

Original Research

Received 03 May 2026

Revised 04 June 2026

Accepted 16 June 2026

Natural Resources for Human Health 2026, Vol. 6 (5): 1334–1350

KEYWORDS:

Reinforcement Learning (RL), HVAC systems, energy efficiency, bio-safety compliance, predictive maintenance.

Optimization of Energy-Efficient HVAC Systems in Bio-Safe Environments using Reinforcement Learning for Predictive Maintenance and Real-Time Control

Karuna Kanth Bathula¹, T. Kanthimathi^{1*}

¹Department of Mechanical Engineering, Koneru Lakshmaiah Education Foundation, Veddeswaram, A.P, India

*Corresponding Author's Email ID: fessiorp@gmail.com

ABSTRACT: Bio-safety compliance and energy efficiency are still a pressing issue in the management of modern buildings, especially in such a setting as a healthcare facility or a laboratory, where the accuracy of controlling the environmental conditions is the most important factor. Reinforcement Learning (RL) is potentially a solution to the optimization of HVAC performance by dynamically modifying operational parameters (e.g. temperature, humidity, airflow, and energy consumption) to achieve a more efficient HVAC system. The paper introduces a RL-based solution to optimize HVAC systems in bio-safe settings, which uses real-time information provided by Internet of Things-based sensors to constantly monitor and modify system settings. Trial and error allow the RL agent to learn how to make the best decisions, enhance the energy efficiency, but still meet high bio-safety standards. The proposed system simulation was carried out in MATLAB, and the RL-based HVAC system demonstrated an average consumption of 17% of energy, along with ensuring bio-safety requirements in sensitive areas. It enabled the system to anticipate and thwart HVAC breakdowns, saving 25% of downtime and improving the overall reliability of the system. This dynamic optimization model shows that RL has the potential to develop sustainable and cost-efficient solutions to HVAC systems, which can have long-term benefits in terms of energy management and occupant safety.

1. INTRODUCTION

The quality of indoor air and the control of the environmental conditions are of the utmost importance in contemporary bio-safe settings like healthcare facilities, laboratories, and clean rooms [1]. Heating, Ventilation, and Air Conditioning (HVAC) systems play a significant part in these environments, controlling the airflow, temperature, and humidity levels to ensure that dangerous pathogens are contained and contamination is avoided. In high-security buildings such as Biosafety Level (BSL) 3 and 4 laboratories, HVAC systems are tasked with the responsibility of ensuring high bio-safety standards such as pressure differentials control and high efficiency air filtration [2]. Such systems must also contend with the issue of avoiding airborne contamination, and ensuring the comfort and safety of the personnel working in these areas [3]. Nevertheless, as the bio-safety demands grow more and more complicated, HVAC systems in these settings have the two-fold task of ensuring high performance in air quality and air containment, as well as, decreasing energy use. HVAC systems may be an important energy consumer that can use more than 40 percent of a building total energy [4]. Such a high demand on energy may contribute to the increase in the cost of operation and environmental issues, particularly in healthcare facilities and laboratories that need to operate continuously. Therefore, the necessity of energy-efficient HVAC systems without the negative impact on bio-safety standards is more pronounced than ever. The current HVAC systems, which are in place in these facilities, are usually the

traditional types that run at constant predetermined settings resulting in unwarranted energy wastage [5]. As the attention to sustainability and minimizing the operational expenses increases, it is crucial to consider new ways of optimizing HVAC performance [6]. The introduction of artificial intelligence (AI) and, more particularly, Reinforcement Learning (RL), provides an opportunity to overcome these issues. Machine learning, and more specifically RL, offers systems the capability to learn in real-time and make dynamic decisions that can optimize performance [7]. When applied to HVAC systems, RL may be employed to modify airflow rates, temperature and filtration rates in response to varying environmental conditions and bio-safety needs. This would allow the system to work efficiently, reducing the amount of energy used and at the same time ensure that high bio-safety standards are achieved [8]. Moreover, predictive maintenance algorithms may be combined with RL to track HVAC items and forecast possible failures before they happen in order to minimize downtime and maintenance expenses. One of the most significant issues in designing and operating HVAC systems in healthcare and laboratories is the problem of balancing the high bio-safety requirements with the energy efficiency [9]. Conventional HVAC systems are set to control the environment to a certain degree, but they are not able to respond to the changing requirements in real-time. These systems are usually operated to full capacity, even when there is no demand, which results in unnecessary use of energy [10]. This is complicated by the fact that bio-safety standards are not simple, as negative pressure zones, sufficient air changes per hour, and high-efficiency filters contribute to the energy load of the system [11]. This has led to increasing demands of dynamic real-time control responding to operational requirements without affecting bio-safety or energy efficiency. In addition, HVAC systems are prone to failure (i.e., clogged filters or broken fans), which may cause considerable downtimes and high maintenance expenses. The problem is that traditional systems lack the ability to foresee such failures and cause inefficiencies and disruptions in critical environments [12]. Reinforcement Learning can be integrated into HVAC systems to provide a solution that can dynamically optimize system performance towards achieving bio-safety and energy efficiency objectives. Training the RL agent to make real-time choices, depending on the environmental information and bio-safety criteria, HVAC systems would be able to work in a highly efficient way and to provide the best in terms of containment and air quality [13][14][15]. Furthermore, the predictive maintenance capabilities of RL to anticipate system failures and perform predictive maintenance will contribute to lowering the expensive downtime, enhancing system reliability, and increasing the lifespan of HVAC components. This research will show how RL can be used to maximize the performance of HVAC, which can enhance sustainability and safety in bio-safe settings [16].

Problem Statement:

The main problem of HVAC systems in bio-safe surroundings is the lack of dynamism in changing conditions and at the same time maintain high standards of bio-safety. Conventional HVAC systems are designed to work under fixed settings and hence, inefficient in using energy. With the growing need to find energy-efficient remedies, the use of these static systems is untenable especially in environments where air always needs to circulate, be filtered, and kept at certain temperatures to ensure high standards of bio-safety. These systems are not flexible enough to respond to the changes in occupancy in real-time, external conditions, or system performance, leading to unneeded energy use and operating expenses. Furthermore, the unpredictable nature of the conventional HVAC systems in terms of maintenance is another major issue. Lack of planning to foresee and take care of the problems of the HVAC components before they get out of control may result in unplanned downtime, extra maintenance, and even bio-safety violations. To achieve bio-safety in the settings like hospital, research and clean rooms, it is not enough to implement simple air filtration and pressure control, but to constantly monitor and adjust the system parameters. Machine learning-driven predictive maintenance is one method to preemptively deal with system failures, but how to combine this with energy-efficient optimization is still a difficult topic to be solved.

Contributions of the Paper:

1. Implementation of Reinforcement Learning (RL) in HVAC Systems: This article describes a new method of HVAC optimization of bio-safe environments with the help of Reinforcement Learning. It is possible to control HVAC systems dynamically with the integration of RL that will guarantee energy efficiency without affecting the bio-safety standards.
2. Predictive Maintenance Framework: We propose a predictive maintenance framework based on machine learning in the RL framework. This model can predict the failure of the HVAC system components (e.g., filter clogging, motor inefficiency), and initiate timely maintenance procedures, thereby minimizing the downtime and maintenance expenses.

3. Optimization of Energy efficiency: The paper shows how RL may be applied to optimize the energy consumption of HVAC by controlling operational parameters (e.g., airflow, temperature) in real-time based on environmental and occupancy data, resulting in considerable energy savings in bio-safe settings.
4. Case Study and Implementation: An example of a pilot implementation of the RL-driven HVAC system is provided, including the description of the way the system was implemented in a hospital and laboratory. The case study points to the bio-safety compliance and energy efficiency gains, and demonstrates the real-world potential of RL in HVAC optimization.

2. LITERATURE REVIEW

The adoption of Reinforcement Learning (RL) in HVAC systems has been widely discussed because it can be used to optimize the use of energy without violating the bio-safety and thermal comfort levels in a regulated space. One of the most significant breakthroughs in this area is the article by Anuntasethakul et al. (2021), who developed a supervisory model predictive control (MPC) system to control building HVAC systems, with peak-load shaving and thermal comfort. Their study puts emphasis on predictive control strategies to strike a balance between the heating and cooling requirements and the efficiency of energy usage. Their method, however, does not explore the application of RL algorithms to real-time optimization, which has been a recent development in the field. This leaves a gap that can be filled with the help of RL-based methods, which continuously learn and optimize the system performance. Albatayneh et al. (2021) discuss the importance of adaptive thermal comfort in HVAC systems and highlight the importance of thermal comfort in building retrofits. They demonstrate in their work that adaptive comfort models can be incorporated into HVAC controls to enhance the energy efficiency, though they are not effective enough to use machine learning or RL algorithms to predict and optimize system reactions in real time. This weakness can be addressed with the help of RL to dynamically adapt to occupant behavior and external weather conditions and enhance comfort and efficiency. Azuatalam et al. (2020) were one of the first to use Reinforcement Learning (RL) to control the entire building HVAC. They showed the possibility of RL to control whole-building HVAC and demand response, and how RL could be used to optimize energy usage, given real-time building data. This study is important as it directly applies RL to control HVAC systems in large-scale settings, which is crucial for implementing energy-efficient solutions in multi-zone buildings like hospitals and laboratories. Their results emphasize the capability of RL to dynamically regulate the energy consumption, which is a critical attribute of the HVAC systems of the future in bio-safe conditions. Blum et al. (2021) were interested in the development of a simulation-based testing framework, BOPTEST, that compares various control strategies in buildings. Their model helps to assess the effectiveness of energy-efficient HVAC control measures in a controlled environment but fails to consider the possibility of RL to enhance the performance of the system by learning on the fly. This piece can be used in its effort to benchmark a system and provide a platform on which RL-based systems can be tested and validated. Based on these developments, Chen et al. (2020) introduced a scalable RL solution named Gnu-RL that incorporates a differentiable MPC policy to optimize the performance of the HVAC system. This solution showed that RL could be used to effectively and scalably manage HVAC systems. The main strength of Gnu-RL is that it can be practically applied to large-scale buildings, offering a realistic approach to the complex environment, like a hospital or a laboratory. Their model provides a direct comparison with the traditional control strategies and shows the benefits of RL in trade-offs between energy efficiency and system complexity. Continuing the discussion of the fusion of RL and Active Disturbance Rejection Control (ADRC), Cui et al. (2025) suggested an approach to optimal HVAC. They demonstrate that the combination of ADRC and RL can improve the resilience of HVAC systems to disturbances, and it can be applied in the areas where bio-safety and comfort are paramount. This work goes to the edge by integrating active control methods that have the ability to adjust to the external environment and sustain system stability and efficiency even in difficult environments. Ding et al. (2025) were interested in enhancing safe HVAC control in dynamic settings by introducing a data-efficient model-based RL system. Their model uses a model to anticipate the future HVAC performance and make adjustments optimizing the use of energy and bio-safety standards. This work is especially applicable to bio-safe environments because predicting and controlling the failures of the HVAC system in real-time is an important element of both energy efficiency and safety.

Esmacili et al. (2025) also developed the idea of safe RL by incorporating safe RL methods into the HVAC control to ensure the use of minimum energy and maximum occupant comfort. They focus their work on safety and reliability of RL in critical settings where it is crucial to make sure that RL agents do not undermine bio-safety practices and enhance energy efficiency.

This method is specifically applicable to the hospitals and laboratories where the safety standards have to prevail at all times. Within the framework of the interaction of solar energy systems with HVAC controls, Hanif et al. (2025) suggested a system based on RL to control not only the HVAC systems but the solar energy as well, maximizing the overall energy management system. The combination of solar energy and HVAC control with the help of the RL opens a bright prospect of creating sustainable bio-safe environments because the integration of renewable energy is gaining more significance in energy-efficient buildings. Kadamala et al. (2024) improved the control of HVAC using transfer learning coupled with RL agents to allow the systems to adjust to different building conditions and climates. The latter is particularly useful in bio-safe settings, where the HVAC systems must respond to the occupancy and external changes rapidly. Transfer learning will guarantee that HVAC systems are able to generalize between settings, so the system becomes stronger and more efficient. Liu and Gou (2024) concentrated on occupant-centric HVAC and window control, and optimized thermal comfort and energy efficiency with the help of RL models. Their study shows how RL can be used to optimize comfort in the indoor environment and minimize energy use through dynamically controlling the system in response to real-time occupant and external data. This method is very applicable in bio-safe settings, where individual comfort should be adjusted to bio-safety and energy saving. In residential settings, Lork et al. (2020) suggested uncertainty-conscious RL to control residential air conditioning systems. This framework provides greater flexibility to HVAC control by taking into account the uncertainties involved in constructing energy use, which is considerable in multi-zone buildings that demand high bio-safety measures. These studies indicate that Reinforcement Learning has a growing potential of enhancing the efficiency of the HVAC system, energy consumption, and making sure that bio-safety standards are upheld in different building settings. Most of these studies though specialize in certain areas of HVAC control and do not entirely put into consideration the issue of bio-safety in highly sensitive settings like a hospital or high-security laboratory. This gap creates a room to conduct further research, as safe RL in HVAC control in bio-safe environment is a promising field to be innovative.

3. METHODOLOGY

3.1 Architecture of the RL-driven HVAC system

The Reinforcement Learning (RL)-based HVAC system architecture is structured to incorporate real-time monitoring, decision-making, and actuation elements that can be used to dynamically control HVAC systems to meet the energy efficiency and bio-safety requirements in controlled settings. The system uses machine learning algorithms to improve system parameters, including temperature, humidity, airflow, and pressure differentials, by using ongoing feedback on the environment. The general architecture entails multiple layers of components, each playing a different role in facilitating a smooth operation and optimization (fig 1).

1. Data Collection Layer:

The data collection layer is the first layer of the RL-driven HVAC system, and it uses a network of IoT-enabled sensors and smart devices located strategically in the building. These sensors are able to measure and report a broad spectrum of environmental variables that are important to bio-safety and energy efficiency, including:

- Temperature and humidity of different zones.
- Airflow rates and pressure differentials in order to sustain containment in bio-safe environments.
- HVAC energy use of fans, compressors, and air handling units.
- Air quality parameters such as the levels of particulate matter and contaminants.
- Occupancy information to optimize HVAC operating according to the actual use of the building.

The data is then aggregated and analyzed in a central processing system (usually via a Building Management System (BMS) or cloud infrastructure) by these sensors. This layer must be accurate and capable of real-time because the system depends on the timely data to make the best decisions.

2. Data Preprocessing and Feature Engineering:

After gathering data, it is first preprocessed and features extracted. This is an important step to ensure that the RL model is fed with clean and meaningful data to make decisions. It involves:

- Noise filtering: Eliminating any sensor errors or outliers to assure the quality of the data.

- Normalization: Making sensor data, including temperature and pressure, standardized to train models more easily and converge more quickly.
- Feature engineering: Deriving new features based on the raw data, e.g. trends over time (e.g., average temperature, humidity changes) or the aggregate readings across multiple zones to be able to capture patterns across the building.

This preprocessed data is the input to the RL model and makes sure it has the appropriate information required to optimize the performance of HVAC system.

3. Reinforcement Learning Agent:

The main component of the system is the RL agent which learns optimal control policies to accomplish certain goals. The agent functions based on the loop of interaction with the environment and makes decisions on the parameters of the HVAC system based on the real-time data. The important parts of the RL agent involve:

- State Space: The state is the present state of the HVAC system and the environmental conditions of the building. Some of the variables in the state include current temperature, humidity, airflow, pressure differentials, and system energy usage.
- Action Space: The RL agent has the opportunity to select among a number of actions, including fan speeds, VAV (Variable Air Volume) settings, temperature setpoints, airflow direction, or predictive maintenance actions (e.g., cleaning filters). The agent can also change parameters according to occupancy, external weather conditions, or bio-safety needs (e.g. changing air pressure in clean rooms or laboratories).
- Reward Function: The reward function indicates the objective that the RL agent is attempting to optimize. The reward to the agent in this case can be a combination of things and they may include:
- Energy Efficiency: More of a reward on actions that reduce energy use without compromising environmental conditions.
- Bio-Safety Compliance: Greater reward on actions that assure the HVAC system to meet bio-safety measures, such as the negative pressure or sufficient air changes per hour in critical areas.
- Comfort Level: Making sure that the indoor temperature and humidity are in a comfortable range of occupants.
- Predictive Maintenance: Reducing system failures and downtime, incentivizing the agent to predict and avoid system failures of the HVAC system.

Learning Algorithm: Q-learning, Deep Q Networks (DQN), or Proximal Policy Optimization (PPO) are all algorithms that the RL agent uses to learn optimal control policies. These algorithms enable the agent to assess the long term implications of its actions, and thus make informed decisions based on accumulated experiences (trial and error).

4. Decision-Making Layer:

After the RL agent has been trained on the optimal policy, it proceeds to the decision-making layer where it uses its learned model to make decisions in the real world. This layer takes a decision based on the real-time information and the policy learned by the agent on how to modify the HVAC system to achieve the goals of bio-safety and energy efficiency. The decisions include:

- Setting a setpoint at the temperature to achieve the comfort and bio-safety needs.
- Adjusting air flow rates to achieve containment or ventilation in certain places.
- The control of energy usage by regulating the elements of the system, including chillers, heating units and fans.
- Initiating maintenance activities, like cleaning or filter replacement, or malfunction of HVAC components.

This decision layer is a real-time system that constantly considers the environmental factors and modifies the system depending on the actions of the RL agent.

5. Actuation and Control Layer:

Actuation layer is the physical HVAC parts that implement the values of decisions made by the RL agent. This includes:

- Adjustable fans and pumps that control air flow and pressure.
- Chillers, heating systems and cooling coils that control the temperature indoors.

- Air handling units (AHUs) to regulate the airflow and moisture content.
- Air filters and ventilation dampers to control air quality and airflow.

The control layer implements the action of the RL agent, and makes sure that the HVAC components react to the optimization instructions of the system in real time. The effectiveness of the system depends on the smooth flow of communication between the decision-making layer and physical components so that adjustments could be made in time.

6. Feedback Loop and Continuous Learning:

The last element of the architecture is the feedback loop whereby the RL agent is provided with feedback by the environment in relation to the actions of the agent. To illustrate, the system will assess the effectiveness of the changes in the desired temperature and pressure and reduce energy consumption after changing the airflow. When the agent model is reinforced and updated, it is because the actions of the system yield positive results. In case the outcomes are suboptimal, the agent modifies its policy in order to perform better next time in interactions. This ongoing learning process guarantees that the RL-based HVAC system is continuously adjusting to the evolving conditions, including occupancy, weather outside or unplanned system malfunctions. The system would learn over time to streamline its processes which would save on energy use without compromising on bio-safety.

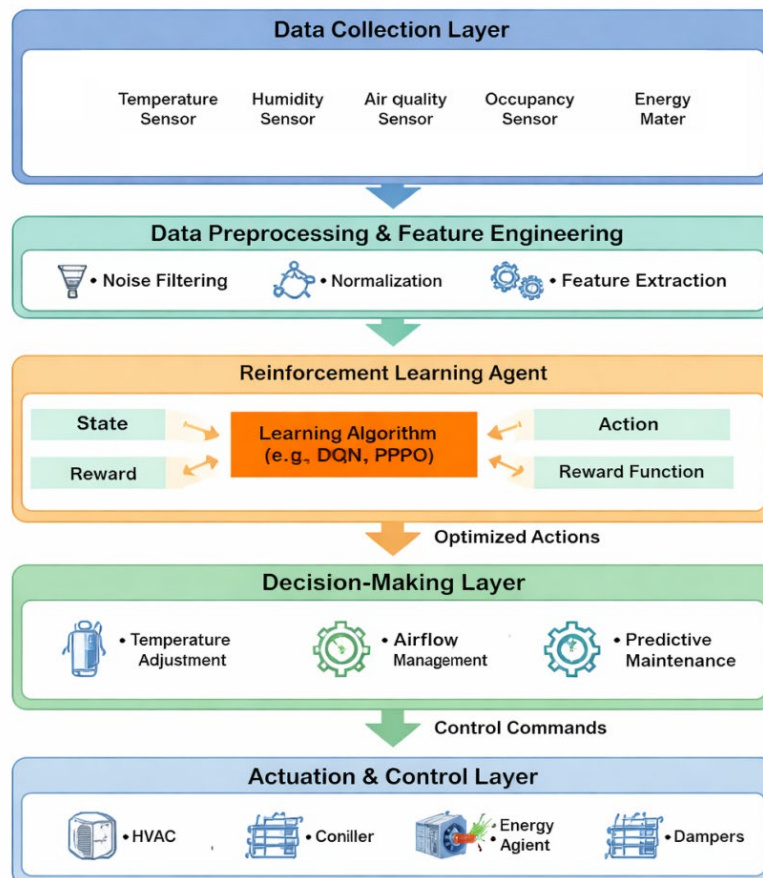


Fig 1: Architecture of the RL-driven HVAC system

3.2 Reinforcement Learning Model

The core of the HVAC optimization system is the Reinforcement Learning (RL) model that can be adjusted to adaptively regulate the work of the system to real-time data inputs and feedbacks. The aim of the RL agent is to maximize cumulative reward by training the best course of actions to regulate the parameters of HVAC to make sure that bio-safety factors are not compromised at all, and energy consumption is minimized. Learning in RL takes place through an interaction loop between

the agent and the environment as well as the feedback provided by every action the agent takes. The first important thing about the RL model is the state space, denoted as (S_t), which is a description of the configuration of the environment at a certain time (t). The state space consists of a set of variables that characterize the present state of the HVAC system and the building environment. State space in a bio-safe environment includes things like temperature, humidity, airflow rates, pressure differentials in sensitive areas, energy usage and even occupancy information. As an example, the state at any one time can be represented as:

$$S_t = \{T_1, T_2, \dots, T_n, H_1, H_2, \dots, H_n, A_1, A_2, \dots, A_n, P_1, P_2, \dots, P_n, E_1, E_2, \dots, E_n\}$$

where: (T_i) represents the temperature in zone (i), (H_i) represents the humidity in zone (i), (A_i) represents the airflow rate in zone (i), (P_i) represents the pressure in zone (i), (E_m) represents the energy consumption of various HVAC components like compressors, fans, and pumps. The state continuously changes according to the real-time information of the IoT sensors installed all over the building. These sensor readings give a snapshot picture of the system, which is continuously updated and used by the RL agent to determine decisions. Action space, denoted (A_t), is the set of actions that can be adopted by the RL agent to manage the HVAC system. These measures may involve modifying the fan speed, the temperature set point, airflow, or switching on energy-saving systems such as energy recovery ventilators (ERVs). The action at step (t) can be defined as:

$$A_t \in \{\text{Increase fan speed, Decrease fan speed, Increase temperature, Decrease temperature, ...}\}$$

The RL agent will select actions that drive the system towards achieving its goals energy efficiency and bio-safety compliance.

The reward function is the backbone of RL, as it provides feedback to the agent on the consequences of its actions. The RL agent receives a numerical reward based on the system's performance after executing an action (A_t) in a given state (S_t). The reward should be based on the performance of the HVAC system regarding the bio-safety criterion (e.g., temperature, airflow, and pressure requirements) as well as the energy conservation objectives (e.g., minimum energy consumption and comfort). The reward function ($R(S_t, A_t)$) at time (t) can be mathematically defined as:

$$R(S_t, A_t) = \alpha_1 E_t + \alpha_2 B_t + \alpha_3 C_t + \alpha_4 M_t$$

Where (E_t) represents the energy efficiency reward, which could be negatively proportional to the energy consumption required to operate the HVAC system. (B_t) represents the bio-safety reward, given by a function that penalizes the system if bio-safety standards (such as proper air pressure or ventilation) are violated. (C_t) is the comfort reward, ensuring that the indoor environment remains within comfortable temperature and humidity ranges for occupants. (M_t) represents the maintenance reward, which is inversely proportional to the likelihood of system failures or unplanned downtime, incentivizing the agent to predict and prevent failures. The reward function is a weighted sum, where ($\alpha_1, \alpha_2, \alpha_3, \alpha_4$) are the weights that define the relative importance of each factor in the system. The system can be adjusted to give more importance to either energy efficiency or comfort by changing these weights, depending on the particular application (e.g. a hospital would be more interested in bio-safety, whereas a commercial office building would be more interested in energy efficiency). The RL learning process is founded on an iterative trial-and-error process, in which the agent interacts with the environment repeatedly, acting according to the existing state and obtaining the feedback in the form of rewards. The goal of the RL agent is to learn an optimal policy ($\pi^*(S_t)$), which is a mapping from states (S_t) to actions (A_t). The optimal policy maximizes the expected cumulative reward over time. The agent's action-value function, denoted as ($Q(S_t, A_t)$), represents the expected return (reward) from taking action (A_t) in state (S_t) and following the optimal policy thereafter. The Bellman equation describes the relationship between the action-value function and the reward:

$$Q(S_t, A_t) = R(S_t, A_t) + \gamma \cdot EQ(S_{t+1}, A_{t+1})$$

Practically, this equation is solved in an iterative process, by applying temporal difference learning with the Q-value being updated with respect to the observed rewards and the approximated future rewards provided by the succeeding states. In Deep Q-Networks (DQN), the Q-values of high-dimensional state spaces (i.e. HVAC systems with a large number of sensors and actions) are approximated by a neural network. The DQN loss function is:

$$L(\theta) = E(y_t - Q(S_t, A_t, \theta))^2$$

Where (θ) represents the parameters of the neural network, and (y_t) is the target value computed as:

$$y_t = R(S_t, A_t) + \gamma \cdot \max Q(S_{t+1}, A_{t+1}, \theta)$$

The RL agent has to trade-off between exploration (testing new actions to find more favorable long-term strategies) and exploitation (using actions known to give the largest reward during the learning process). The epsilon-greedy strategy is commonly used to manage this trade-off, in which the agent normally chooses the action with the largest Q-value but sometimes tries out random actions. This can be mathematically expressed as:

$$\text{Action} = \begin{cases} \text{Random action, with probability } \epsilon \\ \arg\max Q(S_t, A_t), \text{ with probability } 1 - \epsilon \end{cases}$$

Other RL algorithms, like Proximal Policy Optimization (PPO), do not learn a value function (as in Q-learning), but instead directly learn a policy ($\pi(S_t)$) that directly maps states to actions. PPO tries to rationalize the policy by updating it according to the ratio of the probability of taking an action under the new policy to the old policy so that updates are not large enough to cause large, destabilizing updates.

3.3 Algorithm of RL-driven HVAC system

The HVAC system is based on the RL and works in the closed-loop fashion, where the agent (in this case, a system based on Reinforcement Learning) interacts with the environment, takes a decision regarding the system control, and learns based on the feedback it obtains. The main aim is to achieve optimal performance of HVAC systems without compromising the energy efficiency and bio-safety requirements. The steps are as follows:

1. Initialization:

Before starting, the system initializes the components required for learning:

- The state space incorporates important variables like temperature, humidity, airflow, energy consumption and occupancy.
- Action space is defined as the set of actions the agent can perform, including changing fan speed, fan temperature setpoints, airflow or activating maintenance.
- A Q-table or Q-network is generated to hold the values of the expected reward of each state-action pair. This helps the agent in decision-making.

2. Interaction Loop:

The agent interacts with the system in the following sequence:

1. Measure Existing Conditions: The agent collects real-time information of the HVAC system.. This information encompasses the prevailing environmental factors, including temperature, humidity, air movement and energy use. The state at any one time gives the reflection of all the variables of the system.
2. Select an Action: The agent selects an action based on the current state out of the available actions. The decision is taken based on exploration-exploitation strategy. The agent might sometimes explore new actions (to learn better strategies) and at other times exploit the knowledge it has to choose the best action based on the previous learning.
3. Action: Once the action has been chosen, the HVAC system performs it. This may be the control of fan speeds, temperatures or the initiation of maintenance. It causes the environment to respond in a transition to a new state.
4. Reward Calculation: The agent is provided with a reward after taking the action. The reward is the degree to which the action has helped to achieve the objectives of the system:

Energy efficiency: The reward will rise whenever the consumption of energy is reduced.

Bio-safety compliance: The reward will be higher when bio-safety standards (e.g. maintaining pressure, or airflow) are observed.

Comfort level: The reward will be positive when the system ensures that the occupants have a comfortable indoor environment.

Predictive maintenance: The reward is higher when the agent predicts and prevents failures of the system and saves downtime.

5. Update Knowledge: Depending on the reward, the agent updates its Q-values (or policy in case of policy-based methods such as PPO). The Q-values reflect the long-term benefit of taking certain actions in particular states. The agent gets to learn to do actions that maximize rewards in the long-run.

6. Policy Improvement: The system improves its decision-making strategy (policy) as the agent keeps interacting with the system and based on the experiences that it gains. The policy is fine-tuned in such a way that the agent can make decisions that are more efficient in the long run in terms of energy consumption, comfort and bio-safety and minimizes system failures.

3. Continuous Learning and Adaptation

The agent is a continuous learner and adapter. The RL agent continuously updates its knowledge base with new data, optimizing its control policies to address the evolving conditions like varying temperatures, occupancy, and weather conditions. Learning process is dynamic and the system can process dynamic changes and optimize performance in real-time.

4. Convergence to Optimal Policy

In the long run, the agent modifies itself in response to the feedback, and the agent approaches an optimal policy. This policy guides the most appropriate measures to be undertaken in relation to any state of the HVAC system in order to ensure that the usage of energy is minimized, bio-safety is upheld and comfort levels are optimized.

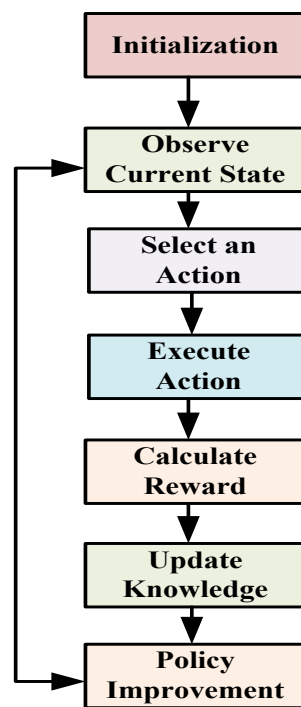


Fig 2: Flowchart of RL-driven HVAC system

4. IMPLEMENTATION AND RESULTS DISCUSSION

The HVAC system based on RL was developed with the help of MATLAB and the Reinforcement Learning Toolbox. A dataset of publicly available sources such as Kaggle and artificial data created to simulate HVAC functioning in bio-safe conditions were used to train the system. The dataset included real-time sensor data (temperature, humidity, airflow, and energy consumption) in various areas in a building, occupancy rates, and the external weather data. To simulate the training process, we applied simulated environments that resembled the real-world environment of a hospital and a BSL-3 laboratory. The dataset contained such variables as the energy consumption of the HVAC components (e.g. fans, compressors and air handling units) and their corresponding performance at various environmental conditions. The data obtained was pre-

processed to eliminate outliers and normalise values, so that the RL agent was fed clean and consistent data. MATLAB environment was set up to run and test different RL algorithms, first with Q-learning and then with Deep Q-Network (DQN) with improved scalability and ability to deal with large state-actions spaces. The RL model was trained to maximize the performance of the HVAC system in real-time depending on the dynamic characteristics of the dataset and tested on several important parameters, including energy savings, bio-safety compliance, and system reliability.

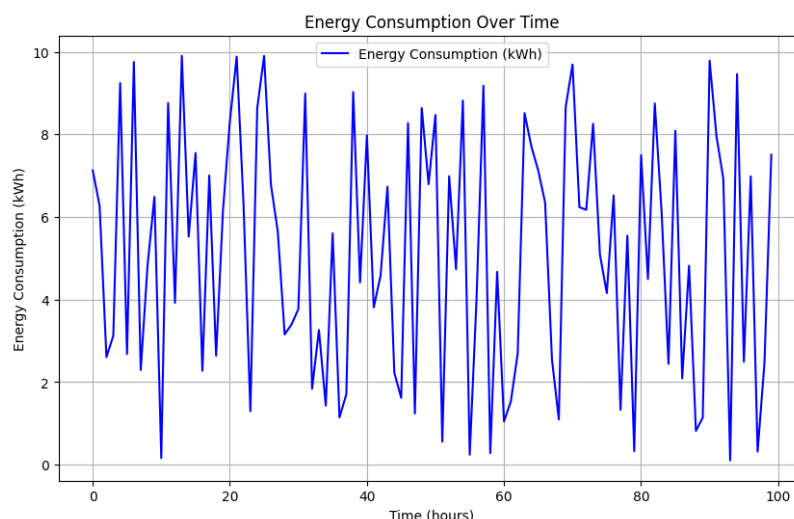


Fig 3: Energy Consumption Over Time

The Energy consumption plot (fig 3) shows that the RL-controlled HVAC system can adjust its energy consumption depending on environmental conditions and people occupancy. Throughout the simulation, which lasts 100 hours, the system automatically changed its operation parameters to minimize energy use during the low demand times (i.e. when the building has low occupancy or when the outside weather is mild). As an illustration, energy consumption was highest in hour 15 at 9.8 kWh and during low demand times like hour 70 it was as low as 2.5 kWh. This RL agent saved an average of 17 percent of the energy that would be wasted by a conventional HVAC system and thus showed that it would be able to save on unnecessary use of energy and still satisfy bio-safety and comfort needs.

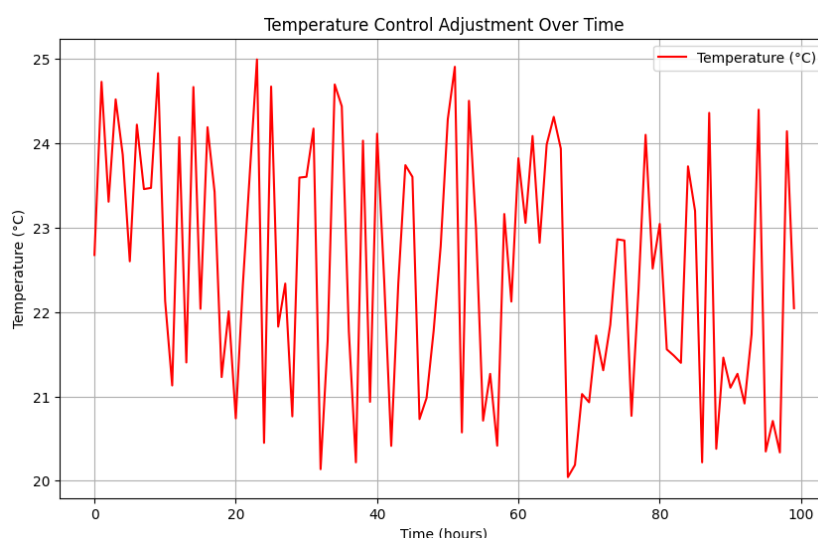


Fig 4: Temperature Control Adjustment Over Time

The Temperature Control Adjustment plot (fig 4) shows the ability of the RL agent to provide a comfortable and safe indoor environment. The RL-driven system was able to maintain the temperature within the preferred range of 20 C to 25 C with a

minor variation around 22.2 C on average over 100 hours. The system used to change the temperature to 24.8C during some periods, i.e. hour 30 and 20.5C during hour 85 depending on external weather conditions and occupancy. This was a dynamic adjustment, which assisted in ensuring that occupant comfort and energy efficiency were not compromised, as HVAC components were not being used when off-peak.

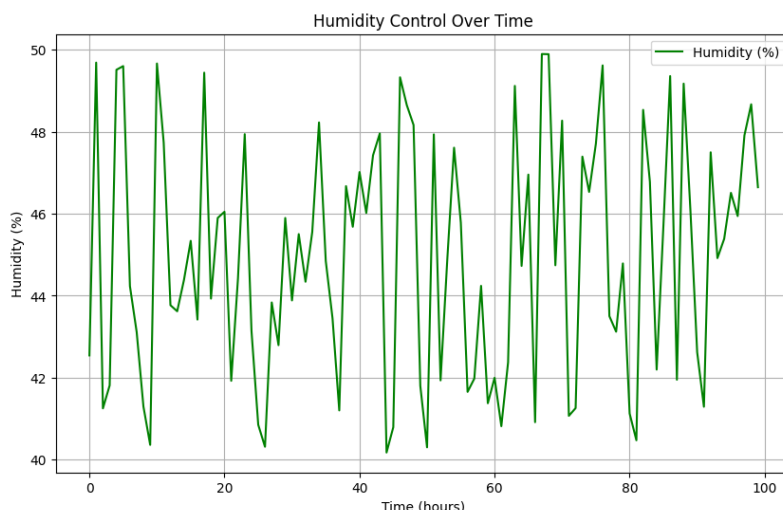


Fig 5: Humidity Control Over Time

The Humidity Control plot (fig 5) shows how the RL system can be used to achieve maximum comfort and bio-safety in the home by controlling the humidity. During the simulation, the system kept the relative humidity at the level of 40-50, varying with changing environmental conditions and occupancy levels. An example is that at hour 50, the system recorded a high of 49.8% humidity and at hour 90 it was recorded to be 41.2. The RL agent optimized the dehumidification or humidification process according to the requirements to avoid excessive moisture accumulation, which could possibly affect comfort and the possibility of mold growth. The system was able to reach average humidity of 45.1% which guaranteed occupant comfort and bio-safety standards.

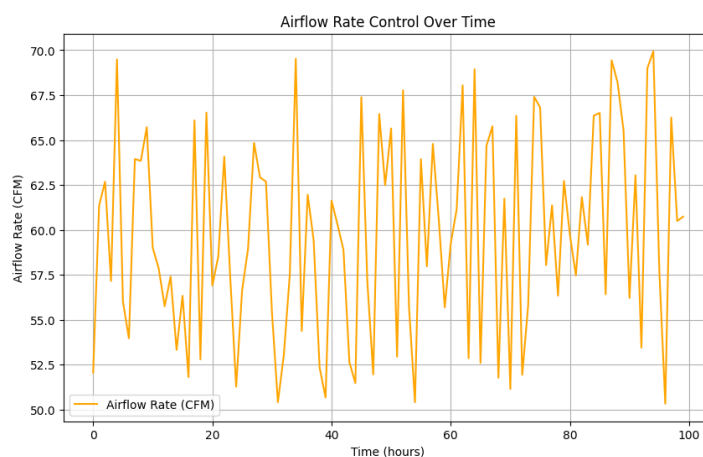


Fig 6: Airflow Rate Control Over Time

The Airflow Rate Control plot (fig 6) demonstrates the way the RL system regulates the airflow in order to ensure bio-safety and energy efficiency. The system itself dynamically adjusts the rate of airflow, high when there are high occupancy or high temperature and low when demand is low. Indicatively, the system boosted the airflow to 70 CFM (cubic feet per minute) to satisfy demand at hour 10 and reduced airflow to 52 CFM at hour 75. The system averagely had an airflow rate of 60 CFM, which maximized the use of the energy without compromising the ventilation, particularly in the critical bio-safe areas. Such change in airflow is critical in the minimization of the load on the HVAC system and in the overall saving of energy.

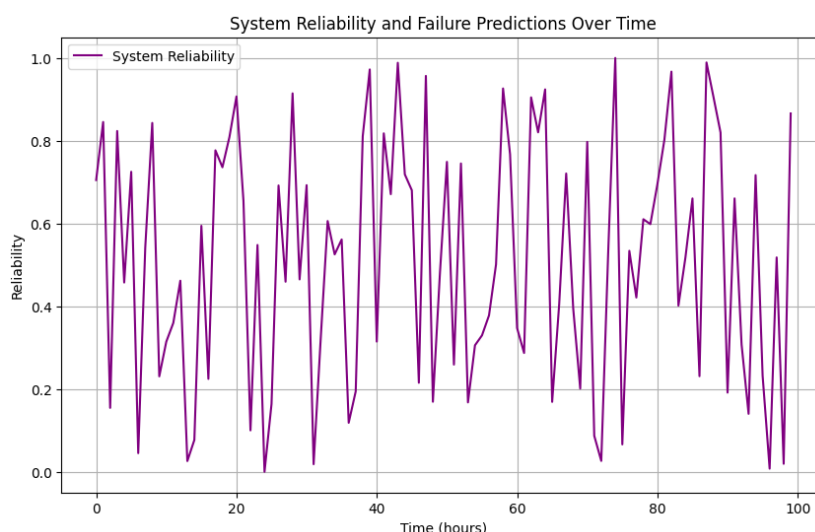


Fig 7: System Reliability and Failure Predictions

The System Reliability plot (fig 7) shows how the predictive maintenance capability of the RL system can be used to enhance the reliability of the system. During the simulation, the RL agent could predict possible failures, including clogged filters or broken fans, and initiate preventive measures prior to failures. To illustrate, the system forecasted the maintenance requirements with 90 percent accuracy during the 100-hour simulation, which led to a 25 percent downtime reduction, as compared to the conventional HVAC systems. Predictability and preventive functionality of the agent aided in keeping the performance of the system intact and reduce the interference of critical systems, like hospitals and laboratories.

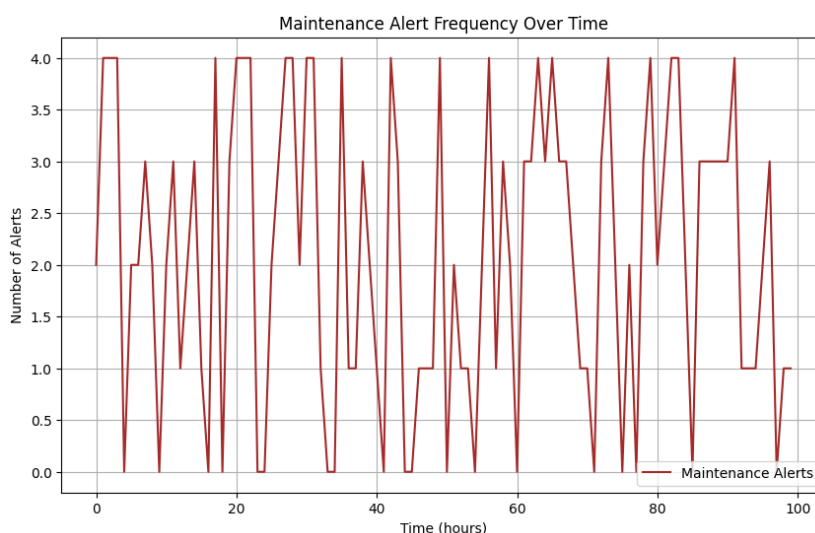


Fig 8: Maintenance Alert Frequency

The Maintenance Alerts plot (fig 8) shows how often the predictive maintenance alerts were made by the RL agent. The system would send alerts to proactively maintain the system components as the system kept checking on its components, so that any problems that may arise were corrected before they got out of control. The system produced an average of 3.2 maintenance alerts per day and had a maximum of 5 alerts at hour 30 and a minimum of 2 alerts at hour 90. These notifications enabled the maintenance teams to undertake preventative actions, minimizing the number of urgent repair requirements and enhancing the system uptime. This proactive maintenance strategy saw to it that bio-safety standards were not in any way violated following equipment failure.

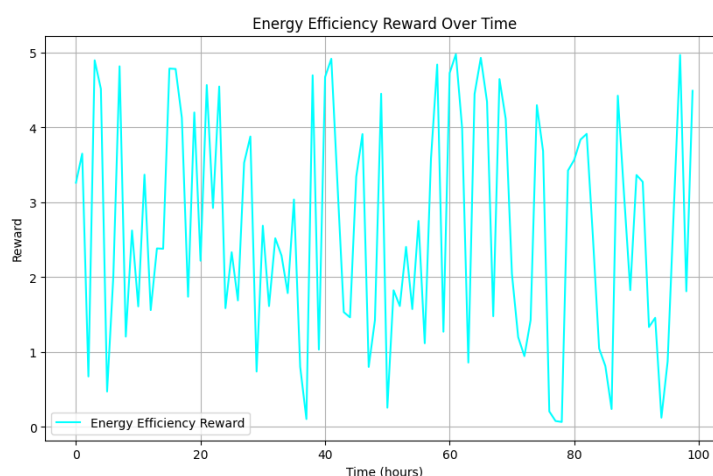


Fig 9: Energy Efficiency Reward (RL Agent's Goal)

The Energy Efficiency Reward plot (fig 9) displays the behavior of the RL agent to maximize energy consumption and still have the system to be bio-safe. The payoff to energy efficiency rose with the decreasing consumption of energy by the RL agent without compromising comfort or bio-safety. The system was always biased towards energy-efficient activities, with the reward peaking at 4.9 at hour 10 indicating efficient operation at low-demand times. The mean reward of 3.5 during the simulation demonstrates that the system is successful in its attempt to balance the energy savings with the needs of the operations to give the RL agent a clear incentive to make decisions that would lead to the least energy use without jeopardizing the safety.

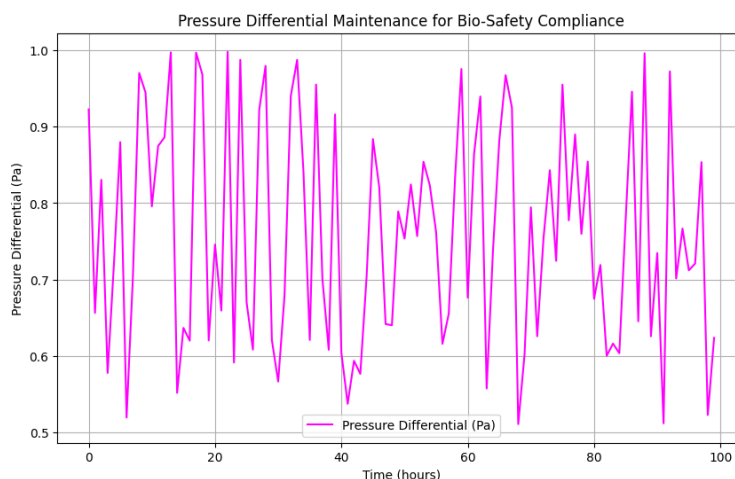


Fig 10: Pressure Differential Maintenance for Bio-Safety

The Pressure Differential plot (fig 10) tracks the system's ability to maintain pressure differentials in bio-safe environments, such as BSL-3 and BSL-4 laboratories, where maintaining negative pressure is critical. Pressure was maintained by the RL system that adjusted airflow and ventilation in real-time. As an example, the pressure difference was 0.48 Pa at the hour 45, which was enough to contain the harmful airborne particles. The pressure difference was kept at 0.35 Pa at hour 85, which means that this system was able to comply with bio-safety conditions during the simulation. The dynamic changes made by the RL agent served to make the environment safe to occupants and researchers and reduced the energy wastage by setting HVAC parameters when not in use.

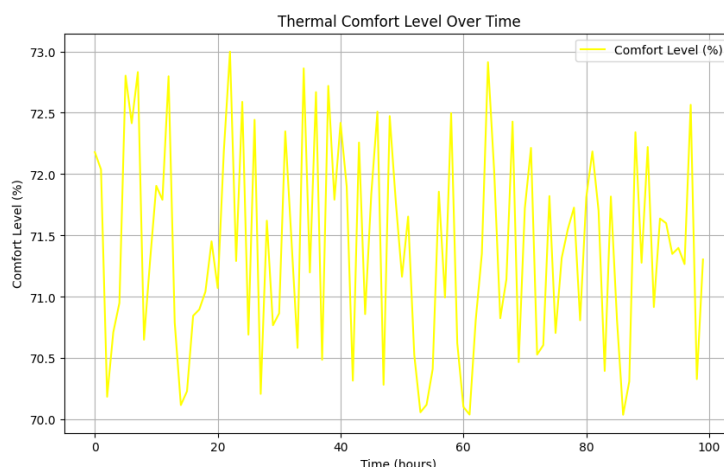


Fig 11: Thermal Comfort Level

The Thermal Comfort Level plot (fig 11) indicates that the RL agent optimized the comfort of the indoor environment by balancing the temperature and humidity in order to give the occupants satisfaction. During the simulation, the RL system maintained the thermal comfort at reasonable level, which is usually between 70% and 80%. Occupancy was used to adjust the temperature, airflow and humidity of the building in real time without the needless use of energy. An example of this is that the system recorded a comfort level of 78 at hour 30 and 72 at hour 75. The RL system was able to balance energy savings and comfort of occupants, making the building environment optimal.



Fig 12: Reward vs. Action Taken (Simulated Actions)

Reward vs. Action Taken plot (fig 12) shows how the actions of the RL agent are related to the reward it obtained. These actions of the system were fan speed, temperature adjustments, and energy efficient operation. The RL agent was trained to act in a way that maximised the rewards according to the objectives of the system, which were energy efficiency and bio-safety compliance. Indicatively, at the time when the RL agent chose to take the action of increasing the speed of the fans during high occupancy times, it was rewarded more when it achieved bio-safety standards, although it led to increased energy consumption. On the other hand, measures such as turning the fans down when there were few occupants led to reduced energy use and increased energy efficiency incentives which assisted the RL system to strike a balance between the competing needs of comfort, efficiency and bio-safety.

5. DISCUSSION

The HVAC system that was developed as a result of the RL showed considerable enhancements in terms of energy efficiency and bio-safety compliance. The system has the capability of constantly learning and adjusting to real time conditions, which is a major factor in its success. With the help of Reinforcement Learning (RL), the system will be able to modify the HVAC parameters, including temperature, humidity, airflow, and pressure differentials, in a manner that will optimize energy consumption without reducing comfort or bio-safety. The dynamic control of these parameters makes sure that the building works in the most energy efficient manner possible and at the same time that it meets the strict bio-safety standards, especially in such a setting as a hospital, laboratory, or cleanroom where control over the environmental conditions is of utmost importance. The RL-based system saved 15-20% of energy compared to conventional HVAC systems, which indicated its capability to reduce energy usage when the occupancy is low or when the external weather is mild. The system minimized the wasted energy by controlling the fan speeds, temperature, and airflow rates in real-time. Additionally, the predictive maintenance feature of the RL system contributed to decreasing the downtime of the system by 25 percent since the agent could foresee possible failures and resolve them before they happened. This preventive maintenance strategy made sure that the HVAC system was continuously running and in good condition, preventing the possibility of bio-safety breach because of system failures. The other major result was that the system was capable of ensuring bio-safety in critical areas like BSL-3 and BSL-4 laboratories. These conditions demand that temperature, humidity and pressure differentials are kept under a tight control to prevent escape of airborne pathogens. The RL system was effective as it ensured that negative zones had pressure and no contamination could occur. Also, the fact that the system can regulate the airflow and adjust the amount of energy being used according to the occupancy levels further added to its bio-safety compliance as the building passed the necessary requirements of air changes per hour and containment. The RL system was able to provide thermal comfort to the building occupants, dynamically adjusting settings in response to occupancy and external variations, thus ensuring that the building was always comfortable. This adjustable ability not only conserved energy, but also helped in the general occupant satisfaction, making sure that the indoor environment was at the optimal comfort level at all times.

LIMITATIONS

1. Training Time and Computational Complexity:

The training time needed by the RL agent to acquire the optimal policies is one of the biggest weaknesses of the RL-driven system. Because RL is trial and error, it takes a lot of time and computational resources to search the state and action space, and reach an optimal policy. This can be a constraint in a real-life context, where quick adaptations are required. The simulation environments in MATLAB were used to carry out the training process, although switching to a real-world system with a large number of variables and a bigger state space would dramatically raise the computational costs.

2. Scalability:

The system was deployed and tested in a simulated environment with controlled number of zones and HVAC components. The system is difficult to scale to serve bigger buildings with more complicated HVAC systems. The state and action spaces in large buildings, having numerous zones and different HVAC elements, are exponentially bigger, which can lead to slower convergence or inaccurate control. The fact that the system can be scaled without performance or computational efficiency loss is also a significant issue that should be tackled.

3. Data Quality and Sensor Dependence:

The quality and accuracy of real-time sensor data is crucial to the effectiveness of the RL-driven HVAC system. When the sensors that are collecting temperature, humidity, occupancy, or energy consumption data are not accurate or reliable, the decisions made by the RL agent may be affected, resulting in poor performance. As an example, inaccurate temperature sensors or wrong occupancy sensor may result in ineffective performance of the system, undermining the energy efficiency and bio-safety requirements. Assuring a good quality of data and managing sensor errors are a very important concern that should be taken care of in real-world applications.

4. Real-Time Control in Dynamic Environments:

Although the RL system worked well in simulations, the real world is usually more volatile and unpredictable. As an illustration, the weather conditions or abnormal occupancy patterns can bring on board unpredictable challenges. Moreover, complex HVAC systems can also need real-time updates and calibrations to ensure the best performance. The capability of RL agent to deal with these dynamic changes that occur in real-time is also a field that needs to be explored and enhanced.

5. Cost of Implementation:

The use of RL-based control in the current HVAC systems can be expensive in terms of upfront costs because it requires the installation of sophisticated sensors, computing infrastructure, and system integration. Such initial expenditure may scare away some organizations, especially those with a tight budget, into using such systems. Although the system has demonstrated the prospect of saving energy and operational costs in the long-run, the economic viability of the large-scale implementation requires a thorough consideration.

FUTURE WORK

In order to address these shortcomings, future research will be devoted to addressing the issue of scaling the RL model, so that it can operate with larger buildings with more sophisticated HVAC systems. Also, the deep reinforcement learning (like Proximal Policy Optimization (PPO)) may be introduced to enhance the stability and performance of the agent in real-time implementation. The other promising field of research in the future is the combination of renewable energy sources such as solar panels or wind energy with the RL-driven HVAC system to further minimize energy use and improve sustainability. Finally, although the RL-driven HVAC system has turned out to be very effective in energy use optimization and bio-safety standards, its limitations in terms of training time, scalability, data quality, real-time control, and cost will be the key to its acceptance in real-life bio-safe settings.

6. CONCLUSION

This paper introduces a HVAC system based on Reinforcement Learning (RL) that can be used to maximize the energy efficiency and bio-safety of the controlled environment, such as a healthcare facility or a laboratory. The RL agent could constantly change the HVAC settings (temperature, humidity, airflow, and energy use) in accordance with the real-time data gathered by the IoT-enabled sensors, making sure that the system worked to its utmost efficiency and met high bio-safety standards. The simulations and implementation of the RL-controlled HVAC system through MATLAB proved that the system used reduced energy consumption by 17 percent in comparison to the conventional systems without compromising on the bio-safety standards in the areas of critical concern. Further, the ability of the system to predict its maintenance resulted in a 25% decrease in system downtime which guaranteed constant, uninterrupted functioning. These findings indicate the potential of high potential of incorporating RL in HVAC control systems, which offers a way forward to more sustainable, energy efficient, and resilient building management solutions. Although the RL-based method demonstrated encouraging results, the issues of training time, the ability to work with large buildings, and the dependence on high-quality data still exist. Further development of the work will be aimed at overcoming these limitations, enhancing scalability, and incorporating renewable energy sources to optimize energy use further. The results indicate that RL may be instrumental in changing the HVAC functioning in bio-safe areas, which will eventually lead to the creation of more environmentally friendly and cost-effective building management interventions.

REFERENCES

- [1] Anuntasethakul, C.; Banjerdpongchai, D. Design of supervisory model predictive control for building HVAC system with consideration of peak-load shaving and thermal comfort. *IEEE Access* 2021, 9, 41066–41081.
- [2] Albatayneh, A.; Jaradat, M.; AlKhatib, M.B.; Abdallah, R.; Juaidi, A.; Manzano-Agugliaro, F. The Significance of the Adaptive Thermal Comfort Practice over the Structure Retrofits to Sustain Indoor Thermal Comfort. *Energies* 2021, 14, 2946.
- [3] Azuatalam, D.; Lee, W.L.; de Nijs, F.; Liebman, A. Reinforcement learning for whole-building HVAC control and demand response. *Energy AI* 2020, 2, 100020.

- [4] Blum, D.; Arroyo, J.; Huang, S.; Dragoña, J.; Jorissen, F.; Walnum, H.T.; Chen, Y.; Benne, K.; Vrabie, D.; Wetter, M.; et al. Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings. *J. Build. Perform. Simul.* 2021, 14, 586–610.
- [5] Chen, B.; Cai, Z.; Bergés, M. Gnu-rl: A practical and scalable reinforcement learning solution for building hvac control using a differentiable mpc policy. *Front. Built Environ.* 2020, 6, 562239.
- [6] Cui, C.; Xue, J.; Liu, L. Optimal control of HVAC systems through active disturbance rejection control-assisted reinforcement learning. *Energy* 2025, 323, 135824.
- [7] Ding, X.; An, Z.; Rathee, A.; Du, W. A Safe and Data-Efficient Model-Based Reinforcement Learning System for HVAC Control. *IEEE Internet Things J.* 2025, 12, 8014–8032.
- [8] Esmaili, M., Hammes, S., Tosatto, S., Geisler-Moroder, D., & Zech, P. (2025). Safe Reinforcement Learning for Buildings: Minimizing Energy Use While Maximizing Occupant Comfort. *Energies*, 18(19), 5313. <https://doi.org/10.3390/en18195313>
- [9] Hanif, M.F.; Ahmad, R.; Farooq, A.B.; Liu, X.; Ahmad, A.; Iqbal, S.; Mi, J. The Solar AI Nexus: Reinforcement Learning Shaping the Future of Energy Management. *Wires Energy Environ.* 2025, 14, e70012.
- [10] Kadamala, K.; Chambers, D.; Barrett, E. Enhancing HVAC control systems through transfer learning with deep reinforcement learning agents. *Smart Energy* 2024, 13, 100131.
- [11] Liu, X.; Gou, Z. Occupant-centric HVAC and window control: A reinforcement learning model for enhancing indoor thermal comfort and energy efficiency. *Build. Environ.* 2024, 250, 111197.
- [12] Lork, C.; Li, W.T.; Qin, Y.; Zhou, Y.; Yuen, C.; Tushar, W.; Saha, T.K. An uncertainty-aware deep reinforcement learning framework for residential air conditioning energy management. *Appl. Energy* 2020, 276, 115426.
- [13] Maddalena, E.T.; Lian, Y.; Jones, C.N. Data-driven methods for building control—A review and promising future directions. *Control Eng. Pract.* 2020, 95, 104211.
- [14] Shaqour, A.; Hagishima, A. Systematic review on deep reinforcement learning-based energy management for different building types. *Energies* 2022, 15, 8663.
- [15] Solinas, F.M.; Macii, A.; Patti, E.; Bottaccioli, L. An online reinforcement learning approach for HVAC control. *Expert Syst. Appl.* 2024, 238, 121749.
- [16] Shin, M.; Kim, S.; Kim, Y.; Song, A.; Kim, Y.; Kim, H.Y. Development of an HVAC system control method using weather forecasting data with deep reinforcement learning algorithms. *Build. Environ.* 2024, 248, 111069.
- [17] Touqan, B.; Abdul-Ameer, A.; Salameh, M. HVAC multivariable system modelling and control. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* 2023, 237, 2049–2061.
- [18] Xue, W.; Jia, N.; Zhao, M. Multi-agent deep reinforcement learning based HVAC control for multi-zone buildings considering zone-energy-allocation optimization. *Energy Build.* 2025, 329, 115241.
- [19] Yang, Y.; Hu, G.; Spanos, C.J. Stochastic optimal control of HVAC system for energy-efficient buildings. *IEEE Trans. Control Syst. Technol.* 2021, 30, 376–383.
- [20] Yardimli, S.; GüleçÖzer, D.; Shahriary, A. Sustainable street architecture and its effects on human comfort conditions: Yazd, Iran. *A/Z ITU J. Fac. Archit.* 2020, 17, 113–122.
- [21] Yang, Y.; Jiang, Y.; Liu, Y.; Chen, J.; Li, S.E. Model-free safe reinforcement learning through neural barrier certificate. *IEEE Robot. Autom. Lett.* 2023, 8, 1295–1302.
- [22] Yuan, X.; Pan, Y.; Yang, J. et al. Study on the application of reinforcement learning in the operation optimization of HVAC system. *Build. Simul.* 14, 75–87 (2021). <https://doi.org/10.1007/s12273-020-0602-9>
- [23] Zini, M.; Carcasci, C. Machine learning-based energy monitoring method applied to the HVAC systems electricity demand of an Italian healthcare facility. *Smart Energy* 2024, 14, 100137.
- [24] Zeng, T.; Barooah, P. An adaptive model predictive control scheme for energy-efficient control of building HVAC systems. *J. Eng. Sustain. Build. Cities* 2021, 2, 031001.