BURSA TEKNİK ÜNİVERSİTESİ

SMART HVAC SYSTEM

DEEP REINFORCEMENT LEARNING FINAL PROJECT

23435004013-MELIKA JIBRIL SEID

30/06/2025 | BURSA TEKNIK UNIVERSITESI

TABLE OF CONTENTS

INTRODUCTION

Heating, Ventilation, and Air Conditioning (HVAC) systems are critical components in building infrastructure, responsible for maintaining indoor environmental quality by regulating temperature, humidity, and air quality. These systems are essential for ensuring occupant comfort and energy efficiency, which are increasingly important in the context of global energy consumption and environmental impact. HVAC systems are designed to ensure thermal comfort for building occupants by controlling indoor climate conditions [1] They account for a significant portion of global energy consumption and carbon emissions, making their optimization crucial for energy efficiency and environmental sustainability[2]. HVAC systems typically include components such as air conditioners, heaters, ventilation ducts, and thermostats, which work together to maintain desired indoor conditions. The systems are often controlled using various algorithms and models to optimize performance and reduce energy consumption[3].

Reinforcement Learning (RL), including Q-learning and Deep Q-Networks (DQN) and their variants, has been applied to optimize HVAC control by framing it as a Markov Decision Process (MDP). These methods aim to improve system performance by learning optimal control policies that balance energy efficiency with occupant comfort. Managing the large state space and ensuring tractability in RL applications is a significant challenge, necessitating the development of efficient algorithms and models [3], [4]. Techniques such as variable-length rolling horizon optimization and stochastic optimization are used to handle uncertainties in building energy management, enhancing the flexibility and efficiency of HVAC systems . The integration of energy storage systems with HVAC control can further reduce operating costs and improve energy management during peak electricity price periods[1]

DEEP Q-NETWORK (DQN)

Deep Q-Network (DQN) is an advanced reinforcement learning algorithm that extends Q-learning, a fundamental reinforcement learning algorithm. It was developed by DeepMind in 2015 and has since been applied in various control problems, including HVAC systems. Unlike traditional Q-learning, which uses tables to store Q-values, DQN employs deep neural networks (DNNs) to approximate the Q-value function, enabling it to handle high-dimensional and continuous state spaces where tabular Q-learning would be impractical [3], [5], [6], [7]. To enhance the stability of the learning process, DQN incorporates two key techniques:

Separate Target Q-network: It uses a separate target Q-network alongside the primary Q-network. The parameters of the Q-network are periodically copied to the target network to stabilize the learning process[5], [7]

Experience Replay Buffer: DQN utilizes an experience replay buffer to store past experiences. During training, samples are randomly drawn from this buffer, which helps mitigate issues related to correlated data and improves convergence[5], [7]

DUELING DEEP Q-NETWORK (D3QN)

D3QN, or Double Dueling Deep Q-Network, is an enhanced version of the Deep Q-Network (DQN) algorithm, primarily used in reinforcement learning. It improves upon DQN by addressing issues of overestimation during Q-value learning and enhancing stability and convergence speed. D3QN combines the principles of Double DQN and Dueling DQN, which include using two identical Q-networks and introducing a value function and an advantage function to accelerate convergence speed. This combination enhances the algorithm's performance, stability, and reliability, particularly in complex environments. D3QN demonstrates superior optimization performance, stability, and reliability compared to DQN, especially in HVAC system control projects. It shows a more robust efficacy in improving the system's coefficient of performance (COP) over time, indicating better control capability and adaptability. While DQN is effective, D3QN's additional mechanisms provide a significant advantage in handling diverse and complex control tasks[2]. In energy-saving applications, D3QN can achieve better results from the early stages of deployment compared to DQN, which may require more exploration time before converging on an optimal policy[7]

PREVIOUS WORKS

The complex problem of balancing tenants thermal comfort while minimizing energy consumption has been a hot topic of research in recent years.Tsenis et al.s SmartKlima used DQN and double DQN to optimize the heating conditions of a residential building and their model was able to achieve strong convergence under certain conditions [8]. Ghane et al presented a model-free, offline reinforcement learning approach for optimizing thermostat control in HVAC systems. Their Dueling double DQN (D3QN) network , trained using historical data from HomeLab 1, was able to achieve an 18.66% reduction in energy use compared to traditional rule based methods [5]. Liu et al developed a novel reinforcement learning algorithm that combines DQN and XGBoost for HVAC and window control optimization. Their model improved indoor thermal comfort duration by 24% and reduced air conditioning runtime by 24.7% compared to baseline models [6]. McKee et al. employed a model-free Deep Reinforcement Learning to minimize energy costs in HVAC operations. They utilized DQN for reinforcement learning control objectives and explored Deep Deterministic Policy Gradient (DDPG) for continuous action spaces. Their DRL controller achieved a 43.89% cost reduction compared to traditional methods in a simulated environment [9]. Han et al. used a deep forest based deep Q-network (DF-DQN) to optimize cooling water systems. Their DF-DQN model avoided unnessary explorations and converged faster and has a better energy-saving effect in the early stages compared to DQN [7]. Barrett et al. applied tabular Q-Learning and Bayesian inference to optimize a thermostat and were able to achieve a 10% cost reduction compared to the programmable control method [4]. Yu et al. proposed a knowledge-based reinforcement learning control approach for cooling towers in HVAC systems. Their DQN model utilizes a knowledge based exploration strategy, reducing training time and improving control performance. Their model demonstrates lower Integral Absolute Error (IAE) and Integral Square Error (ISE) than the Proportional Integral (PI) controller [10]. Wang et al. proposed a DQN-based building energy management (BEM) method to reduce operating costs while ensuring occupant comfort. Their method integrates a third-order thermal model to accurately simulate dynamic temperature changes in buildings. Simulation results confirm the method's effectiveness in enhancing energy flexibility and reducing costs during peak electricity price periods [1]. In another research, Wang et al. explored classical and deep reinforcement learning methods for HVAC control. They highlighted the importance of state space choices in high-dimensional observation environments and identified DQN as a suitable method for HVAC control tasks due to its balanced performance [3]. Finally, Qin et al. did a comparative study of DQN and dueling double DQN (D3QN) for HVAC system optimization. Their study highlights D3QN's superior optimization effectiveness and stability across various HVAC projects compared to DQN. Their proposed D3QN structure includes two hidden layers with 64 and 12 neurons, respectively[2].

CURRENT WORK

In this study, three deep reinforcement learning models were evaluated for autonomous HVAC control: a vanilla DQN model, a custom Dueling Q-Network (DuelingQNet) and a Dueling Double DQN model inspired by prior research (D3QNP). Both models were trained for 10,000 episodes using the same environment and reward shaping. All the source code, as well as trained models can be found in the github link below:
Github Link: https://github.com/m3likaj/HVAC-Q-Learning

ENVIRONMENT

The main source and inspiration for the environment design was the paper by E. Barrett and S. Linder, titled "Autonomous HVAC Control, A Reinforcement Learning Approach" [4]. Based on the fixed values and formulas in this paper, an office environment was designed. While the office is located in Istanbul, it was modeled using weather data from the year 1991. The dimensions of the office are 3×3 meters. According to one-year weather data for Istanbul obtained from the Energy Plus [11] website, the temperature ranges between -8ºC and 31ºC. The working hours of the office are defined as every week day between 8:00 and 18:00. Holidays were calculated based on pythons holiday library. The presence or absence of people in the office is detected by a sensor that checks and returns the status every minute. The air conditioner in the room is capable of both heating and cooling. The environment was coded to comply with the Gym format, and a separate class was defined to perform Bayesian inference.

STATE AND ACTİONS

The air conditioner, which acts as the agent, receives 3 types of data from the environment:

- the office temperature (room temperature / rt),
- the outside temperature (outside temperature / ot),
- and the estimated time until people arrive (time to occupancy / tto).

After receiving this data, it can perform 3 actions:

- Turn on heating (Heat On),
- Turn on cooling (Heat Off),
- Turn off the HVAC (Therm Off).

During the learning process, the air conditioner performs one of these three actions every minute.

CONSTANTS

These are the constant values used to calculate heat change and model the environment.

Name	Value
Surface Area	○ $Surface\ Area = 54\ m^2$

Volume	○ $Volume = 27\ m^3$
Heater Output	○ $Heater\ output = 500\ W$
Heat Capacity of Air	○ $Heat\ Capacity\ of\ Air = 718\ J/kgK$
Air Density	○ $Air\ Density = 1.3\ \ Kg/m^3$
U-values for Ceiling, Floor and Wall	○ $U_{Ceiling} = 0.4$ ○ $U_{Floor} = 0.5$ ○ $U_{Wall} = 0.6$

FORMULAS

To design the environment, the following formulas were taken from E. Barrett and S. Linder's paper and converted into functions. Additionally, a Bayesian inference class was defined to predict working hours. The relevant functions, the formulas they use, and their purposes are as follows:

Name of Function/Class	Function	Formula
calculate_heat_transfer	Calculates the heat lost or gained through the walls, ceiling and floor	○ $Heat\ Transfer = U_{Value} \times Surface\ Area \times (rt - ot)$ ▪ rt = room temprature ▪ ot= outdoor temprature ○ $Heat\ Transfer_{Total=}\ Heat\ Transfer_{Ceiling} + Heat\ Transfer_{Floor} + Heat\ Transfer_{Wall}$
calculate_change_in_temp	Calculates the change in the room's temprature	○ $Temprature\ Increase = \frac{(Heater\ Output - Heat\ Transfer_{Total})\times Time(\text{sec})}{Heat\ Capacity\ of\ Air} \div (Air\ Density \times Volume)$
get_outside_temp	Returns the outside temprature at	○ $Outdoor\ Temp\ Change\ per\ min = \frac{Temp\ at\ T - Temp\ at\ T+1}{60}$

	the given minute	
class BayesianOccupancyPredictor	Returns the propability of the room being occupied	○ $P'(s = s \vert a, s = s) = \frac{P(S = S \vert a, S = S)\times Expc+1}{Expc'}$ ▪ $Expc = experience\ counter$ (tecrübe sayacı) ○ $P(X \vert Y) = \frac{P(Y \vert X)\times P(X)}{P(Y)}$

MODEL ARCHITECTURES

Three different DQN architechtures were used in this research, a normal (vanilla) DQN, and two D3QN models (A custom D3QN and D3QNP). D3QNP is a simpler version of D3QN that matches the architecture in [2]. The details of each model are given in the table below

Feature	DuelingQNet	DuelingQNetP	NormalizedQNet
Architecture Type	Dueling DQN	Lightweight Dueling DQN	Vanilla DQN
Input Normalization	No	No	Yes
Feature Layers	Linear(3→128) + ReLU	Linear(3→64) → Linear(64→12) + ReLU	Linear(3→64) + ReLU
Value Stream	Linear(128→128) → ReLU → Linear→1	Linear(12→1)	None (not dueling)
Advantage Stream	Linear(128→128) → ReLU → Linear→3	Linear(12→3)	None
Output	Combined Q(s,a) via dueling formula	Same	Linear(64→64) → Linear(64→3)
Parameter Count	High	Moderate	Low/Moderate
Model Depth	Deepest	Shallow	Medium
Target Usage	Soft update every 50 episodes	Same	Same

TRAINING LOOP

Training typically took around 8-10 hours. Different techniques were utilize to optimize the training including shuffling the days at the beginning of each year to reduce overfitting and decreasing the batch size to speed up training, and clipping the rewards to stabilize learning. The training parameters are listed in the table below.

Component	DuelingQNet	DuelingQNetP	Normal QNet
Learning Rate	1e-5	1e-4	1e-5
Optimizer	Adam	Adam	Adam
Target Update Freq	Every 50 episodes	Same	Same
Batch Size	256	Same	Same
Experience Replay Size	100,000	Same	Same
Warmup Steps	7200	5000	Same
Episode Count	10,000	same	Same
Reward Scaling	reward * 0.01 then clipped to [-1, 1]	Same	Same
ε-start	1.0	same	same
ε-end	0.01	0.2	0.01
ε-decay	0.9999	0.9995	0.9999

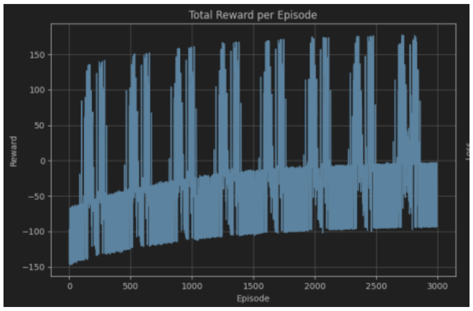
REWARD FUNCTİON

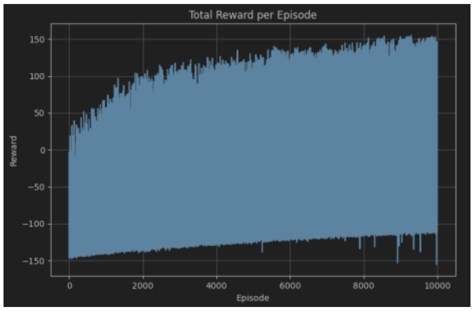
- Turning on heating when the room is already warm or turning on cooling when the room is already cold → **-5**
- Turning on the HVAC when there is more than one hour until someone arrives → **-10**
- When a person is present and the room temperature is outside the tolerance range → **-15**
- When a person is present and the room temperature is within the optimal range (target ± tolerance) → **+30**

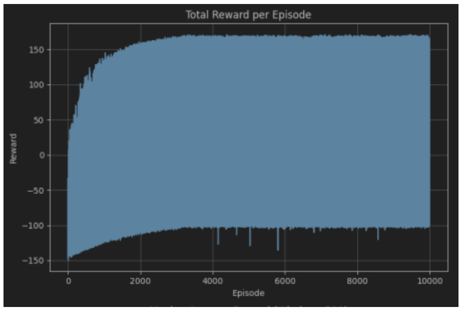
RESULTS

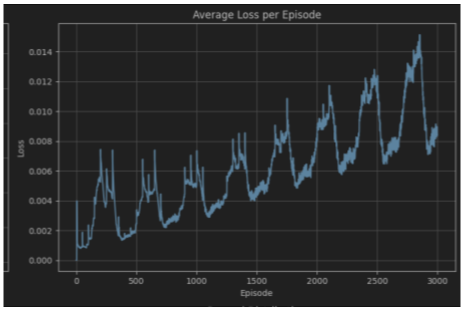
The results of the 3 models are listed in the table below. The results confirm the superiority of D3QN over DQN when it comes to optimizing HVAC projects and affirm the architechture proposed by Qin et al. in [2] as it was the best-performing-model. The results show that the D3QNP model outperforms DuelingQNet in terms of learning speed, stability, and overall performance. Specifically, D3QNP achieved higher and more consistent total rewards per episode, with a sharper and more concentrated reward distribution around optimal values. In contrast, DuelingQNet exhibited slower convergence, broader reward distribution, and greater variance in early training episodes. This suggests that the architectural and training optimizations adopted from the reference paper—such as smaller, more efficient networks and better value-advantage separation—enhanced learning efficiency. Overall, the findings support adopting the D3QNP structure for more effective HVAC policy learning in future work.

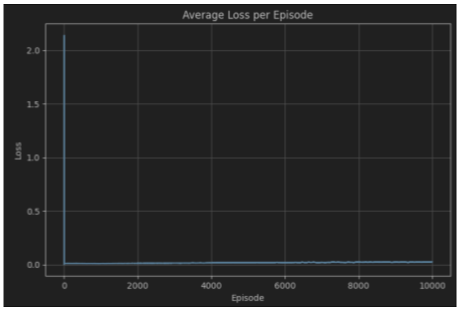
Vannila DQN **D3QN** **D3QNP**

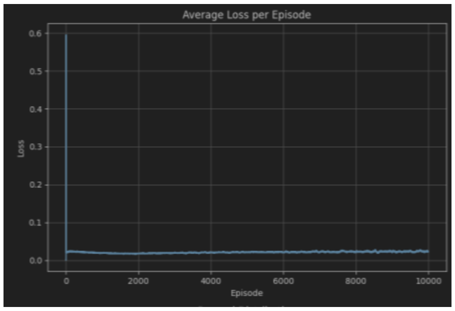


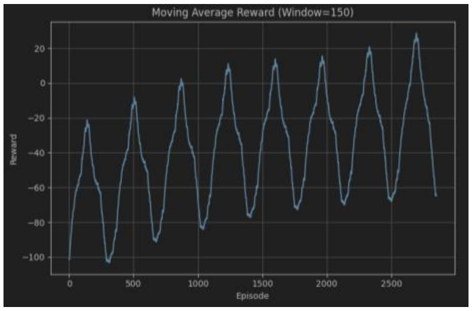


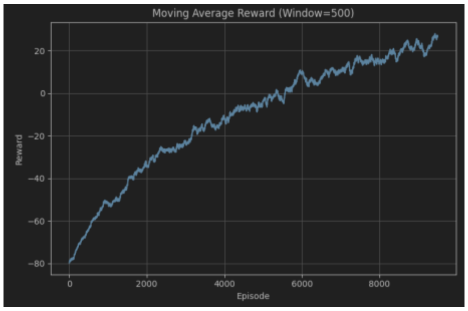


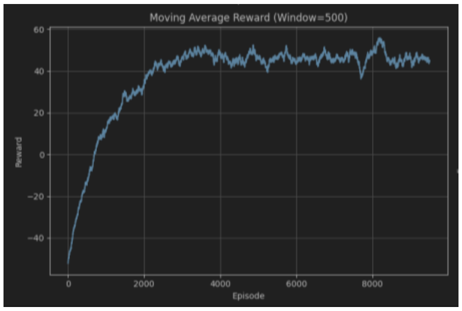


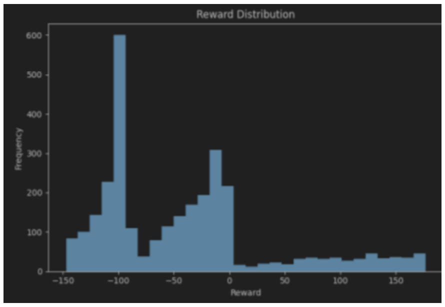


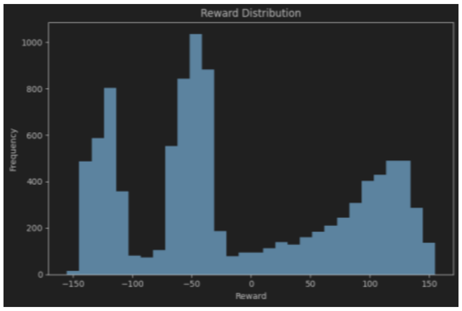


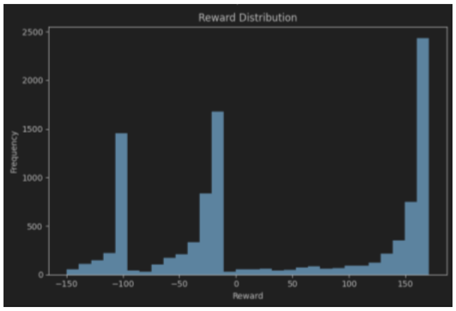












CONCLUSION

The comparison between DQN and D3QN highlights the advancements in reinforcement learning algorithms that enhance performance and stability. D3QN's integration of Double DQN and Dueling DQN principles provides it with a significant edge in complex environments, making it a more suitable choice for applications requiring high stability and efficiency. Future research could explore further enhancements to these algorithms, potentially incorporating new techniques to improve learning speed and adaptability in even more challenging scenarios.This exploration emphasizes the importance of selecting appropriate Q-network structures to maximize the performance of reinforcement learning algorithms in HVAC applications, ultimately enhancing energy efficiency and system reliability.

REFERENCES

[1] S. Wang, X. Chen, L. Bu, B. Wang, K. Yu, and D. He, “A DQN-Based Coordination Method of HVAC and Energy Storage for Building Energy Management,” in *2023 IEEE 7th Conference on Energy Internet and Energy System Integration, EI2 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 4891–4896. doi: 10.1109/EI259745.2023.10512933.

[2] H. Qin, T. Meng, K. Chen, and Z. Li, “A comparative study of DQN and D3QN for HVAC system optimization control,” *Energy*, vol. 307, Oct. 2024, doi: 10.1016/j.energy.2024.132740.

[3] M. Wang, J. Willes, T. Jiralerspong, and M. Moezzi, “A Comparison of Classical and Deep Reinforcement Learning Methods for HVAC Control,” Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.05711

[4] E. Barrett and S. Linder, “Autonomous HVAC Control, A Reinforcement Learning Approach,” vol. 9286, A. Bifet, M. May, B. Zadrozny, R. Gavalda, D. Pedreschi, F. Bonchi, J. Cardoso, and M. Spiliopoulou, Eds., in Lecture Notes in Computer Science, vol. 9286. , Cham: Springer International Publishing, 2015, pp. 3–19. doi: 10.1007/978-3-319-23461-8_1.

[5] S. Ghane *et al.*, “Real-World Implementation of Offline Model-Free Reinforcement Learning for Thermostat Control,” in *2025 IEEE International Conference on Mechatronics, ICM 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ICM62621.2025.10934768.

[6] X. Liu and Z. Gou, “Occupant-centric HVAC and window control: A reinforcement learning model for enhancing indoor thermal comfort and energy efficiency,” *Build Environ*, vol. 250, Feb. 2024, doi: 10.1016/j.buildenv.2024.111197.

[7] Z. Han *et al.*, “Deep Forest-Based DQN for Cooling Water System Energy Saving Control in HVAC,” *Buildings*, vol. 12, no. 11, Nov. 2022, doi: 10.3390/buildings12111787.

[8] T. T. Tsenis, G. Kapsimanis, and V. Kappatos, “SMARTCLIMA: Reinforcement learning residential thermostat-less heating control system,” in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2021*, Institute of Electrical and Electronics Engineers Inc., Oct. 2021. doi: 10.1109/ICECCME52200.2021.9591000.

[9] E. McKee *et al.*, “Deep reinforcement learning for residential hvac control with consideration of human occupancy,” in *IEEE Power and Energy Society General Meeting*, IEEE Computer Society, Aug. 2020. doi: 10.1109/PESGM41954.2020.9281893.

[10] Z. Yu, X. Yang, F. Gao, J. Huang, R. Tu, and J. Cui, “A Knowledge-based reinforcement learning control approach using deep Q network for cooling tower in HVAC systems,” in *Proceedings - 2020 Chinese Automation Congress, CAC 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 1721–1726. doi: 10.1109/CAC51589.2020.9327385.

[11] U.S. Department of Energy, “Energy Plus.” Accessed: Jun. 03, 2025. [Online]. Available: https://energyplus.net/weather-location/europe_wmo_region_6/TUR/TUR_Istanbul.170600_IWEC