23)$$

$$\text{Building 24635:} \quad PT = -0.519 \times OT + 41.4 \quad (24)$$

$$\text{Building 116:} \quad PT = -0.746 \times OT + 40.8 \quad (25)$$

$$\text{Building 357:} \quad PT = -1.056 \times OT + 36.0 \quad (26)$$

$$\text{Building 3313:} \quad PT = -0.886 \times OT + 39.9 \quad (27)$$

A rise in pipe temperature as outdoor temperature declines is confirmed by the negative slope coefficients. The visualisation in **Figure 5** makes clear that these linear mappings are emblematic of the reactive control philosophy that remains standard across the sector: supply temperature is varied strictly in response to the immediate outdoor temperature reading, without any mechanism for anticipating near-future conditions or for incorporating the distinct thermal fingerprints of individual buildings. Against this incumbent system, we benchmark our proposed pairing of an adaptive, context-aware transformer environment with a DQN-based control methodology.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

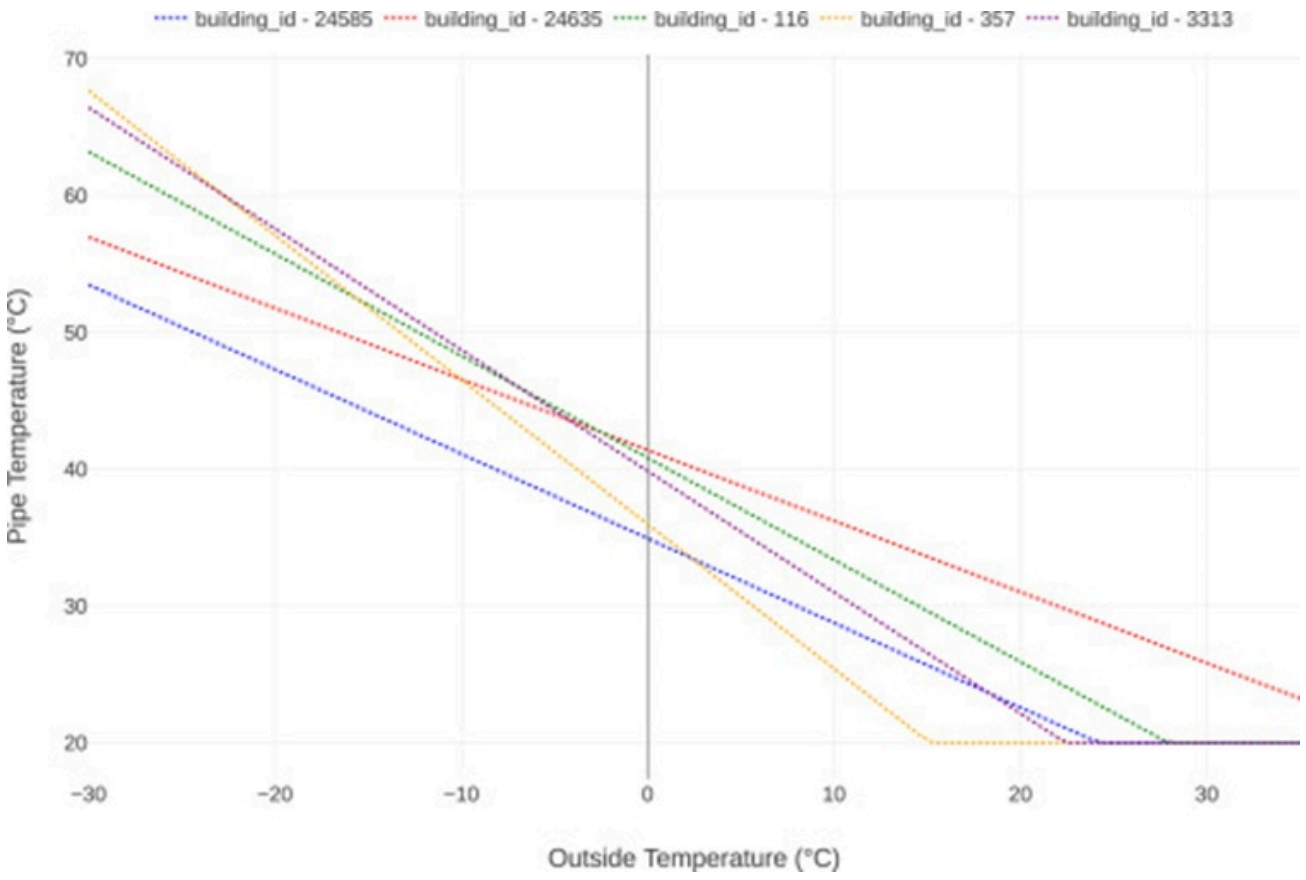


Figure 5. Traditional outdoor temperature compensation curves for five Stockholm buildings showing the linear relationship between outdoor temperature and pipe temperature. Each dotted line represents the current industry-standard control strategy for each building, demonstrating the reactive approach that serves as the baseline for comparison with the proposed DQN framework.

Training and test setup for adaptive context-aware transformer

Our earlier adaptive context-aware transformer architecture [8] serves as the foundation and is experimentally extended to broaden its scope to a wider array of building control contexts. Three primary signals — Indoor Temperature, Outdoor Temperature, and Pipe Temperature — are fed into the model, which is tasked with projecting indoor temperatures over a 6-hour horizon while concurrently learning and operationalising building-specific contextual characteristics. The rationale for selecting a 6-hour forecast horizon stems from domain specialists in building control systems. It is consistent with the documented thermal inertia behaviour of residential properties in district heating schemes. According to these building control specialists, the thermal response lag typical of residential construction falls within the 4–8 hour range; thus, the effects of a heating control intervention take a commensurate interval to become fully apparent indoors. The 6-hour horizon endows the reinforcement learning agent with adequate lead time to formulate anticipatory control decisions, yet falls within the bounds of the transformer model's trustworthy prediction range. This forecast window matches the thermal inertia figures reported for the typical Swedish and Finnish residential building stock: temperature responses become measurable after 2–3 hours, while thermal equilibrium following a control adjustment is reached only after 6–8 hours.

A thorough sweep of hyperparameter combinations, conducted as a grid search across numerous architectural variants, yielded the optimal model configuration recorded in **Table 3.**

Table 3. Hyperparameter configurations and optimal values for the context-aware transformer model.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Hyperparameters	Optimal value	Search range
Transformer Layers	2	[2, 4, 6, 8]
Dimension (D)	64	[32, 64, 128, 256]
Attention Heads	4	[4, 8, 16]
Feed-Forward Dimension	128	[128, 256, 512]
Dropout Rate	0.2	[0.1, 0.2, 0.3]
Batch Size	256	[128, 256, 512]
Embedding Size Factor	10	[5, 10, 15]

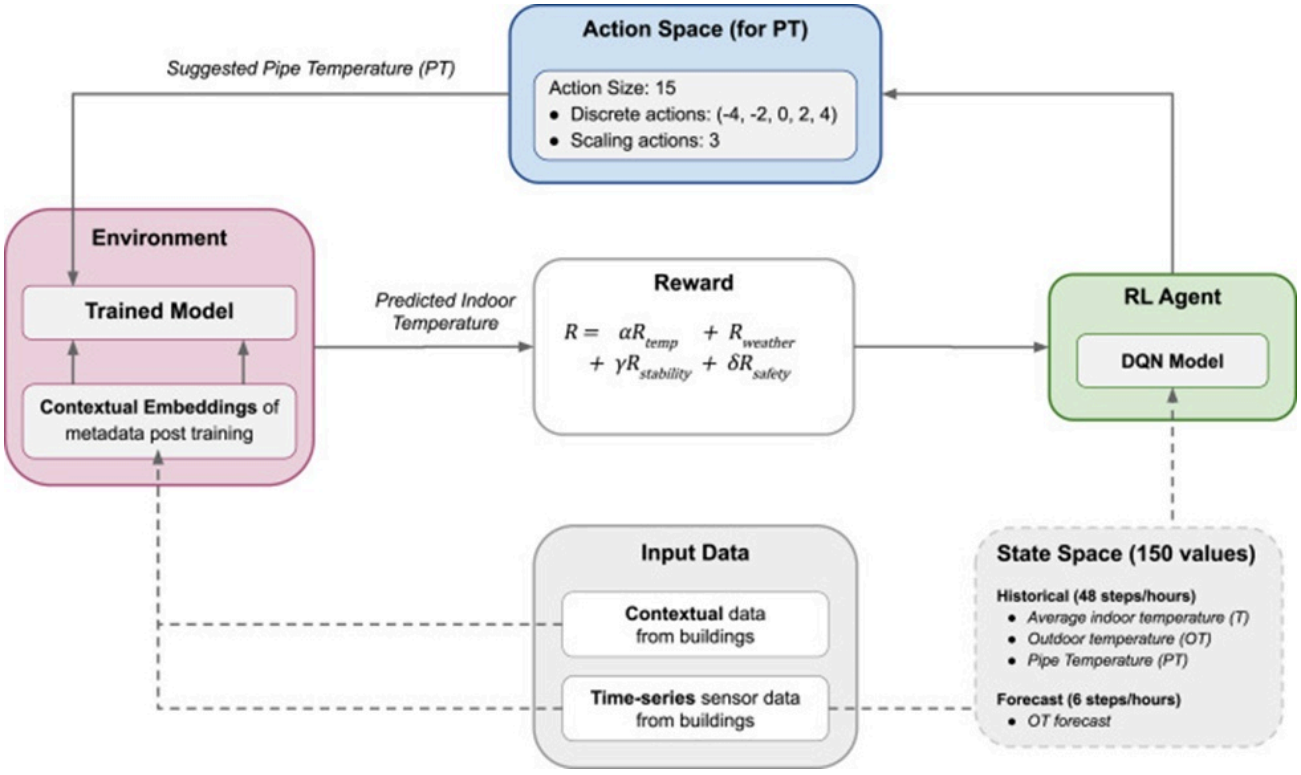


Figure 6. High-level proposed Reinforcement Learning framework for district heating control. The system integrates an adaptive context-aware transformer environment with historical sensor data and weather forecasts to optimize pipe temperature settings through a DQN model.

The adaptive context-aware transformer features 2 encoder layers. Inside each layer reside 4 multi-head attention mechanisms that process 64-dimensional embeddings, alongside 128-dimensional feed-forward sublayers activated by ReLU. Among the inputs, normalised temperature readings (indoor, outdoor, pipe) are scaled to the interval [−1, 1]; temporal markers (year, month, week, day, hour) take the form of positional embeddings; and the six contextual variables (Land, City, Rise, ADI, Floors, Building Type) are passed through individual embedding layers that map each to a 10-dimensional representation. These six embeddings are then concatenated into a single 60-dimensional vector, which is projected to occupy the same 64-dimensional space as the transformer. The loss function used during training is mean squared error, with optimisation handled by Adam (β_1 = 0.9, β_2 = 0.999), gradient clipping at a threshold of 1.0, and early stopping triggered after 10 consecutive epochs of stagnant validation loss. The evaluation protocol is rigorous: a

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

held-out test set encompassing 15% of the total data is reserved. This subset spans several seasons and a diversity of building types from both countries, thus providing a stringent test of how well the model generalises. The yardstick for model performance is its capacity to furnish accurate 6-step-ahead predictions without losing awareness of building-specific context. Prediction accuracy is quantified through the metrics RMSE, MAE, and R^2.

Reinforcement learning - DQN agent

With the adaptive context-aware transformer fully trained as an environment, the next step is to construct a reinforcement learning setup in which a DQN agent assumes responsibility for heating control across five residential buildings in Stockholm. The division of labour illustrated in **Figure 6** has the transformer environment handling building-specific contextual embeddings alongside indoor temperature forecasts, while the DQN agent works exclusively with time-series data. This design choice renders the agent building-agnostic, enabling it to transfer across building types and geographical settings.

What the DQN agent receives as input is strictly temporal: 48-step historical sequences for Indoor Temperature (T), Outdoor Temperature (OT), and Pipe Temperature (PT), augmented with a 6-hour outdoor temperature forecast. Meanwhile, contextual metadata (Land, City, Rise, ADI, Floors, Building Type) are processed independently by the transformer environment via its embedding mechanism, enabling the system to produce reliable indoor temperature forecasts regardless of building category or national context. The separation of these two streams allows the DQN agent to learn general control strategies, while the environment provides fine-tuning specific to each building.

Both instantaneous observations and a forward-looking outdoor temperature forecast are encoded in the DQN's state space. The transformer environment exploits its contextual embeddings to generate high-fidelity temperature projections tailored to the specific building, and the DQN agent, in turn, bases its control action selection on these projections. The action space consists of 15 possible moves, formed by the cross-product of five discrete step sizes (−4, −2, 0, 2, 4) and three scaling multipliers, which together afford precise, graded control over how much to adjust the temperature.

Training extends across 4000 episodes. In each episode, building data flows through the environment, which leverages its trained contextual building embeddings to issue indoor temperature forecasts. The DQN agent then applies an action to the Pipe Temperature, drawing on both the time-series state representation and the indoor temperature forecasts provided by the environment, thereby committing to a control decision. The environment subsequently evaluates these Pipe Temperature actions against the building’s particular characteristics to determine the reward signal.

Structurally, the DQN network comprises a 9-dimensional input layer (aggregating current indoor/outdoor/pipe temperatures and the 6-hour outdoor forecast), two hidden layers of 256 neurons apiece fitted with ReLU activations and a dropout rate of 0.3, and a 5-dimensional output layer producing Q-values aligned with the permissible actions. The training regimen incorporates experience replay, with a memory buffer archiving state transitions, actions, and rewards. Every 5 episodes, the target network is refreshed. Exploration is governed by an ϵ-greedy policy, initialised at $\epsilon = 0.8$ and annealed down to 0.05. The chosen optimiser is Adam (learning rate = 0.001), and gradient clipping (threshold = 1.0) is applied to promote stable learning. Selection of the top-performing model is based on cumulative reward alongside temperature control effectiveness observed across the various building types.

For validation of the RL-DQN agent, the most recent year of Stockholm test data is used. This validation strategy capitalises on the transformer’s universal applicability — any building type can be accommodated through its contextual embeddings — while the DQN agent restricts itself to applying learned control strategies rooted purely in temporal patterns. The outcome is dependable temperature regulation that holds up across heterogeneous building stock and varying weather, demonstrating the integrated system’s generalisability.

Training duration and computational complexity

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

The hardware deployed for the experimental runs consisted of NVIDIA RTX 3090 Ti and RTX 4090 graphics processing units, each with 24 GB of VRAM. Bringing the adaptive context-aware transformer environment to a trained state, including the hyperparameter optimisation exercise outlined in **Table 3,** took 120 hours of nonstop computation across 148 buildings. The subsequent DQN agent training, spanning 4000 episodes and incorporating both experience replay and periodic target network updates, occupied 218.6 hours. When it comes to inference, the transformer environment takes 2.8 seconds to deliver a prediction, while the DQN agent takes 1.9 seconds to select an action; the combined latency per control decision thus amounts to 4.7 seconds. In district heating control practice, where decision cycles are refreshed every 15–60 minutes, a 4.7-second inference time is entirely permissible. As a point of contrast, the traditional outdoor temperature compensation approach resolves control decisions in microseconds, simply by indexing into precomputed lookup tables running on standard building management hardware.

Results and Discussion

What follows is arranged into two major blocks. The opening segment concentrates on quantifying how accurately the adaptive context-aware transformer model predicts indoor temperatures when applied to a heterogeneous collection of buildings situated across Sweden and Finland; this covers both offline evaluation and live online deployment. The subsequent segment turns to the district heating control proficiency exhibited by the DQN agent, benchmarking its pipe temperature regulation against that of standard controllers. The closing analysis addresses energy optimisation outcomes, measuring the magnitude of energy reduction across distinct building typologies and seasonal contexts. As a body of evidence, these results confirm not only the prediction model's precision but also the integrated control framework's capacity to reconcile occupant comfort with lower energy expenditure.

Adaptive context-aware transformer (Indoor temperature prediction) environment

Discussed here is the appraisal of our adaptive context-aware transformer model when deployed across numerous residential buildings in Sweden and Finland. The assessment draws on test datasets covering multiple geographic zones and building configurations, with the forecasting task set to a 6-hour horizon (6 time steps). Contextual embeddings that were refined during the training phase confer an advantage, allowing the model to generalise more effectively to unfamiliar building categories and locales.

Within the Swedish building set **(Table 4)**, the model exhibits considerable predictive strength irrespective of context. From a geographical standpoint, the strongest figures emerge from Götaland (RMSE: 0.1958, R^2: 0.7187), trailed by Svealand (RMSE: 0.2191) and Norrland (RMSE: 0.2245). Breaking results down by building class reveals Mid Rise properties as top performers, attaining an RMSE of 0.1817 and R^2 of 0.6956. City-level inspection highlights Skellefteå as a high point (RMSE: 0.1487), while the remaining urban centres post consistently respectable accuracy, generally keeping RMSE readings under 0.23.

Finnish buildings **(Table 5)** mirror these behavioural trends across different analytical slices. Residential stock yields dependable accuracy (RMSE: 0.2256), and the geographic split between Norrland (RMSE: 0.2829) and Österland (RMSE: 0.3029) points to steady, if slightly varied, performance. **Figure 7** enriches the picture by mapping RMSE distributions over Finnish cities, revealing location-dependent variation in accuracy while confirming that RMSE remains bounded below 0.35 in all cases. The fact that performance levels shift across building typologies and geographies speaks directly to the model's capacity for adapting to localised building traits and regional contexts.

Table 4. Evaluation metrics averaged for all buildings in Sweden for each context of Lands, Rise, City.

Context		**Prediction Horizon 6 h (6 steps)**		
Empty Cell			**MAE**	**RMSE**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Lands	Norrland	0.6858	0.1971	0.2245
	Svealand	0.6720	0.1883	0.2191
	Götaland	0.7187	0.1866	0.1958
Rise	Low Rise	0.6666	0.2283	0.2439
	Mid Rise	0.6956	0.1677	0.1817
	High Rise	0.7142	0.1760	0.2135
City	Skellefteå	0.6628	0.1078	0.1487
	Luleå	0.7283	0.2092	0.2548
	Stockholm	0.6864	0.2113	0.2064
	Västeras	0.6223	0.1974	0.2261
	Karlstad	0.7012	0.1922	0.2317
	Skövde	0.7838	0.1911	0.2233
	Trollhattan	0.6143	0.1996	0.2240
	Gothenburg	0.7002	0.1960	0.1989
	Malmö	0.7295	0.2116	0.2040

Table 5. Evaluation metrics averaged for all buildings in Finland for each context of Lands, Building Type, and Rise.

Context		**Prediction horizon 6 h (6 steps)**		
Empty Cell			**MAE**	**RMSE**
Lands	Norrland	0.5938	0.2468	0.2829
	Österland	0.5538	0.2669	0.3029
Building Type	Residential building	0.5697	0.2164	0.2256
	Care building	0.4497	0.3379	0.3777
	Commercial building	0.7018	0.2163	0.2754
Rise	Low Rise	0.6527	0.2326	0.2618
	Mid Rise	0.5215	0.2732	0.2968
	High Rise	0.5473	0.2647	0.3203

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

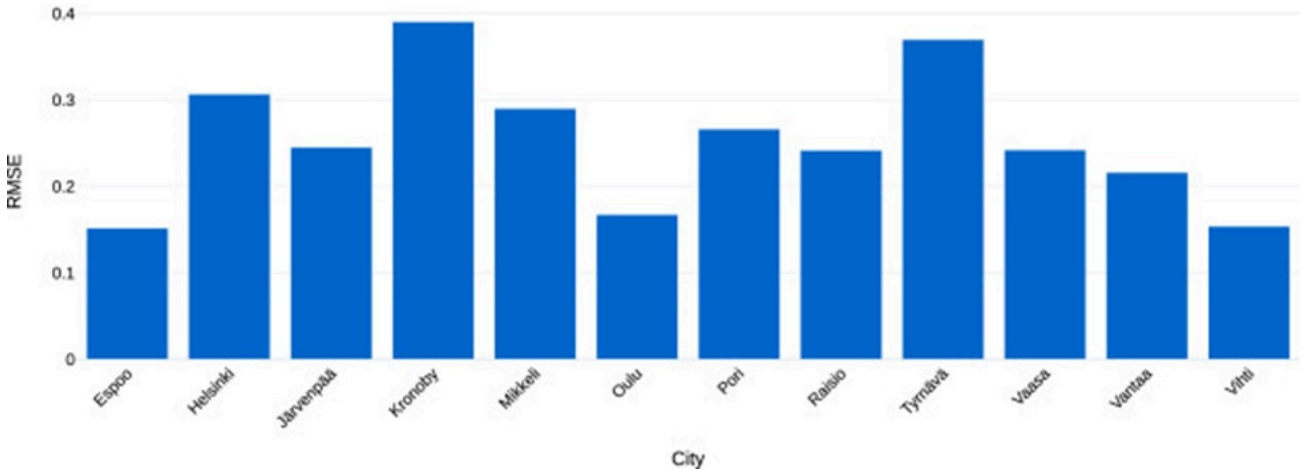


Figure 7. Average RMSE of indoor temperature predictions for buildings in Finland by city.

What emerges from these findings is a clear demonstration that training with contextual meta-data embeddings pays dividends in indoor temperature forecasting, endowing the model with the capability to discern and adjust to the individual thermal identities of buildings scattered across varied geographies. The uniformly modest error metrics registered across diverse settings further endorse the model's readiness for real-world application in building temperature regulation.

Adaptive context-aware transformer on real-world online prediction

A real-world test of practical viability and suitability as a reinforcement learning environment was conducted through live deployment in Stockholm, where the adaptive context-aware transformer was trialled on five buildings originating from the training corpus. This online prediction apparatus has been functioning for close to two months, uninterruptedly issuing indoor temperature forecasts informed by live building telemetry, and has sustained strong predictive accuracy (RMSE values falling between 0.12 and 0.23 for the five buildings). **Figure 8** captures the model's behaviour under genuine operating conditions, overlaying actual and forecast indoor temperatures for a mid-rise Stockholm building over a four-day stretch in March 2025.

Close alignment between forecasts and observed temperatures is maintained throughout, with the model faithfully reproducing both gradual drifts and transient fluctuations alike. Accuracy proves resilient even when the temperature trace passes through local highs and lows, confirming that the model can contend with the messiness of real thermal dynamics.

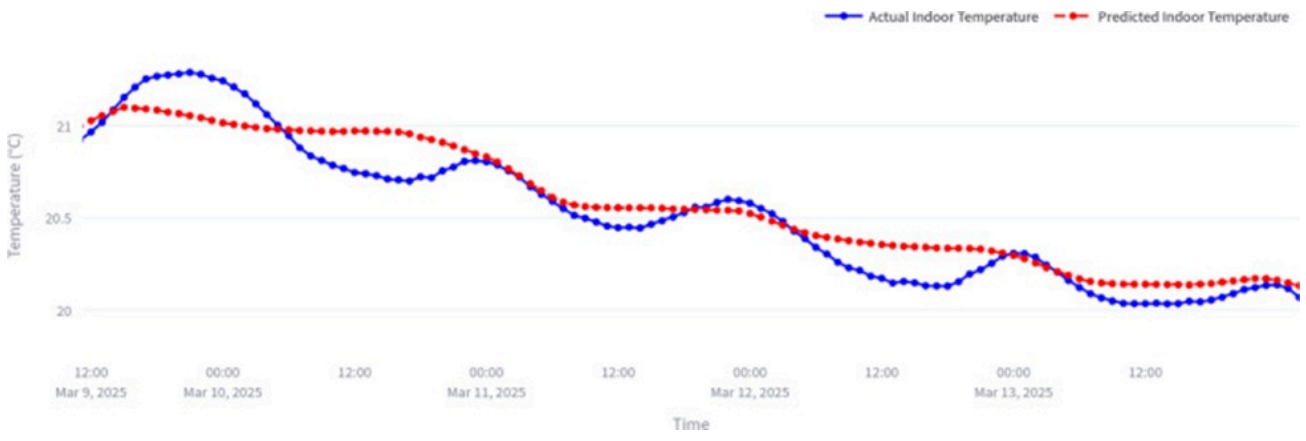


Figure 8. Comparison between actual and predicted indoor temperatures during real-time online deployment: Real-time 6-hour forecasting results for a mid-rise residential building in Stockholm.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

That the model performs so robustly under live conditions carries direct implications for reinforcement learning: it establishes the adaptive context-aware transformer as a credible environment simulator upon which DQN agents can be trained. The persistently small prediction residuals (all staying beneath 0.23 RMSE) indicate that the environment delivers forecasts — and therefore feedback signals — of sufficient accuracy to guide the RL agent's decision cycle. Because predictive fidelity holds up across different buildings and fluctuating temperature regimes, the model constitutes a solid foundation for learning optimal PT control strategies. For the DQN training process, this is critical, as the eventual quality of the learned control policies is a direct function of how accurately the environment communicates the consequences of actions through its temperature projections.

DQN (RL)

After building confidence in the adaptive context-aware transformer as an environment simulator — validated both offline and through live deployment — the spotlight shifts to the DQN agent's effectiveness in steering building heating plant.

The evidence speaks to the strength of the integrated design philosophy, whereby the agent exploits high-fidelity temperature forecasts to roll out efficient heating control policies. The assessment spans the twin dimensions of control quality and energy efficiency.

District heating prediction (PT)

Relative to the legacy control infrastructure, the DQN agent's pipe temperature management for district heating delivers clear and substantial improvements. A side-by-side comparison of pipe temperature trajectories under the original and DQN-driven regimes, covering an eight-day winter window (December 30, 2021–January 7, 2022), is given in **Figure 9a**. Under DQN governance, pipe temperature follows a far more dynamic path, adjusting to shifting conditions and avoiding the near-static modulation that characterises the original system. Whereas the legacy controller confines PT to a tight 30–32 °C band with minimal variance, the DQN agent pursues a markedly more responsive course, its pipe temperature commands varying from 39 °C up to 49 °C depending on building-side requirements.

Confidence in this control strategy is bolstered by the indoor temperature forecasts produced by the adaptive context-aware transformer and plotted in **Figure 9b**: the indoor environment remains anchored near 21 °C while outdoor conditions swing between 3 °C and 8 °C. This outcome validates the agent's dual capacity to safeguard occupant comfort while implementing a more assertive and variable PT control policy.

Several illustrative system responses during the winter period are highlighted by **Figure 9.** Noteworthy among them: (1) A sharp outdoor temperature descent on December 31st, from −4 °C to −12 °C, prompts a pre-emptive lift in pipe temperature by the DQN agent (42 °C → 46 °C), whereas the original system remains static at roughly 47 °C; (2) The system's adaptive reaction to a warming trend on January 2nd (outdoor temperatures climbing from −8 °C to −4 °C) is visible in the DQN agent's decision to ratchet pipe temperature down to 40 °C, a signature of energy-saving behaviour; (3) During January 4–5th, temperature stability is preserved despite outdoor oscillations between −6 °C and −10 °C, the DQN system holding indoor temperature to within 0.5 °C of the 21 °C target by dynamically modulating pipe temperature across the 41–44 °C range. These snapshots collectively illustrate the system's anticipatory capacity and its adeptness at responding to environmental variation without sacrificing thermal comfort.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

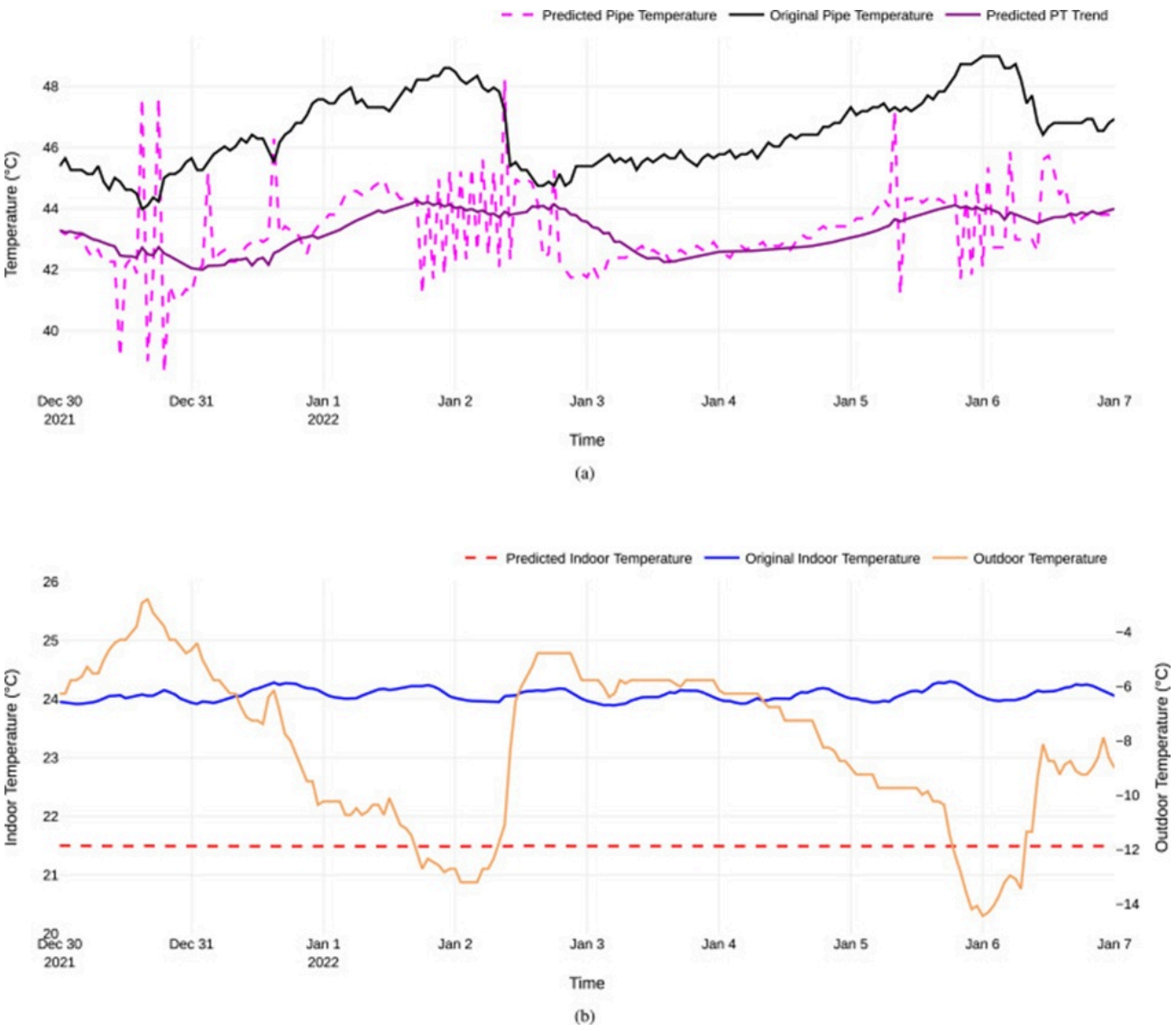


Figure 9. Illustration of actual vs. predicted Indoor temperature values and Pipe temperature values. (a) Comparative analysis of supply pipe temperature control strategies during eight-day winter period (December 30, 2021–January 7, 2022) in Stockholm residential building, showing original control system maintaining constant 46–48 °C operation versus Deep Q-Network agent achieving dynamic 39–49 °C temperature adjustments with overall trend toward reduced energy consumption while maintaining indoor temperature stability at 21 °C target despite outdoor temperature fluctuations between −12 °C and −4 °C., (b) Indoor temperature comparison during winter conditions (Dec 2021–Jan 2022). The original system (solid blue line) maintains higher indoor temperatures while the DQN agent consistently targets an optimal 21.5 °C setpoint (dashed red line), with outdoor temperature variations shown in orange.

A per-building dissection of temperature control outcomes is assembled in **Table 6**, which pits the original baseline controller against the DQN predictive controller for the five Stockholm buildings. The DQN agent draws indoor temperatures closer to the 21 °C reference while dramatically compressing variability: indoor temperature standard deviations shrink from the 0.23–0.87 °C range seen under the original scheme to just 0.02–0.33 °C under DQN control, representing a gain in temperature steadiness that reaches as high as 92%. To take a concrete case, building 24635's indoor temperature fluctuated by a mere ±0.02 °C, set against ±0.25 °C in the baseline. The DQN framework further demonstrates building-specific calibration through distinct pipe temperature strategies spanning 30.23 °C (building 3313) to 48.89 °C (building 357), whereas the legacy system showed far weaker responsiveness. This degree of regulatory

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

precision substantiates the agent's capability to deliver thermal comfort via adaptive heating strategies tuned to each building's particular thermal signature.

District heating energy optimization

The DQN agent's more alert and flexible control logic translates into material energy reductions, the scale of which varies by building category and time of year. **Figure 10** juxtaposes energy consumption profiles, illustrating the mechanism by which the DQN agent realises these savings through superior PT management. The original system's energy draw stays comparatively flat, hovering around 180–200 kW; the DQN-managed counterpart, by contrast, displays a more dynamic consumption signature, dropping to the 130–170 kW band under favourable conditions without compromising indoor comfort. The figures collated in **Table 7** attest to the consistency of energy savings across building classes. Mid-rise properties (MR) register the most pronounced gains — Building 313 hits 14.85% savings — a result linked to the favourable ratio of thermal mass to envelope surface area that yields more predictable thermal behaviour. High-rise buildings (HR) cluster in the 10%–11% range. The low-rise representative (26394-LR) posts a steady 7.40% improvement.

When the data are sliced by season **(Figure 11)**, distinctive optimisation patterns come into focus. Winter months deliver savings reliably in the 5%–15% window for all building classes, with Building 648 touching approximately 14%. Springtime witnesses the strongest efficiency peaks, Building 313 notably clearing the 20% threshold. Summer brings greater dispersion and, for select buildings, a reduction in savings; Building 648 slips into negative territory, pointing to a net increase in energy use. Autumn sees a broad recovery, as most buildings re-enter the 8%–15% savings corridor.

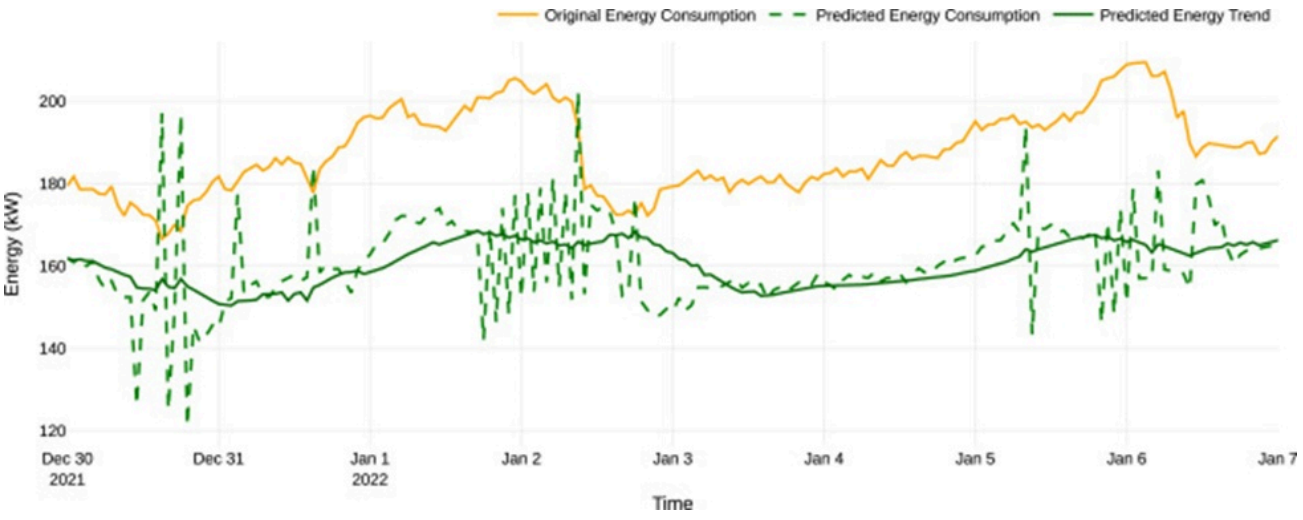


Figure 10. District heating energy consumption comparison between conventional control (solid orange line) and DQN-optimized control (dashed green line) with trend line (solid green).

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press



Figure 11. Seasonal distribution of district heating energy savings across different building types in Stockholm: Analysis of five buildings with varying characteristics (LR - Building 26394, MR - Building{24635, 313}, and HR - Building{116, 648}) showing distinct performance patterns across winter, spring, summer, and fall seasons.

Table 6. Temperature control analysis across five Stockholm residential buildings showing indoor and pipe temperature statistics (mean ± standard deviation) for original baseline control versus Deep Q-Network predictive control, demonstrating thermal stability and control precision while maintaining thermal comfort requirements.

Stockholm buildings (ID)	**Pipe temperature (°C)**		**Indoor temperature (°C)**	
Empty Cell	**Predicted**	**Original**	**Predicted**	**Original**
24585 – Low Rise	36.10 ± 13.84	28.16 ± 3.44	21.15 ± 0.30	23.74 ± 0.87
24635 – Mid Rise	31.53 ± 10.52	35.48 ± 0.68	21.31 ± 0.02	23.60 ± 0.25
116 – High Rise	34.90 ± 8.88	38.24 ± 8.17	22.08 ± 0.24	24.25 ± 0.55
357 – Mid Rise	48.89 ± 5.00	51.20 ± 3.98	21.21 ± 0.28	23.08 ± 0.23
3313 – High Rise	30.23 ± 10.11	31.48 ± 9.01	21.73 ± 0.33	22.91 ± 0.77

Table 7. Quantitative energy consumption analysis and savings performance across five Stockholm residential buildings categorized by building height (Low Rise: 1–3 floors, Mid Rise: 4–6 floors, High Rise: 7+ floors), showing original energy consumption in megawatt-hours, predicted consumption under Deep Q-Network control, and percentage energy savings achieved while maintaining thermal comfort requirements.

Stockholm buildings (ID)	**Energy Savings (%)**	**Energy Consumption (MWh)**	
Empty Cell	Empty Cell	**Predicted**	**Original**
26394 – Low Rise	7.40	169.6	183.16
24635 – Mid Rise	9.06	88.58	97.4
313 – Mid Rise	14.85	71.16	83.57

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

116 – High Rise	11.08	71.5	80.41
648 – High Rise	10.03	30.69	34.11

The emergence of negative savings during summer, Building 648 being the clearest example, can be traced to the DQN agent's propensity for deploying more frequent corrective interventions during periods when heating loads are minimal, in contrast to the traditional controller's largely passive stance. In summer, when space heating demand nearly vanishes, the agent's pursuit of precise temperature regulation via micro-corrections results in marginally higher energy consumption than a passive baseline. This is worsened by the operational reality that district heating plants run less efficiently under very low load factors. Such findings hint at the desirability of future refinements that embed seasonally adaptive control logics, capable of modulating behaviour according to the wide annual swing in heating requirements.

When considered in their entirety, the results confirm that the DQN agent strikes an effective balance between comfort maintenance and energy parsimony, extracting meaningful savings through more sophisticated control logic while continuing to satisfy indoor temperature objectives. The spread of outcomes across building typologies and seasons further reinforces the system's versatility in accommodating varied operating conditions while reliably delivering energy performance gains.

Results insights and limitations

The evidence assembled through this investigation confirms that the fusion of adaptive context-aware transformer models with deep reinforcement learning constitutes an effective methodology for improving district heating performance in residential contexts. Building-specific behavioural signatures were reliably encoded by the context-aware transformer environment via its contextual embedding mechanisms, with predictive fidelity measured by root mean square error falling between 0.18 °C and 0.32 °C across a heterogeneous collection of building classes and geographic settings. Drawing on these projections, the Deep Q-Network agent formulated intelligent operational strategies that produced concrete energy reductions in the range of 7.40% to 14.85% across the different building categories, while indoor thermal conditions were confined to a band of ±0.5 °C around the target setpoints. Mid-rise structures emerged as particularly strong beneficiaries, and maximum seasonal efficiency gains surpassed the 20% mark during the spring period. Field validation spanning two national contexts lends credence to the claim that machine-learning-driven approaches possess international relevance for strengthening the efficiency and environmental performance of district heating networks. On balance, the results affirm the promise held by context-aware predictive control paradigms to unlock considerable energy optimisation without sacrificing occupant thermal welfare, thereby advancing the built environment's broader efficiency ambitions and reinforcing agendas for sustainable urban transformation.

The decision to broaden the model's geographic remit to include Finnish buildings supplies additional evidence of the framework's capacity to transfer across contrasting building traditions and climatic regimes. The addition of building type as a supplementary contextual variable empowered the model to distinguish the operational signatures of residential blocks, commercial premises, and care institutions, adjusting its output to reflect their respective occupancy schedules and thermal expectations. This demonstration of cross-national validity strengthens the argument in favour of implementing context-aware control solutions across a more diverse set of district heating infrastructures.

The coupling of building-specific contextual embeddings with the DQN architecture represents a consequential departure from the capabilities of conventional control methodologies. The transformer architecture learns and then operationalises the particularities of each building, allowing the system to converge on control strategies markedly more sophisticated than those attainable through legacy approaches. Because temporal dynamics and building-level particularities are jointly modelled, the DQN agent acquires the capacity to make decisions informed by each structure's distinctive thermal identity, producing regulatory interventions that are simultaneously more exact and more resource-efficient.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Analysis of the DQN-based controller's operational record brought to light several consistent motifs in how district heating energy optimisation manifests. The magnitude of energy savings, as captured quantitatively through **Eq. 22**, diverged considerably according to building category and seasonal phase. Mid-rise properties captured the largest proportionate reductions (touching 14.85%), an outcome that intimates such structures may offer near-ideal preconditions for advanced control deployment. Disaggregating performance by season threw springtime efficiency maxima into sharp focus, one case exceeding 20% in savings, whereas the summer interval introduced greater inconsistency and, in a minority of instances, a net rise in energy draw. Such seasonal dependency implies the necessity of control policies differentiated by time of year, most pressingly for low-demand intervals during which the dynamics of the system undergo substantial alteration. The performance gap separating high-rise from mid-rise buildings further indicates that structural factors extending beyond the contextual features currently modelled may exert influence over the efficacy of control. It is critical to note that across all of these outcomes, indoor temperatures were held to within ±0.5 °C of the 21 °C reference, a clear sign that comfort objectives and energy-saving imperatives were successfully balanced — a non-negotiable condition for any real-world deployment. The adaptive scaling apparatus that governs the magnitude of control actions preserved operational steadiness through widely varying conditions, while the safety corridor of 20–70 °C consistently averted thermal overstressing of the heating hardware.

These observations carry weighty consequences for the evolutionary path of district heating systems, above all in countries such as Sweden where they constitute backbone infrastructure. Energy savings spanning 7%–15% across building categories signal a sizeable opportunity to depress aggregate energy consumption without eroding the thermal experience of building occupants. Regions marked by similarly cold climatic profiles — Norway, Canada, Alaska, and Russia provide illustrative examples — may well register comparable outcomes. Climatically distinct settings, however, invite separate scrutiny: Mediterranean contexts could diverge in energy savings potential owing to briefer heating seasons, whereas continental climatic conditions might furnish circumstances conducive to even better predictive accuracy. The contextual embedding framework does accommodate climatic adaptation, yet producing dependable baselines for tropical or subtropical environments would necessitate locally sourced training datasets and dedicated validation protocols. As individual nations advance along trajectories toward renewable energy milestones, integrated control architectures of the kind developed here stand to occupy a central role in the fine-tuning of district heating operations and the underpinning of sustainable urban growth.

A set of constraints conditions the practical roll-out and scalability of what we propose. Buildings for which the available historical record spans fewer than 6–12 months are susceptible to degraded forecast quality, given the context-aware transformer environment's dependency on a sufficient volume of training data to characterise building-specific thermal behaviour. Expanding the system to encompass thousands of buildings surfaces computational obstacles arising from the requirement for centralised model training, and the contextual embedding strategy may not transfer economically across numerous national boundaries marked by differing construction norms. Latency in communication networks and the challenge of data synchronisation could erode real-time control fidelity within geographically dispersed system architectures. Among the avenues worth exploring as countermeasures are federated learning schemes that decentralise model training, hierarchical control topologies that group buildings of shared typology, and transfer learning methodologies that accelerate the onboarding of buildings not seen during training.

Practical implementation and real-world applications

The journey from research-stage verification to operational deployment demands engagement with a range of practical challenges. The live Stockholm trial offers concrete proof of system feasibility under real operating conditions, with predictive accuracy maintained without interruption across a two-month observation window and root mean square error values held consistently under 0.23 °C. The system's architectural design permits interoperability with incumbent building management platforms via standardised communication interfaces, thereby enabling phased introduction alongside legacy control apparatus. The design philosophy is intrinsically compatible with expansion across district heating networks. By virtue of the centralised transformer training paradigm, insights can be transferred across building typologies, collapsing the data prerequisites for newly connected installations from the historically typical 2–3 years of

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

archival records to 6–12 months when the buildings are of a comparable category and located within the same geographic region. Deployment variants hosted in cloud environments unlock remote surveillance and actuation functionalities, a capability that becomes essential for district heating operations of large scale that extend across multiple geographic nodes.

Conclusion

In summation, the work reported here establishes that situating an adaptive context-aware transformer as an environment for reinforcement learning constitutes a promising strategic direction for the optimisation of district heating serving residential buildings. The marriage of high-quality forecasting with carefully tuned control policies yields meaningful energy reductions that do not come at the expense of occupant comfort, and this holds across a diverse spectrum of building typologies and environmental circumstances. The evidentiary record assembled supports the broader case for machine-learning-centred approaches as instruments capable of strengthening both the operational efficiency and the long-term sustainability of district heating systems, thereby making a contribution to overarching energy optimisation objectives within the built environment.

Future research will seek to evolve the adaptive context-aware framework so that the control strategies it generates become more sensitive to the seasonal cadence of district heating loads and to evolving occupancy and usage schedules. This programme of work will encompass experimentation with dynamic reward formulations whose relative weightings shift according to seasonal context, as well as an expansion of the contextual feature set to encompass a richer characterisation of building attributes. With respect to real-world integration of the RL-based control system, transition protocols will be designed that allow for incremental assimilation alongside functioning legacy controllers, creating conditions in which performance can be rigorously evaluated without placing operational safety at risk.

Acknowledgements

None

Conflict of interest

None

Financial support

This work has been supported by the Industrial Graduate School Collaborative AI & Robotics funded by the Swedish Knowledge Foundation Dnr: 20190128 and in collaboration with industrial partner EcoGuard AB.

Ethics statement

None

Received: 30 Sep 2021 Revised: 04 Dec 2021 Accepted: 21 Dec 2021
Published online: 15 June 2022

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Werner S. District heating and cooling in Sweden. Energy. 2017;126:419-29.
2. Kensby J. Buildings as Thermal Energy Storage [master's thesis]. Universidade Tecnica de Lisboa (Portugal); 2015.
3. Mangold M, Österbring M, Wallbaum H. Handling data uncertainties when using Swedish energy performance certificate data to describe energy usage in the building stock. *Energy Build*. 2015;102:328-36. https://doi.org/10.1016/j.enbuild.2015.05.045.
4. Wang Z, Hong T, Piette MA. Building thermal load prediction through shallow machine learning and deep learning. Appl Energy. 2020;263:114683.
5. Li L, Su X, Bi X, Lu Y, Sun X. A novel Transformer-based network forecasting method for building cooling loads. Energy Build. 2023;296:113409.
6. Liu X, Ren M, Yang Z, Yan G, Guo Y, Cheng L, et al. A multi-step predictive deep reinforcement learning algorithm for HVAC control systems in smart buildings. Energy. 2022;259:124857.
7. Nagy A, Kazmi H, Cheaib F, Driesen J. Deep reinforcement learning for optimal control of space heating. arXiv preprint arXiv:1805.03777. 2018.
8. Kalidindi SSV, Banee H, Karlsson H, Loutfi A. Adaptive context embedding for temperature prediction in residential buildings. In: ECAI 2024. IOS Press; 2024:4727-33.
9. Werner S. District Heating and Cooling. Elsevier; 2013.
10. Lund H, Østergaard PA, Chang M, Werner S, Svendsen S, Sorknæs P, et al. The status of 4th generation district heating: Research and results. Energy. 2018;164:147-59.
11. Gustafsson J, Delsing J, van Deventer J. Improved district heating substation efficiency with a new control strategy. Appl Energy. 2010;87(6):1996-2004.
12. Carli R, Cavone G, Ben Othman S, Dotoli M. IoT based architecture for model predictive control of HVAC systems in smart buildings. Sensors. 2020;20:781.
13. Serale G, Fiorentini M, Capozzoli A, Bernardini D, Bemporad A. Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities. Energies. 2018;11(3):631.
14. Ahmad T, Chen H, Huang Y. Short-term energy prediction for district-level load management using machine learning based approaches. Energy Procedia. 2019;158:3331-68.
15. Wang Z, Hong T. Reinforcement learning for building controls: The opportunities and challenges. Appl Energy. 2020;269:115036.
16. Weinberg D, Wang Q, Timoudas TO, Fischione C. A review of reinforcement learning for controlling building energy systems from a computer science perspective. Sustain Cities Soc. 2023;89:104351.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

17. Kensby J, Trüschel A, Dalenbäck JO. Potential of residential buildings as thermal energy storage in district heating systems–results from a pilot test. Appl Energy. 2015;137:773-81.

18. Ni Z, Zhang C, Karlsson M, Gong S. A study of deep learning-based multi-horizon building energy forecasting. Energy Build. 2024;303:113810.

19. Tian C, Li C, Zhang G, Lv Y. Data driven parallel prediction of building energy consumption using generative adversarial nets. Energy Build. 2019;186:230-43.

20. Kalidindi SSV, Banaee H, Karlsson H, Loutfi A. Indoor temperature prediction with context-aware models in residential buildings. Build Environ. 2023;244:110772.

21. Wei T, Wang Y, Zhu Q. Deep reinforcement learning for building HVAC control. In: Proceedings of the 54th Annual Design Automation Conference 2017. 2017:1-6.

22. Al Sayed K, Boodi A, Broujeny RS, Beddiar K. Reinforcement learning for HVAC control in intelligent buildings: A technical and conceptual review. J Build Eng. 2024;110085.

23. Nagy Z, Yong FY, Frei M, Schlueter A. Occupant centered lighting control for comfort and energy efficient building operation. Energy Build. 2015;94:100-8.

24. Deng J, Eklund M, Sierla S, Savolainen J, Niemistö H, Karhela T, et al. Deep reinforcement learning for fuel cost optimization in district heating. Sustain Cities Soc. 2023;99:104955.

25. Mocanu E, Mocanu DC, Nguyen PH, Liotta A, Webber ME, Gibescu M, et al. On-line building energy optimization using deep reinforcement learning. IEEE Trans Smart Grid. 2018;10(4):3698-708.

26. Dalamagkidis K, Kolokotsa D, Kalaitzakis K, Stavrakakis GS. Reinforcement learning for energy conservation and comfort in buildings. Build Environ. 2007;42(7):2686-98.

27. Langer L, Volling T. A reinforcement learning approach to home energy management for modulating heat pumps and photovoltaic systems. Appl Energy. 2022;327:120020.

28. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.

29. Mason K, Grijalva S. A review of reinforcement learning for autonomous building energy management. Comput Electr Eng. 2019;78:300-312.

30. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529-33.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press