Chapter

Machine Learning Techniques in Indoor Environmental Quality Assessment

Mohan Kumar Gajendran, Ijaz Fazil Syed Ahmed Kabir, Sudhakar Vadivelu, Eddie Yin-Kwee Ng and Ravi Chandra Thota

Abstract

This chapter provides a comprehensive exploration of the evolving role of machine learning in Indoor Environmental Quality (IEQ) assessment. As urban living spaces become increasingly enclosed, the importance of maintaining optimal IEQ for human health and well-being has surged. Traditional methods for IEQ assessment, while effective, often fail to provide real-time monitoring and control. This gap is increasingly being addressed by the integration of machine learning techniques, allowing for enhanced predictive modeling, real-time optimization, and robust anomaly detection. The chapter delves into a comparative analysis of various machine learning techniques including supervised, unsupervised, and reinforcement learning, demonstrating their unique benefits in IEQ assessment. Practical implementations of these techniques in residential, commercial, and specialized environments are further illustrated through detailed case studies. The chapter also addresses the existing challenges in implementing machine learning for IEQ assessment and provides an outlook on future trends and potential research directions. The comprehensive review offered in this chapter encourages continued innovation and research in leveraging machine learning. for more efficient and effective IEQ assessment.

Keywords: indoor environmental quality, machine learning, IEQ assessment, predictive modeling, supervised learning, unsupervised learning, reinforcement learning, real-time control, anomaly detection, residential buildings, commercial buildings, future trends

1. Introduction

This chapter delves into the integration of Indoor Environmental Quality (IEQ) assessment with Machine Learning (ML), emphasizing the critical role of optimal IEQ in urban living spaces. It explores the impact of IEQ on occupant health, comfort, and productivity [1], and examines how ML can revolutionize IEQ assessment, offering significant advantages for building management, urban planning, and public health sectors.

IEQ encompasses the conditions within buildings that affect human health and comfort, including air quality, thermal comfort, acoustical comfort, and visual comfort [2]. These elements, as illustrated in **Figure 1**, significantly influence occupant well-being and productivity [4]. The scope of IEQ spans various building types, from residential to commercial and institutional spaces like schools and hospitals [5]. Monitoring of IEQ involves diverse methods, ranging from manual controls to automated systems, and adheres to standards set by organizations like the EPA and ASHRAE [6].

Accurate and efficient assessment of IEQ is essential for maintaining healthy indoor environments. It aids in optimizing HVAC systems for energy efficiency and comfort [7] and addresses health concerns such as Sick Building Syndrome (SBS) and Building-Related Illnesses (BRI) [8]. Traditional IEQ assessment methods often fall short in real-time monitoring and managing the dynamic nature of indoor environments [9].

Machine learning, a transformative branch of artificial intelligence, is increasingly applied in IEQ assessment. ML offers advanced capabilities in data processing, predictive modeling, pattern recognition, anomaly detection, and real-time adaptive control [10]. Through various techniques, such as supervised, unsupervised, and reinforcement learning, ML enhances the accuracy and efficiency of IEQ assessments [11].

IEQ is a complex, multifaceted concept integral to occupant well-being. It spans a wide range of parameters, from physical to biochemical and psychological aspects [12]. The trend towards energy-efficient building designs has intensified the need to focus on maintaining IEQ [7], which includes factors like air quality, lighting, thermal

Figure 1.
Factors Influencing IEQ [3].

conditions, and acoustics, as well as emerging considerations such as ergonomics, esthetics, and electromagnetic field quality, as defined by entities like ASHRAE, the EPA, and WHO [13–15].

Each aspect of IEQ, critically impacting human health and comfort, is depicted in **Figure 1**. For instance, poor air quality can lead to a range of health issues, from allergies to severe conditions like cancer [1]. Inadequate thermal conditions can diminish comfort and reduce productivity [2], while neglect in acoustical and visual comfort can result in stress, cognitive impairments, and long-term mental health issues [16, 17].

To illustrate further, air quality is influenced by various pollutants, both internal and external to the building, with effects ranging from minor discomfort to serious conditions like asthma and cancer [18]. Thermal comfort, a subjective but crucial metric, hinges on multiple environmental and personal factors [2]. Acoustical comfort, dependent on sound characteristics, is vital for mental well-being [16]. Visual comfort, shaped by lighting and visual elements, affects everything from eye strain to mood and circadian rhythms [17].

The integration of ML into IEQ assessment represents a significant advancement. ML's ability to analyze complex datasets and discern patterns enables a more nuanced understanding and control of indoor environments. This chapter aims to explore the applications of ML in enhancing various facets of IEQ, from air quality analysis to thermal comfort prediction. The integration of ML in IEQ assessment not only promises improved real-time monitoring and proactive management but also opens doors to personalized environmental controls tailored to individual needs and health requirements.

As we delve into the specifics of applying ML to IEQ, the chapter will address both the technical advancements and the challenges, including ethical considerations. Our goal is to provide a comprehensive overview of how ML can be leveraged to improve indoor environmental quality, ultimately contributing to the enhanced health, comfort, and well-being of occupants in urban living spaces.

2. IEQ assessment: Traditional methods and their limitations

The assessment of Indoor Environmental Quality (IEQ) plays a pivotal role in ensuring environments that support the health, comfort, and productivity of building occupants. While traditional methods of IEQ assessment have been the cornerstone of this field, they exhibit significant limitations, impacting their overall effectiveness and reliability. In this section, we will expand upon these traditional methods, underscoring their constraints and paving the way for the introduction of machine learning techniques as a more robust solution.

2.1 Traditional assessment methods

2.1.1 Spot measurements

- *Description*: This method involves taking discrete measurements of various IEQ parameters such as temperature, humidity, carbon dioxide levels, particulate matter, and light intensity at a specific moment using specialized instruments [19].

- *Advantages*: The primary benefits include simplicity in execution and the immediate availability of results, making it a convenient option for quick assessments.

- *Limitations*: However, spot measurements may not accurately represent ongoing exposure conditions, as they overlook temporal and spatial variability. The accuracy of these measurements is also heavily dependent on the calibration and precision of the instruments used [20].

2.1.2 Long-term monitoring

- *Description*: This approach utilizes sensors or data loggers to continuously record IEQ parameters over extended periods. It is designed to capture daily and seasonal variations, providing a more comprehensive view of the indoor environment [21].

- *Advantages*: Offers a broader and more detailed perspective on IEQ, capturing long-term trends and fluctuations.

- *Limitations*: The challenge lies in the significant resources required for data management and analysis. Additionally, there is a potential for data errors and inconsistencies due to sensor calibration issues or environmental interference [22].

2.1.3 Occupant surveys

- *Description*: This method involves collecting subjective responses from occupants about their perceptions and satisfaction with the indoor environment [23].

- *Advantages*: Occupant surveys can uncover issues that might be overlooked by physical measurements, providing valuable insights into the human aspect of IEQ.

- *Limitations*: However, these surveys are prone to response biases and may not accurately reflect the actual IEQ conditions. The qualitative nature of this data also makes it challenging to quantify or systematically analyze [24].

2.1.4 Physical inspections

- *Description*: Physical inspections involve thorough examinations of the building by professionals to identify visible or detectable issues such as water leaks, mold growth, or insufficient ventilation [25].

- *Advantages*: They are particularly effective in identifying overt and tangible problems within a building.

- *Limitations*: The limitation of physical inspections lies in their inability to detect latent or intermittent problems. They also require significant expertise and can be time-consuming and labor-intensive [26].

2.2 Limitations and the need for advanced methods

The traditional methods discussed above share a common drawback: their limited ability to provide real-time, continuous data. They often fail to capture the dynamic and interactive nature of various IEQ parameters, leading to an incomplete or skewed understanding of the indoor environment. Furthermore, these methods generally lack predictive capabilities, which are essential for proactive IEQ management. They can also be resource-intensive, both in terms of time and financial investment [27].

2.3 Emerging advancements and the role of machine learning

In response to these limitations, emerging advancements in sensor technology and data analytics, particularly machine learning (ML), present promising solutions. Machine learning techniques, with their capability to process vast amounts of data, enable continuous monitoring, real-time analysis, and predictive modeling of IEQ parameters. This paradigm shift aims not only to enhance the accuracy and comprehensiveness of IEQ assessments but also to streamline the process, reducing the demand on resources. The integration of ML in IEQ assessment represents a significant step towards more intelligent, adaptive, and occupant-centered building management.

However, it is important to note that the full-scale adoption of these advanced technologies is still in progress. Further research and development are necessary to establish their effectiveness, ease of use, and integration into existing building management systems. As the field of IEQ assessment evolves, the potential of machine learning to transform this domain is immense, offering a pathway towards more sustainable, healthy, and productive indoor environments [27, 28].

3. An overview of machine learning techniques

Machine learning, as a pivotal subfield of artificial intelligence, has seen notable evolution since its inception. This discipline, coined by Arthur Samuel in 1959, initially revolved around the idea of giving "computers the ability to learn without being explicitly programmed" [29]. Samuel's work with checkers-playing programs laid the groundwork for subsequent progress in this field. The following years saw pioneering efforts from researchers like Rosenblatt who, in 1958, introduced the concept of the perceptron, an early neural network that could classify linearly separable patterns [30]. At the same time, Nilsson and Minsky's explorations in learning machines added depth to the understanding of how computational models could replicate cognitive processes [31].

The 1970s and 1980s introduced the concepts of decision trees [32], a fundamental machine learning approach which paved the way for more sophisticated algorithms like Random Forests. This era also witnessed the emergence of the kernel methods, which includes the Support Vector Machine (SVM) - a powerful classification tool that marked a significant shift in the direction of machine learning [33]. The advent of the Backpropagation algorithm in the 1980s sparked a renewed interest in neural networks [34]. However, the full potential of this method wasn't realized until recent years, due to the limitation in computational capabilities and lack of large-scale, labeled datasets.

The 1990s saw a move towards data-driven approaches, prompted by the increasing availability of digital data. The introduction of ensemble methods such as boosting and bagging brought forth a new phase in machine learning, improving model robustness and predictive power [35]. The turn of the millennium marked the era of 'big data', pushing the boundaries of machine learning further. The development of more sophisticated algorithms, like deep learning, could now be propelled by the explosion of data and advancements in computational capabilities [36]. The seminal AlexNet model, developed by Krizhevsky, Sutskever, and Hinton, for the ImageNet competition in 2012, demonstrated the incredible potential of deep learning in practice, revolutionizing the field and driving a resurgence in neural networks [37].

Today, machine learning has permeated various sectors, revolutionizing healthcare, finance, environmental science, and many more fields. From a theoretical concept to a practical tool, machine learning has become an integral part of scientific research and technological development.

3.1 Fundamentals of machine learning

In this subsection, we will delve into the core concepts that form the foundation of machine learning, encompassing its definition, the nature of the data it utilizes, the problems it addresses, and the metrics used to evaluate its performance. Machine learning, a subset of artificial intelligence, enables computers to learn autonomously by training mathematical models on data, thereby going beyond the explicit programming used in traditional computing [38]. Unlike classical programming that relies on pre-defined rules, machine learning algorithms identify patterns within data, offering a unique capacity for experiential learning.

Data is essential for machine learning performance. It can vary in type—numerical, categorical, or a blend—and often requires preprocessing steps such as data cleaning and missing data imputation to ensure model quality [38]. Machine learning can address multiple problem types, including but not limited to regression, classification, clustering, anomaly detection, and recommendation systems [39].

Evaluating the effectiveness of a machine learning model involves various metrics. In classification tasks, commonly used metrics include accuracy, precision, recall, F1-score, and ROC-AUC [40]. For regression tasks, mean absolute error, mean squared error, and R-square are commonly applied [38]. The choice of evaluation metric is contingent on the specific problem and objectives, making a thorough understanding of these elements vital for developing robust machine learning models.

3.2 Relevance of machine learning in environmental data analysis

Environmental systems are inherently complex, characterized by high variability and unpredictability. Traditional statistical models often struggle to capture the nuances of these systems due to their linear nature and the assumption of independence among predictors. Machine learning, on the other hand, offers the flexibility to model non-linear relationships and consider interaction effects among multiple variables, making it adept at capturing the complexity inherent in environmental systems [41].

Machine learning's strength lies not just in its ability to model complex systems, but also in its ability to handle unstructured data. Unlike traditional models that require structured tabular data, machine learning algorithms can handle a variety of data types, including text, images, and sound. For instance, images from remote

DOI: http://dx.doi.org/10.5772/intechopen.114012

sensing can be used in land use classification or change detection tasks using convolutional neural networks, a type of deep learning algorithm [42]. Text from social media posts can be analyzed using natural language processing techniques to gauge public sentiment towards environmental issues or to detect early signs of natural disasters [43].

Time-series analysis is another critical aspect of environmental data analysis. Environmental data is often collected over time, leading to a temporal sequence of observations. Machine learning plays a crucial role in the analysis of these time-series data. Recurrent neural networks (RNN) and Long Short Term Memory (LSTM) networks, types of deep learning algorithms, have shown significant promise in modeling temporal dependencies in environmental data, such as weather forecasting and prediction of air quality indices [44].

Lastly, the importance of feature engineering cannot be understated in the context of environmental data analysis. Feature engineering is the process of creating new features or modifying existing ones to improve model performance. While machine learning algorithms can automatically learn features in some cases (particularly in deep learning), domain knowledge can often guide more effective feature creation. For instance, in the case of predicting rainfall, features such as the season, geographical location, historical weather patterns, and other atmospheric conditions might be crucial. Incorporating such features in the model could potentially improve the model's predictive performance [38, 45].

3.3 Supervised, unsupervised, and reinforcement learning: a comparative analysis

This subsection aims to delve deeper into the nuances of supervised, unsupervised, and reinforcement learning, providing a comparative analysis across multiple dimensions.

Supervised Learning: Supervised learning algorithms learn from labeled training data to predict outcomes for unforeseen data [46]. The concepts of bias and variance are central to understanding the performance of these algorithms [47]. Bias refers to the error due to the model's assumptions in the learning algorithm, while variance pertains to the error from the model's sensitivity to fluctuations in the training set. An optimal balance between the two is crucial to avoid overfitting (high variance) and underfitting (high bias), both of which harm the model's predictive accuracy on new data. This balance is often achieved through methods such as cross-validation and hyperparameter tuning. Ensemble methods, including bootstrapping and bagging, and boosting methods, aim to enhance model performance by combining several weak learners to form a stronger overall model [38].

Unsupervised Learning: Unlike supervised learning, unsupervised learning algorithms discern patterns in data without pre-existing labels, primarily through clustering and dimensionality reduction techniques [48]. Concepts such as data density and similarity measures are instrumental in these techniques, and specific algorithms, such as hierarchical clustering and DBSCAN, utilize these concepts [49]. Handling outliers, determining the appropriate number of clusters (e.g., via the elbow method or silhouette score), are other critical aspects of unsupervised learning [50].

Reinforcement Learning: In reinforcement learning, an agent learns to perform actions in an environment to maximize a reward through the process of trial and error [51]. This learning paradigm is defined in terms of a Markov decision process, where the agent's actions in a certain state determine the probabilities of transitioning to other states [52]. The exploration versus exploitation dilemma is central to reinforcement learning, guiding the balance between trying new actions and exploiting known ones [53]. Techniques like Q-learning, SARSA, policy gradient methods, and more recent advances in deep reinforcement learning, provide frameworks to navigate this space [54].

Comparative Analysis: The three learning paradigms vary significantly in terms of data requirements, computational resources, interpretability, susceptibility to noise and outliers, and the level of human supervision required. For instance, supervised learning often requires large labeled datasets, which can be resource-intensive to collect, while unsupervised and reinforcement learning can work with unlabeled data. Supervised models are generally more interpretable than unsupervised and reinforcement learning models, making them suitable for applications where interpretability is a concern [55]. However, each of these learning paradigms has its strengths and weaknesses, and their appropriate usage largely depends on the specific problem at hand.

Understanding the subtleties of these paradigms is crucial for their effective application in environmental data analysis, given the complexity and high dimensionality of the data typically involved in this field [56].

4. Machine learning for enhanced IEQ assessment

In this section, we explore the practical applications of machine learning in assessing and enhancing Indoor Environmental Quality (IEQ) across a range of settings—residential, commercial, and specialized environments. We will examine how these algorithms contribute to optimized living conditions in homes, facilitate real-time environmental adjustments in commercial buildings, and ensure stringent IEQ compliance in specialized settings like healthcare facilities and data centers. Through these case studies, the section aims to provide an empirical substantiation of the machine learning methodologies discussed earlier, demonstrating their transformative potential in diverse IEQ contexts.

4.1 Residential buildings: Improving IEQ with machine learning

Machine learning has become a transformative force in improving Indoor Environmental Quality (IEQ) in residential buildings, offering significant advancements in Indoor Air Quality (IAQ), energy management, and occupancy detection.

Machine learning algorithms such as stochastic gradient boosting and support vector machines have demonstrated effectiveness in the domain of IAQ. These algorithms play a pivotal role in offering accurate predictions concerning room occupancy based on air quality indicators [57, 58]. Moreover, they are capable of efficiently classifying factors affecting IAQ under various conditions [59]. Other innovative approaches, including decision trees and hidden Markov models, have been applied to account for both real-time and anticipated occupancy states. These models have shown the capability to rectify sensor errors, thereby ensuring the reliability of IAQ parameters [60, 61].

Focusing on the challenges of sensor failures in pollutant monitoring, a significant study evaluated three Machine Learning (ML) algorithms – Multi-layer Perceptron (MLP), K-Nearest Neighbor (KNN), and Random Forest (RF) – for their effectiveness in classifying sensor readings as faulty or not. These algorithms demonstrated their superiority over standard statistical methods by creating better separation boundaries and utilizing contextual information. Numeric results from a 20-fold cross-validation displayed high average Area Under Curve (AUC) scores for each pollutant: 0.96 for MLP, 0.97 for KNN, and 0.97 for RF, indicating their robust performance in detecting faulty sensor readings [62].

Another key aspect of machine learning in IEQ is its application in energy efficiency. Studies have validated sensor data using techniques like artificial neural networks (ANNs) and Bayesian Networks, particularly focusing on U.S. commercial and residential buildings. Analyzing data from various sensors, these studies found that the root-mean-square error (RMSE) was generally less than 10% of the sensor's mean value for most sensors. However, the energy sensor data showed higher RMSEs, often exceeding 200%. Bayesian Networks, despite requiring longer training times, yielded lower errors compared to ANNs. Comparative RMSEs for ANNs and Bayesian Networks were 17.7% vs. 13.5% for liquid pressure, 5.0% vs. 2.1% for humidity, 1.9% vs. 1.1% for temperature, 0.03 vs. 0.02 for water flow, and 0.89 vs. 0.85 for energy consumption, highlighting the potential of these methods in enhancing sensor accuracy and energy efficiency [63].

In the field of residential energy management, machine learning, particularly Random Forest algorithms, has shown impressive efficacy in predicting home occupancy based on thermostat data. This was evident in a study evaluating various machine learning models, including heuristic baselines, traditional classifiers, and sequential models like recurrent neural networks (RNNs), for their ability to predict home occupancy. The Random Forest algorithm was noted for its high accuracy across different time horizons and efficient training for individual edge devices. Key numeric results included training times of 0.38 seconds for the Random Forest and 10 seconds for the RNN, with inference times of 0.001 seconds and 0.01 seconds, respectively, underscoring the potential of machine learning in adaptive thermal control in residential buildings [64, 65].

Further advancements in residential IEQ have been made using ensemble techniques and ANNs for the precise prediction of heating and cooling loads, especially in air-conditioned residential buildings in warm climates. One particular study focused on reducing energy consumption through energy-efficient building design, employing machine learning methods to estimate heating and cooling loads based on physical building characteristics. The performance of these methods was assessed using metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), with significant improvements noted in cooling load prediction. The lowest MAE was 0.390, and the highest R^2 value was 0.996, indicating the model's high accuracy and efficiency [66, 67].

The integration of machine learning with smart home technologies has opened new avenues in IEQ management. For example, machine learning algorithms, used alongside IoT sensors, have enabled real-time monitoring and predictive analysis of indoor air pollutants and noise levels. Techniques such as fuzzy logic have been utilized to assess the efficacy of HVAC systems. These multifaceted applications of machine learning provide a robust foundation for enhancing IEQ in residential buildings and pave the way for a holistic approach to residential IEQ, marking a strong.

4.2 Commercial buildings: Real-time IEQ control with machine learning

Managing Indoor Environmental Quality (IEQ) in commercial buildings presents unique challenges due to varied occupancy patterns and the multifunctional nature of these spaces. The deployment of machine learning techniques has emerged as a key solution, offering real-time adaptability and predictive power for IEQ management.

Machine learning's versatility in real-time monitoring, classification, and predictive modeling is evident from studies that utilized algorithms like decision tree, random forest, and Support Vector Machine (SVM) for air quality and noise level predictions. Notably, SVM showed significant effectiveness in air quality prediction, with its accuracy improving from 75–95% after data quality adjustments. In contrast, decision tree and random forest algorithms achieved accuracies of 76% and 82%, respectively. For noise data classification, SVM achieved a remarkable 98% accuracy, consistently outperforming other models over a six-year evaluation period. The critical role of data quality and normalization in improving prediction accuracy was highlighted, with CO and SO2 sensors achieving high accuracies, while PM2.5 sensor accuracy was comparatively lower [68].

Despite these advancements, gaps in the literature are evident. Some studies have developed forecast models for CO2 buildup without elaborating on real-time IEQ monitoring [69]. Others have explored innovative areas like integrating machine learning with EEG signals for predicting IEQ conditions based on occupants' brainwave patterns [70]. Additional studies employed decision trees, hidden Markov models, and stochastic gradient boosting algorithms for immediate and future state occupancy prediction, demonstrating machine learning's potential for real-time IEQ control [57, 60]. The advantages of machine learning over traditional statistical approaches were also noted in IAQ sensor reading classification [62].

Specialized applications in commercial settings have seen the use of neural network modeling for identifying contaminant source positions and Sparse Spectrum Gaussian Process Regression (GPR) for real-time air quality prediction. These applications highlight machine learning's adaptability to various commercial environmental challenges [71, 72]. In educational and enterprise environments, studies have focused on ventilation rate control using IoT protocols and employing machine learning for air quality forecasting, further underscoring the diverse applicability of these technologies [73, 74].

Empirical evaluations often favor neural network-based models, using metrics such as RMSE, R2 score, and error rate. The ongoing advancements in these techniques, particularly in algorithm development and sensor calibration, promise enhanced precision and efficiency in monitoring indoor pollutants [75].

While machine learning offers promising opportunities for real-time IEQ control in commercial buildings, further research is needed for its effective implementation. Future research directions may explore the integration of machine learning with biometric signals for personalized environmental control [76].

4.3 Specialized environments: Strict IEQ regulation through machine learning

The challenge of maintaining optimal Indoor Environmental Quality (IEQ) in specialized environments, notably hospitals, has been rigorously researched. In such contexts, regulating IEQ is critical, impacting patient health, recovery, and satisfaction. Recognizing the limitations of conventional methods, researchers are increasingly turning to machine learning algorithms for more advanced computational approaches to these complexities.

Machine learning's application in hospitals has been pivotal in enhancing patient satisfaction, particularly concerning IEQ. A significant study at King Abdullah University Hospital (KAUH) in Jordan utilized Support Vector Machines (SVM) with a linear kernel and K-Nearest Neighbors (K-NN) to predict patient satisfaction. The SVM model, with a configuration of C = 0.01, proved more effective than other SVM kernels and K-NN. This effectiveness is evidenced by SVM's Pearson R-square value of 0.8948 with a P-value of .0001, surpassing K-NN's Pearson R-square of 0.2984. The study involved gathering self-reported data and field monitoring of environmental indicators within patient rooms at KAUH, using the same dataset for both training and testing. These outcomes highlight machine learning's role in providing insights that impact healthcare outcomes, particularly in enhancing IEQ to improve patient well-being [77].

Data collection in these environments often involves a mix of methodologies, including self-reported data, field monitoring, and sensor networks, creating diverse datasets for machine learning training and validation. For example, Elnaklah et al. combined sensor-generated data and human-reported satisfaction metrics to amass a comprehensive dataset [78].

Several machine learning models have shown effectiveness in regulating IEQ in hospital settings. An Autoregressive Hidden Markov Model (ARHMM) developed for a laboratory equipped with a sensor network demonstrated an average estimation accuracy of 80.78% in predicting occupancy patterns. This model, utilizing data from passive infra-red (PIR) sensors, CO2 concentration sensors, and relative humidity (RH) sensors, was benchmarked against classical Hidden Markov Models (HMM) and Support Vector Machines (SVM).

Bayesian networks have also been used to assess the risk of symptomatological complaints related to poor IEQ in Intensive Care Units (ICUs). A study in nine ICUs in João Pessoa, Brazil, analyzed temperature, noise, lighting, and air quality data, along with professional interviews. The Bayesian model revealed a 42.2% probability of physical symptoms and a 45.3% probability of psychological symptoms from environmental discomfort among ICU employees, with environmental temperature identified as a significant impacting factor [79, 80]. These probabilistic graphical models provide insightful perspectives on potential risk areas in IEQ within hospital settings.

Notably, there are gaps in existing literature, with several studies investigating IEQ through questionnaires and field studies without employing machine learning. For example, a study examining apartment and office buildings did not leverage machine learning, representing a missed opportunity for data-driven insights [81]. This highlights the significant potential for integrating machine learning techniques in future studies to provide more accurate, actionable recommendations for IEQ regulation in hospitals.

5. Conclusions

The chapter has provided an in-depth exploration of the integration of Machine Learning (ML) techniques into Indoor Environmental Quality (IEQ) assessment. IEQ, an aspect critical to human well-being, is intricately linked with numerous parameters including air quality, thermal, acoustical, and visual comfort. The chapter delineated the limitations of traditional methods in IEQ assessment, especially the lack of real-time data and predictive capabilities, and suggested machine learning as a potent solution for augmenting both the scope and accuracy of such evaluations. The implications are far-reaching, impacting residential buildings, commercial structures, and specialized environments such as healthcare facilities.

While machine learning appears to offer a robust approach to managing and enhancing IEQ, there are questions that remain unanswered. For instance, the ethical considerations of data collection and usage, particularly in specialized environments like healthcare facilities, warrant attention. Similarly, the interpretability of complex machine learning models and the implications for regulatory compliance are subjects requiring more in-depth study.

5.1 Future research directions

Data Ethics and Privacy: As machine learning increasingly intersects with IEQ assessment, ethical considerations around data collection and use will become more pressing. Future studies could focus on the ethical guidelines that ought to govern this intersection.

Interpretable Machine Learning Models: Future work could concentrate on developing more interpretable models without sacrificing predictive power, which is crucial for gaining regulatory approval and social acceptance.

Adaptive Systems: Research could delve into the development of adaptive machine learning systems that can evolve in real-time to meet the dynamic nature of IEQ elements, thereby offering more precise control mechanisms.

Personalization: Given the inter-individual differences in comfort and health responses, studies focusing on personalized IEQ assessment and control algorithms could offer a more nuanced approach to managing indoor environments.

Longitudinal Studies: To robustly assess the impact of machine learning on IEQ over time, future research could employ longitudinal study designs, perhaps integrated with natural experiments in real-world settings.

Cross-Sector Collaboration: As machine learning and IEQ assessment are topics with multidisciplinary ramifications, partnerships between AI researchers, environmental scientists, healthcare professionals, and policy-makers are essential for holistic solutions.

Resource Optimization: Machine learning algorithms capable of balancing both IEQ and energy efficiency would be a significant advancement, addressing the practical constraints of implementing ML in building management systems.

Global Standards: The development and adoption of global standards for machine learning in IEQ assessment could further facilitate interoperability and effectiveness across diverse building types and geographical locations.

Author details

Mohan Kumar Gajendran[1†], Ijaz Fazil Syed Ahmed Kabir[2†], Sudhakar Vadivelu[3], Eddie Yin-Kwee Ng[2*] and Ravi Chandra Thota[4]

1 Division of Energy, Matter and Systems, University of Missouri-Kansas City, Kansas City, MO, USA

2 School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore

3 Department of Aerospace Engineering, Toronto Metropolitan University (TMU), Toronto, ON, Canada

4 Department of Computer Science, University of Missouri-Kansas City, Kansas City, MO, USA

*Address all correspondence to: mykng@ntu.edu.sg

† These authors contributed equally.

References

[1] Mujan I, Anelković AS, Munćan V, Kljajić M, Ružić D. Influence of indoor environmental quality on human health and productivity – A review. Journal of Cleaner Production. 2019;**217**:646-657 ISSN 0959-6526

[2] Ma N, Aviv D, Guo H, Braham WW. Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. Renewable and Sustainable Energy Reviews. 2021;**135**: 110436 ISSN 1364-0321

[3] The Atmospheric Fund. Improving Indoor Environmental Quality in Multi-Unit Residential Buildings. Toronto, ON: The Atmospheric Fund; 2019. Available from: https://taf.ca/custom/uploads/2019/05/Improving-Indoor-Environmental-Quality-in-Multi-Unit-Residential-Buildings-May-2019.pdf. [Accessed: December 23, 2023]

[4] Alhorr Y, Arif M, Katafygiotou M, Mazroei A, Kaushik A, Elsarrag E. Impact of indoor environmental quality on occupant well-being and comfort: A review of the literature. International Journal of Sustainable Built Environment. 2016;**5**(1):1-11 ISSN 2212-6090

[5] Mannan M, Al-Ghamdi SG. Indoor air quality in buildings: A comprehensive review on the factors influencing air pollution in residential and commercial structure. International Journal of Environmental Research and Public Health. 2021;**18**(6):3276. ISSN 1660-4601

[6] Kim H. Methodology for Rating a Building's Overall Performance Based on the ASHRAE/CIBSE/USGBC Performance Measurement Protocols for Commercial Buildings. Texas, USA: Texas A&M University; 2012

[7] Pedersen E, Borell J, Li Y, Stålne K. Good indoor environmental quality (ieq) and high energy efficiency in multifamily dwellings: How do tenants view the conditions needed to achieve both? Building and Environment. 2021; **191**:107581 ISSN 0360-1323

[8] Kishi R, Norbäck D, Araki A. Indoor Environmental Quality and Health Risk toward Healthier Environment for All. Vol. 1. Singapore: Springer;

[9] Persily A. Challenges in developing ventilation and indoor air quality standards: The story of ashrae standard 62. Building and Environment. 2015;**91**: 61-69 ISSN 0360-1323. Fifty Year Anniversary for Building and Environment

[10] Tien PW, Wei S, Darkwa J, Wood C, Calautit JK. Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality – A review. Energy and AI. 2022;**10**:100198 ISSN 2666-5468

[11] Fard ZQ, Zomorodian ZS, Korsavi SS. Application of machine learning in thermal comfort studies: A review of methods, performance and challenges. Energy and Buildings. 2022; **256**:111771 ISSN 0378-7788

[12] Mujeebu MA. Indoor Environmental Quality. London, UK, Rijeka: IntechOpen; 2019 ISBN 978-1-78985-252-3

[13] Laue J. Ashrae 62.1: Using the Ventilation Rate Procedure. 2018. Available from: https://www.csemag.com/articles/ashrae-62-1-using-the-ventilation-rate-procedure/ [Accessed: July 29, 2023]

[14] U.S. Environmental Protection Agency. Indoor Air Quality (iaq). Available from: https://www.epa.gov/indoor-air-quality-iaq [Accessed: July 29, 2023]

[15] World Health Organization. Environmental Health. Available from: https://www.who.int/europe/health-topics/environmental-health#tab=tab_1 [Accessed: July 29, 2023]

[16] Mujan I, Licina D, Kljajić M, Čulić A, Anelković AS. Development of indoor environmental quality index using a low-cost monitoring platform. Journal of Cleaner Production. 2021;**312**:127846 ISSN 0959-6526

[17] Kaushik A, Arif M, Ebohon OJ, Arsalan H, Rana MQ, Obi L. Effect of indoor environmental quality on visual comfort and productivity in office buildings. Journal of Engineering, Design and Technology, ahead-of-print(ahead-of-print). 2021;**21**:1746-1766

[18] Sarbu I, Sebarchievici C. Aspects of indoor environmental quality assessment in buildings. Energy and Buildings. 2013; **60**:410-419 ISSN 0378-7788

[19] Choi J-H, Loftness V, Aziz A. Post-occupancy evaluation of 20 office buildings as basis for future ieq standards and guidelines. Energy and Buildings. 2012;**46**:167-175 ISSN 0378-7788. Sustainable and healthy buildings

[20] Kim H et al. Field-test of the new ashrae/cibse/usgbc performance measurement protocols: Intermediate and advanced level indoor environmental quality protocols. ASHRAE Transactions. 2012;**118**:58

[21] Yao H, Cheng X, Shen W, Lv Y, Li A, Shen X. Sampling method for long-term monitoring of indoor environmental quality in residential buildings. Building and Environment. 2022;**215**:108965 ISSN 0360-1323

[22] Loomans MGLC, Mishra AK, Kooi L. Long-term monitoring for indoor climate assessment – The association between objective and subjective data. Building and Environment. 2020;**179**: 106978 ISSN 0360-1323.

[23] Roa CD, Schiavon S, Parkinson T. Targeted occupant surveys: A novel method to effectively relate occupant feedback with environmental conditions. Building and Environment. 2020;**184**: 107129 ISSN 0360-1323.

[24] Rasheed EO, Byrd H. Can self-evaluation measure the effect of ieq on productivity? A review of literature. Facilities. 2017;**35**(11/12):601-621

[25] Lai ACK, Mui KW, Wong LT, Law LY. An evaluation model for indoor environmental quality (ieq) acceptance in residential buildings. Energy and Buildings. 2009;**41**(9):930-936 ISSN 0378-7788

[26] Faqih F, Zayed T. Defect-based building condition assessment. Building and Environment. 2021;**191**:107575 ISSN 0360-1323

[27] Licina D, Wargocki P, Pyke C, Altomonte S. The future of IEQ in green building certifications. Buildings and Cities. London, United Kingdom: Ubiquity Press; 2021;**2**(1):907-927. DOI: 10.5334/bc.148

[28] Broday EE, Gameiro MC, da Silva. The role of internet of things (iot) in the assessment and communication of indoor environmental quality (ieq) in buildings: A review. Smart and Sustainable Built Environment. 2023; **12**(3):584-606

[29] Samuel AL. Some studies in machine learning using the game of checkers. IBM

Journal of Research and Development. 1959;**3**(3):210-229

[30] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review. 1958;**65**(6):386

[31] Nilsson NJ. Artificial intelligence prepares for 2001. AI Magazine. 1983; **4**(4):7-7

[32] Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. Oxfordshire, United Kingdom: Taylor & Francis;1984. pp. 1-358. ISBN: 9780412048418. Available from: https://books.google.com/books?id=JwQx-WOmSyQC

[33] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;**20**: 273-297

[34] Hinton GE et al. Learning distributed representations of concepts. In: Proceedings of the Eighth Annual Conference of the Cognitive Science Society. Vol. volume 1. Amherst, MA: Taylor & Francis; 1986. p. 12

[35] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences. 1997;**55**(1):119-139

[36] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;**521**(7553): 436-444

[37] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems. 2012;**25**:84-90

[38] Gajendran MK, Rohowetz LJ, Koulen P, Mehdizadeh A. Novel machine-learning based framework using electroretinography data for the detection of early-stage glaucoma. Frontiers in Neuroscience. 2022;**16**: 869137. ISSN 1662-453X

[39] Gajendran MK, Khan MZ, Khan MA, Khattak. Ecg classification using deep transfer learning. In: In 2021 4th International Conference on Information and Computer Technologies (ICICT). Hawaii, USA: IEEE; 2021. pp. 1-5

[40] Khan MZ, Gajendran MK, Lee Y, Khan MA. Deep neural architectures for medical image semantic segmentation: Review. IEEE Access. 2021;**9**: 83002-83024

[41] Habibi S. Smart innovation systems for indoor environmental quality (ieq). Journal of Building Engineering. 2016;**8**: 1-13 ISSN 2352-7102

[42] Parkinson T, Parkinson A, de Dear R. Continuous ieq monitoring system: Context and development. Building and Environment. 2019;**149**: 15-25 ISSN 0360-1323

[43] Finch KC, Snook KR, Duke CH, King-Wa F, Tse ZTH, Adhikari A, et al. Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. Natural Hazards. 2016;**83**(1):729-760

[44] Qi H, Xiao S, Shi R, Ward MP, Chen Y, Wei T, et al. Covid-19 transmission in mainland China is associated with temperature and humidity: A time-series analysis. Science of the Total Environment. 2020;**728**: 138778 ISSN 0048-9697

[45] Barandas M, Folgado D, Fernandes L, Santos S, Abreu M, Bota P, et al. Tsfel: Time series feature extraction library. SoftwareX. 2020;**11**:100456 ISSN 2352-7110

[46] Cunningham P, Cord M, Delany SJ. Supervised Learning. Berlin Heidelberg, Berlin, Heidelberg: Springer; 2008. pp. 21-49 ISBN 978-3-540-75171-7

[47] Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences. 2019;**116**(32): 15849-15854

[48] Ghahramani Z. Unsupervised Learning. Berlin Heidelberg, Berlin, Heidelberg: Springer; 2004. pp. 72-112 ISBN 978-3-540-28650-9.

[49] Saxena A, Prasad M, Gupta A, Bharill N, et al. A review of clustering techniques and developments. Neurocomputing. 2017;**267**:664-681 ISSN 0925-2312

[50] Askari S. Fuzzy c-means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. Expert Systems with Applications. 2021; **165**:113856 ISSN 0957-4174

[51] Sutton RS, Barto AG. Reinforcement Learning: An Introduction. Massachusetts, London, England: MIT Press; 2018

[52] Puterman ML. Chapter 8 markov decision processes. In: Stochastic Models, volume 2 of Handbooks in Operations Research and Management Science. Amsterdam, Netherlands: Elsevier; 1990. pp. 331-434

[53] Jung D, Choi Y. Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. Minerals. 2021;**11**(2):148. ISSN 2075-163X

[54] Gupta S, Singal G, Garg D. Deep reinforcement learning techniques in diversified domains: A survey. Archives of Computational Methods in Engineering. 2021;**28**(7):4715-4754

[55] Casert C, Vieijra T, Nys J, Ryckebusch J. Interpretable machine learning for inferring the phase boundaries in a nonequilibrium system. Physical Review E. 2019;**99**(2):023304

[56] Zhong S, Zhang K, Bagheri M, Burken JG, April G, Li B, et al. Machine learning: New ideas and tools in environmental science and engineering. Environmental Science & Technology. 2021;**55**(19):12741-12754

[57] Dela Cruz JC, Amado TM, Hermogino JQ, Andog MLC, Corpuz FT, Ng JRT, et al. Machine Learning-Based Indoor Air Quality Baseline Study of the Offices and Laboratories of the Northwest and Southwest Building of Mapúa University – Manila. In: 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). Shah Alam, Malaysia: IEEE; 2020

[58] Fritz H, Tang M, Kinney K, Nagy Z. Evaluating machine learning models to classify occupants' perceptions of their indoor environment and sleep quality from indoor air quality. Journal of the Air amp; Waste Management Association. 2022;**72**(12):1381-1397

[59] Saad SM, Shakaff AYM, Saad ARM, Yusof AM, Andrew AM, Zakaria A, et al. Analysis of feature selection with Probabilistic Neural Network (PNN) to classify sources influencing indoor air quality. In: AIP Conference Proceedings, Author(s). Penang, Malaysia: American Institute of Physics (AIP): 2017

[60] Ryu SH, Moon HJ. Development of an occupancy prediction model using indoor environmental data based on machine learning techniques. Building and Environment. 2016;**107**:1-9

[61] Deepa R, Raj KM, Balaji N, Durgadevi K. Machine learning based estimation of room occupancy using non-intrusive sensors. In: 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT). Chennai, India: IEEE; 2022

[62] Seibert V, Araújo R, McElligott R. Sensor Validation for Indoor Air Quality using Machine Learning. In: Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2020). Pelotas, RS, Brazil: Sociedade Brasileira de Computação - SBC; 2020

[63] Smith MK, Castello CC, New JR. Machine learning techniques applied to sensor data correction in building technologies. In: 2013 12th International Conference on Machine Learning and Applications. Miami, FL, USA: IEEE; 2013

[64] Huchuk B, Sanner S, O'Brien W. Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. Building and Environment. 2019; **160**:106177

[65] Sadeghi A, Younes SR, Young WA, Weckman GR. An intelligent model to predict energy performances of residential buildings based on deep neural networks. Energies. Basel, Switzerland: MDPI; 2020;**13**(3):571. Availbale from: https://www.mdpi.com/1996-1073/13/3/571. ISSN: 1996-1073. DOI: 10.3390/en13030571

[66] Goyal M, Pandey M. A systematic analysis for energy performance predictions in residential buildings using ensemble learning. Arabian Journal for Science and Engineering. 2020;**46**(4): 3155-3168

[67] Asif S, Saba T. Machine learning for exploration of the factors affecting the building energy usage: A case study. Applied and Computational Engineering. 2023;**2**(1):694-699

[68] Shah SK, Tariq Z, Lee J, Lee Y. Real-time machine learning for air quality and environmental noise detection. In: In 2020 IEEE International Conference on Big Data (Big Data). Atlanta, GA, USA: IEEE; 2020

[69] Rastogi K, Lohani D. An Internet of Things Framework to Forecast Indoor Air Quality Using Machine Learning. Singapore: Springer; 2020. pp. 90-104

[70] Kim J, Kim H, Hong T. Automated classification of indoor environmental quality control using stacked ensembles based on electroencephalograms. Computer-Aided Civil and Infrastructure Engineering. 2020;**35**(5): 448-464

[71] Vukovic V, Tabares-Velasco PC, Srebric J. Real-time identification of indoor pollutant source positions based on neural network locator of contaminant sources and optimized sensor networks. Journal of the Air amp; Waste Management Association. 2010; **60**(9):1034-1048

[72] Arun G, Rathi S. Real Time Air Quality Evaluation Model using Machine Learning Approach. Journal of Information Technology and Digital World. Coimbatore, India: Inventive Research Organization; 2022;**4**(1):23-33. DOI: 10.36548/jitdw.2022.1.003

[73] Tagliabue LC, Cecconi FR, Rinaldi S, Ciribini ALC. Data driven indoor air quality prediction in educational facilities based on IoT network. Energy and Buildings. 2021;**236**:110782

[74] Iskandaryan D, Ramos F, Trilles S. Air quality prediction in smart cities

using machine learning technologies based on sensor data: A review. Applied Sciences. 2020;**10**(7):2401

[75] Saini J, Dutta M, Marques G. Indoor air quality prediction systems for smart environments: A systematic review. Journal of Ambient Intelligence and Smart Environments. 2020;**12**(5): 433-453

[76] Maaloul K, Brahim L. Comparative analysis of machine learning for predicting air quality in smart cities. WSEAS Transactions on Computers. 2022;**21**:248-256

[77] Wedyan M, Ali H, Abdullah M. Toward predicting Patient's satisfaction of indoor environmental quality in Jordanian hospitals using SVM and K-NN machine learning algorithms. In: 2022 13th International Conference on Information and Communication Systems (ICICS). Irbid, Jordan: IEEE; 2022

[78] Elnaklah R, Fosas D, Natarajan S. Indoor environment quality and work performance in "green" office buildings in the Middle East. Building Simulation. Berlin, Germany: Springer Nature; 2020; **13**(5):1043-1062. DOI: 10.1007/s12273-020-0695-1. ISSN: 1996-8744

[79] Han Z, Gao RX, Fan Z. Occupancy and indoor environment quality sensing for smart buildings. In: 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings. Graz, Austria: IEEE; 2012

[80] de Araújo Vieira EM, da Silva LB, de Souza EL. The influence of the workplace indoor environmental quality on the incidence of psychological and physical symptoms in intensive care units. Building and Environment. 2016; **109**:12-24

[81] Roulet C-A, Bluyssen PhM, Cox C, Foradini F. Relations between perceived indoor environment characteristics and well being of occupants at individual level. In: Healthy Buildings 2006. Lisbon, Portugal: LESO-PB, EPFL; 2006. Available from: http://infoscience.epfl.ch/record/78593