Environmental and Climate Technologies
2025, vol. 29, no. 1, pp. 527–539
https://doi.org/10.2478/rtuect-2025-0036
https://content.sciendo.com

Data Challenges in AI-Driven HVAC Systems: A Critical Analysis and Future Directions

Dalia Mohammed Talat Ebrahim ALI[1*], Violeta MOTUZIENĖ[2]

[1,2]*Lithuanian Energy Institute, Breslaujos gatvė 3 LT-44403 Kaunas, Lithuania*
[2]*Vilnius Gediminas Technical University, Sauletekio Av. 11, LT-10223 Vilnius, Lithuania*

Received 19.03.2025; accepted 27.08.2025

Abstract **– Integrating Artificial Intelligence (AI) into heating, ventilation, and air conditioning (HVAC) systems is a promising approach that helps enhance energy efficiency in buildings, which leads to cost savings and provides environmental benefits. However, the effective performance of the AI models depends not only on the model design but also on the data quality, reliability, size, availability, and management. This paper analyses recent studies that apply AI models, specifically Deep Learning and Hybrid models, to achieve energy efficiency in HVAC systems in buildings from a data perspective, examining various aspects of data management. This analysis aims to provide insights into data-related challenges in AI-driven HVAC systems and propose strategies to overcome them, ensuring more accurate, efficient, and reliable models. The findings reveal that combining multiple data types can enhance model performance and generalizability. The findings also indicate that data quality is overlooked by researchers in many studies, where only 31 % of the analysed papers discussed quality issues, reflecting that it is not yet a standard practice in this field. Additionally, this analysis highlights the scarcity of reliable and audited data. Therefore, and in response to this issue, this paper recommends accessible and reliable data resources that can be employed in AI applications for HVAC systems in buildings.**

Keywords **– Building management systems; deep learning; energy efficiency; machine learning.**

1. INTRODUCTION

Heating, Ventilation, and Air Conditioning (HVAC) systems account for over 30 % of worldwide energy consumed in buildings, making them a critical target to optimize [1]. The integration of Artificial Intelligence (AI) in these systems has become a key strategy for improving building energy efficiency. AI-driven approaches, particularly Deep Learning (DL) and hybrid models, offer promising solutions by enabling intelligent control, fault detection, and predictive maintenance [2], [3]. These technological advances save energy, reduce costs, and reduce carbon footprints [4]. However, while AI-based HVAC systems management solutions have significant potential, their effectiveness depends largely on the quality and availability of data [5]. Challenges related to data availability, reliability, and pre-processing remain significant [6]. Moreover, the lack of standardized data management, privacy concerns, and limited real-world datasets hinder the training of robust models [7].

*Corresponding author.
E-mail address: dalia.ali@lei.lt

Several studies have discussed the role of data in AI-driven HVAC optimization, highlighting key challenges that align with this paper's focus. Zhou *et al.* [8] reviewed the application of data mining and machine learning techniques in building energy management. Their work highlights the importance of data-driven strategies but does not extensively discuss data quality issues, which this paper aims to discuss [8]. Similarly, Xiao *et al.* [9] investigated the impact of data pre-processing and feature selection on HVAC energy consumption prediction models, demonstrating that different smoothing methods and training set sizes significantly affect model accuracy. Their findings highlight the need for robust data handling practices but do not explore the broader challenges of data availability and real-world applicability [9]. Additionally, Huang *et al.* [10] compared real and simulated datasets for building defect detection and concluded that models trained on simulated data struggle to generalize to real-world conditions. This finding is particularly relevant to this study, as it highlights the limitations of relying on simulated data for AI applications in HVAC. A study by Albatayneh [11] highlighted that many countries lack up-to-date and reliable data on residential building energy consumption, forcing researchers to depend on small-scale surveys or incomplete measurements More recently, Mukhtar *et al.* [12] emphasized that reproducibility in HVAC and Machine Learning (ML) research is often undermined by poor dataset reporting, accessibility, and the absence of publicly available data.

While prior reviews have examined AI applications in HVAC for energy efficiency, discussions of data challenges are typically brief and confined to subsections. This study aims to address this gap by utilizing insights from prior original research studies to conduct a more thorough investigation into the data challenges associated with AI in HVAC applications. It emphasizes data-related challenges and their impact on model accuracy.

This paper comprehensively analyses recent studies employing DL and hybrid AI models designed to achieve energy efficiency in HVAC systems from a data-centric perspective. The following sections present the methodology of the literature review and data extraction, followed by results on data availability, data type, quality, sources by building type, and data split, before concluding with key insights. This study aims to identify deficiencies in data management practices in the application of AI for HVAC systems in buildings, propose solutions to these challenges, and recommend accessible and reliable data resources.

This paper enhances the understanding of data-related challenges in AI-driven HVAC applications and provides recommendations for optimizing models' performance.

2. METHODOLOGY

This section discusses the approach taken to select relevant studies for this analysis. The study selection followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [13]. A review protocol was not registered, but the search strategy (keywords, databases, and search fields) and selection criteria are reported to ensure transparency and reproducibility. A visual summary of the articles used in the review and selection process is provided by the PRISMA flowchart in Fig. 1. The literature search was conducted using the keywords 'Machine Learning' and 'HVAC' in the 'Title, Abstract, Keywords' fields across two major scientific databases: SCOPUS and Web of Science (WoS) Core Collection. The initial search retrieved 1130 articles (637 from SCOPUS and 493 from WoS). After removing duplicates, 715 articles remained for screening. After screening the articles' titles and abstracts, 202 articles remained. Lastly, 36 articles were assessed for eligibility and included in the analysis based on inclusion and exclusion criteria shown in Fig. 2.

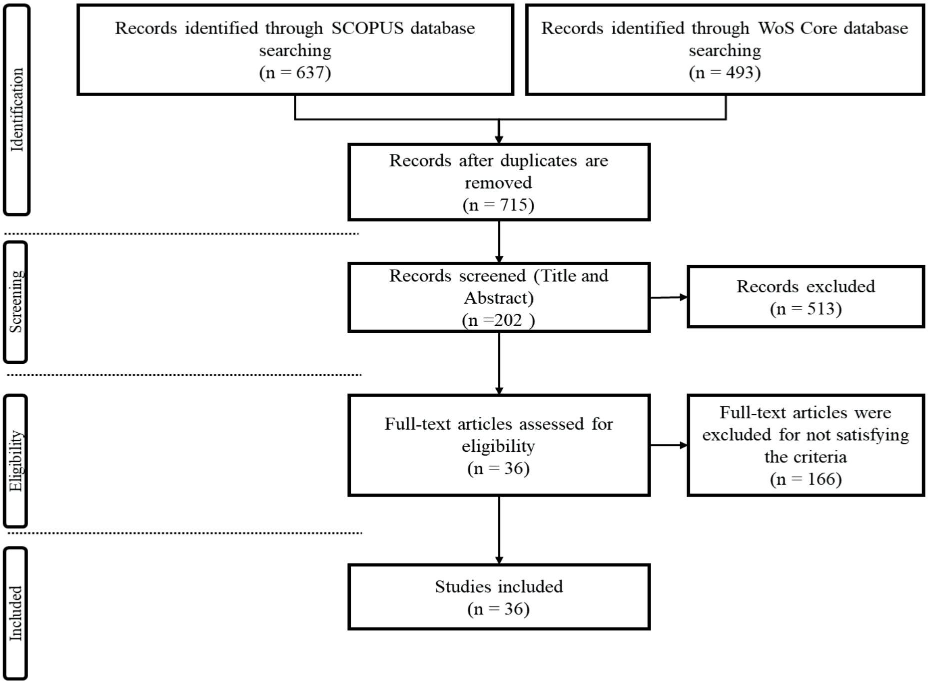

Fig. 1. PRISMA Chart.

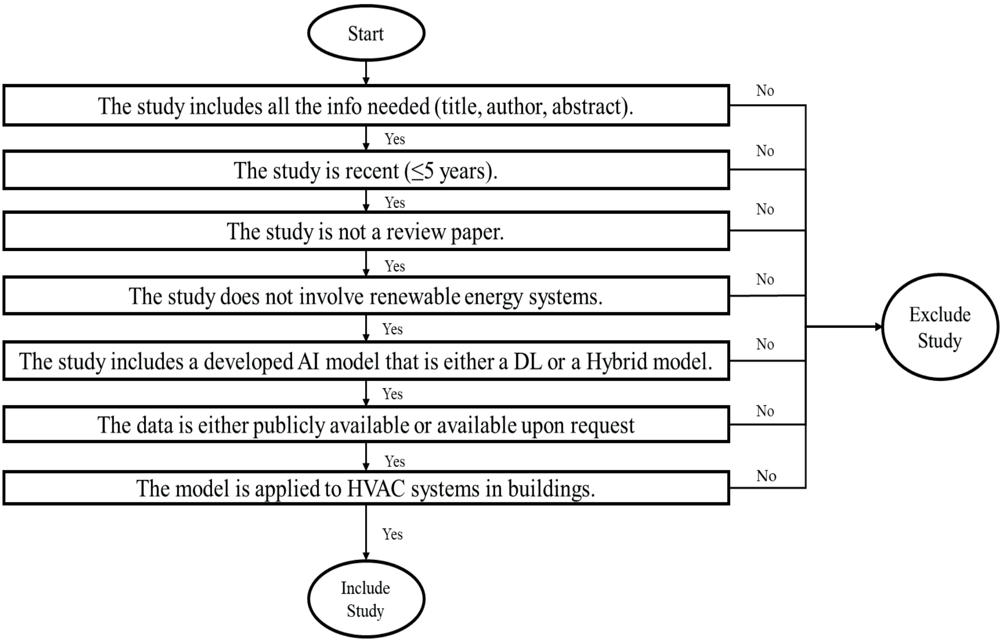

Fig. 2. Literature selection criteria.

- Only studies published within the last five years with complete bibliographic information were included to reflect recent advancements in the field. The criteria for inclusion were as follows:
- Studies must focus on HVAC systems in buildings and develop DL or Hybrid models. This choice was made because DL and hybrid approaches have shown significant potential in improving HVAC efficiency and require further evaluation [2].
- Review papers were excluded to focus on original research that develops and applies AI models rather than summarizing existing work.
- Renewable energy studies were excluded to avoid data challenges specific to storage and grid integration, ensuring a focused analysis on HVAC energy management.
- The analysed studies' data must be publicly available or accessible upon request. This criterion was applied to ensure transparency and reproducibility in AI model development.

After screening the selected papers, a structured analysis was conducted to extract data type, quality, availability, source, building type, data split, size, and data granularity. To enhance clarity, Fig. 3 illustrates the desired extracted data attributes.

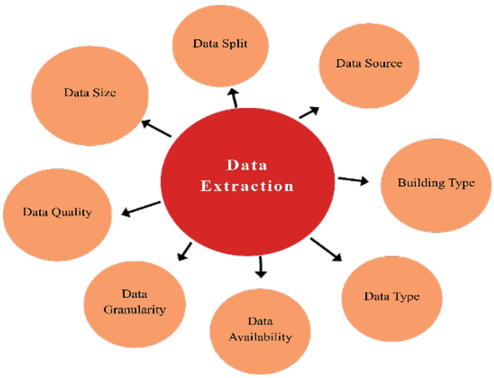

Fig. 3. Desired data extracted from the studies.

The characteristics depicted in Fig. 3 were chosen for their significance and impact on the development and training of AI models in HVAC systems. Data source, building type, and data type provide context regarding applicability. Furthermore, attributes like data size, granularity, and availability affect robustness and accuracy, while data availability and data split affect productibility and generalization [14]–[16]. This selection aligns with common practices in reviews and research related to AI and energy, ensuring a structured overview of data issues and potential avenues for improvement.

However, we were unable to derive meaningful conclusions, patterns, or insights regarding the data size and granularity from the collected dataset. Our analysis did not reveal any clear trends. Since we opted not to compare model performance, assessing their impact remains challenging. Even so, the reviewed papers exhibit diverse data sizes and granularities, ranging from large datasets with over 272 160 measurements [17] to smaller ones with 720 samples per fault class [18].

Granularity also differed, with most studies using hourly data [19], [20], while some reported different time scales [21], [22]. These differences highlight inconsistencies in dataset structures.

3. RESULTS

3.1. Data Availability

As presented in Fig. 1, 81 papers out of 166 were excluded from the studies in the full-text article assessment due to the data being unavailable publicly or upon request. This highlights a significant lack of transparency and credibility in many studies. Data accessibility is important as it ensures transparency, enables the replication of scientific results, and encourages data reuse by researchers and scientists. In the 36 papers analysed, 67 % of the datasets are available upon request, while 33 % are publicly available. Despite the data being available in both cases, the datasets available upon request pose some challenges, such as limited accessibility, where access will depend on the author's willingness to share and how fast they respond. Some researchers may need to fill out forms, agreements, or terms of use. This can make it challenging and time-consuming to get access to the data. On the other hand, publicly available datasets provide transparency and ease of access, allowing researchers to instantly download the datasets, encouraging more collaboration. This highlights the need for more open data source sharing and easy access to the data used in the studies to accelerate the research process for other researchers and practitioners.

While open data provides transparency and benefits researchers, it is also important to point out why some authors and data providers may have some constraints when it comes to data sharing. These constraints include privacy concerns, particularly when occupant-related data is involved, intellectual property restrictions, and the additional time and resources required to prepare the datasets for public release. Additionally, authors may fear data misuse or loss of competitive advantage [23], [24]. Addressing both user needs and data providers' constraints is necessary to advance open data practices in HVAC and building energy research.

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) [25] emphasizes that scientific knowledge should be "as open as possible," while recognizing that certain restrictions may be necessary to protect human rights, confidentiality, and intellectual property. By encouraging scientists to develop tools and methods for managing data, UNESCO aims to maximize data sharing and advance science [25].

Available Audited Data Sources

Open data sources, like Kaggle [26] for instance, provide datasets in different fields, including HVAC systems. This encourages collaboration and provides value to researchers who need data to train AI models. However, not all datasets on these platforms are audited or are reliable. Some datasets are incomplete and unclear, which is a problem, especially for HVAC systems, where building occupancy, climate, or maintenance activities can greatly affect energy use but are often not included.

TABLE 1. RECOMMENDATION OF AUDITED DATA SOURCES

Data Source Name	Description
[27] American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)	The majority of ASHRAE data sets are derived from research projects funded by ASHRAE. Technical committees and experts closely monitor these projects. The methods used to collect, process, and analyse the data are reviewed to ensure scientific credibility and accuracy.
[28] U.S. Department of Energy's Office of Scientific and Technical Information (OSTI)	Datasets provided by the US Department of Energy's Office of (OSTI) are generally reviewed and validated.

The AI modelling process will be affected if the data used to train an AI model is inaccurate or contains gaps. This can result in AI models making poor decisions, ultimately defeating the purpose of using AI for energy efficiency [9].

Reliable and audited data for AI applications in HVAC systems remains scarce. In order to solve this problem, we provide in Table 1 easily available and reliable data sources that can aid in the creation of strong and efficient AI models for building energy management.

3.2. Data Type

In AI-driven HVAC applications, data typically falls into three categories: real-world data [29], historical data gathered from past observations [30] and simulated/synthetic data generated through models or simulations to supplement real-world scenarios [31]. Table 2 and Fig. 4 present the number of studies from the reviewed papers using different data types.

TABLE 2. THE DATA TYPES USED IN DIFFERENT STUDIES

Data Type	Ref
Historical Data	[30], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61]
Real-Time Data	[29], [32], [33], [34], [62]
Synthetic/Simulated Data	[31], [32], [33], [34], [37], [46], [47], [52], [53], [57], [63]

Historical data has the most significant number of papers, which could indicate that this data type is easily accessible or widely available. This could also reflect its reliability for model training. However, AI models trained only on historical data impose limitations in adapting to dynamic or unforeseen conditions. Only five studies used real-time data. This is likely due to their complexity, latency issues, and high computational requirements.

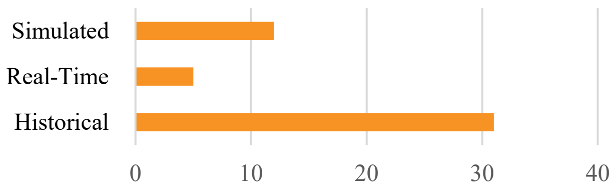

Fig. 4. Number of papers by data type.

Twelve studies used simulated (or synthetic) datasets. These allow the testing of AI models when data is inaccessible, limited, or requires extensive pre-processing. They are valuable for hypothesis testing and prototyping but could be less relevant to real-world complexities.

Combining multiple data types may help reduce challenges and provide deeper insights, improving the generalizability of the AI model as demonstrated in one study [32]. This research uses [32]. DL models that operate on different time scales, combining historical data and real-time data to predict room conditions and adjust controls, which helps the system reduce uncertainty and improve accuracy in different situations. Similarly, another study [33] utilizes EnergyPlus-based simulations in combination with real-world weather data, demonstrating that reinforcement learning-based HVAC control benefits from the synergy of simulated and real-time environmental data. The hybrid approach allows for training robust models capable of handling dynamic scenarios [33]. Moreover, the integration of computational simulations with operational data in [34] shows improved energy efficiency, reinforcing the importance of multiple data sources in refining AI models for real-world deployment.

These studies show that the integration of different data sources leads to more robust AI models. By including multiple data types like historical trends, real-time variations, and simulated scenarios, models can better adapt to new environments, reducing prediction errors and improving overall HVAC efficiency.

3.3. Data Quality

The quality of the data used in AI models highly influences their performance. Quality issues, such as noise and missing values, can affect model training and reduce accuracy [36], [39]. This section discusses the quality issues addressed in the analysed papers. The result of our analysis demonstrated that 69 % of the analysed studies did not discuss quality issues. This suggests that many researchers do not prioritize data quality as a key factor in their analysis and that it is not yet a standard practice.

Several studies address handling missing data. In study [36], the authors mentioned using imputation to fill in missing values, while another study [39] discussed having quality issues, including missing data and class imbalance problems. In study [43], linear interpolation was applied to fill in the missing quarters of hours for outdoor temperature and humidity data.

The authors in [46] mention noise due to mismatches in the sampling rate, a critical issue in HVAC systems with sensors operating in varying methods. The authors in [58] identify problems such as incorrect entries and mismatches in historical data, highlighting how poor-quality data affects the model's prediction performance. Both studies [44] and [60] address pre-processing steps to enhance data quality, such as interpolation, data cleaning to handle missing data, and outlier removal.

While data quality issues can have a significant impact on AI model performance, many studies do not report them, suggesting a gap in standard practice. However, those that do acknowledge these challenges employ various methods such as imputation, interpolation, data cleaning, and outlier removal to improve data reliability and increase model accuracy.

3.4. Data Sources by Building Type

The building types to data sources help to explore the patterns and challenges associated with data sources specific to the building type when applying ML models for energy-efficient HVAC systems. Table 3 shows that in residential buildings, simulated data, such as TRNSYS and Ecotect, appeared more frequently than real-world data. This could be due to privacy concerns and the high costs of retrofitting older buildings. However, given the small sample size, this observation should not be generalized to all residential contexts. On the other hand, for commercial buildings, the reviewed studies utilized structured and empirical datasets, such as large-scale surveys, measured building data, and ASHRAE projects; this enables more precise modelling.

Educational institutes and offices have leveraged diverse data sources from advanced data collection tools, such as Building Automation System (BAS), Building Management System (BMS), Computer Aided Monitoring System (CAMS), Internet of Things (IoT) sensors, district heating monitoring tools, thermal cameras, and other environmental sensors [38], [44], [60], [62].

TABLE 3. BUILDING TYPE AND DATA SOURCES

Ref.	Building Type	Data Source
[55]	Commercial	Commercial building in Singapore.
[59]	Commercial	Empirical data from a study by Borda *et al.* [44]
[41]	Commercial	ASHRAE Project RP-1312
[49]	Commercial	Survey of 50 Large commercial buildings, and measured data from one large commercial building.
[34]	Data Centre	Not specified
[62]	Educational Institute	The data source includes real-time frames captured by a camera for clothing classification.
[38]	Educational Institute	BAS, CAMS, IoT sensors, and a district heating energy consumption monitoring tool.
[29]	Educational Institute	Sensors (temperature, humidity, PIR) and thermal cameras in meeting rooms.
[46]	Educational Institute	BMS.
[40]	Laboratory	ASHRAE Project RP-1043
[56]	Not specified	MZVAV-2 dataset from the ASHRAE project 1312
[58]	Not specified	Historical maintenance data from a firm.
[36]	Not specified	FLEXLAB dataset from Lawrence Berkeley National Laboratory.
[31]	Not specified	BIMs of real-world buildings.
[45]	Not specified	ASHRAE Project RP-1043.
[47]	Not specified	EnergyPlus building template "5ZoneAirCooled", Weather data collected from Lester B. Pearson International Airport in Toronto.
[50]	Not specified	BIM data from CERTH's nZEB Smart Home in Greece, Passive House database, SRI assessment data.
[51]	Not specified	Historical operation data of the ACS.
[54]	Not specified	ASHRAE Project RP- 1312
[57]	Office	Simulated data from TRNSYS software.
[60]	Office	55 sensing devices
[33]	Office	Co-simulation environment based on Python and EnergyPlus.
[44]	Office	BMS of the Energy Center.
[48]	Office	EnergyPlus simulation model calibrated with actual building data.
[32]	Office	Simulations with rule-based control in a physics-based simulator.
[53]	Office	Generated using EnergyPlus simulation software.
[43]	Production Plant	HVAC parameters collected through Modbus protocol, outdoor environment parameters from OpenWeatherMap.Indoor environment parameters from sensors, smart meters, and supporting variables.
[30]	Residential	Simulated using TRNSYS software.
[42]	Residential	Not specified
[64]	Residential	Custom-designed sensor nodes that collect image and acoustic energy data.
[37]	Residential	Simulated data using Ecotect software.
[63]	Residential	Simulated data using Ecotect software.
[52]	Residential Commercial	BMS, energy meters, weather data, occupancy data, and historical consumption patterns.
[61]	Smart Buildings	BAS
[39]	Smart Buildings	ASHRAE RP-884 dataset
[35]	Sport Facility	BAS, CMMS, IoT vibration sensors, and IoT electric meters.

These advanced tools and sensors demonstrate the potential of innovative BAS applications. However, educational buildings are often controlled environments, which could limit these studies to educational buildings and make them less relevant to other building types. Smart buildings have also used advanced data collection tools such as BAS and other real-time monitoring systems [39]. Their advanced technology and infrastructure make them inherently suitable for ML implementation. Nevertheless, this indicates a data availability gap in buildings that lack the same technologies and capabilities.

Building types such as laboratories, manufacturing plants, data centres, and sports facilities were less frequently addressed. Furthermore, the unspecified building types in some studies, such as [45], [47], [50], [51], which used datasets from ASHRAE projects, Building Information Modeling (BIM), National Laboratory, and EnergyPlus, could limit the applicability and relevance of the studies. This could make it difficult to gain a proper understanding or adopt the presented solutions.

The diversification of data sources in older and less well-equipped buildings should be encouraged, as well as the standardisation of practices for documenting building types and data sources. This will improve the adaptability, scalability, and reliability of ML models across different building types, promoting more effective energy-efficient solutions.

3.5. Data Split

The data-splitting approaches for AI model training and testing vary across the studies. As shown in Fig. 5, the most common splits include splitting 80 % of the data for training and 20 % for testing [5], [17] or splitting 70 % for training and 30 % for testing [66], [67]. Some studies applied time-based splitting, where June to July data was used for training and August for testing [20]. Other studies, such as [19], [64], [68], did not report the data splits, which could undermine their credibility.

Differences in data splitting approaches across studies underscore the importance of selecting a careful method for reliable model evaluation. Transparent reporting of these choices is essential to improve the interpretability and trustworthiness of AI research.

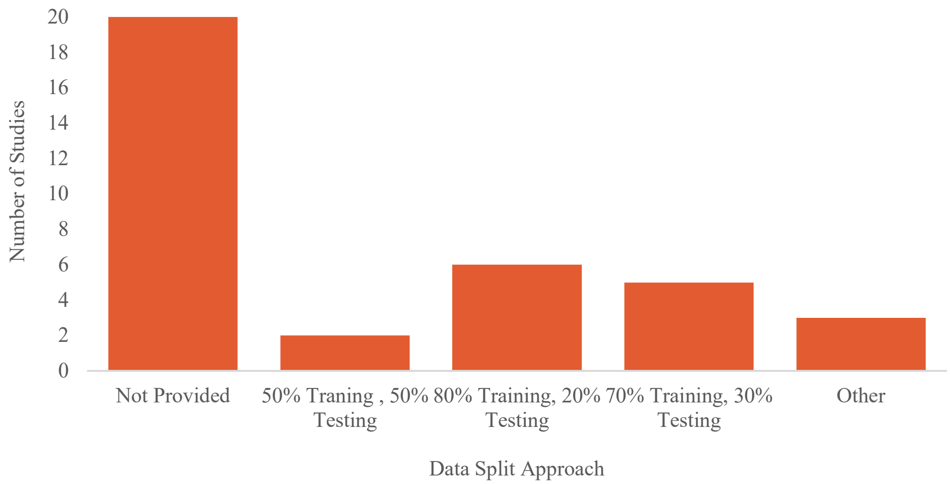

Fig. 5. Data split approaches in the studies.

4. Conclusion

This review highlights several key challenges and opportunities. First, the lack of publicly available, shared, and standardized datasets in HVAC and building energy is a major limitation, hindering research progress and the ability to replicate and validate AI models designed for buildings. Increasing open data sharing and improving access to audited datasets would accelerate advancements in this area. Moreover, combining building-specific data types, such as historical data on consumption patterns, real-time data (e.g., HVAC sensor data), and simulated data from tools like TRNSYS or EnergyPlus, can improve model generalizability and provide deeper insights. Another key challenge in HVAC-focused AI research is that data quality issues are often overlooked, with many studies failing to report noisy data caused by faulty sensors, missing values (such as sensor readings), or inconsistencies (like calibration errors). Addressing these issues can improve the reliability of HVAC AI models.

Among the residential buildings' studies examined, simulated data were often employed due to privacy concerns. However, given the limited number of cases, this observation should be interpreted with caution. Additionally, buildings with advanced data collection technologies, such as educational institutions and smart offices, have an advantage in data availability, while older and less equipped buildings have significant gaps. Encouraging data diversification and improved documentation practices can improve model adaptability, scalability, and reliability across different building types.

Furthermore, many of the available datasets are incomplete. A transparent and standardized way of reporting data would enhance the trustworthiness of AI-based energy management solutions. In addition, differences in data splitting approaches among studies highlight the importance of selecting a careful method and providing clear reporting to ensure reliable model evaluation and comparability.

Addressing these challenges through improved data availability, quality, diversity, verification, and transparent reporting will contribute to more robust and scalable AI solutions, ultimately advancing energy-efficient and sustainable building management. The study is limited in that it does not undertake an in-depth performance comparison on models trained on different data sources, as this was not in the intended scope of the study.

References

[1] Zhou S. L., Shah A. A., Leung P. K., Zhu X., Liao Q. A comprehensive review of the applications of machine learning for HVAC. *DeCarbon* 2023:2:100023. https://doi.org/10.1016/j.decarb.2023.100023
[2] Ali D. M. T. E., Motuzienė V., Džiugaitė-Tumėnienė R. AI-Driven Innovations in Building Energy Management Systems: A Review of Potential Applications and Energy Savings. *Energies* 2014:17(17):4277. https://doi.org/10.3390/en17174277
[3] Motuzienė V., Bielskus J., Lapinskienė V., Rynkun G. Office building's occupancy prediction using extreme learning machine model with different optimization algorithms. *Environmental and Climate Technologies* 2021:25(1):525–536. https://doi.org/10.2478/rtuect-2021-0038
[4] Shahrabani M. M. N., Apanaviciene R. An AI-Based Evaluation Framework for Smart Building Integration into Smart City. *Sustainability (Switzerland)* 2024:16(18):8032. https://doi.org/10.3390/su16188032
[5] Moayedi H., Mosavi A. Suggesting a Stochastic Fractal Search Paradigm in Combination with Artificial Neural Network for Early Prediction of Cooling Load in Residential Buildings. *Energies (Basel)* 2021:14(6):1649. https://doi.org/10.3390/en14061649
[6] Albatayneh A. The Share of Energy Consumption by End Use in Electrical Residential Buildings in Jordan. *Environmental and Climate Technologies* 2022:26(1):754–766. https://doi.org/10.2478/rtuect-2022-0058
[7] Tejani A. AI-Driven Predictive Maintenance in HVAC Systems: Strategies for Improving Efficiency and Reducing System Downtime. *International Journal of Advancements in Science & Technology* 2024:2:6–19.

[8] Zhou X. L., Du H., Xue S., Ma Z. J. Recent advances in data mining and machine learning for enhanced building energy management. *Energy* 2024:307:132636. https://doi.org/10.1016/j.energy.2024.132636
[9] Xiao Z. W. *et al.* Impacts of data preprocessing and selection on energy consumption prediction model of HVAC systems based on deep learning. *Energy and Buildings* 2022:258:111832. https://doi.org/10.1016/j.enbuild.2022.111832
[10] Huang J. J. *et al.* Real vs. simulated: Questions on the capability of simulated datasets on building fault detection for energy efficiency from a data-driven perspective. *Energy and Buildings* 2022:259:111872. https://doi.org/10.1016/j.enbuild.2022.111872
[11] Albatayneh A. The Share of Energy Consumption by End Use in Electrical Residential Buildings in Jordan. *Environmental and Climate Technologies* 2022:26(1):754–766. https://doi.org/10.2478/rtuect-2022-0058
[12] Mukhtar A., Hadwiger M., Wotawa F., Schweiger G. Reproducibility of Machine Learning-Based Fault Detection and Diagnosis for HVAC Systems in Buildings: An Empirical Study. Jul. 2025, [Online]. Available: http://arxiv.org/abs/2508.00880
[13] Moher D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Jul. 01, 2009, Public Library of Science. https://doi.org/10.1371/journal.pmed.1000097
[14] Himeur Y., Ghanem K., Alsalemi A., Bensaali F., Amira A. Artificial Intelligence based Anomaly Detection of Energy Consumption in Buildings: A Review, Current Trends and New Perspectives. *Applied Energy* 2021:287:116601. https://doi.org/10.1016/j.apenergy.2021.116601
[15] Ogundiran J., Asadi E., Gameiro da Silva M. A Systematic Review on the Use of AI for Energy Efficiency and Indoor Environmental Quality in Buildings. *Sustainability* 2024:16(9)3627. https://doi.org/10.3390/su16093627
[16] Khan W., Liao J. Y., Walker S., Zeiler W. Impact assessment of varied data granularities from commercial buildings on exploration and learning mechanism. *Applied Energy* 2022:319:119281. https://doi.org/10.1016/j.apenergy.2022.119281
[17] Taheri S., Ahmadi A., Mohammadi-Ivatloo B., Asadi S. Fault detection diagnostic for HVAC systems via deep learning algorithms. *Energy and Buildings* 2021:250:111275. https://doi.org/10.1016/j.enbuild.2021.111275
[18] Chen Z., Xiao F., Guo F. Similarity learning-based fault detection and diagnosis in building HVAC systems with limited labeled data *Renewable and Sustainable Energy Reviews* 2023:185:113612. https://doi.org/10.1016/j.rser.2023.113612
[19] Lei L., Shao S. L. Prediction model of the large commercial building cooling loads based on rough set and deep extreme learning machine. *Journal of Building Engineering* 2023:80:107958. https://doi.org/10.1016/j.jobe.2023.107958
[20] Seo B., Yoon Y., Lee K. H., Cho S. Comparative Analysis of ANN and LSTM Prediction Accuracy and Cooling Energy Savings through AHU-DAT Control in an Office Building. *Buildings* 2023:13(6):1434. https://doi.org/10.3390/buildings13061434
[21] Nelson W., Culp C. FDD in Building Systems Based on Generalized Machine Learning Approaches. *Energies (Basel)* 2023:16(4):1637. https://doi.org/10.3390/en16041637
[22] Lavanya R., Murukesh C., Shanker N. R. Microclimatic HVAC system for nano painted rooms using PSO based occupancy regression controller. *Energy* 2023:278(A):127828. https://doi.org/10.1016/j.energy.2023.127828
[23] Tenopir C. *et al.* Data sharing by scientists: Practices and perceptions. *PLoS One* 2011:6(6). https://doi.org/10.1371/journal.pone.0021101
[24] Bakar A. A., Yussof S., Ghapar A. A., Sameon S. S., Jørgensen B. N. A Review of Privacy Concerns in Energy-Efficient Smart Buildings: Risks, Rights, and Regulations. *Energies (Basel)* 2024:17(5):977. https://doi.org/10.3390/en17050977
[25] UNESCO. UNESCO Recommendation on Open Science. [Online]. [Accessed 17.03.2025]. Available: https://www.unesco.org/en/open-science/about
[26] Kaggle. [Online]. [Accessed 18.03.2025]. Available: https://www.kaggle.com/
[27] ASHRAE. [Online]. [Accessed 18.03.2025]. Available: https://www.ashrae.org/
[28] U.S. Department of Energy. [Online]. [Accessed 18.03.2025]. Available: https://www.osti.gov/
[29] Lavanya R., Murukesh C., Shanker N. R. Microclimatic HVAC system for nano painted rooms using PSO based occupancy regression controller. *Energy* 2023:278(A):127828. https://doi.org/10.1016/j.energy.2023.127828
[30] Gharsellaoui S., Mansouri M., Refaat S. S., Abu-Rub H., Messaoud H. Multivariate features extraction and effective decision making using machine learning approaches. *Energies (Basel)* 2020:13(3):609. https://doi.org/10.3390/en13030609
[31] Duan K., Suen C. W. K., Zou Z. Robot morphology evolution for automated HVAC system inspections using graph heuristic search and reinforcement learning. *Automation in Construction* 2023:153:104956. https://doi.org/10.1016/j.autcon.2023.104956
[32] Chen E. X., Han X., Malkawi A., Zhang R. Y., Li N. Adaptive model predictive control with ensembled multi-time scale deep-learning models for smart control of natural ventilation. *Building and Environment* 2023:242:110519. https://doi.org/10.1016/j.buildenv.2023.110519

[33] Yu L., Xu Z., Zhang T., Guan X., Yue D. Energy-efficient personalized thermal comfort control in office buildings based on multi-agent deep reinforcement learning. *Building and Environment* 2022:223:109458. https://doi.org/10.1016/j.buildenv.2022.109458

[34] Sarkar S. *et al.* Sustainability of Data Center Digital Twins with Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 2024:23832–23834. https://doi.org/10.1609/aaai.v38i21.30580

[35] Bouabdallaoui Y., Lafhaj Z., Yim P., Ducoulombier L., Bennadji B. Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. *Sensors* 2021:21(4):1044. https://doi.org/10.3390/s21041044.

[36] Taheri S., Ahmadi A., Mohammadi-Ivatloo B., Asadi S. Fault detection diagnostic for HVAC systems via deep learning algorithms. *Energy and Buildings* 2021:250:111275. https://doi.org/10.1016/j.enbuild.2021.111275.

[37] Moayedi H., Mosavi A. Double-Target Based Neural Networks in Predicting Energy Consumption in Residential Buildings. *Energies (Basel)* 2021:14(5):1331. https://doi.org/10.3390/en14051331

[38] Metsä-Eerola I., Pulkkinen J., Niemitalo O., Koskela O. On Hourly Forecasting Heating Energy Consumption of HVAC with Recurrent Neural Networks. *Energies (Basel)* 2022:15(14):5084. https://doi.org/10.3390/en15145084

[39] Rehman S. U., Javed A. R., Khan M. U., Awan M. N., Farukh A., Hussien A. PersonalisedComfort: a personalised thermal comfort model to predict thermal sensation votes for smart building residents. *Enterprise Information Systems* 2022:16(7):1852316. https://doi.org/10.1080/17517575.2020.1852316.

[40] Yan K., Zhou X. K. Chiller faults detection and diagnosis with sensor network and adaptive 1D CNN. *Digital Communications and Networks* 2022:8(4):531–539. https://doi.org/10.1016/j.dcan.2022.03.023.

[41] Chen Z., Xiao F., Guo F. Similarity learning-based fault detection and diagnosis in building HVAC systems with limited labeled data. *Renewable and Sustainable Energy Reviews* 2023:185:113612. https://doi.org/10.1016/j.rser.2023.113612

[42] Ma H., Xu L., Javaheri Z., Moghadamnejad N., Abedi M. Reducing the consumption of household systems using hybrid deep learning techniques. *Sustainable Computing: Informatics and Systems* 2023:38:100874. https://doi.org/10.1016/j.suscom.2023.100874

[43] Segala G., Doriguzzi-Corin R., Peroni C., Gerola M., Siracusa D. EECO: An AI-Based Algorithm for Energy-Efficient Comfort Optimisation. *Energies (Basel)* 2023:16(21):7334. https://doi.org/10.3390/en16217334

[44] Borda D., Bergagio M., Amerio M., Masoero M. C., Borchiellini R., Papurello D. Development of Anomaly Detectors for HVAC Systems Using Machine Learning. *Processes* 2023:11(2):535. https://doi.org/10.3390/pr11020535

[45] Shen C., Zhang H., Meng S., Li C. Augmented data driven self-attention deep learning method for imbalanced fault diagnosis of the HVAC chiller. *Engineering Applications of Artificial Intelligence* 2023:117(A):105540. https://doi.org/10.1016/j.engappai.2022.105540

[46] Nelson W., Culp C. FDD in Building Systems Based on Generalized Machine Learning Approaches. *Energies (Basel)* 2023:16(4):1637. https://doi.org/10.3390/en16041637

[47] Tian R. Q., Gomez-Rosero S., Capretz M. A. M. Health Prognostics Classification with Autoencoders for Predictive Maintenance of HVAC Systems. *Energies (Basel)* 2023:16(20):7094. https://doi.org/10.3390/en16207094

[48] Seo B., Yoon Y., Lee K. H., Cho S. Comparative Analysis of ANN and LSTM Prediction Accuracy and Cooling Energy Savings through AHU-DAT Control in an Office Building. *Buildings* 2023:13(6):1434. https://doi.org/10.3390/buildings13061434.

[49] Lei L., Shao S. L. Prediction model of the large commercial building cooling loads based on rough set and deep extreme learning machine. *Journal of Building Engineering* 2023:80:107958. https://doi.org/10.1016/j.jobe.2023.107958

[50] Siddique M. T., Koukaras P., Ioannidis D., Tjortjis C. SmartBuild RecSys: A Recommendation System Based on the Smart Readiness Indicator for Energy Efficiency in Buildings. *Algorithms* 2023:16(10):482. https://doi.org/10.3390/a16100482

[51] Chen S. H. *et al.* A novel machine learning-based model predictive control framework for improving the energy efficiency of air-conditioning systems. *Energy and Buildings* 2023:294:113258. https://doi.org/10.1016/j.enbuild.2023.113258

[52] Ntafalias A. *et al.* Smart buildings with legacy equipment: A case study on energy savings and cost reduction through an IoT platform in Ireland and Greece. *Results in Engineering* 2024:22:102095. https://doi.org/10.1016/j.rineng.2024.102095

[53] Okazawa K. *et al.* Evaluation of Deep Learning-Based Non-Intrusive Thermal Load Monitoring. *Energies (Basel)* 2024:17(9):2012. https://doi.org/10.3390/en17092012

[54] Yan K., Lu C., Ma X., Ji Z., Huang J. Intelligent fault diagnosis for air handing units based on improved generative adversarial network and deep reinforcement learning. *Expert Systems with Applications* 2024:240:122545. https://doi.org/10.1016/j.eswa.2023.122545

[55] Sulaiman M. H., Mustaffa Z. Chiller energy prediction in commercial building: A metaheuristic-Enhanced deep learning approach. *Energy* 2024:297:131159. https://doi.org/10.1016/j.energy.2024.131159

[56] Huang Y., Coursey A., Quinones-Grueiro M., Biswas G. Time-Series Few Shot Anomaly Detection for HVAC Systems. *IFAC-PapersOnLine* 2024:58(4):426–431. https://doi.org/10.1016/j.ifacol.2024.07.255

[57] Soleimani M., Irani F. N., Yadegar M., Davoodi M. Multi-objective optimization of building HVAC operation: Advanced strategy using Koopman predictive control and deep learning. *Building and Environment* 2024:248:111073. https://doi.org/10.1016/j.buildenv.2023.111073

[58] Mohamed A. G., Ghaly J. E., Marzouk M. Revolutionizing semantic integration of maintenance cost prediction for building systems using artificial neural networks. *Journal of Building Engineering* 2024:96:110416. https://doi.org/10.1016/j.jobe.2024.110416

[59] Bian J., Wang J., Yece Q. A novel study on power consumption of an HVAC system using CatBoost and AdaBoost algorithms combined with the metaheuristic algorithms. *Energy* 2024:302:131841. https://doi.org/10.1016/j.energy.2024.131841

[60] Bucarelli N., El-Gohary N. Sensor deployment configurations for building energy consumption prediction. *Energy and Buildings* 2024:308:113888. https://doi.org/10.1016/j.enbuild.2024.113888

[61] Safari A., Kharrati H., Rahimi A. A hybrid attention-based long short-term memory fast model for thermal regulation of smart residential buildings. *IET Smart Cities* 2024:6(4):361–371. https://doi.org/10.1049/smc2.12088

[62] Wei Z., Calautit J. K., Wei S., Tien P. W. Real-time clothing insulation level classification based on model transfer learning and computer vision for PMV-based heating system optimization through piecewise linearization. *Building and Environment* 2024:253:111277. https://doi.org/10.1016/j.buildenv.2024.111277

[63] Moayedi H., Mosavi A. Suggesting a Stochastic Fractal Search Paradigm in Combination with Artificial Neural Network for Early Prediction of Cooling Load in Residential Buildings. *Energies (Basel)* 2021:14(6):1649. https://doi.org/10.3390/en14061649

[64] Jacoby M. *et al.* WHISPER: Wireless Home Identification and Sensing Platform for Energy Reduction. *Journal of Sensor and Actuator Networks* 2021:10(4):71. https://doi.org/10.3390/jsan10040071

[65] HARMONAC. [Online]. [Accessed 18.03.2025]. Available: http://www.harmonac.info/

[66] Moayedi H., Mosavi A. Double-Target Based Neural Networks in Predicting Energy Consumption in Residential Buildings. *Energies (Basel)* 2021:14(5):1331. https://doi.org/10.3390/en14051331

[67] Segala G., Doriguzzi-Corin R., Peroni C., Gerola M., Siracusa D. EECO: An AI-Based Algorithm for Energy-Efficient Comfort Optimisation. *Energies (Basel)* 2023:16(21):7334. https://doi.org/10.3390/en16217334

[68] Ntafalias A. *et al.* Smart buildings with legacy equipment: A case study on energy savings and cost reduction through an IoT platform in Ireland and Greece. *Results in Engineering* 2024:22:102095. https://doi.org/10.1016/j.rineng.2024.102095