DFBI 2025 aims to encourage the international exchange of innovative ideas between researchers from academia and industry. In addition to knowledge dissemination, the conference offers a valuable platform for professional networking, particularly benefiting university professors, graduate students, and postdoctoral researchers.

Review Article

Data-driven approaches to improve Indoor Environmental Quality: A Systematic Literature Review

Kushani Semasinghe[1], Srinath Perera[1], Samudaya Nanayakkara[1], Xiaohua (Sean) Jin[1], Marini Samaratunga[1]

Centre for Smart Modern Construction, School of Engineering, Design, and Built Environment, Western Sydney University, Locked Bag 1797, Penrith NSW 2751, Australia.
Correspondence: k.semasinghe@westernsydney.edu.au

Copyright: Copyright: © 2025 by the authors.

DFBI is an open-access proceedings distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). View this license's legal deed at https://creativecommons.org/licenses/by/4.0/

Abstract

Indoor environmental quality (IEQ) refers to the conditions within a building and is a key factor in occupant comfort. To improve these IEQ, heating, ventilation, and air-conditioning (HVAC) systems are currently being used in indoor environments in buildings. However, inefficient systems to improve IEQ lead to high energy consumption, which is directly connected to building-energy-related CO_2 emissions. Unfortunately, these emissions are inextricably linked to global climatic issues. Data-driven technologies show a positive future in solving climate-related targets. Technologies such as artificial intelligence (AI), machine learning (ML), and digital twin (DT) are being explored to improve energy performance while improving IEQ. Therefore, this paper aims to systematically review the applications of data-driven technologies for improving IEQ and energy efficiency. A comprehensive literature search using the PRISMA method to retrieve publications related to data-driven and IEQ. The findings highlight common areas of data-driven applications in IEQ, including the detection and prediction of IEQ factors, IEQ controls, building energy management, and HVAC system controls. Commonly used ML and AI techniques identified, including deep learning, neural networks, KNN, SVM, decision trees, and multi-objective algorithms. AI in combination with BIM and IoT techniques was employed in the studies to develop DT models for real-time building monitoring. Further, the review highlights barriers to data-driven and digitalisation approaches in buildings that impede the robust approaches to IEQ improvement.

Keywords: Indoor environmental quality, energy efficiency, machine learning, artificial intelligence, digital twin

Highlights

- 152 publications were selected for final analysis.
- Data-Driven technologies including Artificial Intelligence, Machine Learning, and Digital Twin technologies for IEQ and Energy Efficiency
- Key applications include real-time IEQ monitoring, HVAC system control, and energy management.

Cite this article: Semasinghe, K., Perera, S., Nanayakkara, S., Jin, X., & Samaratunga, M., (2025). Data-driven approaches to improve Indoor Environmental Quality: A Systematic Literature Review. Proceedings of Digital Frontiers in Buildings and Infrastructure International Conference Series, Volume 2025, Page range: 121-135

1. Introduction

Most people live in urban areas and work in an office environment (Al Horr et al., 2016), and people spend up to 90% of their lives indoors (Baek, Park, Park, Le, & Chang, 2023). Given this people's lifetime indoors, providing a comfortable, healthy, and affordable indoor environment is crucial for a standard habitable indoor environment (R. Zhao, Sun, & Ding, 2004). To achieve a high level of indoor environment quality (IEQ), previous studies have identified that the human perception of the indoor environment is affected by four fundamental environmental factors: air, acoustics, visual, and thermal environment (Sam Kubba, 2017). Indoor air quality (IAQ) is usually affected by poor air quality, poor lighting quality, moulds and bacteria growth inside the building (S. Kubba, 2010), which can cause severe health conditions such as allergies, respiratory problems, and cognitive function reduction (S. Kubba, 2010). Acoustic quality usually refers to a condition without low or high noise levels or discomfort caused by sound irritation (Vardaxis, Bard, & Persson Waye, 2018). The visual quality depends on factors such as the quality of the indoor environment lighting, visual privacy, outdoor connections, and views (Tekce, Artan, & Ergen, 2021). Thermal environmental quality refers to the thermal comfort level. Extreme thermal environment conditions, such as too cold or too hot, are usually identified as discomfort thermal conditions (Q. Zhao, Lian, & Lai, 2021).

Focusing on these fundamental factors, heating, ventilation, and air conditioning (HVAC) systems, lighting systems, acoustic insulation systems and systems to improve visual comfort are used to maintain IEQ in buildings. These systems are responsible for maintaining a comfortable indoor environment. However, despite maintaining a comfortable indoor environment, these systems heavily influence the energy consumption of buildings. Previous studies highlighted those inefficient systems as the core reason for the significant energy consumption (H. H. Hosamo, 2023), which leads to energy-related climatic issues (United Nations, 2022). Building construction and operation currently report for nearly 40% of global energy-related CO2 emissions (United Nations, 2022). To prevent further global climatic issues, many governmental and inter-governmental bodies endorsed climate targets (IPCC, 2018). Introducing digital technologies is a step towards a greener planet (Ogundiran, Asadi, & da Silva, 2024), and evidence shows that digitalisation technologies are accelerating these climate targets (Opoku, Perera, Osei-Kyei, & Rashidi, 2021). Therefore, globally influenced nations identify the need to embrace the transition towards digitalisation in the construction industry (Xie, Xin, Lu, & Xu, 2023).

1.1 Literature review

Due to Industry 4.0 technological achievements, different industries are being taken over by artificial intelligence (AI) boost (Paige Wenbin Tien, Wei, Darkwa, Wood, & Calautit, 2022). With the innovative methods of data exchanges between the digital and physical world, these data-driven methods became popular in many industries (University of Wollongong Australia, 2025). This AI boost opened up a new society with automation, where many tasks of humans are automated (Coeckelbergh, 2020). The integration of these trending data-driven technologies, such as AI, machine learning (ML), and digital twin (DT), in improving IEQ in buildings while maintaining energy management and energy efficiency (H. H. Hosamo, 2023; H. H. Hosamo, Svennevig, Svidt, Han, & Nielsen, 2022; Van Thillo, Verbeke, & Audenaert, 2022) is a trending topic in the built environment (BE) sector. Artificial intelligence intends to create a synthetic brain called an autonomous agent, focusing on an intelligent program which does tasks like humans (Jo, 2021). AI can learn from the environment and make decisions accordingly to maximise the probability of reaching a given objective (Cerulli, 2023). AI is a human-made product with computation capacity, data availability, and algorithm development capacity to process an intelligent task. (Cerulli, 2023). Machine learning, also known as statistical learning, is an approach that enables computers to learn from data and make predictions without definite programming (LeCun, Bengio, & Hinton, 2015). ML can be identified as a technique that lies between statistics, computer science and AI (Cerulli, 2023). Digital twin solutions have already been explored and deployed in various domains. (Boschert & Rosen, 2016). The DT is viewed as a model which investigates and visualises the condition of a real-world physical object from its digital representation (virtual or digital space) (Crespi, Drobot, & Minerva, 2023). In the BE, the DT is viewed as a key platform for real-time monitoring and control in building systems (Lu, Xie, Parlikad, & Schooling, 2020). DT enables automated controls to adjust HVAC and other services to enhance occupants' comfort by integrating with the Building Automation System (BAS) (Pan et al., 2023).

1.2 Review necessity

Despite the positive impact of these data-driven technologies, the application of these technologies in the built environment sector is a highlighted barrier due to the high level of data requirements (Crespi et al., 2023) to produce meaningful outcomes. However, reviews that discuss ranges, including the building areas, suitable data-driven applications, the real-life applicability of these data-driven applications, and barriers to data-driven applications in real-life building indoor environments are missing. Therefore, critical investigation is needed to determine the suitable data-driven technologies to improve IEQ.

1.3 Objectives, contribution and structure of the present review

To address the research gaps, this study examines the application of AI, ML, and DT in enhancing IEQ in buildings by reviewing previous studies in BE conducted over the past decade. The review is carried out under four objectives: *1. To critically review the data-driven technology involvement in IEQ improvement, 2. To critically investigate the focused IEQ factors in the current literature, 3. To critically investigate data-driven applications for improving IEQ and energy efficiency in buildings, 4. Applicability and barriers of these data-driven applications in real building environments*. The study is presented as follows. In Section 1, the introduction to the selected area of review, in Section 2, the literature review to establish the need for the review objectives, in Section 3, the methodology of the systematic literature search presented, as well as an explanation of the review methods with selection, inclusion and exclusion criteria for the work reviewed. In Section 4, the authors provide in detailed review of results, discussion aligning with the review objectives. Section 5 highlights the real-life applications and barriers of these data-driven applications in improving IEQ. The limitations of the study are discussed in Section 6, and the review is concluded in Section 7.

2. Review Methodology

This review aims to identify the status of data-driven applications in improving IEQ while optimising energy efficiency through a systematic literature review (SLR). To achieve the goal of SLR in this study, the Preferred Method of Systematic Reviews and Meta-Analysis (PRISMA) 2020 protocol is used. The step-by-step process of the SLR using the PRISMA protocol is shown in Figure 1. The PRISMA diagram is 'an evidence-based minimum set of items for reporting in systematic and meta-analysis' (Page et al., 2021). The literature on data-driven applications in improving IEQ was identified in Scopus and the Web of Science databases. Keywords used included "*Artificial Intelligence*", "*Machine Learning*", *"Data-driven"*, *Digital Twin*" and "*Indoor Environmental Quality*", which were mostly used in related studies and covered the core information in this area.

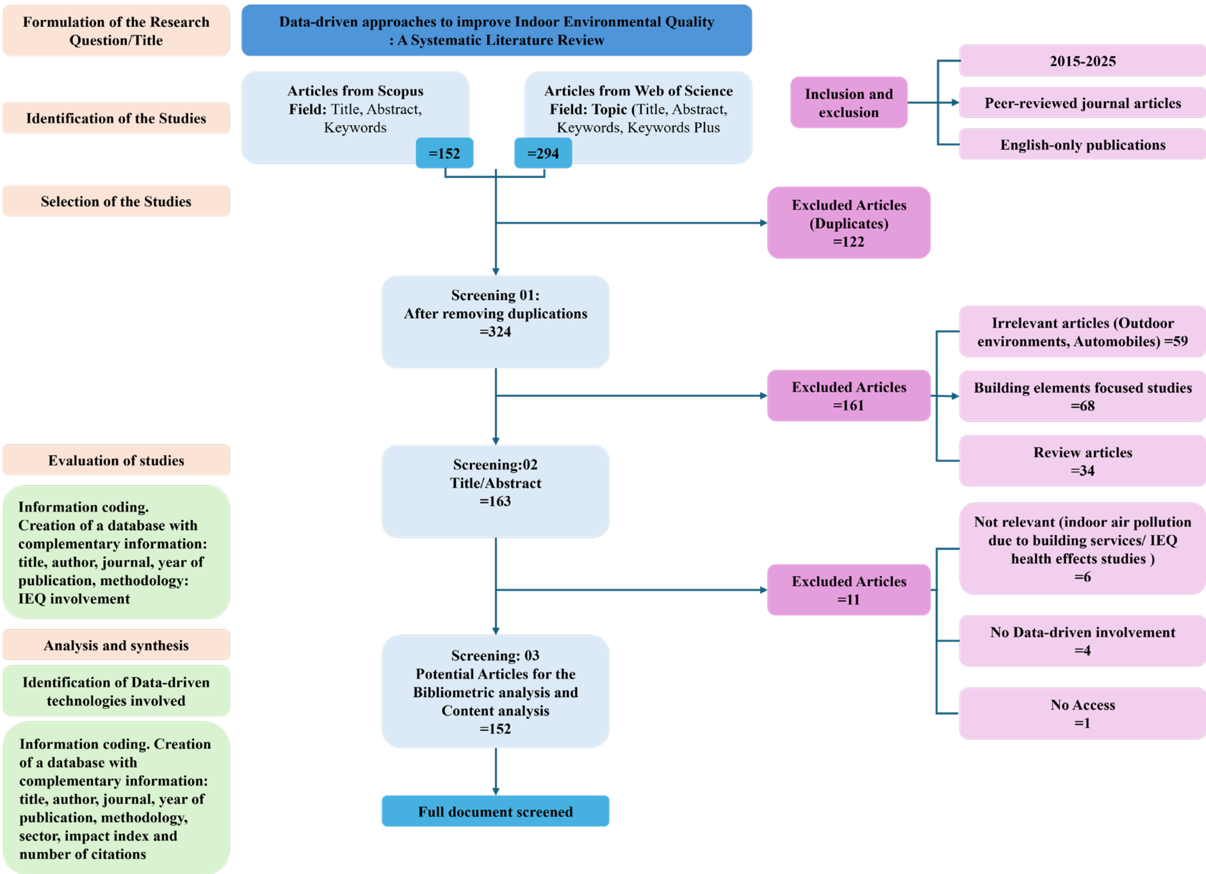

Figure 1: PRISMA diagram

Based on the search results, peer-reviewed journal publications between 2015 and 2025 were selected as introductions of data-driven technologies to the indoor environments of buildings increased in recent years. Additionally, English-only publications were used as inclusion criteria for the language. Initially, the Web of Science and Scopus databases produced a total of 446 records. These records were carefully screened in Screening 01 to eliminate duplicates, resulting in 324 records. The titles and abstract review were conducted as Screening 02, and 163 journal articles were selected, removing irrelevant articles using exclusion criteria '*building elements focused publications*', *irrelevant articles*, and *review articles*. After carefully reviewing the title and abstract, and reading the theoretical framework, findings, and conclusions, 11 articles were removed from the records. Subsequently,152 journal articles were considered for Bibliometric analysis, Content analysis and Full paper review. These articles comprised the AI, ML, and DT applications to improve IEQ in buildings. Figure 1 elaborates the inclusion and exclusion criteria to decide the most relevant and suitable studies from the selected databases. The results of the SLR are discussed in detail in Section 4, aligning with the review objectives. Figure 1 further elaborates on the evaluation of studies and analysis, and synthesis for the SLR.

3. Review results and Discussion

3.1. Bibliometric analysis

i. Analysis of Publication by Country

The selected 152 peer-reviewed journal articles were considered for the bibliometric analysis. As shown in Figure 2, 32 countries were identified, and countries such as China, the USA, South Korea and Italy have relatively higher numbers of publications. With 33 publications from 2016 to 2025, China ranked as the top country, while the USA has only 22 publications, followed by South Korea and Italy, which ranked 3rd and 4th with 14 and 12 publications, respectively. Notably, Oceania countries such as Australia and New Zealand exhibit considerably insignificant studies in the relevant area.

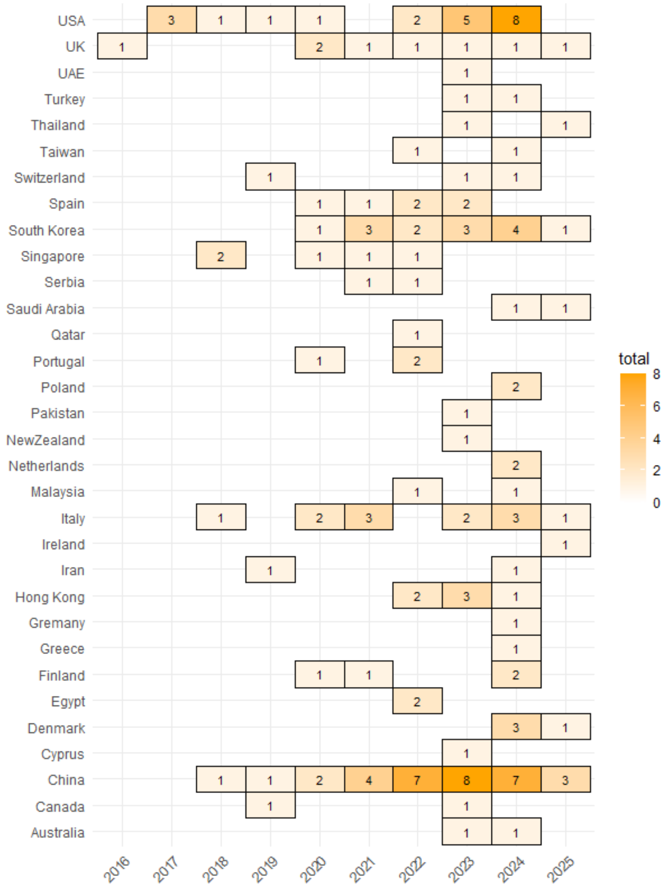

Figure 2: Heat map of the publications by country and year

ii. Analysis of Publication by Year

Journal publications from 2015 to 2025 were selected for the review as introductions of data-driven technologies to buildings' indoor environments have increased in recent years. As shown in Figure 2, the keyword search found a high density of articles from 2018 to 2025. Figure 2 further shows a significant increase in publications from 2022.

3.2. Content Analysis

i. Focused IEQ factors in studies

Based on the comprehensive review of the selected studies, data-driven techniques and approaches are mostly used to improve air quality, thermal aspects, and energy efficiency. As shown in Figure 3, other factors, such as acoustics, lighting, and ergonomics in an indoor environment, received notably less attention from the researchers in the relevant area. This highlights the need for more research on these negligence factors, as the combination of all these factors defines the IEQ.

i. Data-driven applications in studies

Different data-driven approaches were used based on the content analysis of the selected articles. Of the 152 records, 34 papers highlighted AI technology, 27 used data-driven terminology, 10 focused on DT applications, and 75 focused on ML applications (See Figure *4*).

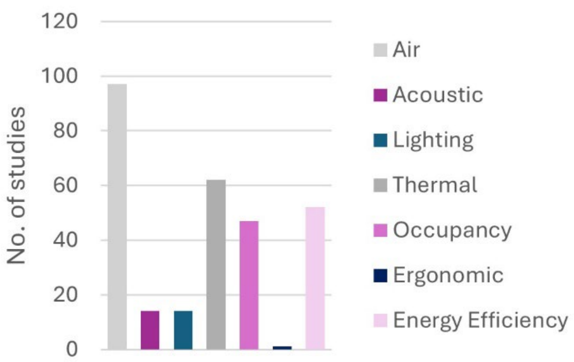

Figure 3: Focused IEQ factor in the selected studies

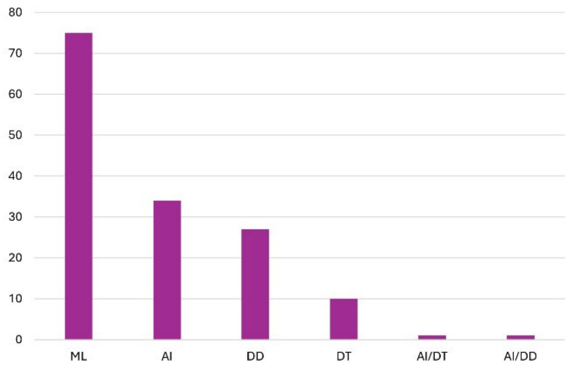

Figure 4: Data-driven involvement in publications

3.3. Full paper review

These Data-driven approaches were broadly used for IEQ condition detection and/or prediction, energy management, energy optimisation, and energy efficiency. The full paper review of the study identified the use areas of the selected articles as IEQ Condition detection/ prediction, IEQ condition control, Energy management, HVAC controls, Real-life Applications and barriers in data-driven applications. Figure 5 was generated using R Studio version 4.2.

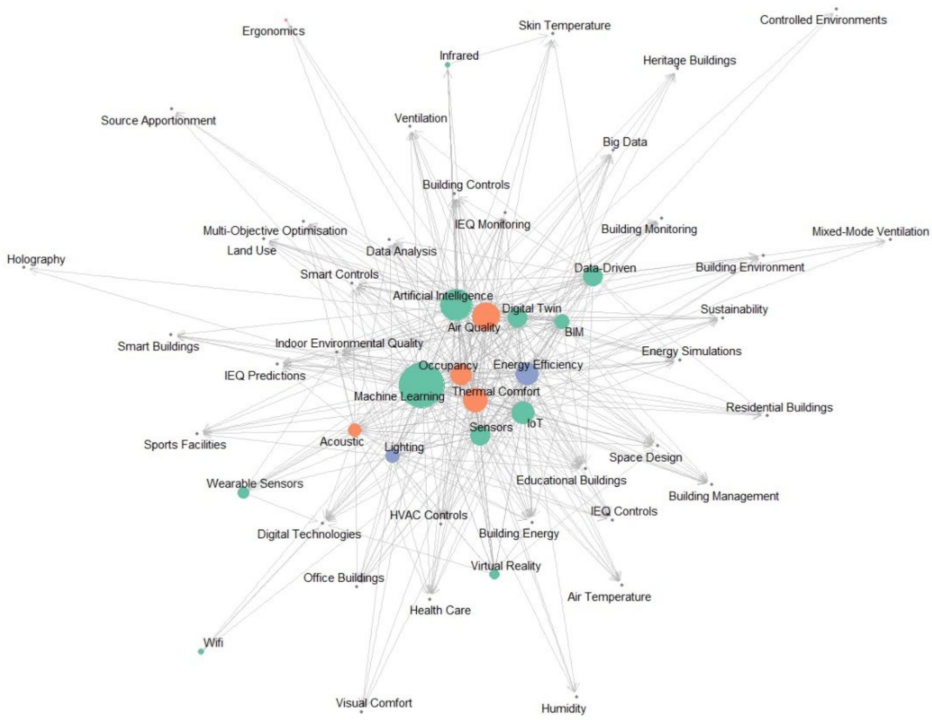

Figure 5: Network visualisation of keywords

This was used to construct and display relationships between data-driven technologies, IEQ parameters, and application types (i.e., prediction, detections or controls). The three group words "*Machine learning*", "*Artificial intelligence*", and "*Air quality*" have the most density of links and networks. Subsequently, "*Thermal comfort*" and "*Occupancy*" were the most connected cluster words. This provides an overview of the current research focus from the IEQ studies selected for the review. Most focused IEQ factors in the research domain are air quality, thermal comfort, and occupancy, despite having other contributing factors which determine the overall IEQ.

i. IEQ Conditions detection/prediction

a) Techniques for IAQ conditions

Data-driven approaches have been applied to IEQ condition detection and prediction in built environments. Symonds et al. (2016) used the ML method for the meta-model development for estimating overheating and air pollution in an England-wide indoor environment using the supervised learning methods feed-forward neural networks (NN) and radial basis function (RBF). Support vector regression (SVR) was compared with the NN system, and NNs showed 50% better overall performance than SVR. Another study used an SVM-based method for the detection and sampling of air particles using false positive detection (Wu et al., 2017). Liu et al. (2018) explored the potential relationship between IAQ and the concentration of airborne culturable fungi using an artificial neural network with ten hidden nodes (ANN-10) and SVM. The results showed that the ANN outperformed the SVM (64% of the accuracy) with an accuracy of almost 90% (SVM<ANN-10<90%). In 2019, Khazaei et al. used an MLP Neural Network to predict CO_2, resulting in a CO_2 concentration of 1465- 1510 ppm when the CO_2 concentration set point in ASHRAE 62 is 1000 ppm. Table 1 presents a part of the articles reviewed based on data-driven approaches and their accuracy.

Table 1: Data-driven and algorithms applications in the studies

Reference	Data-driven approach/ Algorithms	Accuracy/Results
(Symonds et al., 2016)	NN, SVR	NN models gave 50% better overall performance than SVR
(Choi & Yeom, 2017)	Decision Tree (DT), Stepwise	Stepwise= 35.99%, DT= 95.87%
(Wu et al., 2017)	SVM	Sizing accuracy:93%
(Liu et al., 2018)	SVM, ANN	SVM:64%, SVM<ANN-10<90%
(Khazaei, Shiehbeigi, & Kani, 2019)	MLP Neural network	CO2 concentration: 1465-1510PPM (CO2 concentration set point in ASHRAE 62 is 1000PPM)
(Peng, Nagy, & Schlüter, 2019)	GNB, DT, C-SVC, Multi-Layer (MLP)	Energy saving = 4% to 25% as compared to static temperature set points
(Song et al., 2019)	HAGRU neural network	Predictive accuracy: 98.4%
(Jayathissa, Quintana, Abdelrahman, & Miller, 2020)	RF classifier	64%: thermal, 80%: light, 86%; noise
(Pigliautile et al., 2020)	LDA, KNN, DT, NB, SVM, and RF	
(Salamone et al., 2020)	DT, KNN, LDA, RF, NB, SVM,	Average accuracy of DT, LDA, and RF~to 50%, NB=76%, SVM= 76%-84%, KNN= 68%
(P. W. Tien, Wei, Calautit, Darkwa, & Wood, 2020)	CNN-based deep learning	80.62%
(Z. Wang et al., 2020)	PCA, LR and SVM	Prototype accuracy: 99.5%
(Floris, Porcu, Girau, & Atzori, 2021)	DNN-Medium-cost computational models (Bagged tree and GPRI)	Prototype accuracy: 99.5%
(Tagliabue et al., 2021)	BIM, IoT, Digital Twin	
(Abdel-Razek, Marie, Alshehri, & Elzeki, 2022)	KNN, AO-ANN (BP), Decision Tree (DT)	KNN: 99.5% best performance
(Elnour et al., 2022)	NN, SVR, KNN, DT	The NN-based model outperforms other ML models by around 0.06 between the actual and the predicted values
(X. Wang & Dong, 2023)	AI-powered cameras and CNN, Shallow artificial neural network predictive models, ANN	Energy consumption: 0.6% and 29%, Thermal comfort level: 0% to 58.8%, CO2 concentration below 1000ppm, 59.7% for the conventional window control strategy and 89.25% from the proposed model
(Andersen et al., 2024)	SVM, LR, RF, KNN, NB, and XGBoost	Model performance: very well (all 9 metrics in high performing range)

(Banihashemi, Weber, Deghim, Zong, & Lang, 2024)	IoT, RF, XGBoost, and DFNN, CDBLSTM (a hybrid of CNN and bi-directional long short-term memory (Bi-LSTM), BP, SHAP	DFNN and CDBLSTM had high standard deviations of single features, compared to the non-sequential models.
(H. Hosamo & Mazzetto, 2025)	Computer vision-based detection (YOLOv5)	Occupancy-based strategy: 50% of energy saving, Temperature-based control: 37.27% energy saving,
(Sood et al., 2025)	XGBoost, RF, SVM, NSGA-II	over 99%
(Xu et al., 2025)	RF	Building energy consumption= 29.46%, Indoor thermal comfort= 10.46%, Daylighting performance= 65.56%

Note: ANN=Artificial neural network, AO-ANN=Adam optimiser-artificial neural network, BIM=Building information modelling, Bi-LSTM=bi-directional long short-term memory , BP=back-propagation, BT=Bagged tree, CDBLSM=Convolutional deep bidirectional long-term memory, CNN=Convolutional neural network, C-SVM=C-Support Vector Classification, DFNN=Dense feedforward neural network, DNN=Dense neural network, DT=Decision tree, GNB=Gaussian Naïve Bayes , GPRI=Gaussian process regression, HAGRU=Hierarchical attention-gated recurrent unit, IoT=Internet of things, KNN=K-nearest neighbours, LDA=Linear discriminant analysis, LR=Logistic regression, MPL=Multilayer perceptron, MLP=Multi-Layer Classification, NB=Naïve Bayes, NN=Neural network, NSGA-II=Non-Dominant Sorting Genetic Algorithm-II, PCA=Principal component analysis, RF=Random Forest, SHAP=Shapley additive explanations, SVM=Support vector machine, SVR=Support vector regression, XGBoost=Extreme Gradient Boost

Many studies were conducted using data-driven methods to monitor the particulate matter (PM) concentration in the indoor environment. Marques and Pitarma (2020) used IoT to monitor and assess PM. The developed model is a health information system to enhance living environments. Another study developed a practical framework for predicting residential PM concentration using land-use regression and machine learning methods. Woo et al. (2022) used deep learning to forecast the effect of real-time indoor PM 2.5 on peak expiratory flow rates (PEFR) of asthmatic children in Korea. Dai et al. (2023) conducted a study on predicting urban residential indoor PM 2.5 concentration employing a Large-scale spatiotemporal deep learning method. Some authors explored methods of predicting PM with low-cost data and ML techniques (Lagesse et al., 2022). A study in China explored the influencing factors on indoor PM 2.5 of an office building based on statistical and ML methods (Li et al., 2023c). In 2024, Lu et al. (2024) studied enhancing PM 2.5 prediction by applying ML and spatial modelling approaches.

b) Techniques for Thermal comfort conditions

In the review studies, several models for thermal comfort prediction have been proposed and tested. Choi and Yeom (2017) developed a data-driven thermal sensation prediction model to investigate the relationship between the thermal sensation and skin temperatures as a function of local body skin temperatures in an experimental chamber using a DAQ system (Data acquisition system), designed by LabVIEW, which was installed in the desktop computer for the data collection and recording. For accuracy estimation, the decision tree (DT) analysis model (J48) by WEKA (a data mining software) was used. The result showed an accuracy of 95.87% for the DT model. Another study developed a Personal Comfort Model utilising machine learning to predict individuals' thermal preferences using occupant heating and cooling behavior (Kim et al., 2018). The study used several ML algorithms to solve multiclassification complications of an occupant's thermal preference. Algorithms such as classification tree (CTree), Gaussian process classification (GPC), Gradient boosting method (GBM), Kernel support vector machine (kSVM), Random Forest (RF), and Regularised logistic regression (regLR). The results showed a median accuracy of 0.73 for the best performance algorithm, while a Conventional model's accuracy was 0.51. Salamone et al. (2018) developed an integrated method to assess Personal Thermal Comfort using user feedback, IoT and ML techniques. The study tested different ML methods to identify the optimum algorithm, including logistic regression, linear discriminant analysis and non-linear algorithms (K-nearest neighbours, Classification and regression trees, Gaussian naïve bayes, Support vector machine). Classification and regression trees (CART) were most accurate among other tested algorithms. Salamone et al. (2020) studied personal thermal comfort perception in real and virtual environments, assessing the visual stimuli. The machine learning applications they used were DT, KNN, LDA, RF, NB, SVM, and KNN. The results showed that the average accuracy of DT, LDA, and RF is close to 50%, NB of 76%, and SVM of 76%-84%, while KNN shows a performance of 68%. On the other hand, a study researched determining occupant dissatisfaction (Kent et al., 2021) using principal component analysis (PCA) and SVM. The goal was to determine the best IEQ parameter to predict overall space assessment, and the findings showed 90% classification for occupant satisfaction with the overall workplace. However, the study suggested leveraging space designs that ensure privacy and sufficient spaces to mitigate occupant dissatisfaction, irrespective of all other IEQ conditions. Another study on thermal comfort prediction and CO_2 concentration developed the extreme machine learning (EML) model (Hou et al., 2022). The EML model was optimised by the grey wolf optimiser (GWO) and resulted in a reduction of CO2 concentration and PMV within acceptable limits.

c) Techniques for multi-objective optimisation

Martínez-Comesaña et al. (2021) researched the use of optimised multi-layer perception (MLP) neural networks for spatiotemporal estimation of indoor environmental conditions of a building. Again, Martínez-Comesaña et al. (2022) employed an optimised extreme gradient boosting algorithm to assess real-time Indoor temperature, RH, and CO_2 concentration. Multi-objective generic algorithm NSGA-III was employed for estimation and estimated indoor temperature and RH with relative faults below 6% and CO2 levels below 10%. In thermal monitoring, Rida et al. (2023) used machine learning-based human thermo-physiology modelling enhanced with computer vision. Employing this AI-integrated system, they developed a framework that enables contactless human thermal monitoring. Recently, researchers utilised ML applications such as the Instance-based transfer learning method, and SMOTE and SHAP integrated methods are used in thermal comfort prediction studies (Li et al., 2024a & Li et al., 2025).

d) Techniques for occupant detection

In the area of occupant detection in buildings, several studies have been conducted on the use of these data-driven applications. Abdelrahman et al. (2022) used classification algorithms such as K-nearest neighbours (KNN), Hybrid Adam optimiser-artificial neural network-back-propagation network (AO-ANN (BP)), and decision tree (DT) to predict room occupancy based on thermal comfort parameters. The results indicated that the KNN showed the overall best performance with 99.5%. On the other hand, Aliero et al. (2022) tried different ML techniques to predict room occupancy non-intrusively. Another study developed a multimodal framework for smart building occupancy detection (Abuhussain et al., 2024). Anik et al. (2023) attempted to develop smart buildings by creating automated occupant profiles. The study explored the ML-enabled approach for user persona generation in smart buildings. Andersen et al. (2024b) proposed a model to detect occupancy in residential buildings using supervised learning. They explored the model's generalisability using SVM, Logistic Regression, RF, KNN, NB, and eXtreme Gradient Boosting (XGBoost) algorithms. They conclude that the model succeeded with all 9 metrics in the high-performing range. Banihashemi et al. (2024) researched occupancy detection using non-intrusive indoor environmental data and ML techniques in two double-occupied office rooms at a university building in Germany. They employed IoT technology as a method to collect indoor environmental data. RF, XGBoost, and a Dense neural network (dense feedforward neural network (DFNN)) were trained on the data set. To compare the model performance, they used CDBLSTM (convolutional deep bidirectional long-term memory), a hybrid of CNN and bidirectional long short-term memory (Bi-LSTM), and to optimise hyperparameters, they used Random and Bayesian optimisation. They utilised SHAP (Shapley additive explanations) to identify feature importance in the model. The findings showed that DFNN and CDBLSTM had the highest standard deviations of single features compared to the non-sequential models.

e) Digital twin involvement

On the other hand, data-driven technologies such as IoT, BIM (Building Information Modelling) and digital twin received notable attention from researchers in the BE domain. In 2022, Abdelrahman et al. (2022) modelled personal thermal comfort using digital twin technology. They used occupant preference prediction with BIM-extracted spatial-temporal proximity data from the Build2Vec application. Build2Vec was used BIM-extracted graph into a multi-dimensional vector space to extract similarities between different spatial objects and locations. They utilised classification algorithms, K-nearest neighbours (KNN), Hybrid Adam optimiser-artificial neural network-back-propagation network (AO-ANN (BP)), and Decision Tree. The test results showed a 14%–28% accuracy improvement over a conventional model. Desogus et al. (2023) developed a framework to monitor, visualise and assess building thermal comfort using the BIM-based digital twin (DT) model approach. In 2024, Opoku et al developed a BIM and IoT-enabled DT model for indoor condition monitoring indoor environment utilising a university library in Australia (Opoku et al., 2024). Furthermore, Hu and Assaad (2024) developed a BIM-enabled DT framework to autonomously monitor indoor environmental conditions and provide real-time visualisation.

(ii) IEQ Conditions control, Energy management and HVAC controls

In the context of energy management, HVAC controls and IEQ condition control, various data-driven techniques have been proposed and tested. In 2017, Konis and Annavaram developed the occupant mobile gateway to develop an occupant-aware energy management system. They employed a logistic regression model in R machine learning computing software for the data analysis and EnergyPlus software to quantify energy usage, resulting in an error below 10% (Konis and Annavaram, 2017). Another study developed a system to learn individual temperature preferences using NN for occupant-centric building indoor climate controls (Peng et al., 2019), modifying HVAC systems to react automatically to individual temperature preferences of occupants. The used learning models were Gaussian naïve bayes (GNB), Decision tree (DT), C-support vector classification (C-SVC), and Multi-layer (MLP) classification for temperature preferences, providing an energy saving of between 4% and 25% as compared to static temperature set points of HVAC systems. Du et al. (2021) developed a method of identifying suitable air temperature levels in Chinese air-conditioned buildings based on a data-driven model. On the other hand, Floris et al. (2021) proposed an IoT-based smart building (SB) solution for managing the indoor environment. They employed DNN-Medium-cost computational models, bagged tree and Gaussian process regression (GPRI), which resulted in an accuracy of 99.5% of the prototype. Elnour et al. (2022) developed an NN-based control system for optimising BAS (building automation system) and BMS (building management system). The study compared the NN-based model with other models such as SVR, KNN, and DT. The result showed that the NN-Based model performs better than other ML models, with an average root mean square error (RMSE) of around 0.06 between the actual and the predicted values. Utilising the extreme machine learning (EML) model optimised by the grey wolf optimiser (GWO) algorithm, Hou et al. (2022) developed a hybrid prediction model to predict CO_2 level, Thermal comfort level and energy consumption. The results showed 14.34% of energy saving after the model was implemented.

In recent years, researchers have utilised AI techniques for building control systems for energy management. Using AI techniques, Zhu et al. (2022) developed a dynamic sensing and control system for a non-uniform indoor environment. In 2023, Faiz et al. conducted research on energy modelling and predictive control of environmental conditions for BEMS (building energy management system) utilising AI and ML (Faiz et al., 2023). Faulkner et al. (2023) developed a system for fast prediction of indoor airflow distribution. The authors developed this system utilising synthetic image generation AI.

Furthermore, from the reviewed studies, it can be identified that the researchers tried several methods in the indoor environment climate control. Incorporating reinforcement learning techniques, Kim and Moon (2023) developed a system to control indoor climate based on outdoor air PM. Another study in 2023 developed a smart HVAC control system for a university classroom in a subtropical climate using Computer vision techniques (Lan et al., 2023). Li et al. (2023a) developed a calibration approach for HVAC systems at the operational location without removing or disrupting the system's function. The authors utilised Bayesian inference incorporating validation of virtual sensors. The same authors developed an algorithm for smart environmental control using an adaptive multi-objective particle swarm optimiser-grey wolf optimisation algorithm (Li et al., 2023b). They focused on improving IAQ and thermal comfort while enhancing energy efficiency through this system.

4. Real-life applications and barriers

Most developed models and algorithms were created under controlled environmental conditions and various criteria, limiting their applications to selected case studies. Moreover, many of these developments require further validation to achieve broader real-world applicability. The scope of the studies was constrained by the data sample size. Numerous studies suggest that to enhance model accuracy and relevance, larger datasets incorporating diverse environmental conditions must be included. Outdoor environmental conditions are crucial for improving indoor environmental quality (IEQ). The IEQ is influenced by outdoor microclimatic conditions and building morphology. Therefore, more research targeting this area is necessary to enhance IEQ. Many proposed models could be improved by utilising advanced data collection methods that integrate contemporary real-time techniques, such as sensors. This could enable more effective real-time monitoring of the factors affecting IEQ. Additionally, considerations of individual occupancy behaviour and physiological parameters incorporated into the developed models are currently lacking. The reliability of these models is uncertain, as additional guidance is needed for building managers regarding control methods.

Furthermore, these data-driven technologies necessitate large volumes of historical data to train, test, and validate the models, helping to identify patterns in the data and provide more accurate predictions about future indoor environmental conditions. The review reveals that a significant challenge in using these data-driven techniques to develop predictive models lies in the availability of high-quality, reliable historical data. Inaccurate data leads to incorrect or ineffective predictions, since the models cannot learn the correct patterns. The review emphasised that even well- performing data- driven models may falter if they are trained inadequately, inaccurately, or unreliably. All these data-driven technologies are heavily reliant on the data provided, underscoring the need for well-curated, accurate datasets to ensure effective model performance.

5. Limitations of the study

The review limits its scope to studies published only from 2015 to 2025 and considers only peer-reviewed journal articles. Additionally, English-only publications were used as inclusion criteria for the language. Further to that review limits to Scopus and Web of Science databases. This selection process may result in missing important research published outside these inclusion criteria, which could provide valuable context or insight.

6. Conclusion

Improving IEQ conditions inside a building is a crucial factor in occupant comfort. However, maintaining improved IEQ while optimising energy consumption is challenging. The data-driven technologies explored by researchers in the built environment to address these energy management and energy efficiency challenges while increasing IEQ. The review reveals that data-driven technologies have been primarily applied in several key areas within the built environment, such as *air quality monitoring and prediction, thermal comfort prediction, occupant detection and behaviour modelling, energy efficiency and HVAC optimisation, building performance simulation and prediction,* and *smart building and digital twin technologies*.

Review results indicated that ML and AI algorithms were utilised to monitor and predict indoor air quality (IAQ) conditions, including CO_2 concentration, air pollutants, and particulate matter (PM). Additionally, these algorithms were employed to predict and optimise thermal comfort in indoor environments and to non-intrusively detect occupant presence and behaviour. Furthermore, these data-driven technologies were used to enhance heating and cooling systems based on real-time indoor environmental conditions. The review highlighted that these systems increase energy efficiency by learning occupant preferences and automatically adjusting indoor environmental conditions.

Commonly used AI and ML techniques to improve IEQ conditions and energy efficiency were deep learning, neural networks, RF, KNN, decision trees, multi-objective algorithms, XGBoost, and SVM. In addition to ML techniques, BIM and IoT techniques in combination with AI were being used to develop digital twin (DT) models. The DT models create a virtual replica of physical buildings for real-time monitoring and control of indoor environmental conditions. This, in turn, offers real-time visualisation of the indoor environment, helping energy management and increasing IEQ. In addition, EnergyPlus simulation was commonly used for energy modelling.

However, factors such as acoustics, lighting, and ergonomics have received significantly less attention from researchers. This gap underscores the need for further investigation into these overlooked factors, as they collectively contribute to overall IEQ. Furthermore, the review emphasised the necessity of well-curated, accurate datasets to ensure effective model performance. In conclusion, while data-driven technologies hold great potential, their success relies on the quality of the historical data upon which those models are built.

Author Contribution

Conceptualisation, K.S. and S.P.; methodology, K.S.; software, K.S.; validation, S.P. and S.N.; formal analysis, K.S., and S.P.; investigation, K.S.; resources, S.P.; data curation, K.S. and S.P.; writing—original draft preparation, K.S.; writing—review and editing, K.S., S.P., and S.N.; visualisation, K.S.; supervision, S.P., S.N., X.J., and M.S.; project administration, S.P.; funding acquisition, S.P. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgements & Funding

The authors acknowledge the Western Sydney University, Australia for academic and financial support.

Data Availability Statement

The data used to support the analysis, results and findings of this review are available from the corresponding author upon request.

7. References

Abdel-Razek, S. A., Marie, H. S., Alshehri, A., & Elzeki, O. M. (2022). Energy Efficiency through the Implementation of an AI Model to Predict Room Occupancy Based on Thermal Comfort Parameters. *Sustainability (Switzerland), 14*(13). doi:10.3390/su14137734

Al Horr, Y., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M., & Elsarrag, E. (2016). Occupant productivity and office indoor environment quality: A review of the literature. *Building and Environment, 105*, 369-389. doi:https://doi.org/10.1016/j.buildenv.2016.06.001

Andersen, K. H., Johra, H., Schaffer, M., Marszal-Pomianowska, A., Knudsen, H. N., Heiselberg, P. K., & O'Brien, W. (2024). Exploring occupant detection model generalizability for residential buildings using supervised learning with IEQ sensors. *Building and Environment, 254*. doi:10.1016/j.buildenv.2024.111319

Baek, J., Park, D. Y., Park, H., Le, D. M., & Chang, S. (2023). Vision-based personal thermal comfort prediction based on half-body thermal distribution. *Building and Environment, 228*, 109877. doi:10.1016/j.buildenv.2022.109877

Banihashemi, F., Weber, M., Deghim, F., Zong, C., & Lang, W. (2024). Occupancy modeling on non-intrusive indoor environmental data through machine learning. *Building and Environment, 254*. doi:10.1016/j.buildenv.2024.111382

Boschert, S., & Rosen, R. (2016). Digital Twin—The Simulation Aspect. In (pp. 59-74). Switzerland: Springer International Publishing AG.

Cerulli, G. (2023). *Fundamentals of Supervised Machine Learning : With Applications in Python, R, and Stata* (1st 2023. ed.). Cham: Springer International Publishing.

Choi, J. H., & Yeom, D. (2017). Study of data-driven thermal sensation prediction model as a function of local body skin temperatures in a built environment. *Building and Environment, 121*, 130-147. doi:10.1016/j.buildenv.2017.05.004

Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics, 26*(4), 2051-2068. doi:10.1007/s11948-019-00146-8

Crespi, N. l., Drobot, A. T., & Minerva, R. (2023). *The Digital Twin* (1st 2023. ed.). Cham: Springer International Publishing.

Elnour, M., Himeur, Y., Fadli, F., Mohammedsherif, H., Meskin, N., Ahmad, A. M., . . . Hodorog, A. (2022). Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities. *Applied Energy, 318*. doi:10.1016/j.apenergy.2022.119153

Floris, A., Porcu, S., Girau, R., & Atzori, L. (2021). An iot-based smart building solution for indoor environment management and occupants prediction. *Energies, 14*(10). doi:10.3390/en14102959

Hosamo, H., & Mazzetto, S. (2025). Data-Driven Ventilation and Energy Optimization in Smart Office Buildings: Insights from a High-Resolution Occupancy and Indoor Climate Dataset. *Sustainability, 17*(1). doi:10.3390/su17010058

Hosamo, H. H. (2023). *Digital Twin technology toward more sustainable buildings Predictive maintenance, energy optimization and occupants' comfort in non-residential buildings.* (Engineering). University of Agder

Hosamo, H. H., Svennevig, P. R., Svidt, K., Han, D., & Nielsen, H. K. (2022). A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics. *Energy and Buildings, 261*, 111988. doi:10.1016/j.enbuild.2022.111988

IPCC. (2018). Global Warming of 1.5°C. Retrieved from https://www.ipcc.ch/sr15/

Jayathissa, P., Quintana, M., Abdelrahman, M., & Miller, C. (2020). Humans-as-a-sensor for buildings—intensive longitudinal indoor comfort models. *Buildings, 10*(10), 1-22. doi:10.3390/buildings10100174

Jo, T. (2021). *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning* (1st Edition 2021 ed.). Cham: Springer International Publishing AG.

Khazaei, B., Shiehbeigi, A., & Kani, A. (2019). Modeling indoor air carbon dioxide concentration using artificial neural network. *International Journal of Environmental Science and Technology, 16*(2), 729-736. doi:10.1007/s13762-018-1642-x

Kubba, S. (2010). Indoor Environmental Quality. *LEED Practices, Certification, and Accreditation Handbook*, 211-269.

Kubba, S. (2017). *Handbook of green building design and construction : LEED, BREEAM, and Green Globes* (Second edition. ed.). Oxford: Butterworth-Heinemann.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:10.1038/nature14539

Liu, Z., Cheng, K., Li, H., Cao, G., Wu, D., & Shi, Y. (2018). Exploring the potential relationship between indoor air quality and the concentration of airborne culturable fungi: a combined experimental and neural network modeling study. *Environmental Science and Pollution Research, 25*(4), 3510-3517. doi:10.1007/s11356-017-0708-5

Lu, Q., Xie, X., Parlikad, A. K., & Schooling, J. M. (2020). Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance. *Automation in Construction, 118*, 103277. doi:10.1016/j.autcon.2020.103277

Ogundiran, J., Asadi, E., & da Silva, M. G. (2024). A Systematic Review on the Use of AI for Energy Efficiency and Indoor Environmental Quality in Buildings. *Sustainability, 16*(9), 3627. doi:10.3390/su16093627

Opoku, D.-G. J., Perera, S., Osei-Kyei, R., & Rashidi, M. (2021). Digital twin application in the construction industry: A literature review. *Journal of Building Engineering, 40*, 102726. doi:10.1016/j.jobe.2021.102726

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Online), 372*. doi:10.1136/bmj.n71

Pan, Y., Zhu, M., Lv, Y., Yang, Y., Liang, Y., Yin, R., . . . Yuan, X. (2023). Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. *Advances in applied energy, 10*, 100135. doi:10.1016/j.adapen.2023.100135

Peng, Y. Z., Nagy, Z., & Schlüter, A. (2019). Temperature-preference learning with neural networks for occupant-centric building indoor climate controls. *Building and Environment, 154*, 296-308. doi:10.1016/j.buildenv.2019.01.036

Pigliautile, I., Casaccia, S., Morresi, N., Arnesano, M., Pisello, A. L., & Revel, G. M. (2020). Assessing occupants' personal attributes in relation to human perception of environmental comfort: Measurement procedure and data analysis. *Building and Environment, 177*. doi:10.1016/j.buildenv.2020.106901

Salamone, F., Bellazzi, A., Belussi, L., Damato, G., Danza, L., Dell'Aquila, F., . . . Vitaletti, W. (2020). Evaluation of the Visual Stimuli on Personal Thermal Comfort Perception in Real and Virtual Environments Using Machine Learning Approaches. *Sensors, 20*(6). doi:10.3390/s20061627

Song, J. C., Xue, G. X., Ma, Y. P., Li, H., Pan, Y., & Hao, Z. X. (2019). An Indoor Temperature Prediction Framework Based on Hierarchical Attention Gated Recurrent Unit Model for Energy Efficient Buildings. *IEEE Access, 7*, 157268-157283. doi:10.1109/access.2019.2950341

Sood, D., Alhindawi, I., Ali, U., Finn, D., McGrath, J. A., Byrne, M. A., & O'Donnell, J. (2025). Machine learning based multi-objective optimisation of energy consumption, thermal comfort and CO2 concentration in energy-efficient naturally ventilated residential dwellings. *Building and Environment, 267*. doi:10.1016/j.buildenv.2024.112255

Symonds, P., Taylor, J., Chalabi, Z., Mavrogianni, A., Davies, M., Hamilton, I., . . . Macintyre, H. (2016). Development of an England-wide indoor overheating and air pollution model using artificial neural networks. *Journal of Building Performance Simulation, 9*(6), 606-619. doi:10.1080/19401493.2016.1166265

Tagliabue, L. C., Cecconi, F. R., Maltese, S., Rinaldi, S., Ciribini, A. L. C., & Flammini, A. (2021). Leveraging digital twin for sustainability assessment of an educational building. *Sustainability (Switzerland), 13*(2), 1-16. doi:10.3390/su13020480

Tekce, I., Artan, D., & Ergen, E. (2021). An Empirical Study of Visual Comfort in Office Buildings. In (Vol. 200, pp. 319-331). Singapore: Springer Singapore Pte. Limited.

Tien, P. W., Wei, S., Calautit, J. K., Darkwa, J., & Wood, C. (2020). A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions. *Energy and Buildings, 226*. doi:10.1016/j.enbuild.2020.110386

Tien, P. W., Wei, S., Darkwa, J., Wood, C., & Calautit, J. K. (2022). Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality – A Review. *Energy and AI, 10*, 100198. doi:https://doi.org/10.1016/j.egyai.2022.100198

United Nations. (2022). CO2 emissions from buildings and construction hit new high, leaving sector off track to decarbonize by 2050: UN. Retrieved from https://www.unep.org/news-and-stories/press-release/co2-emissions-buildings-and-construction-hit-new-high-leaving-sector

University of Wollongong Australia. (2025). What is Industry 4.0? Retrieved from https://www.uow.edu.au/engineering-information-sciences/industry-4-hub/#:~:text=What%20is%20Industry%204.0?,world%20with%20the%20digital%20world.&text=Advanced%20automation%20and%20robotics%20(including%20collaborative%20robots%20or%20cobots).

Van Thillo, L., Verbeke, S., & Audenaert, A. (2022). The potential of building automation and control systems to lower the energy demand in residential buildings: A review of their performance and influencing parameters. *Renewable & sustainable energy reviews, 158*, 112099. doi:10.1016/j.rser.2022.112099

Vardaxis, N.-G., Bard, D., & Persson Waye, K. (2018). Review of acoustic comfort evaluation in dwellings—part I: Associations of acoustic field data to subjective responses from building surveys. *Building acoustics, 25*(2), 151-170. doi:10.1177/1351010X18762687

Wang, X., & Dong, B. (2023). Physics-informed hierarchical data-driven predictive control for building HVAC systems to achieve energy and health nexus. *Energy and Buildings, 291*. doi:10.1016/j.enbuild.2023.113088

Wang, Z., Wang, J. Y., He, Y. E., Liu, Y. C., Lin, B. R., & Hong, T. Z. (2020). Dimension analysis of subjective thermal comfort metrics based on ASHRAE Global Thermal Comfort Database using machine learning. *Journal of Building Engineering, 29*. doi:10.1016/j.jobe.2019.101120

Wu, Y. C., Shiledar, A., Li, Y. C., Wong, J., Feng, S., Chen, X., . . . Ozcan, A. (2017). Air quality monitoring using mobile microscopy and machine learning. *Light-Science & Applications, 6*. doi:10.1038/lsa.2017.46

Xie, H., Xin, M., Lu, C., & Xu, J. (2023). Knowledge map and forecast of digital twin in the construction industry: State-of-the-art review using scientometric analysis. *Journal of Cleaner Production, 383*, 135231. doi:10.1016/j.jclepro.2022.135231

Xu, S., Chen, Y., Liu, J., Kang, J., Gao, J., Qin, Y., . . . Li, G. (2025). Comprehensive improvement of energy efficiency and indoor environmental quality for university library atrium—A multi-objective fast optimization framework. *Frontiers of Architectural Research, 14*(2), 449-470. doi:10.1016/j.foar.2024.08.010

Zhao, Q., Lian, Z., & Lai, D. (2021). Thermal comfort models and their developments: A review. *Energy and Built Environment, 2*(1), 21-33. doi:10.1016/j.enbenv.2020.05.007

Zhao, R., Sun, S., & Ding, R. (2004). Conditioning strategies of indoor thermal environment in warm climates. *Energy and Buildings, 36*(12), 1281-1286. doi:https://doi.org/10.1016/j.enbuild.2003.07.002

Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and do not reflect the views of the Architecture, Buildings, Construction and Cities (ABC2) Journal and/or its editor(s). DFBI Journal and/or its editor(s) disclaim any responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.