District Heating Optimization in Residential Buildings Using Reinforcement Learning with Adaptive Context-aware Predictive Environment

Sai Sushanth Varma Kalidindi[a,b,*], Hadi Banaee[a], Hans Karlsson[b], Amy Loutfi[a]

[a] *Machine Perception and Interaction Lab, Örebro University, Örebro, 70182, Sweden*
[b] *EcoGuard AB, Örebro, 70225, Sweden*

Abstract

As district heating networks evolve to meet climate-neutral objectives, optimizing their control under heterogeneous building characteristics and dynamic environmental conditions remains a significant challenge. Traditional control strategies often lack the adaptability necessary to account for building-specific dynamics and to ensure real-time adherence to operational safety constraints. In this work, we present an integrated machine learning-based framework that combines an adaptive context-aware transformer model with deep reinforcement learning to address these limitations. The proposed approach introduces an adaptive context-aware transformer as a predictive environment within a Deep Q-Network (DQN) framework, enabling data-driven, building-specific control of district heating systems. Utilizing real-world data from 148 residential buildings across Sweden and Finland, the model incorporates contextual embeddings and temporal features to predict indoor temperature trajectories with high accuracy, achieving root mean square error values between 0.18–0.24 °C for Swedish buildings and 0.26–0.32 °C for Finnish buildings. The DQN agent leverages these predictions to optimize heating control while ensuring compliance with operational safety limits and preserving occupant comfort. Experimental results demonstrate significant energy savings, with mid-rise buildings achieving up to 14.85% reduction in energy consumption, and peak seasonal savings exceeding 20% during spring months. This integrated approach illustrates the potential for substantial energy optimization and reliable indoor climate management in future district heating networks.

Keywords: Residential Buildings; Adaptive Context-aware Transformer; District Heating; Energy Optimization; Reinforcement Learning (RL)

1. Introduction

District heating represents a critical component of sustainable urban energy systems, supplying thermal energy to multiple buildings via centralized production and insulated distribution networks. In Sweden, district heating infrastructure forms the backbone of residential heating, serving approximately 93% of apartment buildings. They account for nearly 62% of the heating market share in residential and service sectors [1]. In 2023, Swedish district heating networks delivered 49.8 TWh of heat, of which residential buildings consumed approximately 31.2 TWh, sufficient to heat over two million homes.

Despite its widespread adoption, the efficient control of district heating systems remains a significant challenge. Conventional control strategies, typically based on outdoor temperature compensation curves and basic feedback mechanisms, often fail to accommodate the heterogeneous thermal characteristics of individual buildings or the dynamic patterns of occupant behavior.. These limitations are further increased by the extreme climatic variability in Sweden, where conventional control strategies fail to accommodate building-specific thermal characteristics as demonstrated by Kensby's analysis of thermal energy storage potential in residential buildings [2] and Mangold et al.'s comprehensive study of energy performance uncertainties in Swedish building stock.

[*] Corresponding author: Sai Sushanth Varma Kalidindi, Email: sushanth.kalidindi@oru.se, Affiliation: Machine Perception and Interaction Lab, Address: Teknikhuset, Fakultetsgatan 1, Örebro University, Örebro, 70182, Sweden

A fundamental requirement for meeting these challenges is the incorporation of context-aware control, wherein building-specific attributes such as geographical location, structural characteristics, vertical rise, apartment density, and functional type are systematically integrated into the control process. In this context, context-awareness refers to the capacity of the control system to model and adapt to the unique thermal behaviors of each building, enabling more accurate prediction of indoor temperature dynamics and facilitating customized control actions. Without such contextualization, traditional models risk oversimplification, leading to inefficiencies, suboptimal comfort conditions, and increased energy consumption.

In parallel with optimizing energy usage and maintaining thermal comfort, it is imperative that any control strategy strictly enforces operational safety constraints. Supply pipe temperatures must be maintained within safe operating bounds to prevent thermal stress, system degradation, and occupant discomfort. Thus, achieving optimization under safety-critical constraints presents an additional, nontrivial challenge that control strategies must explicitly address.

Recent developments in artificial intelligence, particularly Wang et al.'s comprehensive analysis of shallow versus deep machine learning approaches for building thermal load prediction [3] and Li et al.'s novel transformer-based network for cooling load forecasting, offer powerful tools to tackle these complexities [4].

Nevertheless, three key gaps persist in existing approaches: (1) insufficient integration of building-specific contextual information in predictive modeling [5]; (2) inadequate modeling of the complex interactions among indoor, outdoor, and supply temperatures [4]; and (3) limited reinforcement learning strategies that can optimize energy consumption while ensuring comfort and operational safety [6]. In response, this study proposes a novel integrated framework that advances the state of the art by:

- Enhancing our previously developed Adaptive Context-Aware Transformer [7] to serve as an environment simulator for reinforcement learning, enabling accurate, dynamic indoor temperature prediction across diverse buildings in response to varying control strategies.

- Implementing a Deep Q-network (DQN) agent that leverages the adaptive context-aware transformer model as its environment, enabling adaptive control of heat while maintaining safety constraints.

- Moreover, evaluating multi-building district heating control that can generalize across different building types and characteristics while maintaining individual building performance.

The rest of this paper is structured as follows: Section 2 explores the related works; Section 3 describes the data and pre-processing of residential buildings in Sweden and Finland; Section 4 introduces the methodology of adaptive context-aware transformer and reinforcement learning; Section 5 presents the experimental setup for both environment and reinforcement learning; Section 6 showcases the results and findings; and finally, Section 7 presents the discussion and conclusions.

2. Related Work

The intersection of artificial intelligence and energy systems represents a critical frontier for addressing global climate challenges and energy efficiency goals. District heating networks, which supply approximately 12% of global space heating demand, present substantial opportunities for intelligent optimization through advanced artificial intelligence techniques. The integration of machine learning methods in energy systems has emerged as an area for both energy policy and artificial intelligence research communities. Artificial intelligence applications in energy systems have demonstrated transformative potential across multiple domains, including demand forecasting, grid optimization, and building energy management. Deep learning architectures have shown success in capturing complex temporal and spatial patterns characteristic of energy consumption in various environments. The application of reinforcement learning to energy control systems represents a shift from reactive to proactive energy management, enabling systems to learn optimal strategies through interaction with dynamic environments.

District heating control systems have evolved significantly over the past decades, transitioning from basic temperature regulation with control graphs to sophisticated AI-driven solutions. Werner's comprehensive analysis of district heating development demonstrated this evolution [8], while Lund et al. examined the integration of renewable energy sources and waste heat recovery systems [9], showing how technological advancement reflects increasing demands for energy efficiency and sustainability. Gustafsson et al. improved district heating substation efficiency through new control strategies [10], while Carli et al. developed internet-of-things based architectures for model predictive control of heating, ventilation, and air conditioning systems in smart buildings [11]. Serale et al. examined model predictive control for enhancing building and heating, ventilation, and air conditioning system energy efficiency across different climatic conditions and building types [12], while Wang et al. investigated building thermal load prediction through machine learning approaches [3]. Ahmad et al. developed short-term energy prediction methods for district-level load management [13], and Wang and Hong analyzed reinforcement learning opportunities and challenges for building controls [14]. Weinberg et al. provided a comprehensive review of reinforcement learning applications for building energy systems from a computer science perspective [15].

2.1. Traditional HVAC/Central Heating Control Methods

The evolution of district heating control can be traced through three distinct generations of technology, each marking significant improvements in efficiency and control capabilities:

First-generation systems (1970-1990) primarily relied on outdoor temperature compensation curves for supply temperature adjustment as documented by Gustafsson et al. [10]. Werner's comprehensive analysis of district heating systems, while robust, provided only basic functionality and lacked the ability to adapt to building-specific characteristics or occupant behavior patterns. International studies during this period focused mainly on improving the reliability and basic efficiency of these systems [8].

Second-generation systems (1990-2010) introduced Model Predictive Control (MPC), incorporating weather forecasts and building thermal models. Studies by Carli et al. demonstrated that internet-of-things based MPC implementations could achieve 15-25% energy savings compared to first-generation approaches [11]. However, these improvements came with increased computational complexity and the challenge of maintaining accurate building models. Similar findings were reported by Serale et al. in international studies, where MPC implementations showed consistent energy savings across different climatic conditions [12].

Current third generation of control systems has introduced advanced features such as thermal storage utilization detailed by Kenby[16] and dynamic setpoint adjustment. Mangold et al. analyzed energy performance data across different building stocks, revealing how variations in thermal characteristics significantly impact control effectiveness [17]. This generation has also seen the integration of renewable energy sources and waste heat recovery systems which are examined by Lund et al. [9], necessitating more sophisticated control approaches.

2.2. Predictive Models for Energy and Sensor Data

The advent of machine learning has transformed predictive modeling in building energy systems. Early statistical approaches have been replaced by deep learning methods, particularly in handling complex temporal patterns characteristic of district heating systems. Wang et al. investigated building thermal load prediction through shallow and deep machine learning approaches [3]. International research has demonstrated consistent improvements in prediction accuracy using these advanced methods across different climatic zones and building types, as shown by Ahmad et al. in their district-level energy prediction study [13].

Transformer architectures have shown particular promise in this domain, with their self-attention mechanisms proving particularly effective for time series prediction in building energy systems. Ni et al. demonstrated the effectiveness of deep learning approaches for multi-horizon building energy forecasting [18], while Tian et al. showed success with data-driven parallel prediction methods for building energy consumption [19]. While traditional approaches focused primarily on temporal patterns, recent advancements have emphasized the importance of incorporating contextual information. In our previous work, we introduced the concept of context-aware models for indoor temperature prediction, demonstrating their effectiveness across different building types [20]. Building on this foundation, we developed a comprehensive adaptive context embedding architecture that further enhanced prediction accuracy by integrating building-specific characteristics directly into the transformer framework [7]. This architectural advancement validated our earlier findings and showed how contextual information could be systematically incorporated into prediction models. where Tian et al. [19] and Wang et al. [3] showed that the integration of building context and temporal features has emerged as crucial for accurate prediction, particularly in environments with diverse building characteristics and varying weather conditions [19, 3].

Recent developments have focused on handling unique challenges in district heating systems. Wang et al. developed a hybrid approach combining CNN and LSTM architectures achieved a 17% improvement in prediction accuracy compared to traditional methods [3]. This advancement in architectural design has been particularly effective in capturing both spatial and temporal dependencies in building thermal behavior, a finding supported by Wei et al. in multiple international implementations [21].

2.3. Reinforcement Learning Based Control

Reinforcement learning represents a paradigm shift in building control systems, offering adaptive and optimizing capabilities beyond traditional approaches. Al Sayed et al. provided a comprehensive review of deep reinforcement learning applications in HVAC control, highlighting both opportunities and challenges in practical implementations across different geographical and climatic contexts [22].

Recent research has focused on multi-building control strategies. Tian et al. demonstrated the effectiveness of DQN-based approaches in handling system complexity while maintaining individual building performance [19]. This builds on earlier work by Nagy et al. [23] and Wang[14] and Hong, who established the feasibility of deep reinforcement learning for optimal space heating control . International studies have further validated these findings across different building types and climate zones.

Safety and reliability in reinforcement learning-based controllers have emerged as critical research areas. Deng et al. presented a framework for safe reinforcement learning in district heating systems, addressing the crucial need for constraint satisfaction and stability guarantees [24]. This work aligns with international efforts by Weinberg et al. to develop robust control systems that can maintain performance under varying conditions [15].

The integration of predictive models with reinforcement learning has shown significant improvement in prediction accuracy. Mocanu et al. demonstrated effective online building energy optimization through deep reinforcement learning [25], while Dalamagkidis et al. showed successful balancing of thermal comfort with energy savings [26]. These approaches have been validated by Langer and Vollingacross different international contexts, showing consistent performance improvements [27].

2.4. Research Gaps

Despite these advances, three key gaps persist in existing approaches that limit their practical effectiveness. First, insufficient integration of building-specific contextual information in predictive modeling leads to oversimplified control strategies. Although Wang et al. [3] and Ahmad et al. [13] demonstrated machine learning improvements, these methods treat buildings as homogeneous entities without systematically incorporating structural characteristics, geographical location, or functional classifications. Second, inadequate modeling of complex interactions among indoor, outdoor, and supply temperatures constrains control effectiveness. Current approaches focus on isolated temperature relationships rather than capturing dynamic interdependencies. Although Tian et al. [19] and Ni et al. [18] addressed some aspects through advanced forecasting methods, comprehensive thermal interaction modeling remains underdeveloped. Third, limited reinforcement learning strategies that can simultaneously optimize energy consumption while ensuring comfort and operational safety represent a significant constraint. While Al Sayed et al. reviewed various applications [22] and Mocanu et al. demonstrated optimization capabilities [25], existing approaches often prioritize single objectives without adequately balancing competing requirements of efficiency, comfort, and safety. These gaps motivate our integrated approach that combines adaptive context-aware transformer models with deep reinforcement learning to address building-specific adaptation and multi-objective optimization simultaneously.

3. Data preparation

3.1. Raw Data

The data used in this study comprises real-world measurements from buildings across Sweden and Finland. The dataset includes measurements from buildings distributed across three lands in Sweden (Norrland, Svealand, and Götaland) and Finland (Norrland and Österland), with data provided by EcoGuard AB covering periods from 3 to 10 years with 15-minute measurement intervals.

The buildings from both Sweden and Finland, with detailed descriptions, are shown in Table 1. The Swedish buildings, which are all residential, are distributed across three lands: Norrland within the cities of Skellefteå and Luleå; Svealand within the cities of Stockholm, Västeras, Karlstad, and Skövde; and Götaland within the cities of Trollhättan, Gothenburg, and Malmö. The Finnish buildings are more diverse and also include commercial and care

Residential buildings in Sweden and Finland			
Lands	Name of the City (Location)	Type of building	No of buildings
Norrland	Skellefteå	Residential building	6
	Luleå		21
	Oulu	Commercial building	1
	Tyrnävä		1
Svealand	Stockholm	Residential building	16
	Västeras		6
	Karlstad		27
	Skövde		7
Götaland	Trollhattan	Residential building	6
	Gothenburg		19
	Malmö		12
Österland	Vantaa	Residential building Care building Commercial building	8
	Helsinki	Residential building Commercial building	3
	Espoo	Commercial building	7
	Vaasa		1
	Järvenpää		1
	Kronoby		1
	Raisio		1
	Pori		1
	Vihti	Care building	1
	Mikkeli	Commercial building Care building	2
Total number of buildings			148

Table 1: Information on Buildings data from Sweden and Finland

buildings, which are distributed across two lands: Norrland within the cities of Oulu and Tyrnävä; Österland within the cities of Vantaa, Helsinki, Espoo, Vaasa, Järvenpää, Kronoby, Raisio, Pori, Vihti, and Mikkeli.

Three primary sensor types were utilized in each building: *Indoor temperature* sensors installed in each apartment or home, *Pipe temperature* sensors measuring the district heating supply to the building, and *Outdoor temperature* sensors. To obtain a representative indoor temperature for each building, we aggregated readings from all measured spaces within a building to calculate an average value. The final dataset consists of three key measurements per building: the averaged indoor temperature (T), outdoor temperature (OT), and pipe temperature (PT).

For the reinforcement learning framework, we selected a subset of five buildings from Stockholm in Sweden with data availability spanning three years (2020-2023). This subset was chosen to ensure consistent data quality and comparable operating conditions, providing a robust foundation for evaluating our RL control strategies.

3.2. Context-aware data preparation

Besides the raw data, we introduce the following meta-data to characterize the buildings in our dataset through a comprehensive set of contextual attributes reflecting their geographical and structural characteristics. We classify all buildings based on their vertical profile into three categories: Low Rise (LR) for buildings with 1-3 floors, Mid Rise (MR) for buildings with 4-6 floors, and High Rise (HR) for buildings with 7 or more floors.

To provide more understanding of building scale, we introduced the Apartment Density Index (ADI). This metric represents the horizontal spread of buildings, complementing vertical classification. In the Apartment Density Index calculation, "spaces" correspond to apartments in residential buildings, while in non-residential buildings, they represent functional areas (offices, patient rooms, service areas). The ADI is calculated as follows:

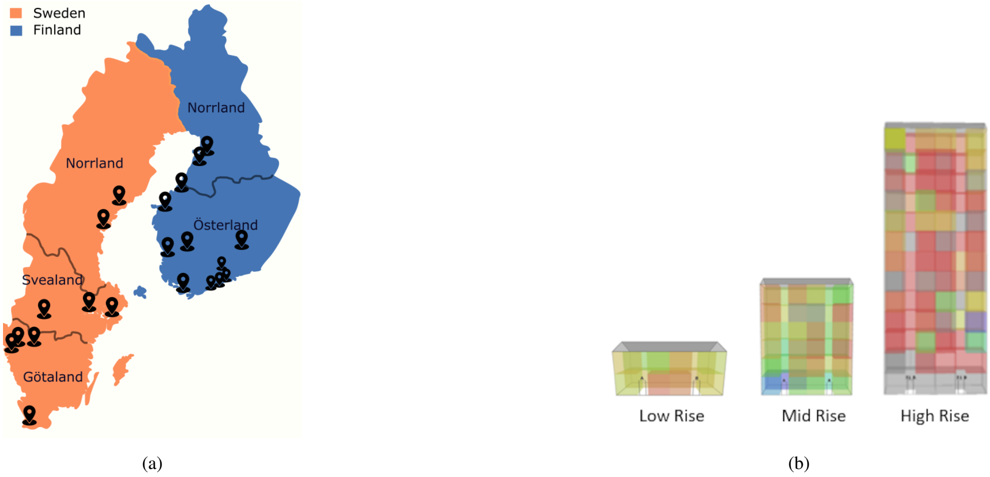
$$\mathrm{ADI} = \frac{S}{F} \tag{1}$$

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

The complete meta-data also includes the specific floor count for each building. Figure 1a presents the geographical distribution of buildings across both countries, with city locations marked. Figure 1b illustrates examples

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
...		...	...	...	...

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



(a)

(b)

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.

of different building rise categories, while Table 2 provides a sample of the meta-data attributes for representative buildings from each location and category.

3.3. Data Pre-processing

The measured multivariate time series data are collected with a time interval of 15 *min*. The multivariate time series data is resampled to hourly intervals to align with the operational characteristics of the district heating control system. For each building, we performed the following preprocessing steps:

- Normalization of temperature values to the range [-1, 1] to ensure uniform scaling across different measurement types and buildings.
- Temporal features (TF) are extracted from timestamps to capture cyclical patterns in heating demand. These features include *Year, Month, Week of the Year, Day of the Month, Day of the Week, and Hour of the Day*, enabling the model to identify and leverage both seasonal trends and daily usage patterns.

The dataset was structured to reflect real-world deployment scenarios, with a 70% allocation for the training set, 30% for the validation set, and 30% of the recent data reserved for the test set. This partitioning strategy ensured robust model development while maintaining appropriate evaluation conditions. For the prediction task of the environment (Adaptive Context-aware transformer model) , the average indoor temperature is the target variable, while outdoor temperature and pipe temperature served as input features. The extracted temporal features were encoded into a positional encoding matrix. A similar pre-processing approach was applied to the data used in the reinforcement learning experiments, maintaining consistency in data handling across both prediction and control tasks. This ensures that our models operate on comparable data representations, facilitating integrated analysis of prediction and control performance.

4. Proposed Approach

The optimization of district heating systems represents a critical challenge in balancing energy efficiency, occupant comfort, and system control. Traditional control approaches rely on reactive mechanisms or simplistic outdoor temperature compensation curves that ignore the unique thermal characteristics of different buildings. This one-size-fits-all approach causes excess energy consumption, inconsistent indoor temperatures, and reduced comfort. What's needed is an intelligent control system that understands each building's distinct thermal behavior, anticipates changing conditions, and makes precise adjustments that maintain comfort while minimizing energy use. To address these challenges, we developed an integrated framework combining deep learning's predictive power with reinforcement learning's adaptive decision making capabilities.

The novelty of this work lies in the integration of adaptive context-aware transformers as an environment simulator within the DQN framework, which fundamentally differs from existing building thermal models used in district heating control. Traditional district heating systems rely on reactive control graphs (outdoor temperature compensation curves) that respond to current conditions without anticipating future changes. In contrast, this integration creates a proactive control paradigm where the transformer environment processes weather forecasts and building-specific characteristics to predict 6-hour temperature trajectories, enabling the DQN agent to make anticipatory control decisions for district heating. This integration provides two key advantages over available building thermal models: (1) proactive rather than reactive control through weather forecast integration, and (2) building-specific adaptation through contextual embeddings versus generic thermal parameters.

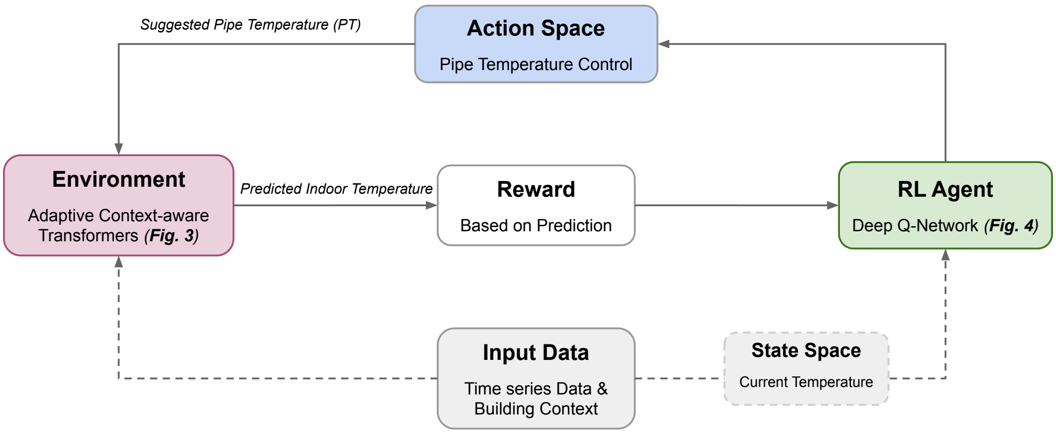


Figure 2: Integrated machine learning framework architecture showing the interaction between adaptive context-aware transformer environment and Deep Q-00agent for district heating control, including data flow paths from time series sensor data and building context through reward calculation mechanisms to pipe temperature control feedback loops.

The proposed approach presents an integrated framework for optimizing district heating in buildings, illustrated in Figure 2. Our system combines an adaptive context-aware transformer model with reinforcement learning to create a closed-loop control system comprising four key components: the Environment (an adaptive context-aware transformer), the Reinforcement Learning Agent (a Deep Q-Network), the Action Space (pipe temperature adjustments), and the Reward (current conditions and forecasts). The Environment processes building-specific features and temperature data, while the RL Agent learns optimal control policies through interaction. In the following sections, we detail the Adaptive Context-aware Transformer architecture, explain its integration with the DQN agent, define the state and action spaces, describe our multi-objective reward function, and demonstrate how these components work together to enable district heating control.

4.1. Adaptive Context-aware Transformer as Environment

Based on our previous model [7], we extend the adaptive context-aware transformer architecture to serve as an environment simulator for reinforcement learning. While the foundational architecture remains similar to our previous work, we make two important extensions: (1) expanding the geographical scope to include Finland alongside Sweden, and (2) incorporating building type as an additional contextual feature.

Our original architecture extended the standard transformer model [28] by adjusting embedding representations based on building-specific characteristics for indoor temperature prediction. For this study, we enhance this model to encode a broader range of contextual meta-data of residential buildings alongside temporal features from multivariate input time series data. The contextual meta-data now includes Land (historical regions of both Sweden and Finland), City, Rise (building height), ADI (Apartment Density Index), Floors, and Building type. The term 'Lands' in this context refers to the traditional historical regions of Sweden and Finland, as illustrated in Figure 1a.Figure 3 illustrates the adaptive context-aware transformer architecture which is modified to use as an environment simulator.

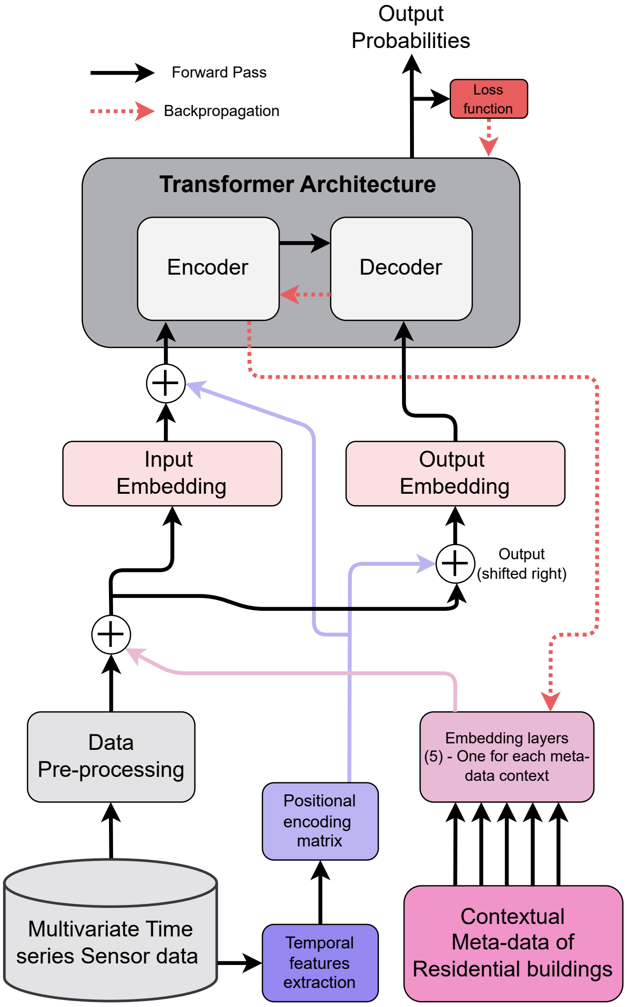


Figure 3: High-level Proposed model architecture [7] with contextual meta-data of residential buildings and Temporal positional encodings.

The temporal features(as detailed in Section 3.3) are then normalized to ensure uniformity. The normalization for each temporal feature TF_i is computed as follows:

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

where TF_i represents the original temporal feature value, min(TF_i) and max(TF_i) denote the minimum and maximum values observed for that feature across the dataset. Aggregating these normalized features results in a temporal positional encoding matrix with each row capturing timestamp time series data.

The core contribution of this work is the integration of the expanded adaptive context-aware transformer model as an environment simulator within the reinforcement learning framework. This creates a predictive control loop where the model generates accurate indoor temperature predictions in response to pipe temperature adjustments. By processing 48 hours of historical data to predict the next 6 hours of indoor temperature, the system enables the reinforcement learning agent to evaluate potential control actions based on their projected outcomes. This forward-looking capability allows the agent to consider both immediate effects and short-term future consequences of its decisions, creating a proactive control strategy that anticipates building thermal responses rather than simply reacting to current conditions.

The contextual meta-data features in table 2 are transformed into dense vector representations through an embedding process as follows:

For each residential building, Given a set of contextual meta-data features $\mathbf{A} = \{A_1, A_2, ..., A_n\}$, each feature A_i is mapped to an embedding vector $\mathbf{e}_i$ using an embedding function E_i, implemented as an embedding layer:

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

The aggregated feature vector $\mathbf{F}$ for an instance is then constructed by concatenating these embedding vectors:

$$\mathbf{F} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n], \tag{4}$$

where n is the total number of features, this vector $\mathbf{F}$, which is an embedded representation, is fed into the transformer model along with the input pre-processed time series, enabling it to learn from a comprehensive and high-dimensional feature space.

Contextual embeddings are fused with temporal features through concatenation and projection at the transformer input. Each contextual data of the building (land, city, rise, ADI, floors, type of building) is converted to a 10-dimensional embedding vector (using the optimal embedding size factor). These six 10-dimensional embeddings are concatenated to form a 60-dimensional contextual vector $\mathbf{F}$ (Equation 4). This contextual vector is then projected to match the transformer's hidden space (64-dimensional) and added to each time step in the 48-hour temporal sequence. This means every timestep carries both temporal information (time patterns) and building information (building characteristics), enabling the model to learn how different building types respond to over time.

Similar to our previous work [7], the embedding layers map variables into a high-dimensional vector space of dimensionality $\boldsymbol{D}$, aligned with the transformer's hidden layers. These embeddings are iteratively refined via gradient descent to capture the underlying contextual patterns of the buildings.

4.2. Reinforcement Learning (DQN)

Reinforcement Learning (RL) represents a paradigm in machine learning where an agent learns to make sequential decisions through interactions with an environment [29]. In this study, we employ Deep Q-Network (DQN), a pioneering deep reinforcement learning algorithm introduced by Mnih et al. [30] that combines Q-learning with deep neural networks to handle high-dimensional state spaces. DQN overcomes the limitations of traditional Q-learning by utilizing a neural network to approximate the Q-function, allowing it to handle complex state-action mappings required for building temperature control.

4.2.1. State space

The state space formulation is critical for capturing the temporal dynamics of building thermal behavior while maintaining computational feasibility. At each timestep t, the state s_t is represented as a 9-dimensional vector that encapsulates both current conditions and future predictions:

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

where:

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

The complete state space $\mathcal{S}$ can be formally defined as:

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

This forward-looking state design allows the agent to develop preemptive control strategies, particularly crucial for buildings with significant thermal inertia where reactive control might be insufficient. The inclusion of future outdoor temperature predictions enables the agent to anticipate and proactively respond to upcoming weather changes, while the bounds on heating reflect physically meaningful constraints of the system.

4.2.2. Action space

The action space balances control granularity with system stability through discrete steps and scaling factors. At each timestep t, the agent selects a base action a_t from the discrete action space $\mathcal{A} = \{-4.0, -2.0, 0, +2.0, +4.0\}$, representing pipe temperature adjustments in °C. The selection of discrete action increments ({-4.0, -2.0, 0, +2.0, +4.0}) was determined through balancing system responsiveness and operational stability. Smaller increments (±1.0°C) provided insufficient control authority for buildings with high thermal inertia, requiring multiple sequential actions to achieve desired temperature changes and resulting in slower response times. Larger increments (±6.0°C or greater) caused excessive temperature oscillations and potential comfort violations due to system overshoot, particularly in buildings with low thermal mass. The scaling mechanism 7 further refines control granularity based on temperature error, effectively creating 15 discrete control levels ranging from ±0.5°C to ±6.0°C. This approach achieves optimal trade-offs between control precision and system stability: fine-grained control when near setpoint temperatures and aggressive correction when significant deviations occur. The final action a_{final} is computed by multiplying the base action a_t with a scaling factor ζ:

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

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

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

This creates 15 possible actions through the combination of five base actions and three scaling factors. For example, a base action of +2.0°C can result in actual adjustments of +1.0°C (ζ = 0.5), +2.0°C (ζ = 1.0), or +3.0°C (ζ = 1.5). This scaling mechanism enables more precise control responses: smaller adjustments when near the desired temperature and larger adjustments when significant corrections are needed.

To ensure safe operation and prevent thermal shock to the heating system, we implement adaptive safety bounds (20-70°C) that dynamically constrain the maximum allowable temperature adjustments based on current system conditions, building thermal state, and operational constraints. The 20-70°C pipe temperature bounds represent the operational limits of district heating systems, where 20°C is the minimum temperature the system can supply and 70°C is the maximum safe operating temperature to prevent thermal stress and system damage. These safety constraints were given to us by experts. These safety constraints are enforced at the agent level during action selection through a clipping mechanism. When the DQN agent selects an action a_t, the final pipe temperature is computed as:

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

ensuring that all control actions remain within safe operational bounds regardless of the agent's policy. This implementation prevents the agent from learning potentially dangerous control strategies while maintaining the flexibility to explore the full safe operating range during training.

4.2.3. Reward Function

The reward function is designed to maintain optimal indoor comfort by keeping building temperatures at a desired set-point of 21°C, while considering multiple control objectives and system stability. The desired indoor temperature set-point was chosen based on standard comfort requirements for residential buildings. We formulate a reward function that considers temperature maintenance, weather responsiveness, and control stability:

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

The temperature maintenance reward (R_{temp}) employs an exponential decay structure to strongly incentivize keeping the indoor temperature close to the 21°C set-point while progressively penalizing larger deviations:

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

where T_{pred} is the predicted indoor temperature. This formulation provides high rewards for maintaining temperatures within ±1.0°C of 21°C while implementing a higher penalty for larger deviations to ensure comfort of the occupants.

The weather responsiveness component ($R_{weather}$) encourages the agent to adjust the pipe temperature proportionally to changes in outdoor temperature:

$$R_{weather} = -|\Delta PT - \lambda \cdot \Delta OT| \tag{12}$$

where ΔPT and ΔOT represent changes in pipe and outdoor temperatures, respectively, and λ is a scaling factor that captures the desired relationship between outdoor temperature changes and control responses.

The stability term ($R_{stability}$) is crucial to preventing rapid temperature fluctuations that could lead to occupant discomfort and system inefficiency:

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

This quadratic penalty on consecutive temperature differences discourages abrupt changes in indoor temperature. By summing over the previous n-1 timesteps, the agent is encouraged to maintain smooth temperature transitions rather than making sudden adjustments. This is particularly important in building heating systems where rapid temperature changes can lead to increased energy consumption and reduced occupant comfort.

Finally, R_{safety} implements a binary reward (+1.0) for maintaining the pipe temperature within safe operational bounds (20-70°C):

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

The weighting coefficients $\alpha = 1.0$, $\beta = 0.5$, $\gamma = 0.3$, and $\delta = 0.2$ were determined through hyper-parameter tuning during the training process. Multiple combinations of these coefficients were evaluated based on the agent's performance in maintaining temperature stability and comfort. The final values were selected based on their effectiveness in balancing the temperature maintenance objective with weather responsiveness and system stability. To ensure stable training dynamics, the total reward is bounded within [-10, 10]. This reward structure effectively trains the agent to maintain the desired 21°C indoor temperature while responding appropriately to weather changes and system constraints.

4.2.4. DQN architecture

The DQN architecture, depicted in Figure 4, implements a comprehensive learning system for building temperature control. The architecture consists of several interacting components that work together to learn and execute optimal control policies. At its core, the system utilizes two neural networks - the DQN Model and Target Model - with identical structures comprising a 9-neuron input layer, two hidden layers of 256 neurons each with ReLU activation functions, and a 5-neuron output layer corresponding to the discrete action space.

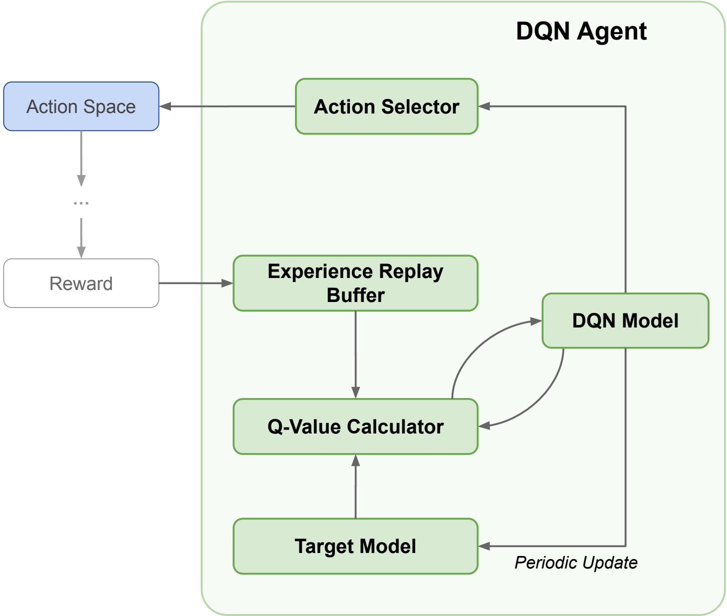


Figure 4: Deep Q-Network (DQN) architecture showing the integration of experience replay, Q-value calculation, and model components.

The learning process begins with the Experience Replay Buffer, which stores tuples of state transitions, actions, and rewards. These experiences are sampled during training to break temporal correlations and improve learning stability. The Q-Value Calculator then processes these samples to compute the temporal difference error, utilizing both the current DQN Model and Target Model to generate stable Q-value estimates.

The Target Model, which is periodically updated from the DQN Model every 5 episodes (shown as "Periodic Update" in the figure 4), serves as a stable reference for Q-value computation. This double Q-learning approach significantly reduces overestimation bias in the learning process. The Action Selector implements an ϵ-greedy strategy, choosing between exploratory random actions and exploiting the learned policy based on the DQN Model's output.

Training is performed using the Adam optimizer with a learning rate of 0.001, with gradient clipping applied to ensure stable updates. The Reward Calculator evaluates the agent's actions based on multiple objectives, feeding this information back to the Experience Replay Buffer to complete the learning loop. The entire system operates within the defined Action Space, which consists of five discrete temperature adjustment actions.

The interaction between these components creates a learning system capable of adapting to varying environmental conditions while maintaining stable and efficient control policies for *Pipe temperature* control.

4.3. Integration of Adaptive Context-aware Transformer Environment and RL Agent

The core idea of our integration approach lies in creating a connection between prediction and control. By combining the adaptive context-aware transformer's ability to model unique building thermal behaviors with the DQN's capacity to learn optimal control strategies, we've developed a system that can simultaneously adapt to individual building characteristics while creating generalizable control policies that perform effectively across diverse building types. This integration creates a comprehensive control system that leverages both predictive modeling and reinforcement learning for building temperature control. The transformer model serves as a differentiable environment simulator, providing the DQN agent with accurate predictions while accounting for building-specific characteristics through contextual embeddings.

At each timestep t, the system operates within the state space defined in Equation 5, with states augmented by building-specific contextual embeddings $F_{building}$ in Equation 4 that encode metadata such as building type and location. The integration process follows a structured interaction loop where the transformer first processes a historical sequence of 48 timesteps to establish the current state and predict future temperature responses:

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

The DQN agent then selects an action a_t from the discrete action space $\mathcal{A} = \{-4.0, -2.0, 0, +2.0, +4.0\}$, representing pipe temperature adjustments in °C. The selected action is constrained by adaptive safety bounds to maintain operation within the valid state space:

$$T^{pipe}_{t+1} = \text{clip}(T^{pipe}_{t} + a_t, 20, 70) \tag{16}$$

The transformer environment processes this action and predicts the resulting indoor temperature trajectory. The system then computes the reward according to Equation 10, evaluating the action's effectiveness across multiple objectives including temperature maintenance, weather responsiveness, stability, and safety constraints.

To enable learning across multiple buildings while maintaining building-specific adaptability, the system employs a shared replay buffer $\mathcal{B}$ that stores experiences as tuples:

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

During training, experiences are sampled from this buffer to update the DQN agent's policy. The inclusion of building embeddings F_{building} in the state representation enables the agent to learn building-specific control strategies while generalizing across similar building types.

This integrated architecture enables effective temperature control across diverse building types while adapting to individual building characteristics and varying weather conditions. The combination of the transformer's predictive capabilities with the DQN's learning abilities results in a control system that can maintain comfortable indoor temperatures while respecting system constraints and optimizing energy usage.

4.4. Evaluation metrics

4.4.1. Evaluation metrics for Adaptive Context-aware Transformer (Environment)

In order to evaluate the prediction performance of the adaptive context-aware transformer, we employed Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) as the evaluation metrics to assess the prediction accuracy of the models. The RMSE and MAE are two widely used metrics in the literature to measure the difference between predicted values and actual values. RMSE takes into account the magnitude of errors and penalizes large errors more heavily than small errors, while MAE treats all errors equally. The R^2 metric measures the proportion of the variance in the dependent variable (i.e., the actual values) that is explained by the independent variable (i.e., the predicted values). An R^2 score of 1 indicates a perfect fit of the model to the data, while a score of 0 means that the model does not explain any of the variability in the data. These metrics provide valuable insights into the performance of time series prediction models and can be used to guide model selection and refinement.

The evaluation metrics are calculated as follows:

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

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

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

where $\hat{y}_i$ is the predicted value, y_i is the actual value, $\bar{y}$ is the output means, and N is the total number of samples in the test data.

4.4.2. Evaluation of RL DQN-agent

The performance of the DQN-agent based control system is evaluated using two key metrics:

1. Control Performance: This metric assesses the agent's ability to maintain desired indoor temperature under varying environmental and operational conditions. Control performance is validated by comparing the predicted indoor temperatures generated by the adaptive context-aware transformer model with actual indoor conditions. A consistent indoor temperature of approximately 21°C, regardless of outdoor fluctuations, indicates robust control.
2. Energy Efficiency: The energy consumption of the system is calculated based on the temperature difference between the pipe supply temperature (T_{pipe}) and the indoor temperature (T_{indoor}). The energy consumption (E) is computed as:

$$E = c \cdot f \cdot \Delta T \cdot t \tag{21}$$

   where:

   - c is the specific heat capacity of water (4.186 kJ/kg·K),
   - f is the assumed water flow rate (2.0 liters per second),
   - $\Delta T = T_{\text{pipe}} - T_{\text{indoor}}$,
   - t is the time interval (in hours).

   Energy values are expressed in kilowatt-hours (kWh), using the conversion factor 1 kWh = 3600 kJ.
   To evaluate energy efficiency, the energy consumed under DQN control is compared against the baseline (original system). Energy savings (S) are calculated as:

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

   where En_{original} and $En_{\text{predicted}}$ represent the energy consumption of the original and DQN-controlled systems, respectively. This metric provides a quantitative measure of how effectively the DQN agent reduces energy usage while maintaining control performance.

5. Experimental setup

5.1. Baseline Traditional Control System

The baseline control system consists of existing outdoor temperature compensation curves currently deployed in district heating networks. This industry standard method establishes a linear relationship between outdoor temperature (OT) and pipe supply temperature (PT) based on the principle that colder outdoor conditions require higher supply temperatures. Each building operates using its linear compensation curve derived from historical operational data. These curves follow the standard linear form $PT = a \times OT + b$, where a represents the temperature slope coefficient that determines the system's responsiveness to outdoor temperature changes, and b represents the base temperature offset (°C) that sets the minimum operating temperature.

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

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

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

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

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

The negative slope coefficients indicate that pipe temperature increases as outdoor temperature decreases. As shown in Figure 5, these linear relationships represent the current industry practice of reactive control that adjusts supply temperatures based solely on current outdoor conditions without anticipating future changes or considering building-specific thermal characteristics. This existing system serves as the benchmark for evaluating our proposed adaptive context-aware transformer environment and DQN-based control approach.

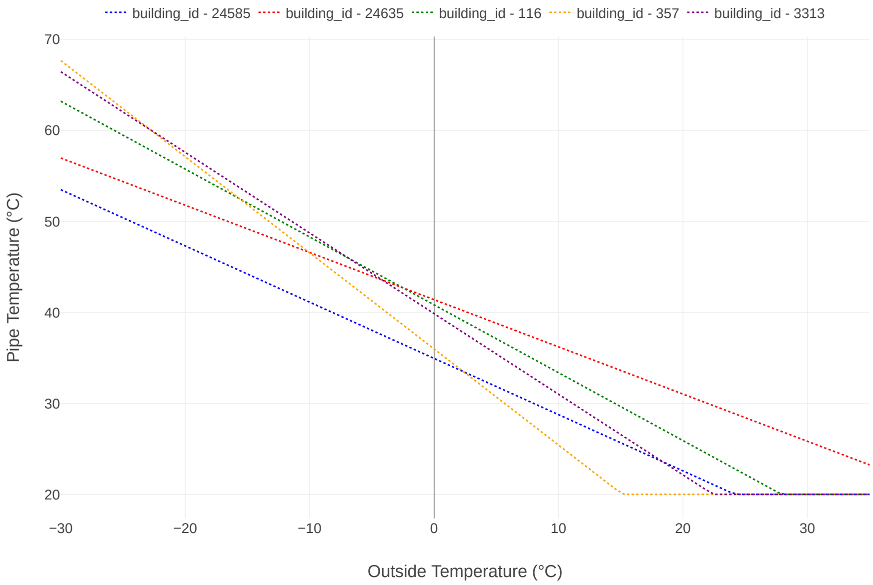


Figure 5: Traditional outdoor temperature compensation curves for five Stockholm buildings showing the linear relationship between outdoor temperature and pipe temperature. Each dotted line represents the current industry standard control strategy for individual buildings, demonstrating the reactive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

5.2. *Training and Test setup for Adaptive Context-aware Transformer*

The experimental implementation extends our previous adaptive context-aware transformer architecture [7] to accommodate a broader scope of building control scenarios. This model is designed to predict indoor temperatures over 6-hour horizons using three primary features: *Indoor Temperature*, *Outdoor Temperature*, and *Pipe Temperature*, while simultaneously learning and utilizing building-specific contextual characteristics. The 6-hour prediction horizon was selected based on expert recommendations from building control systems and aligns with the thermal inertia characteristics of residential buildings in district heating networks. Building control experts indicated that residential buildings typically exhibit thermal response times of 4-8 hours, meaning that heating control actions require this duration to fully manifest their effects on indoor temperatures. The 6-hour horizon provides sufficient lead time for the reinforcement learning agent to make proactive control decisions while remaining within the reliable prediction range of the transformer model. This prediction horizon is consistent with the thermal inertia of typical Swedish and Finnish residential buildings, which show measurable temperature responses within 2-3 hours but require 6-8 hours to reach thermal equilibrium after control adjustments.

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

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

Through extensive hyperparameter optimization, we identified the optimal model configuration through grid search across multiple architectural variations, as detailed in Table 3.

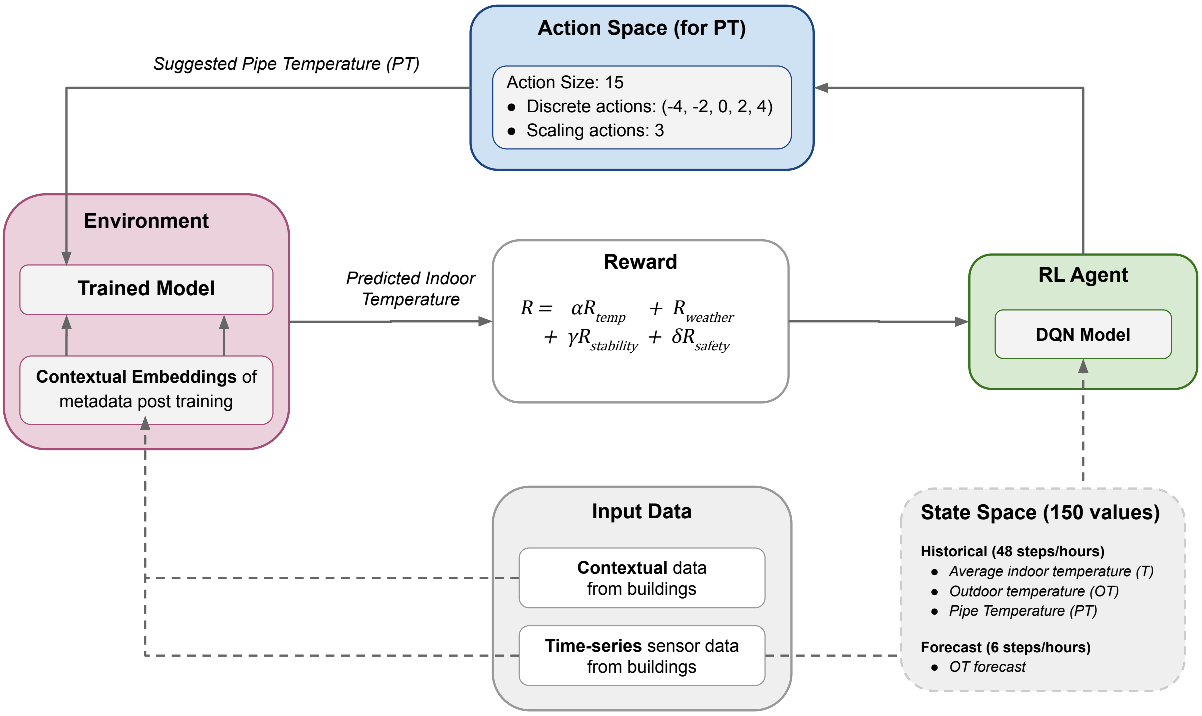

Action Space (for PT)
Action Size: 15
Discrete actions: (-4, -2, 0, 2, 4)
Scaling actions: 3
Suggested Pipe Temperature (PT)
Environment
Trained Model
Contextual Embeddings of metadata post training
Predicted Indoor Temperature
Reward
$R = \alpha R_{temp} + R_{weather} + \gamma R_{stability} + \delta R_{safety}$
RL Agent
DQN Model
Input Data
Contextual data from buildings
Time-series sensor data from buildings
State Space (150 values)
Historical (48 steps/hours)
Average indoor temperature (T)
Outdoor temperature (OT)
Pipe Temperature (PT)
Forecast (6 steps/hours)
OT forecast

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 architecture implements 2 encoder layers, each containing 4 multi-head attention mechanisms operating on 64-dimensional embeddings with 128-dimensional feed-forward networks and ReLU activation. Input features include normalized temperature values (indoor, outdoor, pipe) in the range [-1, 1] and temporal features (year, month, week, day, hour) encoded as positional embeddings and the six contextual features (Land, City, Rise, ADI, Floors, Building Type) are processed through separate embedding layers with 10-dimensional outputs each, concatenated to form a 60-dimensional vector, and projected to match the 64-dimensional transformer space. Training uses mean squared error loss with Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$), gradient clipping at 1.0, and early stopping with 10-epoch patience on validation loss. For testing and validation, we employ a comprehensive evaluation protocol using a held-out test set comprising 15% of the total data. This test set spans multiple seasons and includes diverse building types from both countries to ensure a robust evaluation of the model's generalization capabilities. The model's performance is assessed through its ability to generate accurate 6-step ahead predictions while maintaining building-specific contextual awareness. The testing procedure evaluates the accuracy of temperature predictions with evaluation metrics RMSE, MAE, R^2.

5.3. Reinforcement Learning - DQN Agent

Following the training of an adaptive context-aware transformer as an environment, we will establish a reinforcement learning framework utilizing a DQN agent for heating control across five residential buildings in Stockholm. As shown in Figure 6, while the transformer environment handles building-specific contextual embeddings and predictions of indoor temperature, the DQN agent operates solely on time series data, making it building-agnostic and generalizable across different building types and locations.

The DQN agent's input structure focuses on temporal measurements: 48-step historical sequences of *Indoor Temperature* (T), *Outdoor Temperature* (OT), and *Pipe Temperature* (PT), along with a 6-hour outdoor temperature forecast. The transformer environment separately processes the contextual meta-data (Land, City, Rise, ADI, Floors, Building Type) through its embedding mechanism, enabling accurate indoor temperature predictions for any building type in either country. This separation allows the DQN agent to learn general control strategies while the environment provides building-specific adaptations.

The state space representation for the DQN contains both current conditions and a forecast of outdoor temperature. The transformer environment uses its contextual embeddings to generate accurate temperature predictions for the given building, which the DQN agent then uses to select appropriate control actions. The action space consists of 15 possible actions, with five discrete steps (-4, -2, 0, 2, 4) and three scaling factors to provide fine-grained control over temperature adjustments.

The training process spans 4000 episodes, during which, the environment processes building data and generates indoor temperature predictions using its trained contextual embeddings of buildings. The DQN agent takes action on *Pipe Temperature* from time-series state space and indoor temperature predictions from the environment, selecting control actions. The environment evaluates the *Pipe Temperature* actions using building-specific characteristics to compute rewards.

The DQN network consists of a 9-dimensional input layer (current indoor/outdoor/pipe temperatures plus 6-hour outdoor forecast), two hidden layers with 256 neurons each using ReLU activation and 0.3 dropout, and a 5-dimensional output layer for Q-values corresponding to actions. The DQN network also employs experience replay with a memory buffer that stores state transitions, actions, and rewards. The network uses target updates every 5 episodes and an ϵ-greedy exploration strategy (initial $\epsilon = 0.8$, decaying to 0.05). Training is optimized using Adam optimizer (learning rate = 0.001) with gradient clipping (threshold = 1.0) for stability. The best-performing model is selected based on cumulative reward and temperature control performance across different building types.

The validation of RL-DQN agent uses recent one-year of test set of buildings from Stockholm, leveraging the transformer's ability to handle any building type through its contextual embeddings while the DQN agent applies learned control strategies based solely on temporal patterns. This approach ensures effective temperature control across diverse building characteristics and weather conditions, demonstrating the system's generalization capabilities.

5.4. Training Duration and Computational Complexity

The experimental implementation used NVIDIA RTX 3090Ti and RTX 4090 graphics processing units with 24GB VRAM each. The adaptive context-aware transformer environment training, including hyperparameter optimization from Table 3, required 120 hours of continuous training across 148 buildings. The DQN agent training required 218.6 hours across 4000 episodes with experience replay and target network updates. Inference times are 2.8 seconds for transformer environment predictions and 1.9 seconds for DQN action selection, totaling 4.7 seconds per control decision. The 4.7-second inference time is acceptable for district heating control, where decisions are updated every 15-60 minutes. In comparison, traditional outdoor temperature compensation curves execute control decisions in microseconds using simple lookup tables on standard building management systems.

6. Results

The results are presented in two main parts. First, we evaluate the performance of the adaptive context-aware transformer model for indoor temperature prediction across diverse buildings in Sweden and Finland, including both offline testing and real-world online deployment. Second, we assess the district heating prediction capabilities of the DQN agent, demonstrating its performance in controlling pipe temperature compared to conventional systems. Finally, we analyze the energy optimization outcomes, quantifying energy savings across different building and seasonal conditions. Together, these results validate both the accuracy of our prediction model and the effectiveness of our integrated control approach in maintaining comfort while reducing energy consumption.

6.1. Adaptive Context-aware Transformer (Indoor temperature Prediction) Environment

This section presents the evaluation of our adaptive context-aware transformer model across diverse residential buildings in Sweden and Finland. The model's performance is assessed on test datasets spanning different geographical regions and building characteristics, with a prediction horizon of 6 hours (6 time steps). The model benefits from trained contextual embeddings learned during training, enabling better generalization across different building types and locations.

For Swedish buildings (Table 4), the model demonstrates strong predictive capabilities across different contexts. Geographical analysis shows Götaland achieving the best performance (RMSE: 0.1958, R^2: 0.7187), followed by Svealand (RMSE: 0.2191) and Norrland (RMSE: 0.2245). Among building categories, Mid Rise buildings show particularly strong results with an RMSE of 0.1817 and R^2 of 0.6956. City-wise evaluation reveals higher performance in Skellefteå (RMSE: 0.1487), while maintaining consistent accuracy across other urban areas with RMSE values generally below 0.23.

Context		Prediction Horizon 6 hr (6 steps)		
		RMSE	MAE	R^2
Lands	Norrland	0.2245	0.1971	0.6858
	Svealand	0.2191	0.1883	0.6720
	Götaland	0.1958	0.1866	0.7187
Rise	Low Rise	0.2439	0.2283	0.6666
	Mid Rise	0.1817	0.1677	0.6956
	High Rise	0.2135	0.1760	0.7142
City	Skellefteå	0.1487	0.1078	0.6628
	Luleå	0.2548	0.2092	0.7283
	Stockholm	0.2064	0.2113	0.6864
	Västeras	0.2261	0.1974	0.6223
	Karlstad	0.2317	0.1922	0.7012
	Skövde	0.2233	0.1911	0.7838
	Trollhattan	0.2240	0.1996	0.6143
	Gothenburg	0.1989	0.1960	0.7002
	Malmö	0.2040	0.2116	0.7295

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

Context		Prediction Horizon 6 hr (6 steps)		
		RMSE	MAE	R^2
Lands	Norrland	0.2829	0.2468	0.5938
	Österland	0.3029	0.2669	0.5538
Building Type	Residential building	0.2256	0.2164	0.5697
	Care building	0.3777	0.3379	0.4497
	Commercial building	0.2754	0.2163	0.7018
Rise	Low Rise	0.2618	0.2326	0.6527
	Mid Rise	0.2968	0.2732	0.5215
	High Rise	0.3203	0.2647	0.5473

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

Finnish buildings (Table 5) show comparable performance patterns across different contexts. Residential buildings demonstrate strong prediction accuracy (RMSE: 0.2256), while geographical analysis shows consistent performance in both Norrland (RMSE: 0.2829) and Österland (RMSE: 0.3029). Figure 7 provides a detailed visualization of RMSE values across Finnish cities, showing how prediction accuracy varies by location while maintaining consistent performance below 0.35 RMSE. The variation in performance across different building types and geographical locations highlights the model's adaptability to building types and locations.

These results demonstrate the effectiveness of incorporating and training contextual meta-data embeddings in *Indoor Temperature prediction*, enabling the model to capture and adapt to the unique thermal characteristics of different buildings across various geographical locations. The consistently low error metrics across different contexts validate the model's reliability for practical applications in building temperature control.

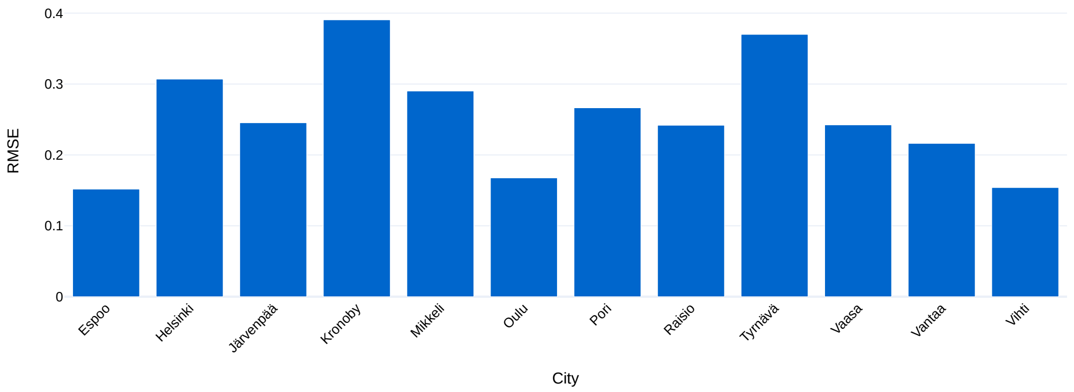


Figure 7: Average RMSE of indoor temperature predictions for buildings in Finland by city.

6.1.1. Adaptive Context-aware Transformer on real-world online prediction

To validate the model's practical applicability and its suitability as an environment for reinforcement learning, we implemented a real-world deployment in Stockholm, testing the adaptive context-aware transformer on five buildings from the training dataset. This online prediction system has been operational for approximately two months, continuously generating indoor temperature predictions based on real-time building data with a strong prediction performance (RMSE ranging from 0.12 to 0.23 across the buildings). Figure 8 illustrates the model's performance in real-world conditions, showing the comparison between actual and predicted indoor temperatures for a mid-rise building in Stockholm over four days in March 2025.

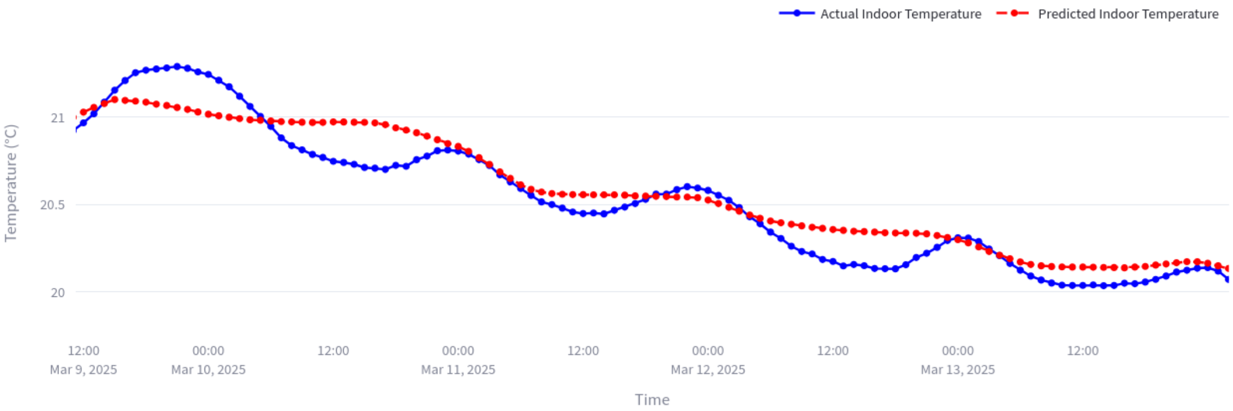


Figure 8: Comparison between actual and predicted indoor temperatures during real-time online deployment: Real-time 6-hour forecasting results for a mid-rise residential building in Stockholm

The predictions closely track the actual temperature patterns, effectively capturing both gradual trends and shorter-term fluctuations. The model successfully maintains prediction accuracy even during temperature peaks and valleys, demonstrating its ability to handle real-world temperature dynamics.

This robust real-world performance is particularly significant for reinforcement learning applications, as it establishes the Adaptive Context-aware transformer as a reliable environment simulator for DQN agent training. The low prediction errors (consistently below 0.23 RMSE) indicate that the model provides accurate predictions and reliable feedback for the RL agent's decision-making process. The model's ability to maintain prediction accuracy across different buildings and varying temperature patterns suggests that it can serve as a dependable environment for learning optimal PT control strategies. This is crucial for the DQN agent's training, as the quality of learned control policies directly depends on the accuracy of the environmental feedback provided by these temperature predictions.

6.2. DQN (RL)

Having established the reliability of an adaptive context-aware transformer as an environment simulator through both offline testing and real-world validation, we now present the performance of the DQN agent in controlling building heating systems.

The results demonstrate the effectiveness of our integrated approach, where the agent leverages accurate temperature predictions to implement efficient heating control strategies. The evaluation spans both control performance and energy efficiency.

6.2.1. District Heating Prediction (PT)

The DQN agent's performance in controlling pipe temperature for district heating demonstrates significant improvements over the existing control system. Figure 9a illustrates the comparison between the original and DQN-controlled pipe temperatures over an eight-day period in winter conditions (December 30, 2021 - January 7, 2022). The DQN agent maintains more dynamic control of the pipe temperature, responding to changing conditions while avoiding the relatively static control patterns seen in the original system. While the original control system maintains PT around 30 - 32°C with minimal adjustments, the DQN agent implements a more responsive strategy with temperatures ranging from 39°C to 49°C based on building needs.

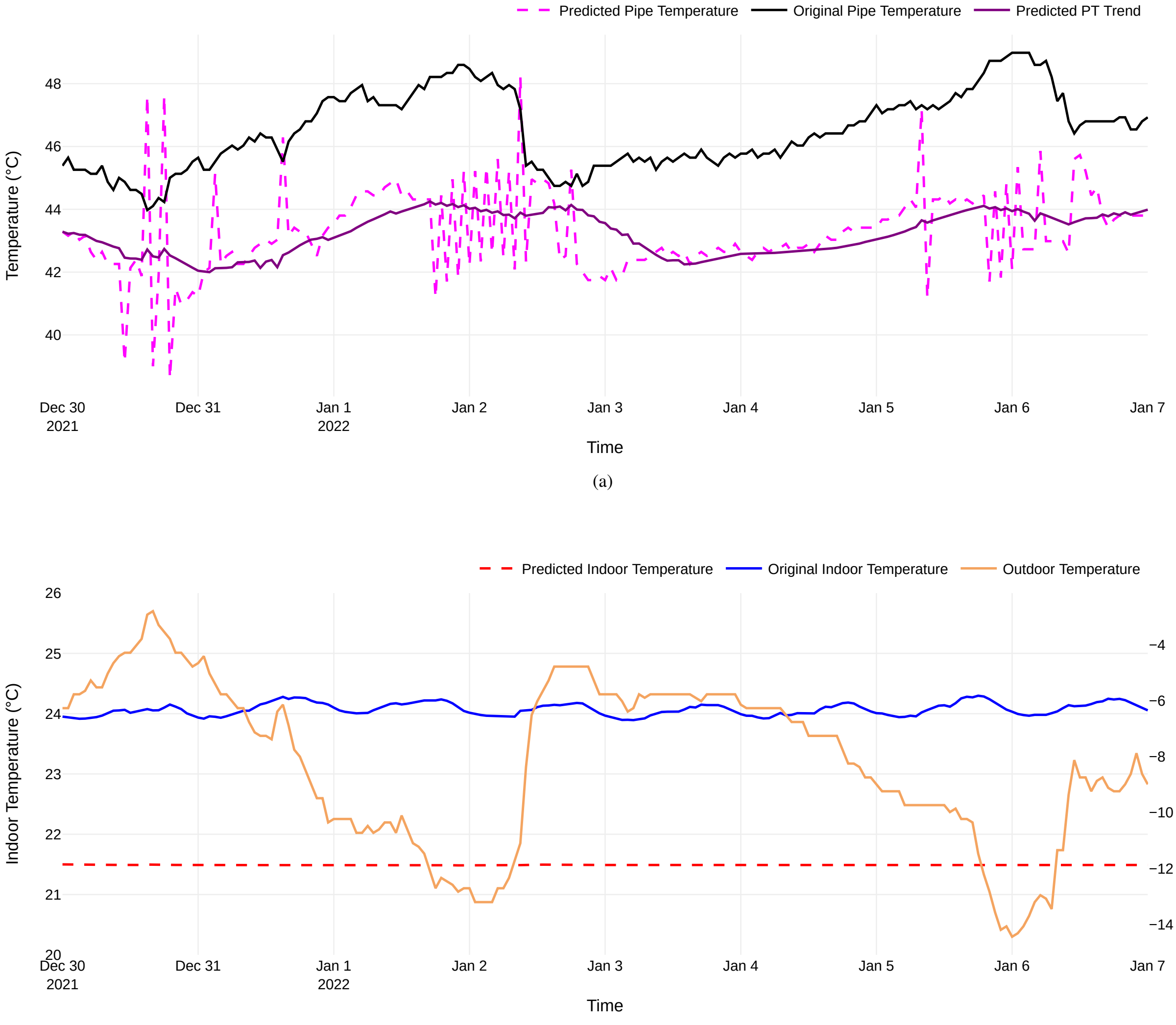
The effectiveness of this control strategy is validated by the indoor temperature predictions from adaptive context-aware transformer model shown in Figure 9b, where the building temperature remains stable around 21°C despite fluctuating outdoor temperatures between 3°C and 8°C. This demonstrates the agent's ability to maintain comfortable conditions while implementing more dynamic PT control strategies.

Figure 9 demonstrates several key system responses during the winter period. Notable events include: (1) December 31st shows a significant outdoor temperature drop from -4°C to -12°C, where the DQN agent proactively increases pipe temperature from 42°C to 46°C while the original system maintains static control around 47°C; (2) January 2nd exhibits the system's adaptive response to warming outdoor conditions (-8°C to -4°C), with the DQN agent reducing pipe temperature to 40°C, demonstrating energy-saving behavior; (3) January 4-5th illustrates temperature stability maintenance, where despite outdoor fluctuations between -6°C and -10°C, the DQN system maintains indoor temperature within 0.5°C of the 21°C setpoint through dynamic pipe temperature adjustments ranging from 41°C to 44°C. These events highlight the system's ability to anticipate and respond to environmental changes while maintaining thermal comfort.

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

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.

Building-specific temperature control performance is detailed in Table 6, which compares original baseline control with DQN predictive control across five Stockholm buildings. The DQN agent maintains indoor temperatures closer to the 21°C target with significantly reduced variability: indoor temperature standard deviations decreased



(a)

(b)

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.

from 0.23-0.87°C (original) to 0.02-0.33°C (predicted), representing up to 92% improvement in temperature stability. For example, building 24635 achieved exceptional stability with indoor temperature variation of only ±0.02°C compared to ±0.25°C in the original system. The DQN approach demonstrates building-specific adaptation through variable pipe temperature strategies, ranging from 30.23°C (building 3313) to 48.89°C (building 357), while the original system showed less responsive control patterns. This precision in temperature control validates the agent's ability to maintain thermal comfort while implementing adaptive heating strategies tailored to individual building thermal characteristics.

6.2.2. District Heating Energy optimization

Stockholm buildings (ID)	Energy Consumption (MWh)		Energy Savings (%)
	Original	Predicted	
26394 - Low Rise	183.16	169.6	7.40
24635 - Mid Rise	97.4	88.58	9.06
313 - Mid Rise	83.57	71.16	14.85
116 - High Rise	80.41	71.5	11.08
648 - High Rise	34.11	30.69	10.03

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.

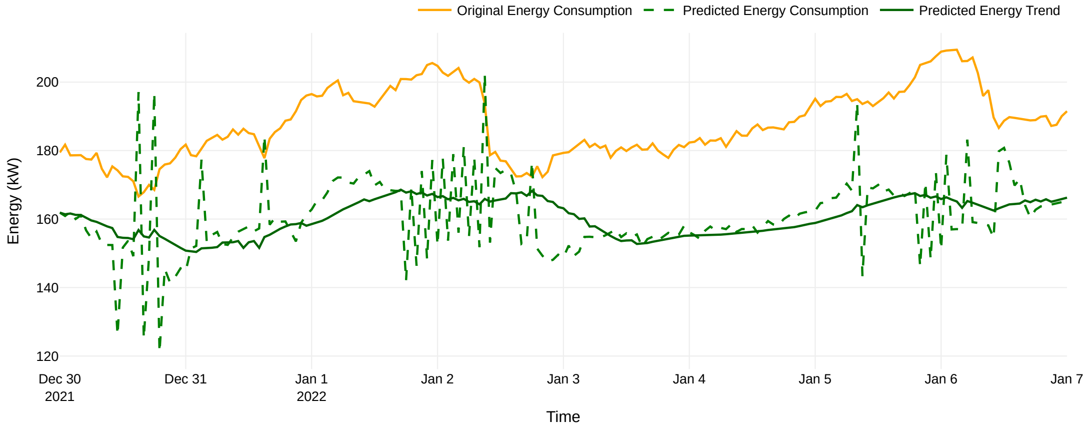


Figure 10: District heating energy consumption comparison between conventional control (solid orange line) and DQN-optimized control (dashed green line) with trend line (solid green).

The DQN agent's more responsive control strategy translates into substantial district heating energy savings across different building types and seasonal conditions. The energy consumption comparison in Figure 10 shows how the DQN agent achieves these energy savings through more efficient PT control. While the original system maintains a relatively constant energy consumption of around 180 - 200 kW, the DQN-controlled system shows more dynamic patterns, reducing consumption to 130 - 170 kW during optimal conditions while maintaining indoor comfort. As shown in Table 7, the system achieves consistent energy savings across different building categories. Mid-rise buildings (MR) show the most significant improvements, with Building 313 achieving 14.85% energy savings, because mid-rise buildings exhibit optimal thermal mass and surface area that provides more predictable thermal response. High-rise buildings (HR) demonstrate similar performance of around 10 - 11%. Low-rise building (26394-LR) maintains stable savings of 7.40%.

The seasonal analysis presented in Figure 11 reveals distinct patterns in energy optimization. In winter, there are consistent savings between 5-15% across all building types, with Building 648 showing the highest efficiency at approximately 14%. During the spring months, Peak energy savings performance, particularly for Building 313, reached over 20%. During summer, more varied performance with some buildings showing reduced savings, notably Building 648, which showed negative savings, indicating increased energy usage. During fall months, energy efficiency recovered with most buildings returning to the 8 - 15% savings range.

The negative energy savings observed in summer, particularly for Building 648, result from the DQN agent implementing more frequent control adjustments during low-demand periods when traditional control remains passive.

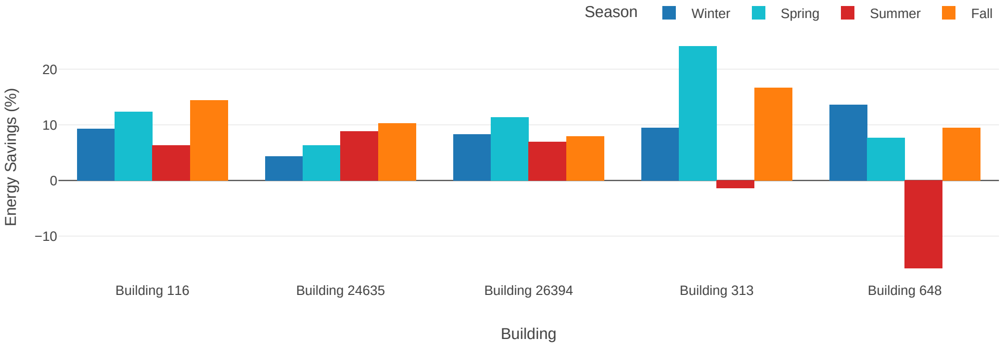


Figure 11: Seasonal distribution of district heating energy savings across different building types in Stockholm: Analysis of five buildings with varying characteristics (LR - Building 26394, MR - Building{24635, 313}, and HR - Building{116, 648}) showing distinct performance patterns across winter,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

During summer months with minimal heating requirements, the agent's precise temperature control through micro-adjustments consumes slightly more energy than passive traditional approaches. This is exacerbated by district heating systems operating less efficiently at very low loads. These results suggest future work should incorporate season-specific control strategies that adapt to varying heating demands throughout the year.

These results demonstrate that the DQN agent successfully balances comfort requirements with energy efficiency, achieving significant savings through more sophisticated control strategies while maintaining desired indoor temperatures. The variation in performance across different building types and seasons highlights the system's ability to adapt to diverse conditions while consistently delivering energy improvements.

7. Discussion

7.1. Results Insights and Limitations

This work demonstrates that integrating adaptive context-aware transformer models with deep reinforcement learning provides an effective approach for optimizing district heating in residential buildings. The adaptive context-aware transformer environment successfully captured building-specific characteristics through contextual embeddings, achieving prediction accuracies with root mean square error values ranging from 0.18°C to 0.32°C across diverse building types and geographical locations. The Deep Q-Network agent effectively leveraged these predictions to implement intelligent control strategies, resulting in substantial energy savings of 7.40% to 14.85% across different building categories while maintaining indoor temperatures within ±0.5°C of desired setpoints. The system demonstrated particular effectiveness for mid-rise buildings and achieved peak seasonal savings exceeding 20% during spring months. The real-world development across multiple countries validates the international applicability of machine learning-based approaches for enhancing district heating efficiency and sustainability. These findings support the viability of context-aware predictive control systems for achieving significant energy optimization while maintaining occupant comfort, contributing to broader energy efficiency goals in the built environment and supporting sustainable urban development initiatives.

The geographical expansion of our model to include Finnish buildings further demonstrates the approach's adaptability to different building types and climate conditions. By incorporating building type as an additional contextual feature, the model successfully differentiated between residential, commercial, and care facilities, adjusting its predictions according to their distinct usage patterns and thermal requirements. This cross-country validation strengthens the case for broader deployment of context-aware control systems in diverse district heating networks.

The integration of building-specific contextual embeddings with the DQN framework represents a significant advancement over traditional control approaches. By learning and utilizing building characteristics through the transformer architecture, the system achieved more optimized control strategies than conventional methods. The models' ability to capture both temporal patterns and building-specific characteristics enables the DQN agent to make informed decisions that account for the unique thermal behavior of each building, resulting in more efficient and precise control.

The DQN-agent based control system's performance revealed several significant patterns in district heating energy optimization. As quantified by Equation 22, district heating energy savings varied notably across building types and seasons. Mid-rise buildings demonstrated the highest energy savings (up to 14.85%), suggesting that these structures may present optimal conditions for advanced control strategies. The seasonal analysis revealed peak efficiency during spring months, with savings exceeding 20% for one building, while summer periods showed more variable performance with occasional increased consumption. This seasonal variation indicates the need for season-specific control strategies, particularly during low-demand periods when system dynamics change considerably. The performance variance observed between high-rise and mid-rise buildings indicates that building-specific factors beyond our current contextual features may influence control effectiveness. Notably, the system maintained indoor temperatures within ±0.5°C of the 21°C setpoint while achieving these energy savings. This demonstrates the successful balance between comfort requirements and energy optimization, a critical consideration for practical deployment. The adaptive scaling mechanism for control actions proved effective in maintaining stable operations across varying conditions, with the safety bounds (20 - 70°C) successfully preventing thermal stress on the heating system.

These findings have significant implications for the future of district heating systems, particularly in countries like Sweden, where such systems form crucial infrastructure. The demonstrated energy savings of 7 - 15% across different building types indicate substantial potential for reducing overall energy consumption while maintaining resident's comfort. Similar cold-climate regions, including Norway, Canada, Alaska, and Russia, may expect comparable results. However, different climatic regions require specific considerations: Mediterranean climates may achieve different energy savings due to shorter heating seasons, while continental climates could potentially achieve better prediction accuracy. The contextual embedding approach enables adaptation to different climates, but establishing reliable reference values for tropical or subtropical zones requires region-specific training data and validation. As countries progress toward renewable energy goals, such integrated control approaches could play a vital role in optimizing district heating networks and supporting sustainable urban development.

Several limitations affect the practical deployment and scalability of our approach. Buildings with less than 6-12 months of historical data may experience reduced prediction accuracy, as the context-aware transformer environment requires sufficient training data to learn building-specific thermal characteristics. Scaling to thousands of buildings presents computational challenges due to centralized training requirements, while the contextual embedding approach may not scale efficiently across multiple countries with diverse architectural standards. Network communication latency and data synchronization could impact real-time control performance in geographically distributed systems. Potential solutions include federated learning for decentralized training, hierarchical control architectures grouping similar buildings, and transfer learning for rapid adaptation to new buildings.

7.2. Practical Implementation and Real-World Applications

The transition from research validation to practical deployment requires addressing multiple implementation considerations. The real-world deployment in Stockholm demonstrates system feasibility under operational conditions, with continuous prediction accuracy maintenance over two-month operational periods achieving root mean square error values consistently below 0.23°C. The system architecture supports integration with existing building management systems through standard communication protocols, enabling gradual deployment alongside conventional control systems. The system design facilitates scalable deployment across district heating networks. The centralized transformer training approach enables knowledge transfer across building types, reducing data requirements for new installations from typically required 2-3 years of historical data to 6-12 months for similar building types within the same geographical region. Cloud-based deployment options support remote monitoring and control capabilities, essential for large-scale district heating operations spanning multiple geographical locations.

8. Conclusions

In conclusion, this work demonstrates that the integration of an adaptive context-aware transformer as an environment for reinforcement learning provides a promising approach for optimizing district heating in residential buildings. By combining accurate prediction capabilities with optimized control strategies, the system achieves significant energy savings while maintaining occupant comfort across diverse building types and environmental conditions. These findings support the viability of machine learning-based approaches for enhancing the efficiency and sustainability of district heating systems, contributing to broader energy optimization goals in the built environment.

In our future work, we aim to extend the adaptive context-aware framework to develop control strategies that better account for seasonal variations in district heating demand and building usage patterns. This includes investigating dynamic reward functions that adjust priorities based on seasonal conditions and expanding the contextual features to capture more building characteristics. For real-world integration of the RL control system, we plan to develop transition methodologies that enable gradual implementation alongside existing control systems, allowing for performance validation without compromising operational safety.

Acknowledgements

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.

CRediT authorship contribution statement

Sai Sushanth Varma Kalidindi: Writing - original draft, data curation, methodology, software, formal analysis. **Hadi Banaee**: Writing, review and editing, visualization, supervision, validation. **Hans Karlsson**: Writing, review, supervision, validation, resources. **Amy Loutfi**: Writing, review, supervision, funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset used in this study contains proprietary data owned by EcoGuard AB. The code for the model architecture and training (excluding the raw data) will be made available upon request.

References

[1] S. Werner, District heating and cooling in sweden, Energy 126 (2017) 419–429.

[2] J. Kensby, Buildings as thermal energy storage, Master's thesis, Universidade Tecnica de Lisboa (Portugal), 2015.

[3] Z. Wang, T. Hong, M. A. Piette, Building thermal load prediction through shallow machine learning and deep learning, Applied Energy 263 (2020) 114683.

[4] X. Liu, M. Ren, Z. Yang, G. Yan, Y. Guo, L. Cheng, C. Wu, A multi-step predictive deep reinforcement learning algorithm for hvac control systems in smart buildings, Energy 259 (2022) 124857.

[5] L. LI, X. SU, X. Bi, Y. LU, X. Sun, A novel transformer-based network forecasting method for building cooling loads, Energy and Buildings 296 (2023) 113409. doi:doi: 10.1016/j.enbuild.2023.113409.

[6] A. Nagy, H. Kazmi, F. Cheaib, J. Driesen, Deep reinforcement learning for optimal control of space heating, arXiv preprint arXiv:1805.03777 (2018).

[7] S. S. V. Kalidindi, H. Banee, H. Karlsson, A. Loutfi, Adaptive context embedding for temperature prediction in residential buildings, in: ECAI 2024, IOS Press, 2024, pp. 4727–4733.

[8] S. Werner, District heating and cooling (2013).

[9] H. Lund, P. A. Østergaard, M. Chang, S. Werner, S. Svendsen, P. Sorknæs, J. E. Thorsen, F. Hvelplund, B. O. G. Mortensen, B. V. Mathiesen, et al., The status of 4th generation district heating: Research and results, Energy 164 (2018) 147–159.

[10] J. Gustafsson, J. Delsing, J. van Deventer, Improved district heating substation efficiency with a new control strategy, Applied energy 87 (2010) 1996–2004.

[11] R. Carli, G. Cavone, S. Ben Othman, M. Dotoli, Iot based architecture for model predictive control of hvac systems in smart buildings, Sensors 20 (2020) 781. doi:doi: 10.3390/s20030781.

[12] G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, A. Bemporad, Model predictive control (mpc) for enhancing building and hvac system energy efficiency: Problem formulation, applications and opportunities, Energies 11 (2018) 631.

[13] T. Ahmad, H. Chen, Y. Huang, Short-term energy prediction for district-level load management using machine learning based approaches, Energy Procedia 158 (2019) 3331–3368. doi:doi: 10.1016/j.egypro.2019.01.967.

[14] Z. Wang, T. Hong, Reinforcement learning for building controls: The opportunities and challenges, Applied Energy 269 (2020). doi:doi: 10.1016/j.apenergy.2020.115036.

[15] D. Weinberg, Q. Wang, T. O. Timoudas, C. Fischione, A review of reinforcement learning for controlling building energy systems from a computer science perspective, Sustainable cities and society 89 (2023) 104351.

[16] J. Kensby, A. Trüschel, J.-O. Dalenbäck, Potential of residential buildings as thermal energy storage in district heating systems–results from a pilot test, Applied Energy 137 (2015) 773–781.

[17] M. Mangold, M. Österbring, H. Wallbaum, Handling data uncertainties when using swedish energy performance certificate data to describe energy usage in the building stock, Energy and Buildings 102 (2015) 328–336.

[18] Z. Ni, C. Zhang, M. Karlsson, S. Gong, A study of deep learning-based multi-horizon building energy forecasting, Energy and Buildings 303 (2024) 113810.

[19] C. Tian, C. Li, G. Zhang, Y. Lv, Data driven parallel prediction of building energy consumption using generative adversarial nets, Energy and Buildings 186 (2019) 230–243.

[20] S. S. V. Kalidindi, H. Banaee, H. Karlsson, A. Loutfi, Indoor temperature prediction with context-aware models in residential buildings, Building and Environment 244 (2023) 110772.

[21] T. Wei, Y. Wang, Q. Zhu, Deep reinforcement learning for building hvac control, in: Proceedings of the 54th annual design automation conference 2017, 2017, pp. 1–6.

[22] K. Al Sayed, A. Boodi, R. S. Broujeny, K. Beddiar, Reinforcement learning for hvac control in intelligent buildings: A technical and conceptual review, Journal of Building Engineering (2024) 110085.

[23] Z. Nagy, F. Y. Yong, M. Frei, A. Schlueter, Occupant centered lighting control for comfort and energy efficient building operation, Energy and Buildings 94 (2015) 100–108.

[24] J. Deng, M. Eklund, S. Sierla, J. Savolainen, H. Niemistö, T. Karhela, V. Vyatkin, Deep reinforcement learning for fuel cost optimization in district heating, Sustainable Cities and Society 99 (2023) 104955.

[25] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, J. G. Slootweg, On-line building energy optimization using deep reinforcement learning, IEEE transactions on smart grid 10 (2018) 3698–3708.

[26] K. Dalamagkidis, D. Kolokotsa, K. Kalaitzakis, G. S. Stavrakakis, Reinforcement learning for energy conservation and comfort in buildings, Building and environment 42 (2007) 2686–2698.

[27] L. Langer, T. Volling, A reinforcement learning approach to home energy management for modulating heat pumps and photovoltaic systems, Applied Energy 327 (2022) 120020.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[29] K. Mason, S. Grijalva, A review of reinforcement learning for autonomous building energy management, Computers & Electrical Engineering 78 (2019) 300–312.

[30] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, nature 518 (2015) 529–533.