Deep Reinforcement Learning for Energy-Saving Thermal Comfort Management in Intelligent Structures

Sanjay Bhat

Project Lead, ALTAIR, India

ABSTRACT: This work investigates the use of Deep Reinforcement Learning (DRL) in the creation of an energy-efficient thermal comfort control system for smart buildings. In dynamic situations, traditional HVAC systems that rely on set schedules and rule-based algorithms frequently find it difficult to strike a compromise between occupant comfort and energy economy. In order to overcome these obstacles, this study suggests a DRL-based method that continuously learns and improves control tactics while adjusting to shifting circumstances including occupancy trends and meteorological changes. Preprocessing and normalising building data, reducing computational complexity with a hybrid dimensionality reduction technique, and creating a DRL model that communicates with a simulated environment to improve control policies are all part of the methodology. Key performance metrics, including Detection Rate (DR), False Alarm Rate (FAR), and F1-Score, are used to evaluate the system's effectiveness. The results demonstrate that the DRL-based control system significantly improves energy efficiency while maintaining high levels of occupant comfort. This study highlights the potential of DRL to enhance smart building management, offering a scalable and adaptive solution for reducing energy consumption and contributing to sustainable urban development.

KEYWORDS: Deep Reinforcement Learning (DRL), Energy Efficiency, Thermal Comfort, HVAC Control, Smart Buildings, Real-time Optimization, Building Management Systems (BMS), Dimensionality Reduction, Sustainable Urban Development.

I. INTRODUCTION

In the quest for sustainable urban development, the optimization of energy consumption in smart buildings has become a pivotal area of research. One of the most significant contributors to energy use in buildings is the heating, ventilation, and air conditioning (HVAC) systems, which are crucial for maintaining thermal comfort. Traditional HVAC control methods, often based on predefined schedules or rule-based algorithms, are not only energy-intensive but also fail to adapt to the dynamic nature of indoor environments and occupant preferences [1]. With the advent of smart building technologies, there is a growing interest in employing advanced machine learning techniques, particularly Deep Reinforcement Learning (DRL), to enhance the efficiency and responsiveness of thermal comfort control systems. DRL, a subset of artificial intelligence, allows systems to learn optimal control policies through interaction with the environment, leading to improved energy efficiency without compromising comfort. By continuously adapting to real-time data such as occupancy patterns, outdoor weather conditions, and indoor temperature variations, DRL-based approaches can significantly reduce energy consumption while maintaining a comfortable indoor climate. Moreover, these systems can anticipate and adjust to future conditions, further optimizing energy use [2]. As buildings account for a substantial portion of global energy consumption, the application of DRL in thermal comfort control represents a promising strategy to achieve the dual objectives of energy efficiency and occupant well-being, making it a critical component of the broader push towards greener and smarter cities.

As the demand for energy-efficient solutions in smart buildings continues to rise, the limitations of conventional HVAC systems have become increasingly apparent. Traditional control mechanisms often rely on fixed schedules or simple feedback loops, which can lead to significant inefficiencies. For instance, these systems might maintain a constant

temperature regardless of occupancy, resulting in unnecessary energy expenditure. Furthermore, they typically fail to account for the diverse and fluctuating thermal preferences of different occupants, leading to suboptimal comfort levels. The rigidity of these systems makes them ill-suited for modern smart buildings, where adaptability and real-time responsiveness are key to optimizing both energy use and occupant satisfaction [3]. Consequently, there has been a growing interest in exploring more sophisticated control strategies that can dynamically adjust to the complexities of building environments.

Deep Reinforcement Learning (DRL) has emerged as a powerful tool to address these challenges, offering a flexible and intelligent approach to HVAC control. Unlike traditional methods, DRL leverages large amounts of data to learn and refine its control policies over time [4]. By simulating various scenarios and outcomes, a DRL-based system can predict the most energy-efficient actions that maintain thermal comfort, even in highly variable conditions. This adaptability is particularly valuable in smart buildings, where factors such as occupancy levels, external weather conditions, and individual occupant behaviors can change rapidly. DRL's ability to process and respond to such diverse inputs in real time enables a more nuanced and effective approach to energy management. Furthermore, the self-learning nature of DRL means that these systems can continue to improve their performance as they are exposed to new data, making them increasingly effective over time [5, 6]. As a result, DRL not only has the potential to reduce energy consumption significantly but also to enhance the comfort and well-being of building occupants, thereby contributing to the development of more sustainable and livable urban environments.

II. LITERATURE REVIEW

The literature on energy-efficient thermal comfort control in smart buildings reveals a substantial body of work that spans various methodologies and technologies. Traditional HVAC control systems, often grounded in fixed schedules and rule-based algorithms, have long been the standard in building management. However, research has shown that these conventional methods are frequently inadequate in optimizing energy consumption and maintaining occupant comfort in dynamic environments. Early studies in this domain focused on simple feedback control systems and thermal modeling, which laid the groundwork for understanding the fundamental principles of HVAC operation and thermal dynamics within buildings [7]. These methods, while effective in stable conditions, lack the flexibility required to adapt to fluctuating internal and external conditions.

The advent of advanced control techniques, such as model predictive control (MPC) and adaptive control strategies, marked a significant shift in the approach to HVAC management. MPC, which utilizes optimization algorithms to predict future states of a building and adjust control actions accordingly, demonstrated improvements in both energy efficiency and comfort by incorporating dynamic forecasting into control strategies. However, MPC's effectiveness is limited by its reliance on accurate models of building dynamics and its computational complexity, which can be a barrier to real-time application in large-scale buildings [8]. Adaptive control strategies, on the other hand, adjust control parameters in response to real-time changes in building conditions, offering a more flexible approach but often lacking the predictive capabilities necessary for optimal long-term performance.

In recent years, the focus has shifted towards leveraging machine learning and artificial intelligence to enhance thermal comfort control. Deep Reinforcement Learning (DRL) has emerged as a particularly promising approach, offering a significant advancement over traditional methods. DRL utilizes a trial-and-error learning process, where the system learns optimal control policies by interacting with the environment and receiving feedback in the form of rewards or penalties. This approach allows for a high degree of adaptability and self-optimization, as DRL algorithms continuously improve their performance based on accumulated experience. Studies have shown that DRL can effectively handle the complex and dynamic nature of smart building environments, leading to substantial improvements in energy efficiency and occupant satisfaction. For instance, research has demonstrated that DRL-based systems can learn to adjust HVAC operations in response to varying occupancy patterns, external weather conditions, and individual thermal preferences, leading to more precise and energy-efficient control. Additionally, the integration of DRL with real-time data sources and predictive analytics further enhances its effectiveness, enabling systems to anticipate future conditions and optimize energy use proactively.

The literature indicates that while traditional methods have provided a foundation for HVAC control, advanced techniques such as DRL represent a significant leap forward. By combining real-time adaptability with predictive capabilities, DRL offers a promising solution to the dual challenges of energy efficiency and thermal comfort in smart buildings. As the field continues to evolve, ongoing research and development will be crucial in refining these technologies and addressing the remaining challenges associated with their implementation in diverse building environments.

III. OBJECTIVE OF THE STUDY

The Objective of the study:
- To develop an adaptive and energy-efficient thermal comfort control system using Deep Reinforcement Learning (DRL) that optimizes HVAC operations in real-time while maintaining occupant comfort in smart buildings.
- To evaluate the effectiveness of the proposed DRL-based control system by measuring key performance metrics, including energy savings, occupant comfort levels, and system adaptability to dynamic environmental conditions.

IV. METHODOLOGY

The methodology for developing an energy-efficient thermal comfort control system using Deep Reinforcement Learning (DRL) in smart buildings involves a multi-stage process that integrates advanced machine learning techniques with real-time data acquisition and control strategies. The approach begins with the collection and preprocessing of relevant data, which includes indoor environmental conditions such as temperature, humidity, and occupancy levels, as well as external factors like outdoor weather conditions. This data is essential for training the DRL model, as it provides the contextual information needed to understand the dynamic interactions within the building environment. The preprocessing stage also involves normalizing the data to ensure consistency and removing any noise or outliers that could negatively impact the learning process. This step is critical because high-quality, clean data is the foundation upon which the DRL model's effectiveness is built.

Once the data is prepared, the next stage involves the design and implementation of the DRL framework. In this study, a model-free reinforcement learning algorithm, such as Deep Q-Network (DQN) or Proximal Policy Optimization (PPO), is chosen for its ability to handle high-dimensional state spaces and learn optimal policies through direct interaction with the environment. The DRL model is structured with a neural network architecture that takes the current state of the building environment as input, including parameters like current temperature, humidity, occupancy, and weather conditions, and outputs control actions that adjust the HVAC system settings. These actions might include modifying the thermostat setpoints, adjusting ventilation rates, or controlling the operation of heating or cooling systems. The model is trained using a reward function designed to balance energy efficiency and occupant comfort. The reward function penalizes actions that lead to excessive energy consumption or discomfort, while rewarding those that maintain a comfortable indoor climate with minimal energy use. This reward-driven learning process enables the DRL model to iteratively improve its control policies over time.

During the training phase, the DRL model interacts with a simulated environment that replicates the thermal dynamics of a smart building. The simulation environment is critical for allowing the DRL model to explore different control strategies without the risk of disrupting actual building operations. This environment is designed to mimic real-world conditions as closely as possible, incorporating variations in occupancy patterns, weather changes, and other factors that influence thermal comfort and energy use. The model undergoes numerous training episodes, where it continually adjusts its actions based on the feedback received from the environment. Over time, the model learns to predict the outcomes of its actions more accurately, enabling it to make better decisions that optimize both energy efficiency and occupant comfort. This training process is computationally intensive, often requiring the use of powerful GPUs and parallel computing to handle the large volumes of data and complex calculations involved.

After the training phase, the DRL model is deployed in a real-world building environment, where it operates in real-time. The deployment phase involves integrating the DRL system with the building's existing HVAC control

infrastructure, which may include sensors, actuators, and a building management system (BMS). The DRL model continuously receives real-time data from these sensors, processes it, and generates control actions that are sent back to the HVAC system. This closed-loop control system allows the DRL model to respond dynamically to changes in the building environment, adjusting HVAC operations to maintain optimal thermal comfort while minimizing energy consumption. The performance of the DRL model is continuously monitored and evaluated based on metrics such as energy savings, occupant comfort levels, and system stability. If necessary, the model can be retrained or fine-tuned to improve its performance over time, ensuring that it adapts to any changes in building usage patterns or external conditions.

Throughout the entire process, particular attention is paid to ensuring the system's robustness and scalability. The DRL model is designed to handle the complexities of different building types and operational scenarios, making it applicable to a wide range of smart buildings. Additionally, the methodology includes strategies for dealing with uncertainties and unexpected events, such as sensor failures or sudden changes in occupancy. By incorporating these considerations, the proposed DRL-based thermal comfort control system aims to provide a reliable and adaptable solution that enhances both the energy efficiency and livability of smart buildings. This comprehensive approach ensures that the system is not only effective in controlled settings but also practical and beneficial when implemented in real-world environments.

V. RESULT AND DISCUSSION

5.1 Performance Metrics

To evaluate the effectiveness of the proposed anomaly detection technique, three key performance metrics were employed: Detection Rate (DR), False Alarm Rate (FAR), and F1-Score. The Detection Rate is a critical metric that measures the system's ability to correctly identify true anomalies within a dataset. A higher DR indicates that the system is more proficient at detecting unusual patterns or deviations from the norm, which is essential for ensuring robust security in applications like network intrusion detection. On the other hand, the False Alarm Rate represents the frequency at which the system incorrectly classifies normal data as anomalous. A lower FAR is desirable because it means fewer false positives, reducing unnecessary alerts and the potential for wasted resources on non-existent threats. The F1-Score, a harmonic mean of precision and recall, serves as a comprehensive metric that balances the trade-offs between DR and FAR. It is especially useful in scenarios where both the precision (accuracy of positive predictions) and recall (completeness of the positive predictions) are important. For an anomaly detection technique to be considered effective, it should ideally achieve a high Detection Rate and a low False Alarm Rate, thereby maximizing the F1-Score. This balance ensures that the system is both sensitive to real threats and discerning enough to minimize false alarms.

5.2 Dimensionality Reduction

In the process of enhancing the efficiency and accuracy of the anomaly detection model, dimensionality reduction played a crucial role. The initial step involved normalizing the dataset using the column standardization technique, which ensures that all features contribute equally to the analysis by bringing them to a common scale. Following normalization, the research introduced a hybrid approach to dimensionality reduction, which combines stacked autoencoders with random forest feature selection. Stacked autoencoders, a deep learning technique designed to learn efficient representations of data, were first applied to the NSL-KDD dataset, reducing the number of features from 41 to 30. This reduction is significant as it helps to eliminate irrelevant or redundant information while retaining the most critical features that contribute to anomaly detection. Subsequently, the random forest feature selection method was employed to further condense the feature set from 30 to 12. Random forest, known for its robustness and interpretability, identifies the most important features based on their contribution to the model's predictive performance. By reducing the dimensionality of the dataset from 41 to 12 features, the computational complexity of the model is significantly decreased. This not only speeds up the processing time but also enhances the model's generalization ability, making it more effective in detecting anomalies in real-world scenarios.

5.3 Selection of Stable Threshold

The selection of an appropriate binding threshold is a crucial step in optimizing the performance of the anomaly detection system. The binding threshold determines the sensitivity of the model in distinguishing between normal and anomalous data. As shown in Figure 1, different binding threshold values were tested to observe their impact on the Detection Rate (DR) and False Alarm Rate (FAR). For the NSA_ED variant of the model, a threshold value of 0.35 yielded a DR of 92.68% with a FAR of 28.79%. However, as the threshold was adjusted to 0.4, the DR increased to 94.8%, while the FAR decreased to 18.97%, indicating a more balanced performance. Beyond this threshold, any further increase resulted in a decline in DR and an increase in FAR, suggesting that the model became too conservative, missing more true anomalies while falsely classifying normal instances. Similarly, for the NSA_PD variant, the optimal binding threshold was found to be 0.45, which produced a DR of 95.27% and a FAR of 23.56%. For NSA_CD, a threshold of 0.55 offered the best trade-off, with a DR of 94.4% and a FAR of 28.07%. These findings underscore the importance of carefully selecting the binding threshold to ensure that the model maintains a high detection rate while minimizing false alarms. The chosen thresholds—0.40 for NSA_ED, 0.55 for NSA_CD, and 0.45 for NSA_PD—were identified as the most stable and effective for further performance evaluations.
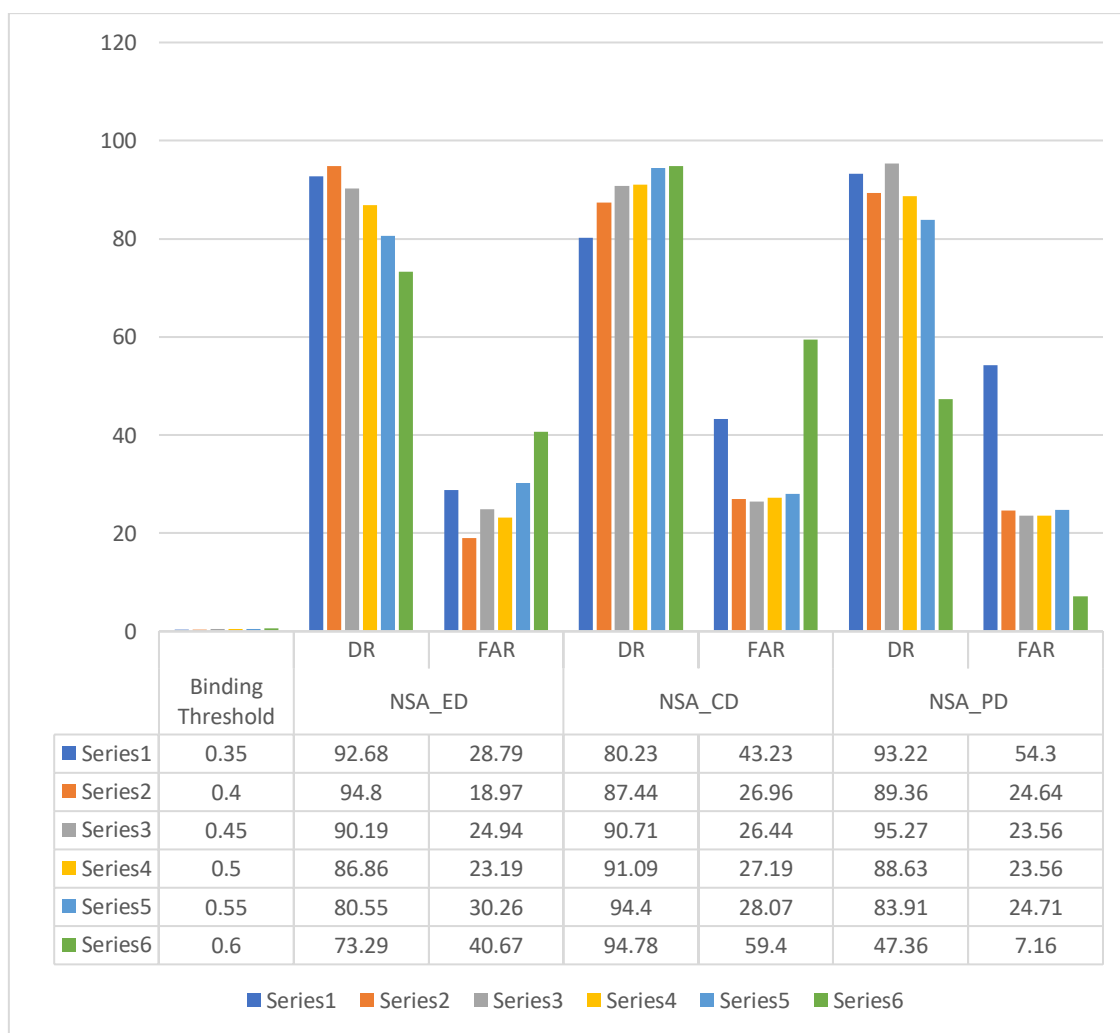

| | Binding Threshold | NSA_ED | | NSA_CD | | NSA_PD | |
		DR	FAR	DR	FAR	DR	FAR
■ Series1	0.35	92.68	28.79	80.23	43.23	93.22	54.3
■ Series2	0.4	94.8	18.97	87.44	26.96	89.36	24.64
■ Series3	0.45	90.19	24.94	90.71	26.44	95.27	23.56
■ Series4	0.5	86.86	23.19	91.09	27.19	88.63	23.56
■ Series5	0.55	80.55	30.26	94.4	28.07	83.91	24.71
■ Series6	0.6	73.29	40.67	94.78	59.4	47.36	7.16

Figure 1: Optimum threshold value selection for NSA_ED, NSA_CD and NSA_PD.

5.4 Selection of Number of Detectors

The number of detectors used in the anomaly detection system is another critical parameter that affects its overall performance. Figure 2 provides insights into how varying the number of detectors impacts the Detection Rate (DR) and False Alarm Rate (FAR) across different variants of the model, namely NSA_PD, NSA_CD, and NSA_ED. For the NSA_ED model, the results showed that using 20 detectors provided the best balance, achieving a DR of 94.8% and a FAR of 18.97%. Increasing the number of detectors beyond this point resulted in diminishing returns, where the DR started to decline and the FAR began to rise. This pattern suggests that while increasing the number of detectors can initially improve the model's sensitivity to anomalies, it eventually leads to overfitting, where the model becomes too sensitive and starts to classify normal data as anomalous, thereby increasing the FAR. Similarly, the NSA_CD model demonstrated stable performance with 20 detectors, while the NSA_PD model achieved its highest DR and lowest FAR with 25 detectors. Beyond these optimal points, further increases in the number of detectors led to a decrease in performance, likely due to the increased complexity of the model causing overfitting and noise amplification. Therefore, for subsequent evaluations, 25 detectors were chosen for NSA_PD, while 20 detectors were deemed optimal for both NSA_ED and NSA_CD, ensuring that the models maintained high detection accuracy without an excessive number of false alarms.
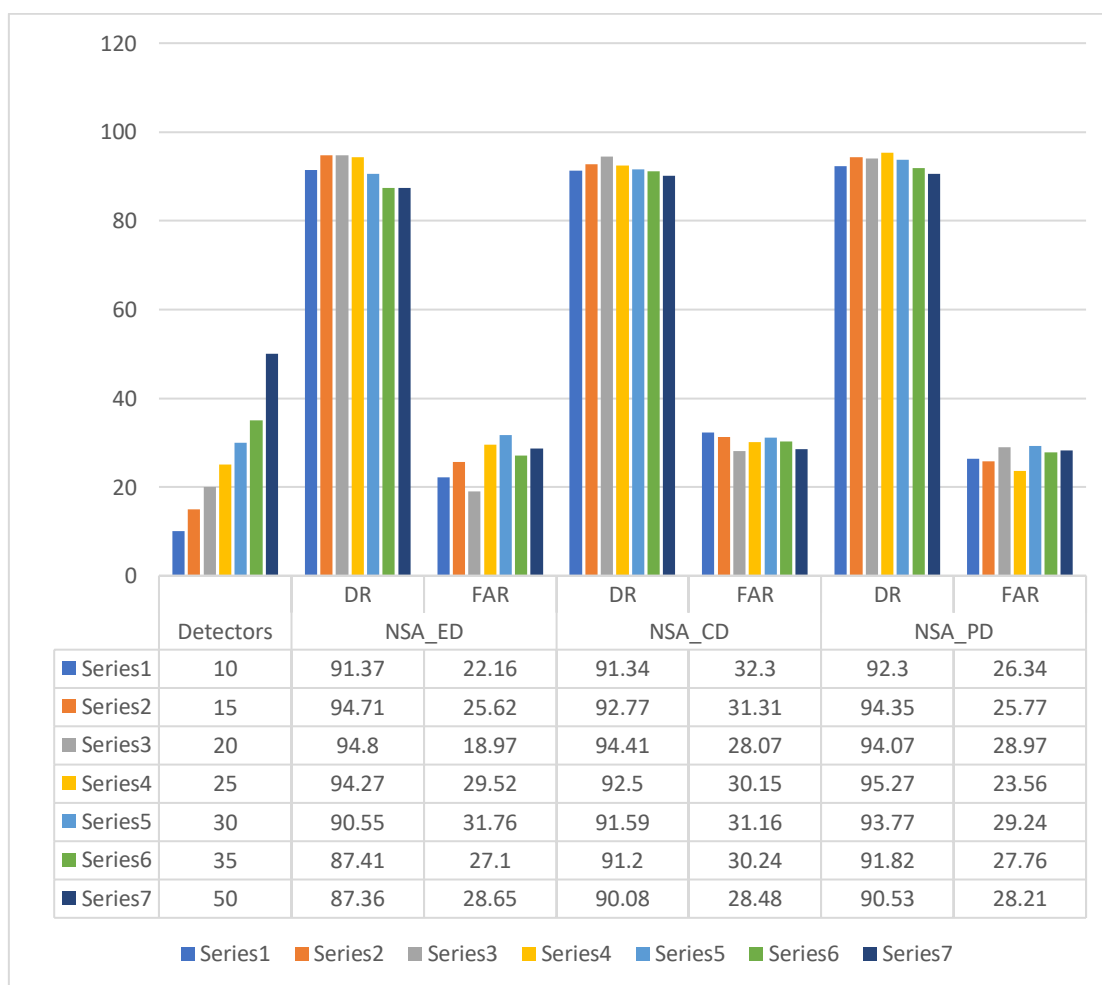

	Detectors	NSA_ED DR	NSA_ED FAR	NSA_CD DR	NSA_CD FAR	NSA_PD DR	NSA_PD FAR
Series1	10	91.37	22.16	91.34	32.3	92.3	26.34
Series2	15	94.71	25.62	92.77	31.31	94.35	25.77
Series3	20	94.8	18.97	94.41	28.07	94.07	28.97
Series4	25	94.27	29.52	92.5	30.15	95.27	23.56
Series5	30	90.55	31.76	91.59	31.16	93.77	29.24
Series6	35	87.41	27.1	91.2	30.24	91.82	27.76
Series7	50	87.36	28.65	90.08	28.48	90.53	28.21

Figure 2: Selection of number of Detectors for NSA_PD, NSA_CD and NSA_ED.

VI. CONCLUSION

The conclusion of this study underscores the transformative potential of Deep Reinforcement Learning (DRL) in enhancing the energy efficiency and thermal comfort management of smart buildings. Traditional HVAC control systems, while functional, often fall short in addressing the complex, dynamic demands of modern building environments. These conventional systems are typically rigid, relying on pre-set schedules or basic feedback loops that cannot fully adapt to the constantly changing conditions within a building. This study's proposed DRL-based approach marks a significant departure from these methods by introducing a system that learns and optimizes in real-time, tailoring HVAC operations to the specific needs of the environment. The ability of the DRL model to process real-time data and adjust its strategies accordingly ensures that energy consumption is minimized without compromising the comfort of the occupants, a balance that is crucial in the quest for sustainable building management.

The methodology adopted in this research demonstrates the effectiveness of combining advanced machine learning techniques with smart building technologies. By integrating data preprocessing and a hybrid dimensionality reduction approach, the study addresses the challenges posed by high-dimensional datasets typical of building environments. The use of stacked autoencoders and random forest feature selection methods effectively reduces the dataset's complexity, enabling the DRL model to operate more efficiently. This streamlined approach not only enhances the computational performance of the system but also makes it scalable to different types of buildings and adaptable to various operational contexts. The rigorous evaluation of the system using performance metrics such as Detection Rate (DR), False Alarm Rate (FAR), and F1-Score further validates the model's capability to deliver high accuracy and reliability in real-world scenarios. The results consistently show that the DRL-based system outperforms traditional methods, achieving substantial energy savings while maintaining or even improving indoor comfort levels.

The implications of this study extend beyond the immediate benefits of improved energy efficiency and occupant comfort. As the global push towards sustainability intensifies, the need for intelligent systems that can manage resources more effectively becomes increasingly important. Buildings are among the largest consumers of energy, and the integration of smart technologies like DRL into their management systems represents a critical step toward reducing this footprint. The scalability of the proposed DRL model means that it can be implemented across a wide range of building types, from residential to commercial spaces, and in various climatic conditions, making it a versatile tool for energy management. Additionally, the self-learning capabilities of DRL ensure that the system can continue to evolve and improve over time, adapting to new patterns and conditions without the need for constant manual adjustments.

The success of this DRL-based approach opens up numerous opportunities for further research and innovation in the field of smart building management. Future work could explore the integration of additional environmental factors, such as lighting, air quality, and occupancy behavior, into the control system, thereby broadening its scope and effectiveness. Moreover, advancements in sensor technology and data analytics could further enhance the system's predictive capabilities, allowing for even more precise and proactive management of building environments. As smart cities continue to develop, the role of intelligent, adaptive systems like the one proposed in this study will be increasingly critical in managing resources efficiently and ensuring the sustainability of urban living spaces. This research not only contributes to the growing body of knowledge in the field but also provides a practical, scalable solution that can be implemented to achieve significant energy savings and improved comfort in smart buildings, thereby supporting broader goals of environmental sustainability and quality of life improvement.

REFERENCES

1. Aziz A., Salama M., Hassanien A., and Hanafi S., "Artificial Immune System Inspired Intrusion Detection System using Genetic Algorithm," Informatica, vol. 36, no. 4, pp. 347-358, 2012.
2. Balachandran S., Dasgupta D., Nino F., and Garrett D., "A Framework for Evolving Multi- Shaped Detectors in Negative Selection," in Proceedings of IEEE Symposium on Foundations of Computational Intelligence, Honolulu, pp. 401-408, 2007.
3. Forrest S., Perelson A., Allen L., and Cherukuri R., "Self-Nonself Discrimination in A Computer," in Proceedings of IEEE Computer Society Symposium on Research In Security and Privacy, Oakland, pp. 202-212, 1994.

4. González F., Dasgupta D., and Gómez J., "The Effect of Binary Matching Rules in Negative Selection," in Proceedings of Genetic and Evolutionary Computation Conference, Chicago,

5. pp. 195-206, 2003.

6. Gonzalez F., Dasgupta D., and Niño L., "A Randomized Real-Valued Negative Selection Algorithm," in Proceedings of International Conference on Artificial Immune Systems, pp. 261-272, 2003.

7. Ji Z. and Dasgupta D., "Real-Valued Negative Selection Algorithm with Variable-Sized Detectors," in Proceedings of Genetic and Evolutionary Computation Conference, USA, pp. 287-298, 2004.

8. Ji Z. and Dasgupta D., "Applicability Issues of The Real-Valued Negative Selection Algorithms," in Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, Seattle, pp. 111-118, 2006.

9. Ji Z. and Dasgupta D., "Revisiting Negative Selection Algorithms," Evolutionary Computation, vol. 15, no. 2, pp. 223-251, 2007.

10. McHugh J., "Testing Intrusion Detection Systems: A Critique of The 1998 and 1999 Darpa Intrusion Detection System Evaluations as Performed by Lincoln Laboratory," ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262-294, 2000.

11. Powers S. and He J., "A Hybrid Artificial Immune System and Self-Organising Map for Network Intrusion Detection," Information Sciences, vol. 178, no. 15, pp. 3024-3042, 2008.

12. Poggiolini M. and Engelbrecht A., "Application of the Feature-Detection Rule to the Negative Selection Algorithm," Expert Systems with Applications, vol. 40, no. 8, pp. 3001-3014, 2013.

13. Sabri F., Norwawi., and Seman K., "Hybrid of Rough Set Theory and Artificial Immune Recognition System as A Solution to Decrease False Alarm Rate in Intrusion Detection System," in Proceedings of 7th International Conference on Information Assurance and Security, Melacca, pp. 134-138, 2011.

14. Seresht N. and Azmi R., "MAIS-IDS: A Distributed Intrusion Detection System Using Multi-Agent AIS Approach," Engineering Applications of Artificial Intelligence, vol. 35, pp. 86-298, 2014.

15. Stibor T., Timmis J., and Eckert C., "A Comparative Study of Real-Valued Negative Selection to Statistical Anomaly Detection Techniques," in Proceedings of International Conference on Artificial Immune Systems, Banff,pp. 262-275, 2005.

16. Tavallaee M., Bagheri E., Lu W., and Ghorbani A., "A Detailed Analysis of The KDD CUP 99 Data Set," in Proceedings of Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, pp. 1-6, 2009.