Artificial Intelligence Review
https://doi.org/10.1007/s10462-022-10286-2

AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives

Yassine Himeur[1,2] · Mariam Elnour[1] · Fodil Fadli[1] · Nader Meskin[3] · Ioan Petri[4] · Yacine Rezgui[4] · Faycal Bensaali[3] · Abbes Amira[5,6]

Abstract
In theory, building automation and management systems (BAMSs) can provide all the components and functionalities required for analyzing and operating buildings. However, in reality, these systems can only ensure the control of heating ventilation and air conditioning system systems. Therefore, many other tasks are left to the operator, e.g. evaluating buildings' performance, detecting abnormal energy consumption, identifying the changes needed to improve efficiency, ensuring the security and privacy of end-users, etc. To that end, there has been a movement for developing artificial intelligence (AI) big data analytic tools as they offer various new and tailor-made solutions that are incredibly appropriate for practical buildings' management. Typically, they can help the operator in (i) analyzing the tons of connected equipment data; and; (ii) making intelligent, efficient, and on-time decisions to improve the buildings' performance. This paper presents a comprehensive systematic survey on using AI-big data analytics in BAMSs. It covers various AI-based tasks, e.g. load forecasting, water management, indoor environmental quality monitoring, occupancy detection, etc. The first part of this paper adopts a well-designed taxonomy to overview existing frameworks. A comprehensive review is conducted about different aspects, including the learning process, building environment, computing platforms, and application scenario. Moving on, a critical discussion is performed to identify current challenges. The second part aims at providing the reader with insights into the real-world application of AI-big data analytics. Thus, three case studies that demonstrate the use of AI-big data analytics in BAMSs are presented, focusing on energy anomaly detection in residential and office buildings and energy and performance optimization in sports facilities. Lastly, future directions and valuable recommendations are identified to improve the performance and reliability of BAMSs in intelligent buildings.

Keywords Artificial intelligence · Big data analytics · Building automation and management system · Deep learning · Evaluation metrics · Computing platforms

✉ Yassine Himeur
yassine.himeur@qu.edu.qa

Extended author information available on the last page of the article

Published online: 15 October 2022

Abbreviations

A3C	Asynchronous advantage actor-critic
AC	Actor-critic
AE	Auto-encoder
AHU	Air handling unit
AMI	Advanced metering infrastructures
ANN	Artificial neural network
AR	Auto-regressive
Adaboost	Adaptive Boosting
BAAKR	Bagged auto-associative kernel regression
BAMS	Building automation and management system
BBN	Bayesian belief networks
BBRT	Bootstrap bagging of regression trees
BDT	Binary decision tree
BESN	Bagged echo state network
BMCDT	Binary multiclass-classification decision tree
BN	Bayesian networks
BNN	Bagging neural network
BPNN	Back propagation neural network
BRT	Bagged regression tree
BSA	Backtracking search algorithm
BT	Bagged tree
BiLSTM	Bidirectional long short-termmemory
CAGR	Compound annual growth rate
CART	Classification and regression tree
CCTV	Closed-circuit televi-sion
CDT	CART decision tree
CIFG	Coupled input and forget gate
CNN	Convolutional neural network
CPS	Cyber-physical system
CRBM	Conditional restricted Boltzmann machines
CRF	Conditional random fields
CRGP	Compact regression Gaussian process
CRT	Completely-random tree
CUSP	Computational urban sustainability platform
ConvLSTM	Convolutional LSTM
DBN	Deep belief network
DBSCAN	Density-based spatial clustering of applications with noise
DCRF	Deep cascade forest
DDPG	Deep deterministic policy gradient
DDoS	Distributed denial of service
DFNN	Deep feed forward neural networks
DL	Deep learning
DNN	Deep neural networks
DQL	Deep Q-learning
DRED	Dutch resi-dential energy dataset
DRL	Deep reinforcement learning
DT	Decision tree
DWT	Discrete wavelet transform

EBT	Ensemble bagging tree
ELM	Extreme learning machine
FCM	Fuzzy C-means
FCRBM	Factored conditional restricted Boltzmann machines
FFNN	Feed-forward neural network
GAM	Generalized additive models
GBM	Gradient boosting machine
GBRT	Gradient boosting regression tree
GLRM	Generalized linear regression model
GPRM	Gaussian process regression model
GRU	Gated recurrent units
GSA	Gravitational search algorithm
GSD	Generated sampled data
HCA	Hierarchical cluster analysis
HEP	Hydro-electric power
HMM	Hidden Markov model
HVAC	Heating ventilation and air conditioning system
ICT	Information and communication technology
IEQ	Indoor environmental quality
Isomap	Isometric feature mapping
KNN	K-nearest neighbors
KPCA	Kernel principal component analysis
LDA	Linear discriminant analysis
LOF	Local outlier factor
LR	Linear regression
LR	Logistic regre-ssion
LSSVR	Least squares support vector regression
LSTM	Long short-term memory
MC	Monte Carlo
MCSVM	Multi-class support vector machines
MDA	Multiple discriminant analysis
MDS	Multidimensional scaling
MLP	Multi-layer perceptron
MLR	Multiple linear regression
MPC	Model predictive control
MSA	Multiplicative season algorithm
MetaFA	Metaheuristic firefly algorithm
NB	Naive Bayes
NILM	Non-intrusive load monitoring
NN-SAE	Neural network-based-supervised auto-encoder
NN	Neural network
OCSVM	One-class support vector machine
PCA	Principal component analysis
PG	Policy gradient
PLS	Partial least square
PPO	Proximal policy optimization
QDA	Quadratic discriminant analysis
QL	Q-leaning
QR	Quantile regression

RBDT	Regression binary decision tree
RBFNN	Radial basis function neural network
RBM	Restricted Boltzmann machines
RF	Random forest
RFT	Regression fitting
RL	Reinforcement learning
RNN	Recurrent neural networks
RT	Regression tree
ResNet	Deep residual network
SAE	Memory-gated RNN-based autoencoders
SAE	Sparse autoencoders
SARIMA	Seasonal autoregressive integrated moving average
SARSA	State-action-reward-state-action
SDN	Software-defined networks
SGPR	Stepwise Gaussian processes regression
SSL	Semi-supervised learning
SSNN	Semi-supervised neural network
SVM	Support vector machine
SVR	Support vector regression
TRL	Tradi-tional reinforcement learning
TSVD	Truncated singular value decomposition
TTS	Text-to-speech
VAE	Variational autoencoders
VHT	Vertical hoeffding tree
WQP	Water quality prediction
XGBM	Extreme gradient boosting machine
XGBoost	EXtreme gradient boosting
mLSTM	Multip-licative LSTM
t-SNE	t-Distributed stochastic neighbor embedding

1 Introduction

1.1 Preliminary

Building automation and management systems (BAMSs) are intelligent systems of both hardware and software, connecting heating ventilation and air conditioning system (HVAC) systems, lighting, security, and other systems to communicate on a single platform. That said, BAMSs deliver crucial information to operators and/or users on the operational performance of buildings, which aim at promoting energy efficiency and optimizing water consumption, enhancing the safety and comfort of the occupants, reducing maintenance costs, extending the life cycle of the utilities, etc (Ippolito et al. 2014). This is possible by networking a plethora of sensors and components responsible for the monitoring and operation of mechanical, security, fire, lighting, HVAC and humidity control and ventilation systems (Su and Wang 2020).

With the broad utilization of information and communication technologies (ICTs), sensing and measurement technologies along with the cloud computing, big data storage and data analytics, conventional BAMSs are being revolutionized. Vast quantities of building

automation and management data are produced, gathered and saved (Sardianos et al. 2020; Himeur et al.). This has offered an excellent opportunity for implementing big data mining and analysis in BAMSs. In this context, as the quantity of data collected in BAMSs is enormous, the "big data" phenomena is surfacing this field and revolutionizing the way we manage data by using AI-big data analytics tools (Quinn et al. 2020, Himeur et al.). Accordingly, with advanced sensing and metering technologies in BAMSs, data split into multiple modalities and many variables can create a comprehensive source of information to analyze. This allows for more targeted analysis, but also means that more powerful, intelligent, and sophisticated tools are needed to identify the most enormous patterns/variables (Muntean et al. 2021). As a consequence, the big data analytics market in the building energy sector is expected to grow at a Compound annual growth rate (CAGR) of 11.28%, during the forecast period, 2021–2026.[1] Data collection in the building industry is becoming all-embracing. This wealth of big data allows informed data-driven decision-making by designers, facilities managers, and owners during building design, operation, and retrofit (Berger et al.). On the other hand, for existing or outdated buildings to make full use of the services offered by the flourishing data analytics market, the necessary enhancement to the existing system for deploying the new technology must be addressed and sorted out (Varlamis et al. 2022). The main challenges of gathering and analyzing data of old buildings are the outdated technologies and the conventional error-prone data collection means (Jia et al. 2019; Al Dakheel et al. 2020). Nevertheless, data analytics can assist in designing and implementing the new system adaptation and existing system renovation (Elnour et al. 2022).

Besides, it is of utmost importance to know the current state of AI-based building automation before presenting the actual study concerning user input, demand, response, energy-saving, and automation. In this respect, it is obvious that AI adds new dimensions to building automation environments by enabling autonomous data analysis for operation optimization. Therefore, many AI-based contributions have recently emerged as key solutions for (i) predicting building occupancy, (ii) forecasting thermal comfort, (iii) boosting energy saving, and (iv) enabling demand-side response (Himeur et al. 2020). Additionally, as mentioned in previous studies, (O'Grady et al. 2021) people can spend up to 90 percent of their lives in buildings; this highlights the importance of user input, behavioral data, and behavioral analytics for optimizing and automating building operations. To that end, a significant research effort is ongoing to develop AI-based behavioral change technologies to promote energy saving in residential and office buildings (Sayed et al., Varlamis et al. 2022), understand consumers demand patterns for successful demand response development (Cruz et al. 2021; Pratt and Erickson 2020), optimizing occupants' thermal comfort (Zheng et al. 2022), transforming water management (Doorn 2021), improving fault detection and diagnosis (Yun et al. 2021), etc. Moreover, AI-based big data analytics are contributing to building automation by making BAMSs self-learning, self-configuring and self-diagnosing, and self-commissioning (Katipamula 2019). Additionally, using AI-based analytics can adapt existing building systems to promote the deployment of BAMSs with fewer investments from building owners.

From another hand, as AI models are very competent to learn common human error patterns, their use in big data analytics is significant. They can (i) detect and resolve possible flaws in datasets, (ii) learn by watching how the operators and users interact with

[1] https://www.researchandmarkets.com/reports/4774956/big-data-analytics-market-in-the-energy-sector.

the analytics programs, and identify anomalies and surface unexpected insights from large-scale datasets fast (Mahmud et al. 2020; Diamantoulakis et al. 2015). In this context, AI models assist operators and users of BAMSs to perform the different tasks related to the big data cycle, among them the operations of collecting, pre-processing, aggregating, storing, analyzing and extracting various kinds of features (Hu and Vasilakos 2016; Bode et al. 2019). Moving on, the integration of AI-big data analytics can (i) optimize energy and operational efficiency, (ii) automate monitoring and control through wireless platforms, (iii) provide quick and better decision making, (iv) smartly control the facility and reduce risk failures, (v) lower life cycle costs, and (vi) increase safety and security measures with ease (Aghemo et al. 2014; Zhou and Yang 2016; Aste et al. 2017).

1.2 Paper contributions

Due to the importance of using AI-big data analytics in BAMSs, a plethora of works have been proposed to (i) address different challenges, (ii) improve and automate building operation, and (iii) optimize building user experience. In addition, different reviews have been introduced to discuss the advances made in this research topic, such as (Zhang et al. 2021; Molina-Solana et al. 2017; Zhao et al. 2020). However, most of them have only focused on addressing one task at a time, e.g., energy management, rather than covering multiple BAMS tasks together (e.g., water management, occupancy detection, comfort optimization, fault diagnosis and anomaly detection (FDAD), etc.) (Sun et al. 2020; Wang et al. 2021; Fan et al. 2018). To that end, we present in this paper a comprehensive systematic survey reflecting the latest developments in the field of AI-big data analytics and their utilization in BAMSs from different perspectives. For example, Zhang et al. (2021) discuss sensor impact verification and evaluation for FDAD in energy systems, while Molina et al. (2017) review the contributions of data science for building energy management issues. Moving on, data mining strategies used for building energy management are overviewed in Zhao et al. (2020). Similarly, in Sun et al. (2020), data-driven techniques for energy prediction in buildings are described. Besides, Wang et al. (2021) focus on studying the practical problems related to implementing ML models for building energy efficiency. It also investigates the commitment of existing studies to comfort and energy saving (i.e., Save energy with/without compromising thermal comfort). Moreover, in Fan et al. (2018), unsupervised data mining methodologies for energy efficiency improvement are analyzed. Lastly, in Pinto et al. (2022), Pinto et al. discuss the roles of transfer learning integration for smart buildings and systems.

To that end, we present in this paper a comprehensive survey reflecting the latest developments in the field of AI-big data analytics and their utilization in BAMSs from different perspectives. Thus, we first introduce a generic taxonomy for classifying AI-big data analytics frameworks based on various criteria, including the learning method, building environment, computing platform, and application or challenge addressed. Typically, an overview of existing works and discussions is presented, highlighting some of the challenges, limitations, and shortcomings. Then, three case studies are presented illustrating the use of AI-big data analytics for critical concerns in the buildings sector, that is, energy efficiency and management, to provide the reader with insight into real-world applications. The optimization of energy consumption in buildings has been a hot research topic recently[2] in

[2] https://www.energy.gov/sites/prod/files/2017/03/f34/qtr-2015-chapter5.pdf.

terms of efficient planning, proper management, and preventive maintenance. Lastly, future directions to ease the use of AI-big data analytics models in BAMSs and improve their feedback are derived. To summarize, the contributions of the presented work are manifold:

- Providing a thorough review covering the general use of AI-big big data analytics in BAMSs and shedding light on their increasing importance for developing efficient and smart BAMSs.
- Presenting a well-designed taxonomy of existing AI-big data analytics frameworks, which helps in understanding intriguing relationships between various concepts and variables in the field. Different criteria have been adopted when analyzing existing frameworks, including the learning method, building environment, computing platform, application, etc.
- Conducting a critical analysis and discussion to (i) extract diverse relevant lessons that are learned from overviewed works; and (ii) highlight open issues and current challenges, among them data scarcity, data benchmarking, security and privacy, scalability and interoperability and real-time big data intelligence.
- Presenting three case studies that describe the use of AI-big data analytics in BAMSs for buildings energy management and optimization, such that the first two case studies demonstrate unsupervised and supervised energy anomaly detection strategies in residential and office buildings, and the third one is about energy and performance optimization in sports facilities.
- Deriving a set of future research and development directions that attract considerable interest in the near and far future, and help in improving the performance and reliability of BAMSs.

Table 1 outlines some of the main differences between the actual review and other survey studies. It also sheds light on some of the main contributions addressed by this review compared to the others in terms of overviewed resources (i.e., ML tools and computing platforms), application scenarios, discussed challenges (i.e., security issues), evaluation metrics, case studies, and proposed future directions (i.e., multimodal data analysis, in-situ sensor calibration in BAMSs, smart building digital twins, blockchain edge analytics, etc.).

1.3 Review methodology

A well-established review methodology is adopted in this paper, where we first conduct a comprehensive literature search in the most popular scientific databases, including Scopus, Elsevier, Wiley, and IEEE. Following, most of the works that deal with the use of AI-big data analytics for BAMSs are included in this study. Many keywords and their combination are then used in the search, e.g., ”building automation and management systems”, ”big data analytics”, ”artificial intelligence”, ”machine learning”, ”deep learning”, ”transfer learning”, ”energy prediction in buildings using machine learning”, ”thermal comfort in building using machine learning”, ”fault diagnosis and anomaly detection in buildings”, ”security in building automation and management systems”, etc. Therefore, research studies introduced between January 2015 and February 2022

Table 1 Comparison of the proposed survey's contributions against other existing related related review studies

	ML tools	Computing platforms	Application scenarios	Security issues	Evaluation metrics	Case studies	Future directions			
							Multimodal data analysis	Blockchain	Edgeanalytics	3D point clouds
Zhang et al. (2021)	✗	✗	FDAD	✗	✗	✗	✗	✗	✗	✗
Molina-Solana et al. (2017)	✓	✓	Energy management	✓	✗	✗	✗	✗	✗	✗
Zhao et al. (2020)	✓	✓	Energy management	✗	✗	✗	✗	✗	✗	✗
Sun et al. (2020)	✓	✗	Energy prediction	✗	✗	✗	✗	✗	✗	✗
Wang et al. (2021)	✓	✗	Energy efficiency	✗	✗	✗	✓	✗	✗	✗
Fan et al. (2018)	✓ (unsupervised)	✗	Energy efficiency	✗	✗	✗	✗	✗	✗	✗
Pinto et al. (2022)	✓ (transfer learning)	✗	Load prediction, system control, occupancy detection, building dynamics	✗	✓	✗	✗	✗	✗	✗
Our paper	✓	✓	Energy management FDAD, IEQ, security and safety, occupancy detection and water management	✓	✓	✓	✓	✓	✓	✓

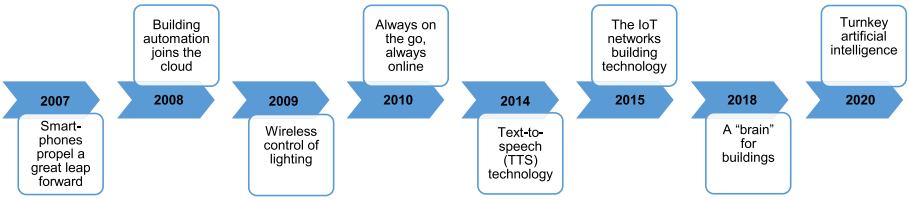

Fig. 1 The significant progress made in the development of BAMSs during the last two decades

are discussed in this framework. This period has arbitrarily been selected to evaluate the recent and pertinent contributions. Typically, this framework discusses English-written peer-reviewed journal articles, conference proceedings papers, and book chapters. The selection process adopted in this review relies on adhering to the specifications of the PRISMA (Moher et al. 2009), which is a practical and efficient approach for writing survey studies. Concretely, a search was performed for the last seven years (January 2015–February 2022). To eliminate duplicate references, a reference manager software was utilized, and only the remaining frameworks have then been considered after filtering them by their titles, keywords, and abstracts.

In addition to reviewing existing AI-big data analytics contributions for BAMSs, three case studies are also included in this article to provide the reader with more explanations about using AI tools in tackling the buildings' energy consumption question in terms of (i) unsupervised energy anomaly detection, (ii) supervised energy anomaly detection, and (iii) energy and performance optimization for sports facilities.

1.4 Organization of the paper

The rest of this paper is structured as follows. Section 2 highlights the significant advances made in the development of BAMSs. Section 3 provides an overview of the interdisciplinary AI-big data analytics research in BAMSs following a well-defined taxonomy. Section 4 evaluates and critically analyses overviewed frameworks to identify the open issues and current challenges. Section 5 presents three case studies that describe the use of AI-big data analytics in BAMSs for energy anomaly detection in residential and office buildings, and energy optimization in sports facilities. Moving on, Sect. 6 presents the future directions for improving the performance of BAMSs. Finally, conclusions and significant findings are summarized in Sect. 7.

2 Evolution of BAMSs

In the last decades, BAMSs have been rapidly developed; indeed, from the 1950s to 1990s, they have been transformed from pneumatics to electronics then open protocols (e.g. BACnets). Moving forward, with the digitization era, BAMSs have further progressed by (i) integrating more powerful and smart technologies, (ii) becoming easier to implement in different kinds of buildings, and (iii) using high-quality softwares to aid the users in getting the most pertinent information from their buildings. Indeed, digitization has significantly

accelerated with the launch of smartphones, where these devices have abruptly replaced mobile communicators, cell phones, etc. and become more practical in many real-world applications as they deliver different benefits, such as supporting apps.

2.1 Progress made during the two last decades

We briefly described in this section the most significant achievements made in the development of BAMSs during the last two decades, which can be summarized as portrayed in Fig. 1. Typically, in 2008, it became possible to virtualize BAMSs in data centers, and hence receives greater security and availability and enables more flexible access to buildings' data. In 2009, WiFi was integrated to BAMSs to help in flexibly and remotely monitoring appliances in households and commercial centers (Wang 2009). Following, in 2010, due to the growing utilization of smartphones and tablet computers in smart city applications to control smart and location-based products, BAMSs have been also positively influenced through developing more sophisticated and portable BAMSs solutions (Aste et al. 2017).

By 2014, audio files had been deployed in BAMSs using the text-to-speech (TTS) technology. The latter enabled to support preventive inspections and maintenance, work contracts, service requests, work contracts, equipment audits, etc (Mundt et al. 2014). In 2016, IoT began to significantly influence the society after that IoT devices found their way into the building management sector, where six billion devices were installed and more than 31 billion were expected in 2020 (Aste et al. 2017). Explicitly, the importance of automation has been increased in both existing and new buildings. Following, in 2018, commercial buildings have progressively evolved into smart buildings and routines have been improved and/or automated in BAMSs, resulting in enhanced comfort and efficiency (Markoska and Lazarova-Molnar 2018; Lee and Karava 2020). Lastly, in 2020, AI joined the spectrum of BAMSs to help in early fire detection, and energy demand prediction. Moreover, it becomes possible to identify behavioral change through the analysis of real-time data. Thus, BAMSs learn from experiences and historical data for automatically adjusting the indoor conditions (Yaïci et al. 2021).

2.2 Big data sources

This section discusses and describes the essential sources of heterogeneous big data used to implement AI-big data analytics in BAMSs. Indeed, developing and accelerating the advance and deployment of BAMSs require the installation of a large number of smart sensors, smart meters, and other measurement devices in the different parts of each building, which helps in (i) increasing the observability of its transient and dynamic events, and (ii) gather actual data related to the diverse functionalities of the building. This will later help the AI-big data analytics in accurately analyzing this data and extracting pertinent features and therefore facilitating the operation and monitoring of all building technology, especially in larger buildings. Figure 2 portrays the overall architecture of a BAMS and its principal data sources. The control module is the central brain of the BAMS, and most of the controllers are built using the industry standard BACnet protocols in addition to Konnex (KNX); an open communication standard for commercial and domestic building automation, LonWorks; a standardized bus system used in centralized and decentralized building automation control (Merz et al. 2018),

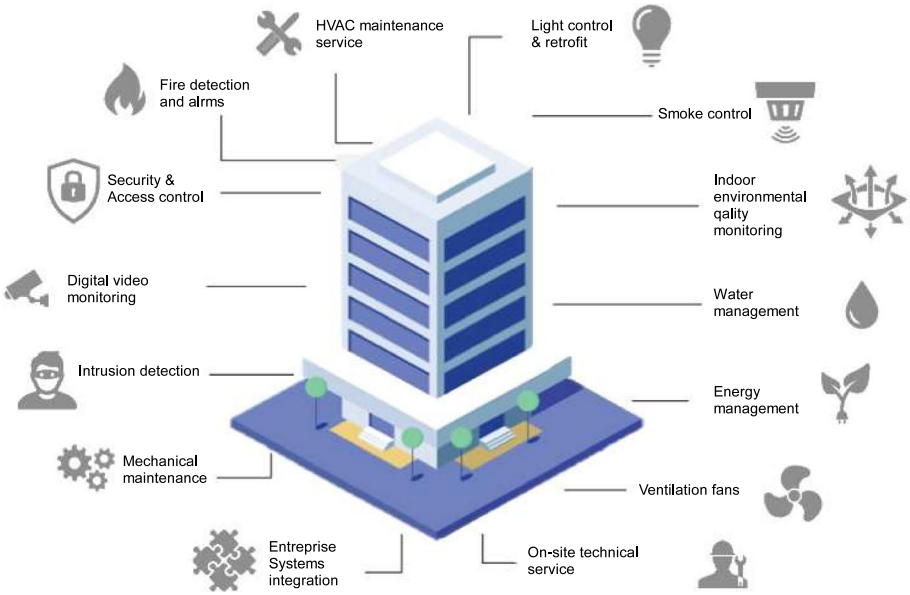

Fig. 2 Principal services of a BAMS system

and Modbus; a network communication protocol for connecting electronic equipment in industrial automation systems. Overall, a BAMS can provide various services to control (i) heating and cooling, (ii) lighting, (iii) security, (iv) access control, (v) fire and life safety, and (vi) elevators and escalators.

3 Overview of AI-big data analytic frameworks

3.1 Overall taxonomy

To understand the challenges related to AI-big data analytics in BAMSs, it is essential to perform a generic taxonomy of existing AI-big data analytics techniques used for monitoring the smart buildings. Specifically, Fig. 3 provides a structured analysis framework that helps in overviewing existing techniques and shedding the light on the organization of the presented framework.

3.2 AI-learning process

The first step in any AI process is system learning. This can take four primary forms: supervised learning, unsupervised learning, semi-supervised, and reinforcement learning. In this section, we present an overview of existing AI learning architectures used to improve the performance of BAMSs.

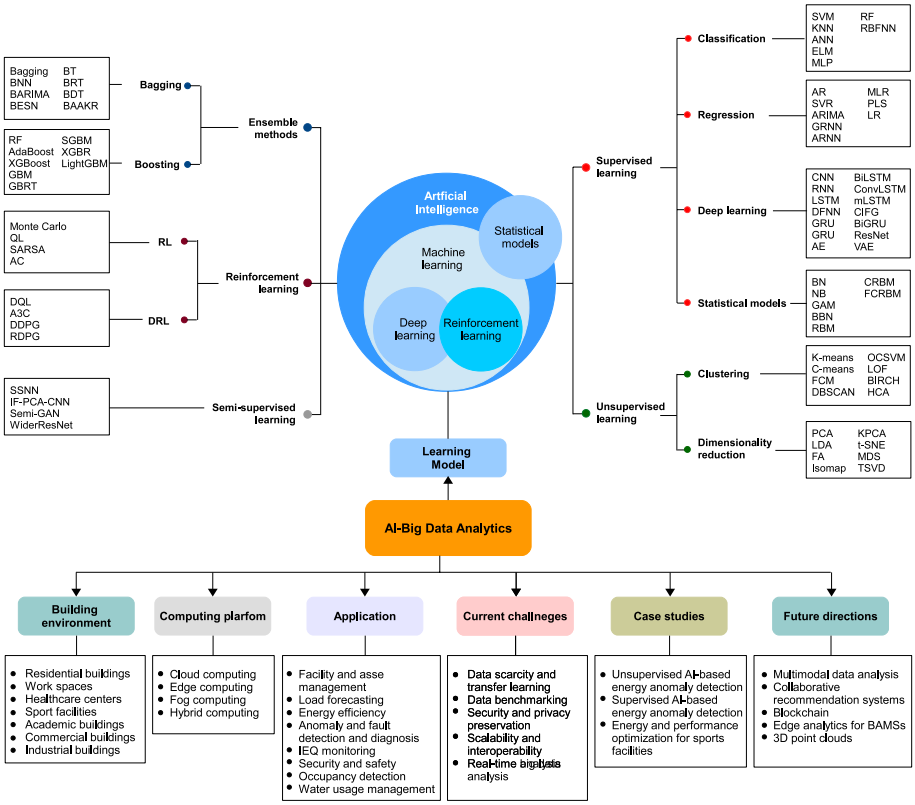

Fig. 3 Taxonomy of existing AI-big data analytics frameworks

3.2.1 Unsupervised learning (U)

Unsupervised learning learns from raw data without prior knowledge and mainly deals with unlabeled datasets. Although it does not need to annotate data as supervised learning, the learning phase can be more computational as all the possibilities are checked. The accuracy is lower since there are no corresponding outputs (labels) (Himeur et al. 2021).

3.2.1.1 U1. Clustering It is a category of ML algorithms used for separating data (e.g. energy consumption observations, ambient conditions, etc.) into different classes or clusters following a specific goal. Clustering algorithms usually pertain to one of the following groups, i.e. hybrid, fuzzy-based, model-based, and density-based approaches. Using the clustering process facilitates the classification tasks when dealing with various problems, such as anomaly detection of energy consumption, indoor environmental quality (IEQ) monitoring and detection of pollutants, detection of abnormal water consumption, etc.

K-means, C-means and fuzzy C-means (FCM) were among the most investigated clustering approaches. They have been applied for non-intrusive load monitoring (NILM) and appliance identification (Ji et al. 2019; Zhang et al. 2020), energy performance evaluation and ranking in working spaces (Sun and Yu 2021), energy efficiency assessment in

industrial buildings (Liu et al. 2018), building management and identification of operating anomalies (when analyzing electricity, gas and water consumption) (Akil et al. 2019), IEQ monitoring (Dogruparmak et al. 2014; Cao et al. 2020; Alghamdi et al. 2020; Roger Rozario 2021), energy forecasting (Tian et al. 2020; Chen et al. 2020; El Motaki et al. 2021), and data sampling for better visualization (Qin and Zhang 2017).

In Culaba et al. (2020), an energy prediction model is introduced using a k-means model to cluster data, and a support vector machine (SVM) is employed to forecast energy consumption. In Himeur et al. (2021), three clustering algorithms, namely one-class support vector machine (OCSVM), density-based spatial clustering of applications with noise (DBSCAN), and local outlier factor (LOF), are used to detect anomalous energy consumption in households by analyzing energy footprints. Besides, in Afaifia et al. (2021), hierarchical cluster analysis (HCA) is implemented to model residential energy consumption and promote energy efficiency. Clustering-based techniques have been used in BAMSs because of their simplicity and relatively computational efficiency. Also, clustering models generally have few parameters to tune. However, they have different limitations that affect their applications in BAMSs, among them the manual selection of the optimal K, dependency on initial values, troubles to cluster data with varying densities and sizes, the need for scaling as the number of dimensions increases, etc. (Li et al. 2018).

3.2.1.2 U2. Dimensionality reduction In diverse ML tasks, dimensionality reduction techniques can be employed to classify data while promoting low computational costs as they first remove irrelevant observations. Accordingly, a plethora of frameworks have been proposed in the literature to explore the applicability of dimensionality reduction schemes in BAMSs. That includes the principal component analysis (PCA), factor analysis, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and multiple discriminant analysis (MDA), isometric feature mapping (Isomap) (Liu et al. 2020), kernel principal component analysis (KPCA) (Abba et al. 2020), t-distributed stochastic neighbor embedding (t-SNE) (Zhan et al. 2020; Lopes et al. 2020), multidimensional scaling (MDS) (Wang 2020, and truncated singular value decomposition (TSVD) (Kalantzis et al. 2021).

For instance, PCA has been utilized for early fault detection and classification (Li and Wen 2014; Cotrufo and Zmeureanu 2016; Chen and Wen 2017; Swiercz and Mroczkowska 2019), IEQ (Mansor et al. 2021), energy consumption prediction (Sha et al. 2019), occupancy detection in buildings (Pal et al. 2019), etc. Moving on, LDA has been employed for thermal comfort evaluation (Gładyszewska-Fiedoruk and Sulewska 2020), sensor-based occupancy detection (Fayed et al. 2019). Using dimensionality reduction in BAMS has gained attention because it can: (i) reduce the storage space and time needed to classify recorded data, (ii) improve the interpretation of the ML models' parameters by removing multicollinearity, and (iii) simplify data visualization (Al-Kababji et al. 2022). However, dimensionality reduction models have some disadvantages. For example, (i) this can result in relevant data loss, (ii) finding linear correlations between variables (as PCA does) can be no appropriate in some scenarios, and (iii) some dimensionality reduction models can fail in classifying variables if the covariance and mean are not sufficient to represent datasets (Himeur et al. 2021; Abdulhammed et al. 2019).

3.2.2 Supervised learning (S)

Supervised learning is applied for the case of labeled energy datasets. Despite its high performance, the necessity of labeled data causes some difficulties in real-world applications.

3.2.2.1 S1. Classification It refers to conventional ML models that attempt to derive some conclusions from the input data given in the training process, and hence aim at predicting the class labels/categories for a new set of data. Classification models have widely been deployed in existing BAMS based big data analytics frameworks to perform different tasks, e.g. energy forecasting, energy balancing, IAQ monitoring, energy optimization, fault and anomaly detection. Typically, SVM, K-nearest neighbors (KNN) (Valgaev et al. 2017), decision tree (DT) (Yu et al. 2010), artificial neural network (ANN) (Moon et al. 2019), multilayer perceptron (MLP) (Haidar et al. 2019), extreme learning machine (ELM) (Salerno and Rabbeni 2018) and logistic regression (LR) (Rehman et al. 2020) are among the famous classification models deployed in BAMSs. Classification models have been used in BAMSs since they are simple to understand, fast and efficient. In addition, they can excel in classifying different kinds of BAMS data if accurately labeled datasets are used in the training process. However, they have a set of limitations. For instance, SVM models are not adequate for non-linear problems, and their performance does not improve if the number of features increases while the number of neighbors "K" is manually selected in KNN. Moreover, poor results are usually obtained with DT algorithms on small datasets, and overfitting can quickly occur (Himeur et al. 2021).

3.2.2.2 S2. Regression It is based on the identification of the relation between two or more energy consumption observations for producing a set of model parameters, that help in predicting and classifying them for different purposes, including energy prediction, anomaly detection, security and privacy preservation, etc. Diverse regression models have been proposed to analyze BAMSs' data, e.g. support vector regression (SVR) (Zhong et al. 2019), linear regression (LR), auto-regressive (AR) models, regression tree (RT) and regression fitting (RFT). Regression models have gained popularity in smart buildings and smart energy systems because most of them are easy to implement and interpret, and efficient to train. Also, they perform remarkably well for linearly separable data. However, it is worthy to note that regression models involve complicated and lengthy procedures of analysis and calculations in addition to assuming the existence of linearity between the dependent and independent variables, which is not always the case in real-world applications (Bilous et al. 2018).

3.2.2.3 S3. Deep neural networks (DNN) It is a subclass of ML that is principally a NN including more than two layers. These DNNs aim at simulating the behavior of the human brain "albeit far from matching its ability", which allows them to learn from large-scale datasets. In addition to the capability of NNs with a single layer for making approximate predictions, DNNs have further benefits via (i) optimizing and refining the classification accuracy when additional hidden layers are considered, and (ii) identifying the most informative features of the data.

DNNs have become the state-of-the-art methods in various ML-based domains, similarly in BAMSs, they are attracting greater attention. They have widely used for energy forecasting, this is the case of recurrent neural networks (RNN), long short-term memory (LSTM) (Gao et al. 2019; Wang et al. 2020), gated recurrent unit (GRU) (Lin et al. 2021), bidirectional LSTM (BiLSTM) (Haq et al. 2021; Ishaq and Kwon 2021), convolutional

LSTM (ConvLSTM) (Syed et al. 2021), multiplicative LSTM (mLSTM) (Krause et al.), bidirectional GRU (BiGRU) (Khan et al. 2020), coupled input and forget gate (CIFG) (Runge and Zmeureanu), deep feed forward neural networks (DFNN) (Marino et al. 2016), and convolutional neural network (CNN) (Li et al. 2017). Moreover, numerous hybrid models have been built by combining the aforementioned models with other deep learning (DL) architectures, such as CNN-LSTM (Alhussein et al. 2020), CNN-BiLSTM (Wu et al. 2021), partial least square (PLS) CNN-BiLSTM (PLS-CNN-BiLSTM) (Wu et al. 2021), CNN-GRU (Sajjad et al. 2020; Wu et al. 2020), conditional random fields (CRF) and RNN (CRF-RNN) (Wytock and Kolter 2013), DFNN-LSTM (Bashari and Rahimi-Kian 2020), radial basis function neural network-CNN (RBFNN-CNN) (Sideratos et al. 2020), etc.

DL models have also been used for other tasks, including smart IEQ monitoring, where different architectures were investigated, such as LSTM (Liu et al. 2020; Janarthanan et al. 2021), GRU (Ahn et al. 2017; Das et al. 2020), BiLSTM (Ma et al. 2019), CNN (Molinara et al. 2020), residual neural network (ResNet) (Zhang et al. 2020), variational autoencoders (VAE) coupled with CNN (VAE-CNN) (Loy-Benitez et al. 2020a), memory-gated RNN-based autoencoders (MG-RNN-AE) (Loy-Benitez et al. 2020b), sparse autoencoders (SAE) (Loy-Benitez et al. 2020)

Occupancy detection in buildings has also received the attention of the DL community through the use of CNN (Zou et al. 2017), RNN (Zhao et al. 2018), LSTM (Mutis et al. 2020), BiLSTM (Feng et al. 2020). Moreover, as some studies have investigated the use of camera imagery (e.g. thermal cameras) to estimate the number of occupants inside buildings, it was rational to use various CNN backbones, which are widely utilized in image classification or image recognition, among them ResNet (Acquaah et al. 2020), VGGNet (Zou et al. 2017), AlexNet (Acquaah et al. 2020) and GoogLeNet (Tien et al. 2020). Using DL models in BAMSs has become a research hot-spot nowadays because of their robustness to natural variations in the data, which is automatically learned. Additionally, their performance significantly improves with increasing the quantity of training data. However, DL models still face different challenges. Typically, DL algorithms require large-scale training datasets to perform better than other ML models. Moreover, their training is computationally expensive as they are built on complex models. Additionally, DL models require expensive GPUs and cloud data centers to run, which increases their deployment cost (Guo et al. 2018; Himeur et al. 2020).

3.2.2.4 S4. Statistical models They refer to mathematical models embodying an ensemble of statistical rules used to generate data samples, predict the relationships between one or diverse random/non-random variables or classify them. widely used statistical models include Bayesian networks (BN) (Singh and Yassine 2018), naive Bayes (NB) (Li et al. 2020), generalized additive models (GAM) (Khamma et al. 2020), bayesian belief networks (BBN) (Bassamzadeh and Ghanem 2017), restricted Boltzmann machines (Elsaeidy et al. 2019), conditional restricted Boltzmann machines (CRBM) (Kang et al. 2020) and factored conditional restricted Boltzmann machines (FCRBM) (Hafeez et al. 2020). In BAMSs, they have been used for different tasks, such as selecting the most energy-efficient primary HVAC systems (Tian et al. 2019), building energy and water retrofitting (Bertone et al. 2018), energy forecasting (Huang et al. 2018), assessing energy efficiency (Grillone et al. 2019), NILM (Verma et al. 2019), gas usage prediction (Pathak et al. 2018), IEQ monitoring (Giovanis 2019), etc. While most statistical models are useful for BAMSs as they have deterministic and stochastic components to mathematically describe the functional relationship between inputs and outputs, they also have pitfalls. In this respect, if recorded data is

biased or faulty, statistical modeling will be misleading. In addition, these kinds of models are hard to apply to heterogeneous data (Agha and Palmskog 2018).

3.2.3 Semi-supervised learning (SSL)

SSL refers to the process of training ML models using a small portion of labeled data along with a large number of unlabeled observations. Then, the ML models should be able to learn and make predictions on new data. It falls between unsupervised learning and supervised learning, which is also considered as a special instance of weak supervision (Van Engelen and Hoos 2020). Although supervised learning techniques are largely utilized in for providing different BMAS services, they can only reach high performance only when they are trained with sufficient labeled data. Otherwise, their performance could drastically decrease if annotated data is insufficient or not accurately labeled. Moreover, annotating data is a challenging, costly, and time-consuming task. In this regard, SSL has been proposed as an alternative solution to address some of these issues.

In BAMS, SSL has been widely used for fault and anomaly detection. For instance, in Fan et al. (2021, 2021), the authors introduce an SSL-based fault detection and diagnosis in air handling units (AHUs) based on a semi-supervised neural network (SSNN), which adopts a self-training strategy. Moving on, in Elnour et al. (2021), Elnour et al. propose an SSL-based data-driven attack detection scheme in HVAC systems to promote security in intelligent buildings. This approach has been developed using an isolation forest and two ML models, i.e. PCA and 1D-CNN (IF-PCA-CNN). While in Li et al. (2021), an SSL-based approach to detect and diagnose chiller faults is presented using a semi generative adversarial network (semi-GAN) model. In the same way, an SSL-based fault identification scheme for building HVAC systems is proposed in Li et al. (2021) using a modified GAN. In Nguyen et al. (2021), an SSL-based load monitoring solution is introduced, which has the ability to (i) augment the data, (ii) transform existing labeled sets, and (iii) train a WiderResNet (the backbone model) on the augmented data. Although SSL is an excellent option for developing AI-big data analytics when labeled data is expensive to obtain, it has some limitations, e.g., the results are not stable, and the performance is lower than that of supervised learning. Typically, the decision boundary might be overstrained if the training dataset does not have the annotated samples required in each class (Lu 2009).

3.2.4 Reinforcement learning (RL)

Reinforcement learning is a field in artificial intelligence that involves an agent that develops the knowledge of the best strategy to follow to accomplish a defined objective by trial and error given the interaction with its environment. Besides the RL agent and the environment, the main elements of an RL system are: (i) the policy, which is a function that defines the action taken by the RL agent in a given time step (i.e., state), (ii) the reward, which defines the result of the action taken by the agent due to its interaction with the environment, and intuitively describes the desired behavior of the agent, and (iii) the value, which indicates the long-term desirability of a set of states/actions given the agent's experience and the likely future rewards (Collins and Cockburn 2020). The agent explores the possible actions to be taken as the learning progresses. Based on the consequence of the actions taken, it opts for actions that maximize the cumulative reward. Reinforcement learning algorithms can be categorized as: (i) traditional RL (TRL) methods in which

tabular (i.e. lookup tables) or conventional value function approximation approaches (e.g., coarse coding, ML algorithms) are used; and (ii) deep RL (DRL), which represents the evolution of the traditional methods where DL models (e.g., deep NNs, CNNs, RNNs) are used to approximate the state and/or action value (Wang and Hong 2020).

3.2.4.1 R1. TRL models It is only efficient to use TRL for simple RL problems where the action-state space can be represented in a tabular form or approximated by a simple function approximation algorithm. Monte Carlo (MC), Q-leaning (QL), State-action-reward-state-action (SARSA), policy gradient (PG), and actor-critic (AC) are examples of TRL approaches. For TRL-based BAMS applications, tabular QL was used for occupancy prediction and HVAC control to optimize the occupant comfort and energy consumption in Barrett and Linder (2015), and controlling the HVAC system and windows for mechanical and natural ventilation in Chen et al. (2018).

3.2.4.2 R2. DRL models Recently, RL has taken advantage of the DL technology to reach phenomenal results. Typically, DL has been combined with RL due to its ability to capture all the intricate details of the knowledge and also perform complicated learning tasks that RL failed in doing so. This has given rise to DRL. In BAMSs and many other research fields, DRL is becoming a significant focus of scientists. The commonly used DRL methods are deep Q-learning (DQL), asynchronous advantage actor-critic (A3C), deep deterministic policy gradient (DDPG), and proximal policy optimization (PPO). A review of DRL applications for intelligent buildings energy management was presented in Yu et al. DQL was used for indoor and domestic hot water temperature control in Lissa et al. (2021) to optimize the home energy management system. In Wang et al. (2017), a DRL-based control system for office HVAC systems using an RNN-based actor-critic approach was presented. In Valladares et al. (2019), double Q-learning was utilized for energy optimization and thermal comfort control, while PPO method was applied in Azuatalam et al. (2020); Chemingui et al. (2020) for controlling the building's HVAC systems for energy and thermal comfort optimization.

AL in all, RL models (TRL and DRL) are utilized for solving very complex problems that can not be fixed using traditional ML or DL models. They can also correct the errors occurring during the training stage. However, exceeding the number of required RL stages can result in an overload of states, and hence reducing the performance of RL models (Ding et al. 2019).

3.2.5 Ensemble methods (E)

Ensemble methods are a class of ML that deploy different aggregation strategies for combining multiple learning models and then achieving better predictive performance compared to the use of a unique learning algorithm.

3.2.5.1 E1. Boosting It implies the gradual development of an ensemble learning using a set of ML models, where every new model occurrence is trained for emphasizing the training occurrences that previous models misclassified. In some applications, boosting can achieve better performance than bagging; however, it often looks after overfitting the training data. Random forest (RF), adaptive Boosting (Adaboost), and eXtreme gradient boosting (XGBoost) were among the most used boosting models for different AI-big data analyt-

ics tasks, such as overall building energy consumption forecasting (Zekić-Sušac et al. 2021; Xiao et al. 2021; Ferdoush et al.; Wang and Chen 2021; Yucong and Bo 2020), heating and ventilation load prediction (Sun et al. 2020), HVAC optimization (Li 2020), Space cooling load forecasting (Feng et al. 2021), load disaggregation and monitoring (Xiao et al. 2021), water monitoring (Somontina et al. 2018; Movahedi and Derrible; Golabi et al. 2020), IEQ monitoring (Mo et al. 2019).

Following, other variants have been then introduced and utilized for performing energy forecasting and load monitoring, IEQ monitoring, water management and occupancy detection in different kinds of buildings, including gradient boosting machine (GBM) (Gong et al. 2020), extreme gradient boosting machine (XGBM) (Gong et al. 2020), gradient boosting regression tree (GBRT) (Nie et al. 2021), LightGBM (Park et al. 2021; Wang et al. 2020).

3.2.5.2 E2. Bootstrap aggregating It is also abbreviated as bagging and refers to the design of a new ML model by aggregating multiple models that have equal weights in the ensemble vote. Every model is trained using a randomly drawn subclass of training data for promoting the model's variance. Various bagging models have been developed, modified and used to perform different tasks in BAMSs. For instance, bagging ARIMA (BARIMA) in de Oliveira and Oliveira (2018) is proposed to conduct a mid-long term load forecasting, while in Khwaja et al. (2015), a bagging neural network (BNN) is developed where the bagging concept is combined with neural networks (NNs) to improve short-term energy prediction. Moving on, in Hu et al. (2020), an enhanced bagged echo state network (BESN) is introduced to forecast energy. In Choi and Hur (2020), a bagging model is developed by setting RF, XGBoost and LightGBMs as the base learners. In Dehalwar et al. (2016), the authors introduce a bagged regression tree (BRT) that has been used for energy forecasting.

Moreover, bagging models have also been employed for water management in buildings using ensemble bagging tree (EBT) (Hasanzadeh Nafari et al. 2016), and thermal evaluation using bagged tree (BT) (Ahmad and Chen 2018), and fault detection using bagged auto-associative kernel regression (BAAKR) (Yu et al. 2017).

Overall, ensemble methods have been used in BAMSs since they can result in better predictive accuracy than individual models in complex systems/models Moreover, they are appropriate for scenarios with linear and non-linear data variables. However, ensembling is less interpretable, and the outputs of ensemble models are complex to explain and predict in most applications. In addition, a wrong selection of the models to be aggregated will arise lower predictive accuracy than individual models. Furthermore, ensemble models are generally computationally expensive and require much storage memory.

3.3 Building environments and their characteristics

Buildings range in size, function, construction, design, and other attributes. Additionally, they present varying levels of potential hazards and risks to the occupants and the surrounding environment. However, buildings are primarily classified based on the utilization purpose that governs occupancy profile, sophistication level, and building design requirements. Building environments are further described in the following subsections.

3.3.1 Residential buildings

Residential buildings are mainly for private occupancy, designed and built for individuals or groups, providing the necessary facilities and utilities to satisfy living requirements. Spaces in residential buildings involve several activities, including sleeping, sitting, conveniences, cooking, dining, and others. Those functions can be in shared spaces or have exclusive rooms per function. They exist in various sizes and have different occupancy rates. A low occupancy density generally characterizes them. Examples of residential buildings are story houses, apartments, terraces, and condominiums. In addition to air conditioning and ventilation systems, lighting, and media equipment, several major appliances are regularly used in residential buildings, such as dishwashers, washers, dryers, refrigerators, freezers, stoves, water heaters, trash compactors, ovens, and others (Estiri 2014).

Residential buildings are typically equipped with simple BAMSs that provide the basic requirements of building management for inhabitants' well-being and comfort. Standard manual control is used for the most part of their BAMSs. For instance, the decentralized control of the indoor environment is driven by the thermal comfort levels of the occupants. Thermal comfort is subjective to outside weather conditions that determine the indoor environment conditioning requirements, which are heating or cooling, humidification or dehumidification, and air ventilation (Do and Cetin 2018).

3.3.2 Office buildings

Office buildings are where people perform routine tasks, execute assignments and jobs for their employers, or provide passive or active, free of charge or remunerated services to the public. Types of workplaces vary in the form and requirements of the work and the variety of tools involved. Hence, they differ in size and the extent of personnel involvement and expertise. Familiar workplaces are office buildings such as law and corporate firms, commercial companies, post offices, banks, courtrooms, and similar places where people are involved in lengthy desk jobs or light-weight activities. Most of the spaces are offices, meeting rooms, or auditoriums of defined capacities. Additionally, they have shared areas such as corridors and lobbies. Office buildings require flexible and technologically-advanced working environments that are safe, healthy, pleasant, durable, and accessible towards promoting the users' comfort, productivity, and working efficiency (Tanabe et al. 2013). It includes the accessibility to natural ventilation and natural lighting sources, and the availability of IEQ control and monitoring. The provision of localized indoor environment control allows users to adjust the air temperature, air movement, and other relevant indoor environment properties according to their preferences. They are characterized by their moderate operation schedules, and fairly regular and established user profiles. Additionally, some workspaces may involve many service recipients (Alsalemi et al.).

3.3.3 Healthcare centers

The indoor environment in healthcare centers is critical for the health, well-being, safety, and comfort of patients, visitors, and the staff, as well as for the medical utilities and services. It has to comply with specific standards related to temperature, infection, and odor control (Salonen et al. 2013). It plays a significant role in the quality of

the provided medical service in terms of the treatment, healing, recovery processes, and the success of the conducted operations and procedures.

The various spaces in healthcare centers have different temperature regulation requirements. For instance, the success of surgical procedures depends in part upon the cold indoor conditions of the operating room to avoid the risks of anesthetic explosions, promote the comfort, productivity, and efficiency of the staff, and conserve the patient's resources (Ellis 1963). On the other hand, burn units are regulated at temperatures between 28 and 33 degrees because burn injuries restrict the ability of patients' bodies to stay warm (Fernández and Pablo 2021). Moreover, healthcare centers require a clean and sterile environment. Hospital-acquired infections are a major threat in healthcare centers (Lobdell et al. 2012). Hence, air ventilation and infection control are essential to control the potential contaminants and other suspended microorganisms, consequently lowering airborne disease risk. Additionally, air ventilation helps dispel odors, which improves the indoor conditions for the patients, staff, and visitors. Moreover, medical waste disposal and management is an essential aspect of the operation of healthcare centers as they are considered one of the main sites for the generation of hazardous waste (Aljabre 2002). The proper management of medical waste is essential to avoid health and environmental risks. Healthcare centers have protocols for the disposal of the generated waste according to their location. Additionally, healthcare centers are obliged to provide adequate security implementations for (i) the safety of patients, the public, and staff, (ii) the privacy and integrity of the patients' data, (iii) the prevention of breaches against the BAMS, (iv) the management of the utilities and equipment, and (v) the prevention of injuries and unwanted occurrences.

3.3.4 Sports facilities

Sports facilities involve areas where individuals or groups engage in physical exercise, participate in athletic competitions, or attend sporting events. Examples of sports facilities are gymnasiums, cultural centers, stadiums, swimming pools, indoor and outdoor tennis courts, squash courts, training halls, and sports arenas. They encompass large and various spaces involving different types of activities. Sports facilities have distinct requirements for air conditioning and ventilation, thermal comfort, and lighting with unique usage and occupancy patterns. They are governed by the type of sports activity, the operating time, the season, and the geographical location of the facility (Trianti-Stourna et al. 1998).

Sports facilities are characterized by the variety of their architectural sophistication and sizes, deployed technologies, and their distinctive energy demand profile compared to other types of buildings (Elnour et al. 2022). For instance, stadiums are the most sophisticated ones, which occupy vast land space. Even though they are often infrequently used, their operation and running costs during a single event are substantial (Aquino and Nawari 2015). Aquatic centers are the second most popular sports facilities that host different water events and tournaments. They encompass other spaces such as changing rooms, shower rooms, and storage rooms.

Sustainability measures and implementation are deployed in sports facilities' design, construction, and operations. They require extensive lighting, air conditioning, broadcasting, surveillance, and security requirements when operated to achieve successful sports events. The proper lighting in the sports facilities ensures good visual conditions. The event's prosperous broadcasting is essential to delivering an entertaining, thrilling, and engaging experience for the athletes and fans. Given the considerable volume of user flow

in sports facilities, emergency evacuation planning, users' entry and exit management, security screening, and preventive measures are among the top priorities in sports facilities management (Hall et al. 2011). Additionally, sports facilities involve extensive body workouts and activity by the users, during which excessive heat and CO2 discharge occur. They demand mainly air conditioning and ventilation, especially for indoor sports events as well as water heating for pools and domestic use to maintain the comfort, health, and well-being of the users. Moreover, sports facilities require constant maintenance, servicing, and overseeing even when not used, such as grass fields, pools, water treatment, and sports equipment.

3.3.5 Commercial buildings

Commercial buildings have at least 50% of their floor spaces for commercial activities (Kiliccote and Piette), such as malls, retail, and food services. Malls and restaurants are typical commercial buildings of various sizes and complexity. They include shops, cafes, kitchens with several commercial appliances, storage rooms, pantries, a refrigerated space, offices, dining areas, and public restrooms. They demand maintaining a clean and well-conditioned environment. For example, in restaurants and coffee shops, compliance with proper food storage and preparation standards is required to reduce the risk of spoiling food and eliminate the risk of incidents jeopardizing the well-being of the users as well as the reputation of the restaurants (El-Sharkawy and Javed 2018).

Air ventilation affects the health and safety of workers and customers and can influence food sanitation levels. The chiefs and the kitchen staff in restaurants are exposed to air pollutants generated from cooking for long periods. Hence, they may suffer potential respiratory and cardiovascular problems in the long run (Juntarawijit and Juntarawijit 2017). Also, they are subjected to high levels of heat generated from cooking activities, decreasing the staff's productivity. Additionally, excessive unpleasant odors or poor air conditioning in restaurants can result in an unpleasant experience for the customers. In addition, malls and shopping centers are commercial buildings where goods or services are sold to customers. They may include ample parking spaces, escalators, elevators, and various outlets such as department stores, food courts, amusement and theme parks, and movie theaters. Safe and comfortable indoor conditions are essential to provide a convenient and enjoyable experience for users and maintain a flourishing business with efficient energy consumption to contain the incurred running and operating costs.

Commercial buildings are famous for their exceptional operating schedule and occupancy patterns. They run for about more than 12 hours all week, and they have peak occupancy during weekends and significant volumes of user flow. They utilize extensive closed-circuit television (CCTV) surveillance, lighting, and air ventilation and conditioning systems. Additionally, fire prevention, suspension, and other security and alarm systems are crucial elements of their management systems to ensure dependable and safe circumstances for the users. Overly, the proper management of commercial buildings is essential to maintain a lucrative operation.

3.3.6 Industrial buildings

Industrial buildings include buildings used for the generation and distribution of power, manufacturing products such as food, apparel, electronics, petrochemicals, construction materials, automobiles, the processing of raw materials, and many others. They have

minimal and relatively low user flow for security purposes, such that they are only accessible to individuals with privileges. However, they involve energy-intensive and delicate machinery. They are generally equipped with sophisticated BAMSs that support the security and the centralized control requirements. Industrial buildings are equipped with robotics, industrial devices, and software-defined production processes. They require a high level of automation, given the nature of the processes involved and the tasks performed. In addition, they may involve delicate processes that are associated with health, social, and environmental risks. Industrial sites and environments can result in air and water pollution due to the generated by-products and the released unwanted toxins of the occurring processes. Hazards from combustion and unstable reactions can lead to highly harmful accidents due to the sudden release of material at high temperatures or pressures (Englund 2007). Additionally, fire hazards are common in industrial facilities, which can endanger the lives of staff and can result in substantial economic losses and environmental implications. Industrial facilities must be safe, secure, and productive. Proper process control, air ventilation, treatment and conditioning, and waste management are crucial to managing the safety and health of the staff as well as the general public and the surrounding environment. Security is an essential dimension in the operation of the BAMSs of industrial buildings.

3.3.7 Academic buildings

Academic buildings are used to conduct teaching activities such as schools, academies, universities, colleges, technical institutes, etc. They encompass classrooms, lecture halls, libraries, student centers, dining halls, laboratories, computer labs, offices, and service areas necessary for the proper functioning of the academic programs. Individuals of various age groups are frequent users of educational facilities, and they engage in multiple types of activities. A convenient and safe environment in academic facilities is an essential requirement for the education process. It affects the well-being and comfort of students, faculty members, and other staff, hence their productivity and working efficiency. A comfortable and safe environment has been identified as an essential element for enhancing the learning of students (Muhammad et al. 2014). Over-heated and poorly ventilated classrooms can result in the discomfort of students and educators, and consequently diverts their attention and affects their abilities to concentrate (Roelofsen). Adequate lighting in the facilities of academic buildings is vital to the comfort and well-being of the students to create an attractive and engaging learning environment and avoid eye strain. Additionally, students' health and well-being are essential for their learning process. The indoor environment influences students' attendance and hence their study. Students need to be in good health to be able to study well. Therefore, spaces in academic buildings should be well conditioned and ventilated to avoid altering users' well-being, spreading airborne diseases spread, and disrupting students' learning. Lastly, a brief summary is presented in Table 2 to compare the characteristics of the different buildings discussed above.

3.4 Computing platforms

3.4.1 Cloud computing

The advancement of cloud computing platforms has opened new opportunities for BAMSs to take control of operations on a large scale. Thus, BAMSs that consist of networked

Table 2 A comparison between the different types of buildings

Building type	BAMS type	Use	Spaces type	Activity type	Size	Occupancy/ Usage	Characteristics	Requirements	Example
Residential	Simple for basic requirements for inhabitants' well-being and comfort, standard manual control	Dwelling	Spaces for sleeping, sitting, conveniences, cooking, dining	- Dwelling, Light weight activities	Small to moderate	Low to moderate	- Light operations, and causal equipment/appliances - Consume relatively substantial amounts of energy and accompanied by peak demand issues	Iinhabitants' well-being and comfort	Houses, apartmentscondominiums
Office	Standard for natural ventilation, natural lighting, local indoor environment control, surveillance and safety system	Business	Offices, meeting rooms, auditoriums, corridors, lobbies	- Routine tasks, services to public - Desk jobs or lightweight activities	Small to moderate	Moderate to high	- Fairly regular and established user profiles	healthy, flexible, durable, productive, efficient, and comfortable working environment	Firms, banks,post offices

Table 2 (continued)

Building type	BAMS type	Use	Spaces type	Activity type	Size	Occupancy/ Usage	Characteristics	Requirements	Example
Heathcare	Advanced for air treatment, ventilation, and temperature regulation, security, surveillance, servicesequipment management	Health services	Treatment rooms, care units, examination rooms, laboratories, blood banks, operating rooms, patients rooms, waiting rooms, corridors, lobbies	- Delicate medical procedures - Lab experiments	Average to large	Moderate to high	- Year-round operation, frequent and considerable users flow- include sophisticated and expensive equipment for diagnostic and treatment-associated with health and environmental risks.	-Health, well-being, safety, and comfort of patients, visitors, the staff, medical utilities and services-comply with specific standards related to temperature, infection, and odor control. - Clean and sterile environment	Hospitals, clinic-shealth centers
Sports	Advanced system for extensive lighting, air conditioning, broadcasting,surveillance, and security	Sporting events and activities	Training rooms, spectators areas, changing rooms, shower rooms, storage rooms, offices, lobbies, corridors	- Sporting activities of different levels and types	Average to large	Moderate to high	-High seasonal usage patterns of high users flow, -Encompass various space types, - Substantial running and operation costs	-Distinct requirements for air conditioning and ventilation, thermal comfort - Specific visual conditions and broadcasting requirements	Stadiums, sports centers,swimming pools

Table 2 (continued)

Building type	BAMS type	Use	Spaces type	Activity type	Size	Occupancy/ Usage	Characteristics	Requirements	Example
Commercial	Inclusive system for closed-circuit television, surveillance, lighting, air ventilation and conditioning	Commerce	Shops, cafes, dining areas, kitchens, storage rooms, offices, lobbies, corridors	- Standard, lightweight activities	Average to large	Moderate to high	Year-round operation with frequent peak periods of considerable users flow	-Clean and well-conditioned environment- Proper air ventilation for maintaining health and safety of workers and customers	Shopping malls, grocery stores,restaurants
Industrial	Sophisticated system for centralized automation and management	Industries	Offices, control rooms, computer rooms, machinery and process rooms, electrical and chemical plants	- Light to heavy weight activities	Average to large	Low to moderate	Involve limited and consistent users flow, energy-intensive and delicate machinery and processes associated with health, social, and environmental risks	- High level of automation - safe, secure, and productive operation - waste management	Factories, power stations
Academic	Inclusive system for surveillance, lighting, air ventilation and conditioning, and security	Education	Classrooms, offices, meeting rooms, dining halls, laboratories, computer labs, lobbies, corridors	- Light to heavy weight activities	Average to large	Moderate to high	- Seasonal usage patterns of moderate to high flow	- A convenient, comfortable and safe environment for students and staff - Adequate lighting	Universities, colleges,schools

sensors and actuators, have been recently adapted to be able to connect to different cloud-based services (Alsalemi et al. 2020). The latter can provide data storage, connectivity, and powerful computing resources. To that end, significant efforts have been devoted to developing cloud-based big data analytics solutions in BAMSs (Bode et al. 2019). For instance, a voice-activated system for remotely monitoring BAMSs using cloud computing is presented in Valenzuela et al. (2013). While in Khattak et al. (2019), the idea of developing vehicular clouds for smart buildings and smart city applications is investigated. Moving on, in Stergiou et al. (2018), the security and privacy concerns along with the efficiency of cloud platforms are analyzed. In Delsing (2017), local cloud IoT automation is studied to promote the use of distributed IoT automation solutions.

Despite the significant effort made during the last decade to promote the use of cloud-services to run BAMSs, some drawbacks are still causing issues to users and operators, among them (i) the increased cost and communication overheads, (ii) the privacy and security concerns, especially when private data is transmitted to a centralized server for processing (Mohamed et al. 2018).

3.4.2 Edge computing

Edge computing it refers to performing data pre-processing, data fusion for different sources and AI-big data analytics at the edge of the network i.e. sensor nodes (Ray et al. 2019). Also, it enables optimizing cloud computing platforms due to its capability to use the processing power of IoT devices for filtering, pre-processing, aggregating and storing IoT sensor data. These tasks can correspondingly be conducted in real-time using convenient analytical tools (Sharma et al. 2018), while cloud platforms perform further enrichment, aggregation and running complex analytics on the filtered data. To that end, the new advances in BAMSs combined with the latest generation of IoT devices make it possible to bring the intelligence and computing tasks to the edge nodes in close proximity to the building's IoT devices (Zakharchenko and Stepanets 2019; Khan et al. 2020). Moreover, a new generation of open software platforms hosted on edge nodes are enabling access to the building data and advanced AI-big analytics deployed on these platforms are providing the technology to create value from this data by transforming data from building environments into actionable information. Various open edge platforms have recently been proposed, e.g. IOTech's Edge Xpert,[3] Echelon SmartServer IoT platform,[4] JENEsys Edge,[5] etc.

3.4.3 Fog computing

Fog computing represents a decentralized computing strategy where data storage, data processing and computing resources are located in the middle layer situated between edge devices and cloud. Typically, IoT smart sensors and submeters periodically collect the data and forward it to a gateway that acts as a fog device (Javadzadeh and Rahmani 2020). In this line, BAMSs can benefit from streaming data over a layer of fog devices (or nodes) to become more connected, where data can be analyzed to detect abnormalities for example, and autonomously react, if authorized, for compensating the problems or fixing the issues. Otherwise, fog nodes will send the convenient requests to the cloud (or services higher up

[3] https://www.iotechsys.com/markets/industries/building-automation/.
[4] https://www.dialog-semiconductor.com/products/industrial-edge-computing/smartserver-iot-edge-server.
[5] https://www.lynxspring.com/technology.

the fog hierarchy) for making further skilled and powerful technical analysis using complex ML models (Ferrández-Pastor et al. 2018; Aazam et al. 2018).

For instance, in some situations that require real-time decision-making, e.g. shut down appliances or equipment before being damaged or adjust crucial process parameters, edge devices or fog nodes can rapidly act with millisecond-level latency, while it is not possible to reach real-time decision making using cloud data centers (Rocha Filho et al. 2018). Therefore, the use of fog computing or edge computing helps avoid potential latency problems, delays an/or network/server down-times that can lead to different kinds of accidents or reduced service optimization and efficiency (Maatoug et al. 2019).

3.4.4 Hybrid computing

Hybrid computing refers to the case when the aforementioned computing architectures, i.e. edge computing, fog computing and cloud computing, are used together to process and analyze data (Himeur et al. 2021; Zhang et al. 2020). In this context, based on the application scenario and computation requirement, some data processing tasks could be made at the edge devices and/or fog nodes, while high-level data processing tasks (e.g. feature extraction, classification, anomaly detection, etc.) could be performed at the cloud data centers (Himeur et al. 2020).

3.5 Applications

3.5.1 Facility and asset management

Facility management to eliminate waste is among the benefits of using AI-big data analytics in BAMSs and can perform in diverse forms. For instance, using an AI strategy, a bathroom supplies monitoring company has saved up to 40% in of the total cost by installing a sensors that collect and send information about the utilization levels of toilet paper rolls and soaps (Gaboalapswe 2019; Sayed et al. 2022). Similar techniques are also be deployed for monitoring sports facilities, commercial buildings, office supplies, and other building necessities (Himeur et al. 2021; Idowu et al. 2016).

For instance, in sports facilities there is an emergency to improve the BAMS services to meet consumer's growing experience needs, and hence, overcome various issues, e.g. poor resource sharing, weak flexibility of response and slow transmission of information, and instability of aero-thermal comfort, which are considerable affecting the end-users' experience and restricting the development of sport venues (Zhong et al. 2020). To that end, a great attention has been put recently to design intelligent BAMS architectures of sport centers. This helps in interconnecting multiple subsystems, improving the interoperability, integrating information, realizing the integration of data application network, and achieves the goal of resource sharing and function upgrading. In Xiao-wei (2020), an AI-big data analytics platform is built using SVM-back propagation neural network (SVM-BPNN) for (i) predicting the end-user flow in the sport facility, (ii) providing recommendations to adjust the service plan, and (iii) improving the overall management and the end-users'

experience. Moving on, in Wan et al. (2021), as the cyber-security is a challenging issue in sports facilities due to the number of spectators and players and the large number of sport events organized, an AI-assisted cyber-physical system (AI-CPS) is integrated to the BAMS for promoting network security and predicting cyber attacks and adversaries.

On the other hand, because developing an appropriate setpoint temperature for the HVAC system is a crucial challenge, the authors in Aparicio-Ruiz et al. (2021) identify such temperature using a KNN-based dynamic adaptive comfort technique. It relies on the idea that occupants' thermal comfort in a building has different acceptability levels, which can be used for learning the comfort temperature corresponding to the average running temperature. Thus, this helps define the adequate range of indoor temperature. While in Carreira et al. (2018), Carreira et al. introduce a framework for tracking building end-users' group preferences, learning from them, and automatically managing HVAC systems. This framework is built by tracking building users using an RFID card, interacting with them on a mobile app, computing setpoints, and sending instructions to the HVAC sub-system over a gateway. Additionally, a K-means algorithm has been used for configuring the setpoint, in line with a prediction based on the current building status.

3.5.2 Load forecasting

In BAMSs, forecasting energy consumption is of significant importance to enable an effective management of energy, in which AI-big data analytics techniques play an essential role. In doing so, load patterns (and ambient conditions) are constantly collected from diverse building smart-meters and then fed into the AI models to predict energy usage. Because of the real-time characteristic of short-term forecasting, it has been more challenging than generic forecasting. Thus, various AI-big data analytics models have been proposed (Chou and Tran 2018; Ahmad and Chen 2018; Seyedzadeh et al. 2018; Fathi et al. 2020). In Pham et al. (2020), a random forests (RF) model is introduced to perform a short-term energy load prediction at an hourly sampling rate in various buildings by using different energy consumption datasets. In Seyedzadeh et al. (2019), the authors investigate the performance of diverse popular ML algorithms to predict buildings heating and cooling energy usage. Accordingly, specific tuning has been carried out for every ML algorithm using two building energy consumption datasets generated in EnergyPlus and Ecotect. In Ribeiro et al. (2018), a transfer learning based load prediction scheme is introduced, where energy consumption data of different buildings are used to forecast the load of a new building. This approach can work with various ML algorithms with pre- and post-processing phases.

In Moon et al. (2018), Moon et al. propose an energy prediction model using diverse ML models, including ANN, SVR, and PCA-factor analysis (PCA-FA). Data from four buildings in an academic institution have been used for evaluating the performance of these models. In Ahmad et al. (2020), an intelligent load prediction scheme is proposed using generated sampled data-based Gaussian process regression model (GSD-GPRM), regression binary decision tree (RBDT), bootstrap bagging of regression trees (BBRT) and binary multiclass classification decision tree (BMCDT). In Idowu et al. (2016), supervised ML algorithms are used to develop a load forecasting model using SVM, regression tree, feed-forward neural network (FFNN), and multiple linear regression (MLR). Moving on, in Ahmad et al. (2018), diverse supervised ML models are implemented to predict energy consumption at short, medium, and long-term levels in different building environments,

namely compact regression Gaussian process (CRGP), binary decision tree (BDT), generalized linear regression model (GLRM) and stepwise Gaussian processes regression (SGPR). In Chou and Ngo (2016), a short-term based energy prediction system is proposed using a seasonal autoregressive integrated moving average (SARIMA) model along with a metaheuristic firefly algorithm-based least squares support vector regression (MetaFA-LSSVR) model. Typically, this framework uses (i) the SARIMA architecture for linearing energy observations, and (ii) the MetaFA-LSSVR model for capturing nonlinear energy patterns.

In Li et al. (2021), a transfer-learning-based ANN scheme is developed to predict short-term energy consumption in information-poor buildings. The efficiency of transfer learning in improving the prediction accuracy has been demonstrated using limited training data. Moving on, in Grolinger et al. (2016), a short-term energy consumption prediction of sports facilities, which is considered as a challenging scenario due to the variations caused by by the hosted events, is performed using NN and SVR. In Zheng et al. (2017), a short-term energy prediction approach using an empirical mode decomposition (EMD)-LSTM-based RNN is proposed with a Xgboost model to select feature patterns based on a feature importance evaluation. In a similar manner, in Haq et al. (2021), a sequential learning-based load forecasting algorithm is developed and used in both residential and commercial buildings. Accordingly, this framework implements a convLSTM integrated with BiLSTM (ConvLSTM-BiLSTM) and compares its performance with various sequential models, including ConvLSTM integrated with BiLSTM, LSTM, auto-encoder (AE), multi-layer Bi-LSTM (MBiLSTM), BiLSTM-AE, GRU, and CNN with multilayer bidirectional GRU (CNN-MB-GRU). Fig. 4 illustrates a flowchart of an energy forecasting system based on AI-big data analytics. Specifically, a short-term energy consumption prediction is performed using EMD-LSTM neural networks with a Xgboost algorithm to extract importance features. Besides, Moradzadeh et al. (2020) propose a heating and cooling load forecasting scheme that is based on MLP and SVR in residential buildings. These models help identify a linear mapping between inputs and outputs. MLP has outperformed SVR in terms of the recall metric, where a recall of 99.93% has been achieved. To summarize, Table 3 compares various AI-Big data analytics frameworks used for energy forecasting, in terms of AI model, forecast horizon, building environment, year of appearance, method description and evaluation metrics.

3.5.3 Energy efficiency

One area that can get immensely benefited from AI-big data analytics is energy efficiency in buildings. This is because of the way a building consumes energy can be quite variable, and is related to various parameters, e.g. the nature of the building, the energy provider, the sources of the energy, the number of devices, the number of end-users/occupants in any building, the behavior of end-users/occupants, etc. Himeur et al. (2021); Fatema et al. (2020). Moreover, comprehending the energy consumption habits of a building is the first and most critical step to achieve energy efficiency. This way, AI-big data analytics and energy efficiency can go hand in hand towards the goal of optimizing energy consumption and reducing the amount of wasted energy without compromising the comfort level of end-users and the level of efficiency and productivity in a company or industry Sardianos et al. (2021).

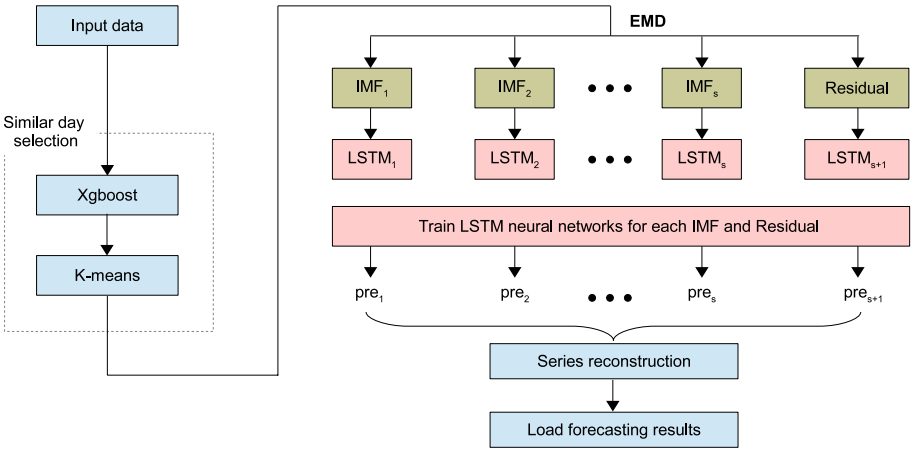

Fig. 4 Flowchart of an energy forecasting system based on AI-big data analytics (Zheng et al. 2017)

In (Yu and Chiller), Yu et al. propose an open IoT cloud-based ML system, namely AI Chiller, to promote energy efficiency in buildings by optimizing the consumption of the HVAC system. An AI-big data analytics scheme based on an RNN-LSTM architecture and to analyzing and fusing BAMS environmental footprints has been developed and combined with a genetic algorithm to achieve 10% savings. In Al-Ali et al. (2017), energy saving in residential buildings in the Gulf region is achieved using IoT, off-the-shelf business intelligence, and big data analytics platforms.

3.5.4 Predictive control and thermal comfort

One solution to save buildings' energy is using model predictive control (MPC). It aims at developing predictive models for (i) simulating input-output interactions; and (ii) helping users to identify optimum control actions that drive the predicted outputs to the desired references. In this context, ML models predict energy demand and simulate MPC control techniques to save energy and optimize end-users' comfort. Typically, these models can provide decision bases for selecting optimal MPC control actions Serale et al. (2018); Mariano-Hernández et al. (2021). In Gao et al. (2020), building thermal comfort control is conducted using RL, where a deep feed-forward neural network (FNN)-based method is introduced. The latter helps predict consumers' thermal comfort before introducing a deep deterministic policy gradients (DDPGs)-based scheme to optimize thermal comfort. In Yang et al. (2020), an MPC approach based on an RNN with nonlinear autoregressive exogenous (NARX) architecture, namely NARX-RNN, is proposed to optimize air conditioning and mechanical ventilation (ACMV) in a hospital office and hence save energy and optimize thermal comfort. Similarly, in Yang et al. (2021), the same methodology is experimentally implemented to control the ACMV systems in office and lecture theatre (LT) testbeds in real-time. In Chen et al. (2020), Chen et al. developed an MPC approach based on deep transfer learning to optimize the HVAC operation in smart buildings.

Moving forward, in Bünning et al. (2020), an RF-based data predictive control (DPC) scheme is proposed using convex optimization and affine functions. This help in controlling

Table 3 A summary of AI-Big data analytics frameworks proposed for energy forecasting in buildings

Ref.	AI model	Forecasthorizon	Building nature	Year	Description	Evaluation metrics			
						RMSE	MAE	MAPE	Others
Skomski et al. (2020)	seq2seq	Short-term	Office	2020	Demonstrate the efficiency of seq2seq RNNs for load prediction using a restricted feature set	✓			nRMSE
Bessani et al. (2020)	Bayesian networks	Short-term	Residential	2020	Handle the volatility and the uncertainty of buildings' loads	✓		✓	nRMSE, MedAE
Ribeiro et al. (2018)	transfer—based MLP, SVR	Long-term	Residential	2020	TL-based trend and seasonal adjustments to predict cross-building load		✓		MSE
Ahmad et al. (2020)	GSD-GPRM, RBDT, BBRT, BMCDT	Short-, long-term	Office	2020	Building load prediction in non-climate sensitive and climate-sensitiveconditions		✓	✓	CV
Moon et al. (2018)	ANN, SVR, PCA-FA	Short-term	Academic	2020	Energy prediction of higher educational institutions	✓	✓	✓	
Zhang et al. (2020)	LSTM, GRU, CIFG	Short-term	Public	2020	Hybrid DL-based energy prediction combined with an interpretation process	✓	✓		CV-RMSE, R^2
Wen et al. (2020)	RNN-GRU	Short-, mid-term	Residential	2020	Achieve well performance with limited input variables	✓	✓	✓	
Park et al. (2020)	XGBoost, RF, DNN	Short-term	Industrial	2020	A Two-Stage energy consumption prediction			✓	CVRMSE

Table 3 (continued)

Ref.	AI model	Forecasthorizon	Building nature	Year	Description	Evaluation metrics			
						RMSE	MAE	MAPE	Others
Khamma et al. (2020)	GAMs	Short-term	Office	2020	Embed domain knowledge and prior understanding of buildings into the prediction model				CVRMSE, NMBE
Somu et al. (2020)	ISCOA-LSTM	Short- , mid-, long-term	Residential	2020	Accurate and reliabale data driven load forecasting	✓	✓	✓	MSE,Theil U1, U2
Liu et al. (2020)	A3C, DDPG, RDPG	Mid-, long-term	N/A	2020	Improve the forecasting accuracy with increasing computation time	✓	✓		R^2, CV
Zhang et al. (2020)	DBN-DEEM	Short-term	Residential	2020	Predict stochastic energy consumption using Cyclic feature (CF) extracted via spectrum analysis	✓	✓	✓	r
Lu et al. (2020)	CEEMDAN-XGBoost	Short-term	Intake towers	2020	Have half of prediction error of XGBoost using real-world data for a period of 8 years				RMSPE, Theil U1, U2
Wang et al. (2020)	stacking model	Short-term	Academic	2020	Building load forecasting using model integration	✓	✓	✓	CVRMSE
Somu et al. (2021)	κCNN-LSTM	Long-term	Academic	2021	Capture the load spatio-temporal features and aid in decision making	✓	✓	✓	MSE

Table 3 (continued)

Ref.	AI model	Forecasthorizon	Building nature	Year	Description	Evaluation metrics			
						RMSE	MAE	MAPE	Others
Yuan et al. (2020)	WNN-cuckoo search	Mid-term	Commercial	2020	Optimally tuning the WNN parameters CS with				DMAPE, AE
				2020	a real-world validation				
Mawson and Hughes (2020)	DFNN, RNN	Mid-term	Industrial	2020	Load forecasting and condition monitoring in manufacturing buildings	✓	✓		
Bui et al. (2021)	LSTM	Long-, and short-term	Residential	2021	Multi-behavior with bottleneck features LSTM for to predict energy consumption	✓	✓	✓	NRMSE
Dun and Wu (2020)	Grey model	Long-term	Residential	2020	Load forecasting of three kinds of buildings, i.e. rural, public and urban buildings			✓	
Khan et al. (2021)	LSTM-KF	Short-term	Residential	2021	Learning to statistical model for ensemble predicting of energy consumption	✓	✓	✓	
Li et al. (2021)	TL-based ANN	Short-term	Residential		Load prediction of information-poor buildings			✓	NTR
Grolinger et al. (2016)	NN-SVR	Short-term	Sport-venues	2016	Load forecasting in a challenging scenario with high variations caused by the hosted events			✓	

Table 3 (continued)

Ref.	AI model	Forecasthorizon	Building nature	Year	Description	Evaluation metrics			
						RMSE	MAE	MAPE	Others
Pinto et al. (2021)	RF, GBR	Short-term	Office	2021	Combine multiple learners to optimize the learning process			✓	

energy consumption and temperature in a room of a real-life apartment. In Yang and Wan (2022), Yang et al., RNN-NARX-based MPC is introduced with instantaneous linearization for ACMV optimization. Table 4 summarizes some of the recent ML-based MPC frameworks described above, their characteristics, and their contributions. Most concentrate on controlling HVAC and ACMV systems in buildings since they consume the most significant proportion of building energy. Thus, this significantly impacts the thermal comfort of buildings' occupants.

3.5.5 Anomaly and fault detection and diagnosis

Failures in building electric networks and devices' operation cycles may result in excessive energy losses and extra costs. To alleviate these issues, AI-big data analytics are a prevalent tool that enables detecting faults and disturbances early enough and predicting maintenance. That is possible by implementing continuous energy consumption monitoring to create "an early warning system" empowered with AI-big data analytics strategies and pattern recognition models to notify the end-users and operators Alsalemi et al. (2020). Accordingly, information related to energy consumption, environmental conditions, and occupancy patterns is fed into the AI black-boxes to identify and classify deviations. Once deviations are classified, their causes are determined before taking appropriate measures for their prevention Himeur et al. (2020), Alsalemi et al. (2020).

In this regard, a plethora of AI-based frameworks have been proposed to develop AI-big data analytics platforms that allow building energy efficiency Himeur et al. (2021). In Himeur et al. (2020), a DNN model and the micro-moment concept have been used to identify energy consumption deviations. A micro-moment rule-based algorithm is employed to extract load features of daily intent-driven energy usage moments. Next, DNN is applied to classify and determine abnormal consumption classes automatically and then compare the performance with various scenarios using conventional ML classifiers, e.g. LR, LDA, NB, SVM, RF, KNN, DT, ensemble classifier, and MLP. Similarly, in Himeur et al. (2021), unsupervised and supervised anomaly detection schemes are introduced to promote energy saving in academic and residential buildings. OCSVM is applied to extract abnormal energy consumption patterns from unlabeled data, while an improved kNN classifier is proposed to process annotated consumption footprints that are benchmarked using the micro-moment concept. In Xu and Chen (2020), a hybrid model using RNNs with quantile regression (QR) is proposed for anomaly detection in residential houses towards improving the performance of the building and reducing energy waste (Table 5).

Additionally, fault detection and diagnosis in HVAC systems, being the most extensively operated equipment, has been covered widely in the literature. For instance, the utilization of the different configurations of deep RNNs is investigated in Taheri et al. (2021) to perform fault detection and diagnosis of common HVAC system faults, such as the malfunction and leakage of valves and dampers and sensor bias faults. A comparative study is presented comparing the performance of a DRNN-based diagnosis approach with RF and GB algorithms. In Yun et al. (2021), a neural network-based supervised auto-encoder (NN-SAE) - which is an auto-encoder with two outputs that are the classification label and the reconstructed signal- is proposed for air handling units fault detection and diagnosis before its validation using ASHRAE experimental data. The two outputs of the NN-SAE have been then processed to determine the diagnosis decision reliability. This approach has been

Table 4 A summary of MPC-based frameworks, their characteristics and contributions in saving energy and optimizing thermal comfort

Work	ML model	Building type	System	Best performance	Control objective
Gao et al. (2020)	FNN + DDPGs	Public buildings	HVAC	4.31%HVAC energy saving	Save energy and improve thermal comfort.
Yang et al. (2020)	ANN-NARX	Office and lecture theater	ACMV	58.5% cooling thermal saving	Energy saving and thermal comfort optimization.
Yang et al. (2021)	RNN-NARX	Office and a lecture theater	ACMV	52% reduction of cooling energy	Save energy and optimize thermal comfort. in experimental testbeds
Chen et al. (2020)	MLP-based transfer learning	Residential buildings	HVAC	MSE=0.16	Optimize energy efficiency and thermal comfort.
Bünning et al. (2020)	RF	Residential buildings	HVAC	24.9% of cooling energy saving	Optimize energy consumption withoutcompromising thermal comfort
Yang and Wan (2022)	RNN-NARX	Office in a hospital	ACMV	26–31.6% cooling energy savings	Save energy and optimize thermal comfort
Li and Tong (2021)	Encoder-decoder RNN	Residential/public buildings	HVAC	4–7% energy saving	Energy saving and smart control of thermal environment
Mtibaa et al. (2021)	CAM- LSTM	Multi-zone buildings	HVAC	MAPE = 0.0872%	Save energy, predict peak power and improve thermal comfort

Table 5 A summary of AI-Big data analytics models used for anomaly and fault detection and diagnosis

Ref.	AI model	Fault			System anomaly	Targeted system	Year	Description	Evaluation metrics			
		Sensor	Component	Actuator					ACC	MAE	MSE	Others
Himeur et al. (2020)	DNN				✓	Building appliances	2020	Classify and determine abnormal building energy consumption.	✓			Precision, F1-score, AUROC curves, confusion matrix
Himeur et al. (2021)	OCSVM, kNN				✓	Building appliances	2021	Classify and determine abnormal building energy consumption.	✓			F1-score
Xu and Chen (2020)	RNN, QR				✓	House	2020	Detect abnormal energy consumption.				PICP, PINAW
Taheri et al. (2021)	RNN	✓	✓	✓		HVAC systems	2021	Investigate various DRNN configurations to perform HVAC system fault diagnosis.	✓			Precision, recall, F1-score, cross entropy, confusion matrix
Yun et al. (2021)	NN-SAE		✓	✓		AHUs	2021	Detect and diagnose common AHUs faults to enable reliable maintenance.				Precision, recall, F1-score

Table 5 (continued)

Ref.	AI model	Fault			System anomaly	Targeted system	Year	Description	Evaluation metrics			
		Sensor	Component	Actuator					ACC	MAE	MSE	Others
Li et al. (2021)	GAN		✓			HVAC systems	2021	Detect and diagnose common HVAC system faults and addressing the issues of labeled data availability and data imbalance.	✓			FPR
Elnour et al. (2020)	AANN	✓				HVAC systems	2020	Perform sensor data validation and fault diagnosis of HVAC system sensor faults using semi-supervised learning.	✓			TPR, FPR, recovery rate, deviation rate, noise reduction rate
Elnour and Meskin	2D CNN			✓		HVAC systems	2021	Diagnose HVAC system single actuator faults using supervised learning.	✓			Precision, recall, F1-score, specificity
Dey et al. (2020)	MC-SVM		✓	✓		HVAC system	2020	Provide automated detection and diagnosis of equipment failures in HVAC systems' terminal units.				Precision, recall

Table 5 (continued)

Ref.	AI model	Fault			System anomaly	Targeted system	Year	Description	Evaluation metrics			
		Sensor	Component	Actuator					ACC	MAE	MSE	Others
Bode et al. (2020)	LR, kNN, CART,RF, NB, SVM, NN		✓			Heat pumps	2020	Investigate fault detection approaches for operational heat pumps using machine learning algorithms.	✓			MCC
Shahnazari et al. (2019)	RNN	✓		✓		HVAC systems	2019	Develop models and a fault diagnosis methodology for HVAC systems.	✓	✓	✓	
Dey et al. (2020)	Clustering		✓	✓		FCUs	2020	Remote detection of fan coil units common faults.				Silhouette indexing, Davies-Bouldin, maximal gap
Han et al. (2019)	LS-SVM		✓	✓		Chiller system	2019	Diagnose common faults in centrifugal chillers.	✓			
Liu et al. (2021)	CNN		✓	✓		Chiller system	2021	Assess transfer learning for fault diagnosis methods in chillers.	✓			Recall, F1-score, precision confusion matrix
Zhu et al. (2021)	DANN		✓	✓		Chiller system	2021	Apply transfer learning for chiller fault diagnosis.	✓			Accuracy improvement degree

Table 5 (continued)

Ref.	AI model	Fault			System anomaly	Targeted system	Year	Description	Evaluation metrics			
		Sensor	Component	Actuator					ACC	MAE	MSE	Others
Han et al. (2020)	kNN, SVM, RF		✓	✓		Chiller system	2020	Assess using ensemble learning for chiller fault diagnosis.	✓			F1-score, confusion matrix
Choi and Yoon (2021)	NN-AE	✓		✓		BAMS	2021	Investigate variants of the AE-based fault diagnosis approach for building automation systems.	✓			Precision, F1-score

compared with conventional ANN and SVM algorithms, and it has been found reliable as it considers undefined situations.

Additionally, a 2D CNN-based HVAC system actuator fault diagnosis is introduced in Elnour and Meskin, in which the system's measurements and control signals were configured into multi-channel images and then processed using the 2D CNN-based diagnosis framework. CNNs are characterized by their high-performance accuracy and powerful capability in learning and realizing complex functions and interdependency from any given data. While in Liu et al. (2021), a CNN-based chiller fault diagnosis method is developed for building energy systems. Additionally, a TRL-based scheme is assessed to investigate the potential of using a pre-trained CNN-based fault diagnosis approach for chillers with different specifications, which is useful when available data is limited in size and/or types of operating conditions/faults captured.

In Dey et al. (2020), a big-data framework is presented to enable automated HVAC system fault diagnosis in large scale buildings in which a feature extraction approach is proposed to reduce the data dimensionality, and a multi-class SVM (MCSVM) algorithm is utilized to develop the diagnosis model. It aims to provide energy savings through preemptive maintenance, behavior analysis, and predictive building identification. In Bode et al. (2020), various ML algorithms are investigated to perform fault detection on a heat pump system, which are LR, kNN, classification and regression tree (CART), RF, NB, SVM, and NNs. This study demonstrates the effect of the data quality and amount and the limitations on the system's features availability (i.e, types of available sensors) on the performance of the developed ML models. While in Han et al. (2020), a supervised hybrid fault diagnosis model using SVM, kNN, and RF is developed for chiller fault diagnosis such that the three models are developed independently to perform fault diagnosis; then the final decision is made based on the plurality voting method. It was found that ensemble learning contributes to diagnostic performance improvements. In Li et al. (2021), a fault diagnosis approach is proposed for the common component faults in the HVAC system of an office building using a modified GAN. The proposed approach enables leveraging the labeled and unlabeled data simultaneously such that it aims to process the unlabeled data and utilize the limited information from the labeled ones to conclude the diagnosis decision.

Those fault diagnosis approaches mainly require sufficient labeled data for training, which can be unavailable or complex, and costly to obtain. Therefore, several studies have developed unsupervised and semi-supervised diagnosis strategies as in Elnour et al. (2020), where an auto-associative neural network (AANN) is utilized for sensor data validation and fault diagnosis in HVAC systems using semi-supervised learning. It demonstrates a compelling performance in sensor error correction, data replacement of unavailable sensors, measurement noise reduction, and sensor inaccuracy correction. Also, it is effective for both single and multiple sensor faults diagnoses. In Zhu et al. (2021), transfer learning is applied to develop a chiller fault diagnosis approach that only requires the system's normal operation data using domain adversarial neural network (DANN). While in Shahnazari et al. (2019), a distributed diagnosis approach using RNNs utilizing the normal system operation data is developed for multiple sensors and actuator faults diagnosis. It is based on developing intercommunicating fault detection and isolation (LFDI) agents for the various HVAC subsystems, i.e., cooling coil, VAV box, etc., and each LFDI agent is composed of two RNN-based models. It demonstrates promising capability in fault diagnosis. However, it is excessively computationally demanding, given the two RNN-based models included in each agent.

Additionally, a multi-level automatic fault detection framework is proposed in Dey et al. (2020) for fan coil units (FCUs). Feature extraction followed by data clustering are

applied to identify faulty and healthy data, and then a clustering-based fault diagnosis model is developed. The least-squares support vector machine (LS-SVM) regression model is used in Han et al. (2019) to develop a chiller fault diagnosis strategy that is validated using ASHRAE data. The proposed approach is compared with two other methods using the SVM algorithm of probabilistic neural networks (PNNs). In Choi and Yoon (2021), a semi-supervised fault diagnosis approach is proposed for building automation systems using NN-based auto-encoders (AEs). An AE is a structure that transfers the input to the latent space then uses the compressed representation to produce a reconstructed version at the output. Variants of the proposed method are investigated: the residual-based approach using the error between the original and the reconstructed signal as the indication of the system status, and the latest space-based approach in which the features of the compressed representation are used for fault diagnosis.

3.5.6 Indoor environmental quality (IEQ) monitoring

IEQ monitoring continues to grow in importance, several works have demonstrated an apparent relationship between the increasing concentration of CO_2 and decreasing cognitive performance (Nejat et al. 2020; Pulimeno et al. 2020). Typically, monitoring ambient IEQ and temperature can reveal valuable information for creating a healthier, more comfortable environment for end-users. It is also a prime opportunity for energy and cost savings (Saini et al. 2020a). Thus it becomes possible, using detailed AI analytics in BAMSs, to identify various environmental problems, including air pollution, where different pollutants may affect the IEQ , such as the cleaning products, cigarettes smoke, perfumes, construction activities, water-damaged building materials, and other types of outdoor pollutants (Saini et al. 2020b). Indeed, albeit these gazes are commonly safe for end-users, their effect on human health can be dangerous if they exceed certain thresholds of exposure. To that end, an intelligent IEQ monitoring system for classifying and recognizing diverse pollutants and measuring their levels is of utmost importance (Wei et al. 2019; Muiruri et al. 2021).

Before the COVID-19 outbreak, IEQ monitoring was not a priority in public buildings, e.g. sport venues, banks, healthcare centers, academic institutions, commercial centers, restaurants, and so on. However, the fast proliferation of the corona virus and its resulting harmful effects have put IEQ in the spotlight as an important component of BAMSs. In Mumtaz et al. (2021), the authors (i) develop an IoT node including various sensors for collecting data, (ii) introduce a NN model for classifying 8 pollutants, and (iii) design an LSTM-based DL model for predicting the concentration of every pollutant and the overall IEQ. In Mad Saad et al. (2017), a pollutant recognition scheme is proposed for IEQ monitoring using different supervised ML algorithms, including MLP, KNN and linear discrimination analysis (LDA). The evaluation has been conducted in a residential building located in a rural area in China, where 5 different indoor air pollutants were considered (combustion activity, presence of chemicals, presence of food and beverages, ambient air, and presence of fragrances). While in Loy-Benitez et al. (2020b), Loy et al. introduce an ML-based scheme for detecting, diagnosing, identifying, and reconstructing abnormal observations of multivariate IEQ data in a subway station. Accordingly, a memory-gated RNN-based autoencoders (MG-RNN-AE) that can process dynamic and sequential IEQ data has been utilized.

In Cruz et al. (2020), an IEQ prediction model is developed using SVM radial basis function (SVM-RBF) and stochastic Gradient Boosting machines (SGBM). The performance of

of these models has been evaluated using root mean squared error (RMSE) and R^2, and a comparison with other ML algorithms has been presented. In Taştan and Gökozan (2019), a real-time IEQ monitoring system is designed using and IoT-based e-nose and diverse ML classifiers, including SVR, generalized regression neural network (GRNN), and extreme learning machine (ELM) with Gaussian kernels. The linear correlation (LC) has been used for evaluating this framework. In Sharma et al. (2021), a cost-effective framework for IEQ prediction is introduced, where MLP and eXtream Gradient Boosting Regression (XGBR) are used for providing real-time measurements of the concentration of pollutants, i.e. CO_2 and particulate matter 2.5 ($PM_{2.5}$) in a set of classrooms. Moving on, an LSTM without using the forget gate (LSTM-wF) is deployed to predict the air quality at a lower complexity and increase the prediction performance.

In Alawadi et al. (2020), an indoor temperature forecasting scheme is proposed, where up to 36 ML models (pertaining to 20 different families) have been deployed. Real-world data gathered for three hours from both smart households and weather station have been used to validate this study. Similarly, in Aliberti et al. (2019), Aliberti et al. propose a smart solution for indoor air-temperature prediction, where a non-linear autoregressive neural network (NN-ARNN) has been utilized to perform short- and medium-term forecasting. This model has been then validated on both a synthetic dataset and real-world data recorded using IoT devices installed in residential buildings.

As CO_2 concentration is appropriate for measuring the IEQ quality due to its over the sensor networks, a set of frameworks have adopted it. For instance, in Taheri and Razban (2021), an ML-based IEQ monitoring approach is proposed by predicting CO_2 concentration in the academic building (campus classrooms) using demand-controlled ventilation. Various ML algorithms have been employed and compared to learn the CO_2 concentration, among them SVM, AdaBoost, RF, Gradient Boosting (GB), LR, and MLP. In a similar way, Kallio et al. (2021) propose a smart approach to forecast office indoor CO_2 concentration by adopting four ML algorithms, i.e. ridge regression (RR), DT, RF, and MLP. Moreover, a baseline to evaluate the indoor CO_2 prediction has been introduced by producing a benchmark dataset covering an entire year. Moving on, In Tagliabue et al. (2021), an ML-based IEQ monitoring approach that relies on measuring the CO_2 concentration in an academic building is proposed. Specifically, an LSTM-based RNN models and IoT sensors have been then used for monitoring the indoor conditions depending on the occupancy patterns.

Other IEQ monitoring frameworks have focused on measuring and predicting other factors, which are recommended in various countries, such as total volatile organic compounds (TVOC), formaldehyde (HCHO), and carbon monoxide (CO). For instance, in Chen et al. (2018), Chen et al. use four ML models, including SVM, Gaussian processes (GP), M5P and backpropagation neural network (BPNN) for predicting CO_2, HCHO, and TVOC in an academic building (in Singapore). In a similar way, in Lagesse et al. (2020), various ML models are utilized for predicting $PM_{2.5}$ in office buildings, i.e. ANN, LSTM, multiple linear regression (MLR), partial least squares regression (PLS), distributed lag model (DLM), and least absolute shrinkage selector operator (LASSO).

Other AI-big data analytics have also been used to perform additional tasks. For example, Loy et al. (2020a) introduce a variational autoencoder (VAE) coupled with convolutional layers (VAE-CNN) model to impute missing IEQ data. Accordingly, two scenarios have been adopted to evaluate the VAE-CNN algorithms: (i) a point-to-point data removal, and (ii) data intervals removing at different sampling rates. While in Kalajdjieski et al. (2020), the capability of generative adversarial networks (GANs) is exploited in

combination with a data augmentation technique for overcoming the class-imbalance issue while monitoring IEQ using large-scale datasets. Table 6 outlines pertinent AI-Big data analytics frameworks introduced for monitoring IEQ and performs a comparison between them, with reference to the AI model, forecast horizon, building environment, year of appearance, method description and evaluation metrics.

3.5.7 Security and safety

Among the major safety concerns in general and in buildings, in particular, are the fire outbreak. The prevention of and the immediate reaction to fires minimizes their consequences in terms of the people's well-being and the financial losses. This includes the minimization of fire incidents potentials, the fast detection and extinguishment of the fire source, the effective execution of emergency evacuation, and the prompt notification of the emergency situation to the concerned authorities. Buildings are usually equipped with conventional fire alarm and extinguishment systems consisting of several sensing devices, including smoke, heat, and flame detectors, automated alarms, and water sprinklers (Zverovich et al. 2016). With the advent of big data algorithms and analytics, the fire safety in buildings can be boosted by employing buildings data to develop frameworks for fire prevention, detection, and suspension. An analysis of the advantage of utilizing ML algorithms for reliable and prompt fire detection was provided in Surya (2017) represented in their distinguished capability of black-box modeling, feature extraction, pattern recognition with high accuracy and reliability. Unlike conventional fire detection systems, ML-based models can be used to detect fire, analyze it effects, assess its risks, predict its behavior utilizing the data collected from the sensing devices.

For example, in Zhang et al. (2021), a model combining a deep belief network (DBN) and a recurrent LSTM neural network (R-LSTM-NN) was proposed for fire hazards prediction in smart cities. The proposed model was used in predicting the air quality that is then used to detect fire outbreaks based on the sensors readings of the IoT system. It shows promising potential in when data records of the IoT system are available for normal operation and scenarios of fire occurrence. In Fu (2020), a comparative study was presented using ML algorithms, namely, DT, KNN, and NNs, to develop classification models to predict failure patterns and to assess the progressive collapse potential for steel framed buildings in fire. The study aimed to develop a reliable fire assessment tool for practitioners and the developed framework demonstrated a satisfactory performance overly. In Sultan Mahmud et al. (2017), another comparative study was conducted for developing an intelligent fire detection system with early notification system in which data mining algorithms such as DTs, Bayesian networks (BayesNet), NNs, and SVM were used to develop data-driven classifiers using supervised learning. Moreover, the proposed smart fire detection system employed an edge detection model to analyze the data collected from the cameras to confirm the fire detection decision. In Huda et al. (2012), an AI-based framework was proposed to assess the thermal condition of electrical installations in buildings to prevent the potential of injuries and fire hazards using infrared images. Raw data was processed using PCA for features selection that are then used to develop an NN-based classifier to determine the condition of electrical equipment.

In Ouache et al. (2021), a fire safety assessment framework was proposed to help predict the potential fire impacts and recommending optimal fire intervention strategies in multi-unit residential buildings using NNs. Supervised learning was used to train a NN based on 5 predictors, among which are the mean of the initial fire detection (i.e., smoke

Table 6 A summary of AI-Big data analtiycs frameworks introduced for IEQ monitoring

Ref.	AI model	Task	Building nature	Year	Description	Evaluation metrics					Others
						RMSE	MAE	MAPE	ACC	F1	
Mad Saad et al. (2017)	KNN, LDA, MLP	Pollutant ecognition	Residential	2017	Recognize 5 different pollutants using supervisedlearning				✓		
Cruz et al. (2020)	SVM-RBF,	IEQ prediction	Academic (labs)	2020	Forecast IEQ and the ideal number of occupants in a lab	✓					R^2
Alawadi et al. (2020)	36 ML models	Temperature prediction	Residential	2020	Compare a set of 36 ML algorithms for temperature prediction	✓					R^2
Moon et al. (2018)	SVR, GRNN, ELM	IEQ monitoring	Residential	2018	Real-Time monitoring of IEQ with an IOT-based e-nose						LC
Taheri and Razban (2021)	SVM, RF, LR, GB MLP, AdaBoost	CO_2 concentration prediction	Academic	2021	Predict CO_2 concentration with the ability to learn nonlinearities connected with the CO_2 data	✓	✓	✓	✓		R^2
Loy-Benitez et al. (2020b)	MG-RNN-AE	IEQ monitoring	Subway station	2020	Detect, diagnose, identify, and reconstruct abnormal observations of multivariate IEQ data	✓					MSE, R^2

Table 6 (continued)

Ref.	AI model	Task	Building nature	Year	Description	Evaluation metrics					Others
						RMSE	MAE	MAPE	ACC	F1	
Kallio et al. (2021)	RR, DT, RF, MLP	CO_2 concentration prediction	Residential	2021	Forecast office indoor CO_2concentration and introduce a benchmark dataset that covers a full year.		✓		✓		
Tagliabue et al. (2021)	RNN	IEQ prediction	Academic		Monitor the ventilation rate using an IoT protocol by analyzing data with ANN communication						MSE, R^2
Aliberti et al. (2019)	ARNN	Temperature prediction	Residential	2019	Predict indoor air-temperature using synthetic data to train an ARNN						MAD, RMSD
Chen et al. (2018)	SVM, GP, M5P, BPNN	Prediction of CO_2, TVOC and HCHO	Academic	2018	Analyze CO_2, TVOC and HCHO footprints for better predicting the IEQ			✓			R^2
Sharma et al. (2021)	MLP, XGBRLSTM-wF	IEQ monitoring	-	2021	Estimate and predict the concentration of pollutants i.e. CO_2 and $MP_{2.5}$	✓	✓	✓			
Lagesse et al. (2020)	ANN, LSTM, MLR PLS, DLM, LASSO	Prediction of $PM_{2.5}$	Commercial	2020	Predict $PM_{2.5}$ for IEQ monitoring in commercial office buildings	✓					

Table 6 (continued)

Ref.	AI model	Task	Building nature	Year	Description	Evaluation metrics					Others
						RMSE	MAE	MAPE	ACC	F1	
Mumtaz et al. (2021)	NN, LSTM	Classification + Prediction	Public	2021	Detect anomalies of IEQ and predict the concentration of each air pollutant	✓	✓		✓	✓	MSE

detector, heat detector, visual, etc.), the action taken to fight fire (i.e., occupant response, fire department, BMS, etc.), and the performance of the BMS in fire detection and extension. The fire impact assessment covered several aspects, which are the occupant response to the incident, fire extension, fire damage, and financial losses. The proposed framework demonstrated a remarkable ability to predict fire impacts accurately, and it represented a promising solution to define and regulate fire safety strategies.

The security of the automation and management system in buildings has become more imperative due to the rapid advancement in the technologies used and the IoT systems. The industry predicts that the IoT market will grow from an installed base of 30.7 Billion devices in 2020 to 75.4 Billion in 2025 (IoT Security Foundation), which will expose them to increased risk of advanced attack vectors. According to Kaspersky Lab, nearly four in ten buildings were targeted by attacks in the first half of 2019 , and it is expected that the impact of cyberattacks on the building and construction industry will be significant in the coming years (Kaspersky). In Elnour et al. (2021), an attack detection framework for false data injection was proposed for a multi-zone HVAC system in office buildings utilizing an isolation forest (IF) algorithm. The operational data of the system's sensor and control command signals were used to develop the detection model using semi-supervised learning. Isolation forests are characterized by the low computational requirement and capability to handle to complex and multi-variate data. They work based on pointing out anomalies using the concept of isolation, which improves the attack detection capability. Feature selection was applied to the raw data, and the study presented a comparative analysis of two models for feature reduction, which are PCA based model and a 1D-CNN-based model.

In de Assis et al. (2020), a security system for industrial IoT was proposed in which a CNN-based classifier was developed to identify distributed denial of service (DDoS) attacks in software-defined networks (SDNs). This system is based on supervised learning from the labeled network data of the IoT system. CNNs are advantageous for their high accuracy and classification performance, and powerful capability in realizing complex interdependency from multi-variate and sophisticated data. While in Aboelwafa et al. (2020), a residual-based attack detection framework was presented in which an NN-based auto-encoder was trained to profile normal system behavior. Then, non-conforming observations are identified as anomalies based on the generated residuals between the input and the output of the AE. Auto-encoders are used to learn the latent feature representation of the system using healthy operational data. They are also used for data dimensionality reduction, noise filtering, information retrieval, etc. In Yahyaoui et al. (2020), a preliminary demonstration of a ML-based intrusion detection system for data protection in healthcare centers was presented. An SVM model was developed using the labeled data of the IoT network and it demonstrated a promising performance in detecting malicious actions launched against the IoT system. The accuracy of the proposed framework was assessed based on the energy consumption in the communication network because attacks result in increased energy usage due to the increased network traffic. Table 7 highlights and compares existing AI-Big data analytics frameworks introduced to ensure security and safety in BAMSs.

3.5.8 Occupancy detection

Occupancy data are collected by various sensors and devices in buildings to help improve the efficiency of the BAMSs in terms of energy utilization and occupants' comfort and well-being. These include cameras, infrared sensors, and carbon dioxide detectors

Table 7 A summary of AI-Big data analytics frameworks proposed to ensure security and safety applications

Ref.	AI model	Task	Building nature	Year	Description	Evaluation metrics			
						ACC	MAE	MSE	Others
Zhang et al. (2021)	DBN, R-LSTM-NN	Fire hazard prediction	Smart cities	2021	Detect fire outbreak based on the sensors readings of the IoT system using supervised learning	✓			TPR, FPR Error rate
Fu (2020)	DT, kNN, NN	Fire safety assessment	Steel framed buildings	2020	Predict failure patterns and progressive collapse potential due to fire	✓			
Sultan Mahmud et al. (2017)	DT, BayesNet NNs, SVM	Fire detection	-	2017	Enhance the fire system detection capability by analyzing system data to conclude the situation	✓	✓		ACC, Precision RAE, Recall, F1
Huda et al. (2012)	PCA, NN	Inspection of electrical installations	Office	2012	Assess the thermal condition of electrical installations based on infrared images using supervised learning	✓			
Ouache et al. (2021)	NN	Fire safety assessment	Residential	2021	Investigate and assess fire protection and intervention strategies				R
Elnour et al. (2021)	IF, PCA, CNN	Attack detection	Office	2021	Detect attacks on the building management system				Precision, Recall

Table 7 (continued)

Ref.	AI model	Task	Building nature	Year	Description	Evaluation metrics			
						ACC	MAE	MSE	Others
de Assis et al. (2020)	CNN	DDoS attack detection	Industrial	2020	Detect DDoS attacks on the SDNs of the IIoT system	✓			Precision, Recall F1 score
Aboelwafa et al. (2020)	NN	FDI attack detection	Industrial	2020	Detect FDI attacks on IIoT systems	✓		✓	FPR, TPR
Yahyaoui et al. (2020)	SVM	Intrusion detection	Hospital	2020	Detect intrusion in IoT systems				Energy

(Sardianos et al. 2020; Sayed et al. 2022). The data can be directly used as inputs to control and regulate some of the buildings' equipment, such as lights, air conditioning, doors, etc. Additionally, scholars and researchers utilize big data analytics to develop approaches for analyzing and processing building occupancy data to facilitate an efficient and reliable overall building management (Sardianos et al. 2020).

In Huang and Hao (2020), a DL-based visual recognition was used to implement an occupancy detection framework utilizing CNNs. The proposed approach was used to determine the number of people present and their location to help operate demand-based HVAC systems more reliably and efficiently. It was found that the proposed approach outperforms the conventional occupancy detection systems in terms of accuracy, precision, robustness, ability to provide occupancy count, and ability of static and dynamic occupancy detection. It also requires hardware and computational considerations as CNNs have a high computational overhead. In Acquaah et al. (2020), a study was presented to estimate the occupancy count based on thermal images using CNNs for feature extraction and SVM for multi-class classification. Two well-known CNN architectures were investigated, which are the 50 layers ResNet (ResNet-50) and AlexNet using transfer learning due to their superior performance in image processing (He et al. 2016; Krizhevsky et al. 2012). In Tien et al. (2021), a computer vision-based occupancy and equipment usage detection framework was proposed to facilitate a demand driven control of the HVAC system in an office room. A multi-class region-based CNN model was developed and deployed to analyze camera images. It can predict the occupancy count, activity type (i.e. sitting, walking, etc.), and equipment usage in real-time.

A thermal-based occupancy detection approach was presented in Zhao et al. (2018) in which data-driven models were developed using SVR and RNN to predict occupancy information using the building's properties, including indoor temperature, towards managing energy use and security monitoring in intelligent buildings. That is, the interaction of the thermal components present in the conditioned space and the necessary part of determining thermal consistency is indicated by the indoor temperature. Supervised learning was used to train the ML models using simulation data generated from Energy Plus such that the target outputs were the occupancy count. The work in Elkhoukhi et al. (2020) combined the IoT technology and Big data analytics to implement real-time occupancy detection such that data of the indoor lighting, temperature, humidity, and CO_2 levels were used to predict the status of the building occupancy. Two models were developed and tested, one using LDA and the other using vertical hoeffding tree (VHT) for offline and online occupancy detection, respectively. Additionally, in Fatema and Malik (2021), feature extraction and correlation analysis were performed on indoor sensors data (i.e., temperature, CO_2 level, light intensity, humidity) and then particle swarm optimization was used to train an NN-based occupancy detection model. It demonstrated improved classification performance compared to conventional NNs optimized using the back-propagation algorithm.

In Wu and Wang, a ML-based model was proposed to improve the operation of the BAMS due to the shortcomings of infrared sensors for stationary occupancy. The model predicted the occupancy status based on multiple statistical features of the signals acquired by the infrared sensors. ML algorithms were investigated, among which SVM demonstrated the best performance due to its ability to capture complex and nonlinear functions, and its efficacy in handling high dimensional data. In Huchuk et al. (2019), a comparative analysis was presented for occupancy forecasting using ML algorithms based on thermostat data. The prediction model was intended to optimize the operation of the air conditioning system, such that both the present and the future occupancy information is taken into

consideration. It was found that RF algorithm outperformed the LR, the Markov model, the hidden Markov model (HMM) and the RNN, which is based on the bagging technique in which multiple models on different subsets of the training dataset are developed, then their predictions are combined to conclude the final output of the RF model. However, occupancy forecasting is only dependable when the building does not exhibit rapid and random fluctuations in the user profile, which is generally the case for residential buildings.

In Razavi et al. (2019), a comparative study was presented for the utilization of supervised ML algorithms such as SVM, RF, KNN, and NNs to estimate and predict occupancy information in residential buildings based on power meters data. It was found that the reliability of occupancy prediction is lower for the larger forecast horizons. In Feng et al. (2020), a DL-based approach is proposed combining a CNN and a bidirectional LSTM network for occupancy detection in houses based on electrical data of advanced metering infrastructures (AMIs). The data essentially contain readings of electric current, voltage, and power that are processed by the CNN for spatial feature extraction. Using supervised learning, the extracted features are then fed to the BiLSTM network to solve a binary classification problem to identify the occupancy condition in real-time. The proposed framework demonstrated improved performance when compared to other ML and DL based models due to its ability to interpret the spatial and contextual features of the data. However, since detailed occupancy information are mostly not recorded and hence not available, supervised learning-based approaches that are based on such a detailing in the data (i.e., people count) can be impractical and difficult to implement using actual building data. While in Pešić et al. (2019), a LSTM network-based framework was proposed to perform occupancy detection and forecasting as well as data analytics based on Bluetooth positioning and WiFi utilization data of the IoT infrastructure in a multi-story residential building. The network data were pre-processed to extract the information of the occupancy of the apartments, then used to develop the LSTM network to predict and forecast occupancy condition and patterns in the different spaces of the building. The proposed work demonstrated the effective fusion of Bluetooth and WiFi data as well as the successful deployment of NN-based data analytics using wireless networks data for occupancy detection application. Table 8 summarizes the relevant AI-Big data analytics frameworks developed to detect occupancy profiles.

3.5.9 Water usage management

Almost all kinds of buildings are users of water, although the cost of water and sewer services varies from area to area and can become a significant expense. Worse, in areas where there is a shortage of water, it is not only a big expense, but an imperative to conserve. Therefore, it becomes of significant importance to bring the monitoring of water levels and switching points of all wet applications in buildings to the BAMS. Furthermore, water monitoring systems can benefit from the advancement of AI and ML technologies for improving their performance (Sun and Scanlon 2019). Typically, by harnessing the power of AI-big data analytics, it is possible to maximize information and data available and hence make better decisions while enhancing service delivery and reducing costs (Rahim et al. 2020).

In this context, using IoT water meters with wireless connectivity (Bluetooth, LoRaWAN, etc), it becomes relatively easy to install water meters within the building. These can be as simple as pulse style meters that can easily be integrated into a BAMS. Moving on, adopting AI-big data analytics to analyze data from water meters has become

Table 8 A summary of AI-Big data analytics models used for occupancy detection

Ref.	AI model	Detection basis	Building nature	Year	Description	Evaluation metrics			
						ACC	MAE	MSE	Others
Huang and Hao (2020)	CNN	Surveillance cameras	Office	2020	Detect the number and location of occupants	✓			Relative error
Acquaah et al. (2020)	CNN, SVM	Thermal cameras	–	2020	Estimate the number of people present based on thermal images	✓			
Tien et al. (2021)	CNN	Vision cameras	Office	2021	Predict equipment use and occupancy count & activity in real time	✓			Precision, Recall, F1
Zhao et al. (2018)	SVR, RNN	Temperature sensor	Office	2018	Detect the number of occupants based on indoor thermal properties				Error rate
Elkhoukhi et al. (2020)	LDA, VHT	Indoor sensors	Office	2020	Predict the status of occupants' presence	✓			
Fatema and Malik (2021)	NN	Indoor sensors	Office	2021	Predict occupancy condition in an office room	✓			TPR, FPR, Precision, Recall, F1, MCC
Wu and Wang	SVM, kNN, DT RF, NN	Infrared sensor	–	2021	Provide accurate predictions of the occupancy status based on motion detectors	✓			
Huchuk et al. (2019)	LR, HMM, MM, RF, RNN	Thermostat data	Residential	2019	Forecast the occupancy information	✓			
Razavi et al. (2019)	kNN, SVM, NN, RF	Energy meter	Residential	2019	Estimate and predict occupancy information	✓			Precision, AUROC

Table 8 (continued)

Ref.	AI model	Detection basis	Building nature	Year	Description	Evaluation metrics			
						ACC	MAE	MSE	Others
Feng et al. (2020)	CNN, BiLSTM	Smart meters	Residential	2020	Predict real-time occupancy status based on data of electrical signals	✓			Precision, Recall, F1, TNR F1, Training time
Pešić et al. (2019)	LSTM	Bluetooth and WiFi devices	Residential	2019	Predict, forecast, and analyze occupancy information using wireless networks data	✓			RMSE and Edit Distance on Real Signals

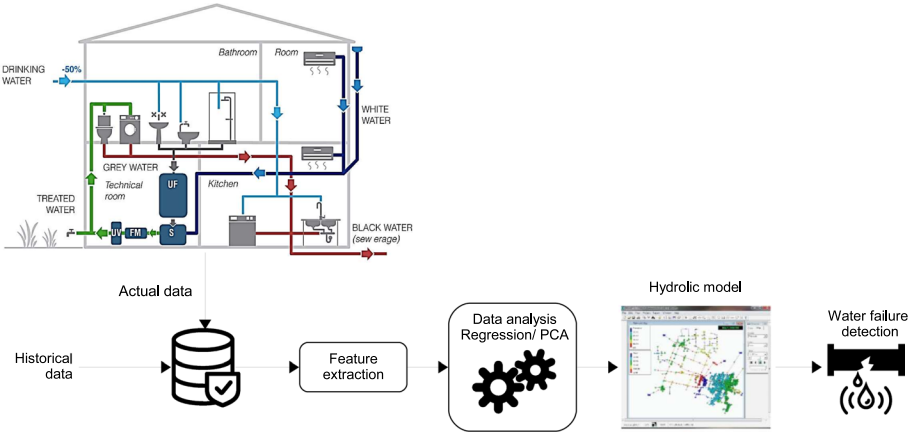

Fig. 5 Flowchart of water usage monitoring system used to detect water leaks and optimize water consumption (Jenny et al. 2020)

crucial for optimizing the management of water resources and sustaining growth and development. Accordingly, various AI-big data analytics frameworks have recently been proposed with the aim of (i) processing complex nonlinear water data, (ii) forecasting water demand, (iii) predicting water meter failures, or (iv) monitoring the quality and temperature of the water. Fig. 5 illustrates the flowchart of a water usage monitoring system used to detect water leaks and optimize water consumption (Jenny et al. 2020).

In Altunkaynak and Nigussie (2017), water demand prediction is conducted by first using multiplicative season algorithm (MSA) to extract pertinent information from water meter records and also capturing periodicity and converting nonstationary signals into stationary signals. Following, their output is fed into an MLP for accurately predicting water demand. The RMSE and Nash-Sutcliffe coefficient of efficiency have been adopted to evaluate the prediction performance of the learning model and its ability to extend prediction lead time. Shine et al. (2018), diverse ML models are used for predicting water consumption in an agricultural building based on analyzing data collected from a remote monitoring system. Thus, RF, ANN, SVM, and CART decision tree (CDT) algorithms were trained to predict water consumption, where a backward sequential variable selection was adopted for excluding variables adding low predictive power along with a hyper-parameter tuning with nested cross-validation for calculating the prediction accuracy for each model. In a similar way, in Smolak et al. (2020), three ML algorithms are implemented and compared for predicting water usage, i.e. RF, SVM, ARIMA. The water consumption data augmented with end-users occupancy patterns were used to improve the prediction accuracy. A novel approach to process and correlate between occupancy and water usage time-series was introduced. This framework was validated on 51 days of water consumption readings and over 7 million occupancy patterns from urban areas.

On the other hand, by using AI-big data analytics, it is also possible to monitor water quality and hence improve water resources management plans. In Chen et al. (2020), 10 ML models are deployed to water quality prediction (WQP). Specifically, DT, NB, LR, LDA, completely-random tree (CRT), KNN, SVM, RF, and deep cascade forest (DCRF) have been trained using water data from a hydro-electric power (HEP) plant, including pH, DO, CODMn, and NH3-N to forecast water quality. The precision, recall, F1 score, and weighed F1-score (wF1) have been selected to evaluate the prediction performance of the ML algorithms. In Roccetti et al.

(2019), Roccetti et al. develop an ML-based classifier, which is personalized for predicting the failure of a water meter. Typically, an RNN model is deployed for (i) processing 15 million of readings collected from 1 million of mechanical water meters, and (ii) extracting relevant patterns representing the complex phenomenon of defective water meters. This has helped in achieving more than 80% accuracy in detecting failures.

In Wang et al. (2018), water demand of urban areas is predicted using gravitational search algorithm (GSA) and backtracking search algorithm (BSA) with ANN with regard to various weather parameters. While in Antunes et al. (2018), four ML models are selected to predict water demand, including ANN, RF, SVM and KNN, through the analysis of real-world data from two Portuguese water utilities. Moving forward, a weighted parallel strategy for combining multiple ML algorithms is introduced to improve the prediction performance. Moreover, additional data related to weather, seasonality, and feature extraction (forecast window of time-series data) are also analyzed. In Nasser et al. (2020), the water demand prediction is performed using an LSTM model based on analyzing data gathered from intelligent IoT water meters. A cloud platform has been used to store water consumption records, enabling near real-time data streaming and storing. The performance of LSTM has been then compared to those of SVR and RF. Similarly, in Du et al. (2021), Du et al. propose an LSTM model that combines discrete wavelet transform (DWT) and PCA to forecast daily urban water demand. Therefore, after smoothing the outliers of water demand time-series, noise components are removed using DWT and pCA. Following, the LSTM network is deployed to predict urban water demand using the outputs of DWT and PCA. Table 9 presents the main AI-Big data analytics frameworks proposed for water management in buildings.

3.6 Evaluation metrics

Evaluation metrics are used to measure the performance of the model in terms of the quality of its output as per what is expected. For AI applications, there are various types of metrics that can be used based on the subject matter. That is, the outputs of an AI model can take two forms, which are categorical variables (Cvars) and quantitative variables (Qvars). For instance, the outputs of classification models represent categorical variables in which the input data are classified into different groups or classes which are characterized by a unique label or value such as detection problems, recommender systems, etc. Each observation can be placed in a single category, and the categories are mutually exclusive. Hence, the performance of the model depends on its ability to correctly classify the observations to their respective categories/groups.

On the other hand, quantitative variables represent numerical values that exhibit quantitative characteristics. Regression models have quantitative variables as outputs such as forecasting and estimation models in which the AI model is used to represent the mapping between the independent variable(s)—i.e., the input(s)—and the dependent variable(s)—i.e., the output(s)—. In this case, the quality of the model is measured by the closeness of the model's outputs to the ideal expected values. Table 10 presents a summary of the common metrics used to evaluate AI models.

Table 9 A summary of AI-Big data analtiycs models for water monitoring

Ref.	AI model	Task	Building nature	Year	Description	Evaluation metrics					Others
						RMSE	MAE	MAPE	ACC	F1	
Altunkaynak and Nigussie (2017)	MLP	Water demand prediction			Predict water demand using MSA-MLP and compare its DWT-MLP	✓					CE
Zubaidi et al. (2018)	GSA-ANN, BSA-ANN	Water demand prediction	Residential	2018	Predict water demand using heuristis algorithms, ANN and weather variables	✓					
Chen et al. (2020)	DT, NB, LR, LDA,CRT, KNN, SVM, RF, CRF	WQP	HEP plant	2020	Predict water quality using different water paranmetersi.e. pH, DO, CODMn, and NH3–N					✓	wF1
Shine et al. (2018)	RF, NNN, SVMCDT	Water consumption prediction	Agricultural	2018	Predict water consumption using a backward sequential variable selection and parameter tuning			✓	✓		
Smolak et al. (2020)	RF, SVM, ARIMA	Water consumption prediction	Residential	2020	Predict water consumption using consumption records and occupancy patterns	✓	✓	✓	✓		

Table 9 (continued)

Ref.	AI model	Task	Building nature	Year	Description	Evaluation metrics					Others
						RMSE	MAE	MAPE	ACC	F1	
Antunes et al. (2018)	ANN, RF, SVM KNN	Water demand prediction	Public	2018	Reliable prediction while no significant anomalies of the data used during training are reported	✓		✓			R^2
Roccetti et al. (2019)	RNN	Predicting water meter failures	-	2019	Predict water meter failures using 15 million of readings				✓	✓	AUC, CM
Nasser et al. (2020)	LSTM	Water demand prediction	Public and residential	2020	Predict energy demand by analyzing data gathered from smart IoT water meters and stored in the cloud	✓	✓	✓			
Du et al. (2021)	LSTM-DWT-PCA	Water demand prediction	Public and residential		The outputs of DWT and PCA are fed into an LSTM network to predict water demand	✓		✓			EVS, R^2

4 Critical discussion and current challenges

A truly smart building combines a BAMS with intelligent data analytics software that offers helpful insights for maintenance, service, and efficiency opportunities. Typically, these tools together offer benefits for building owners, such as: (i) providing a high-level, system-wide big data capture of the entire operations, (ii) ensuring air quality control and a healthier building environment, (iii) saving energy and energy consumption during off-peak or low occupancy periods, (iv) eliminating waste from everyday system usage through intelligent sensor data, (v) offering guidance for performance improvements for individual assets, (vi) addressing equipment that really needs repair and not just those on a fixed schedule that don't need to be serviced, and (vii) offering advanced automation capabilities and actionable results. In addition to these benefits, the cost savings related to the use of smart data analytics can be significant.

However, various gaps specific to each application field of AI-big data analytics are identified. Among them, more effort should be put to efficiently carrying out text analytics on operators' work-order logs; and identifying (i) the information to derive, (ii) the text-mining methods to adopt, and (iii) the efficient approaches to convey the information to the operator and visualize it. Moreover, another challenging issue concerns the use of virtual metering, where a limited number of works were dedicated to virtual meter development (Kim et al. 2021; Wilcox 2020) despite its significance in helping operators for understanding plant-to-zone water and energy flows and ranking their operational decisions, such as identifying and evaluating faults.

The HVAC prognostic and failure prediction is another application that is still very challenging, where limited research activity was conducted to target this challenge. In fact, developing prognostics models is valuable for (i) predicting the time-to-failure, (ii) avoiding global failures in key BAMS components (e.g. boilers, chillers, pumps, fans, etc.), and preventing disruptions in building services. In addition, there are new challenges from emerging BAMSs that need to be addressed, e.g. data benchmarking, big data security and privacy, scalability and interoperability, real-time big data intelligence and knowledge transfer.

4.1 Data quality issues

Usually, raw data gathered from BAMSs can have some data quality problems, including (i) outliers, (ii) noise, (iii) inconsistent data, (iv) duplicate data, and (v) missing values. Data pre-processing techniques are deployed to overcome these issues, such as formatting, cleaning, and resampling. Formatting aims at converting the raw data into appropriate formats to ease the application of ML algorithms, while cleaning refers to removing or replacing missing samples (Zhang et al. 2021). Lastly, resampling can be applied based on the requirements of ML algorithms. Typically, it can be (i) a down-sampling to reduce data redundancy, foster the processing and improve the accuracy; or (ii) an up-sampling that helps increase the amounts of data to train data-hungry ML models, especially DL algorithms (Elnour et al. 2022). To that end, because of the high requirements for data quality set by ML models, developing novel strategies to improve the quality of BAMS recorded data by creating additional data with enhanced quality or augmenting existing datasets is a crucial challenge.

Table 10 A summary of the common evaluation metrics of AI models

Metric	Description	Values type	Application	Formula
Relative error (RE)	The absolute error between the actual and estimate values of a variable to its estimate value	Qvars	Regression	$\lvert(y-\hat{y})/\hat{y}\rvert$
Relative change	The amount of the absolute difference as a fraction of the variable's reference value	Qvars	Regression	$\left(\lvert y-y_{\text{ref}}\rvert/y_{\text{ref}}\right)\times 100\%$
Mean absolute error/difference (MAE or MAD)	Measures the absolute difference between predicted and actual variables describing the same phenomenon.	Qvars	Regression	$\frac{1}{m}\sum_{i=1}^{m}\lvert\hat{y}_i-y_i\rvert$
Mean absolute percentage error (MAPE)	the MAE expressed in percentage.	Qvars	Regression	$\text{MAE}\times 100\%$
Median Absolute Error (MedAE)	Measures the median absolute error between predicted and actual values.	Qvars	Regression	$\text{Median}\left(\lvert\hat{y}_i-y_i\rvert\right)$
Mean squared error/difference (MSE or MSD)	Measures the average difference between the predicted and actual values.	Qvars	Regression	$\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i-y_i)^2$
Root mean squared error/difference (RMSE or RMSD)	The square root of the MSE to interpret the error in the same unit of the variable.	Qvars	Regression	$\sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i-y_i)^2}$
Root mean square percentage error (RMSPE)	Represents the RMSE expressed in percentage	Qvars	Regression	$\text{RMSE}\times 100\%$
Normalized root mean squared error (NRMSE)	Refers to the normalized RMSD to facilitate the comparison between variables with different scales.	Qvars	Regression	$\text{RMSE}/(y_{max}-y_{min})$
R-squared (R^2)	Measures the fit quality of a regression model/function by representing the proportion of the variance for a dependent variable in terms of independent variable(s).	Qvars	Regression	$1-\left(\sum_{i=1}^{m}(\hat{y}_i-y_i)^2/\sum_{i=1}^{m}(\bar{y}_i-y_i)^2\right)$

Table 10 (continued)

Metric	Description	Values type	Application	Formula
Theil U1 index	Measures the relative accuracy between the actual and predicted results.	Qvars	Regression	$\sqrt{\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2 / \sum_{i=1}^{m} y_i^2}$
Theil U2 index	Measures the quality of the predicted results.	Qvars	Regression	$\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2} / \left(\sqrt{\frac{1}{m}\sum_{i=1}^{m} y_i^2} + \sqrt{\frac{1}{m}\sum_{i=1}^{m} \hat{y}_i^2}\right)$
Accuracy (ACC)	Measures the closeness between the predicted values and the targets.	Cvars	Classification	$(TP + TN)/(TP + TN + FP + FN)$
Error rate (ERR)	Measures the proportion of the false predictions in total predictions.	Cvars	Classification	$(FP + FN)/(TP + TN + FP + FN)$
Precision (PPV)	Measures the closeness the set of predicted result	Cvars	Classification	$\mathrm{TP}/(TP + FP)$
Recall or True positive rate (TPR)	Measures the proportion of correct positive (TP) predictions in the true positive class.	Cvars	Classification	$\mathrm{TP}/(TP + FN)$
False-positive rate (FPR)	Measures the proportion offalse positive (FP) predictions in the true negative class.	Cvars	Classification	$\mathrm{FP}/(FP + TN)$
True-negative rate (TNR)	Measures the proportion of correct negative (TN) predictions in the true negative class.	Cvars	Classification	$\mathrm{TN}/(TN + FP)$
False-negative rate (FNR)	Measures the proportion of false negative (FN) predictions in the true positive class.	Cvars	Classification	$\mathrm{FN}/(TP + FN)$
F1-score	Measures the harmonic average of the precision and recall	Cvars	Classification	$2(PPV \times TPR)/(PPV + TPR)$
Matthews correlation coefficient (MCC)	Measures the quality of binary classifications.	Cvars	Classification	$\frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$

Table 10 (continued)

Metric	Description	Values type	Application	Formula
Prediction interval coverage probability (PICP)	Evaluates whether the actual value α is within the prediction interval limits,$\alpha_i = 1$ if it lies within the prediction interval, $\alpha_i = 0$ otherwise	Qvars	Regression	$\frac{1}{N}\sum_{i=1}^{N}$,
Prediction interval normalized average width (PINAW)	Measures the width of the prediction interval with L_i and U_i being the lower and the upper boundaries, respectively, and E is the difference between the maximum and the minimum actual values.	Qvars	Regression	$\frac{1}{NE}\sum_{i=1}^{N}\left(U_i - L_i\right)$
Precision recall curve (PRC)	Used to present the trade-off between precision and recall using different thresholds.	Cvars	Classification	–
Receiver operating characteristic curve (ROC)	Used to present the trade-off between the false positive rate and true positive rate using different thresholds.	Cvars	Classification	–
Area under the ROC (AUROC)	Represents the area under the ROC curve and the greater the value, the better the classification performance	Cvars	Classification	–
Cross validation (CV)	A resampling procedure to evaluate the generalization ability of an AI model	Cvars and Qvars	Classification and Regression	–
Confusion matrix	A tabulated representation of classification of results of the algorithm under study.	Cvars	Classification	–

4.2 Data scarcity and data benchmarking

The different applications of BAMSs necessitate extensive historical data to train the AI-big data analytics, especially those based on DL algorithms before they can be used reliably. However, large-scale data might not be available for some reason or can not be recorded representatively and sufficiently in a short time when we study newly-built environments. Fortunately, the problems addressed within each specific AI-big data analytics task in BAMSs illustrate some similarities. This could be justified by the fact that the different application tasks, despite studying distinct problems, use the same data-driven algorithms which are validated on slightly similar datasets collected from different kinds of buildings and devices. This has opened opportunities for using knowledge transfer and transfer learning to overcome the lack of datasets in some situations, e.g. sports facilities.

On the other hand, collecting and benchmarking data represents the most significant challenge so far when applying AI-big data analytics in BAMSs, especially for the case of large buildings, i.e. sports facilities, commercial centers, industrial buildings. Many tasks require annotated datasets to train AI models and validate them. Indeed, developing and validating new data-driven algorithms require recording and annotating large-scale datasets, especially when using DL algorithms that are notoriously data-hungry (Kučera and Pitner 2018). Improving the performance of BAMSs does not rely only on the selection of AI algorithms but also on the quality and parameters of datasets used to train them.

For instance, labeled and accurate anomaly detection datasets are needed for developing new automatic anomaly detection solutions. Similarly, development and validation of occupancy detection algorithms require repositories of building occupancy profiles with concurrent ground-truth people counts. In this context, to further improve the performance of BAMSs, public and open-access datasets are needed for different application fields (e.g., load forecasting, anomaly and fault detection, demand response, occupant-centric controls, IEQ monitoring, water monitoring, etc.) for assessing the AI-big data analytics algorithms developed by the AI Community (Park et al. 2019; Francisco et al. 2020).

From another hand, successful sustainability strategies to overcome this issue could be via (i) incentivizing buildings/facilities managers for participating in benchmarking campaigns and surveys organized for different application fields, and (ii) encouraging the AI community in organizing data benchmarking competitions and challenges.

4.3 Security and privacy preservation

The nature of data collected in BAMSs introduces new challenges in data analytics, i.e. security and privacy preservation, in which traditional technologies can not deal with. Using encapsulated protocols and IP-based communication, BAMSs are more and more connected to corporate networks and also remotely accessed for management reasons, both for emergency and convenience purposes. However, security and privacy preservation have not been set as a primary concern when designing these protocols. Therefore, most of the BAMSs are being operated with sub-standard or non-existent security implementations, and mainly rely on ensuring security by obscurity. In this line, there has been recently a move to address the shortfalls of security and privacy preserving implementations in BAMSs (Stamatescu et al. 2020; Ashaj and Erçelebi 2020). However, the definition of the new threats against BAMSs, and identification of these threats is still a field that is exceptionally lacking.

Moreover, another critical concern about security in BAMSs is related to the fact that buildings' data is valuable not only for managers and other BAMS competitors, as it is attached to the control of buildings' equipment. Typically, it could be significantly critical to end-users if manipulated. To that end, sharing data in most BAMSs has limited the buildings' intranet. Also, any attempts to extract this data to cloud data centers can result in severe security risks, considerably higher costs for the appropriate security systems or both (Lv et al. 2021; Himeur et al. 2022).

4.4 Scalability and interoperability

An important issue of BAMSs is the inherent lack of scalability and interoperability. This is because each BAMS manufacturer has its own proprietary data protocol that requires the development and maintenance of various processes and integrations. Moreover, BAMS vendors usually have competing products and thus are incentivized to make their data inaccessible to third parties (Png et al. 2019; Tang et al. 2020). Therefore, the interoperability is a legitimate concern for making efficient smart buildings as it refers to the ability of all the systems inside a building to communicate with one another. Specifically, with the actual proliferation of intelligent building technologies designed and manufactured by a plethora of companies, there is a need to make them communicate universally for promoting their deployment inside residential, commercial, industrial, and office buildings (Ozturk 2020; Miori et al. 2019). Put differently, as smart buildings need to meet energy and water efficiency, adequate indoor environmental conditions, high comfort levels, and economic goals set by building managers/users, they require the use of a highly-connected building automation system in which different parts can efficiently communicate with one another and adjust to changes in the environment. However, while numerous buildings are equipped with excellent systems, they often lack a combined monitoring system for lighting, climate control, water monitoring and blinds that could facilitate efficiency measures (Schachinger et al. 2017).

The BAMS community makes great efforts to develop and deploy communication protocols that ensure interoperability, such as Modbus, KNX, LonWorks, and Protocol 3964R (Merz et al. 2018). For instance, LonWorks and KNX are interoperable open standards as they can be used together. However, integration concerns may arise (Tang et al. 2020). The technologies' popularity can hinder interoperability potential in BAMSs, as the example with LonWorks, the market leader in the United States. In contrast, KNX—widely used in Europe—has yet to impact (Merz et al. 2018). Additionally, the installation and operation costs can pose severe integration limitations. Indeed, many contractors or system integrators that provide off-the-shelf solutions are less concerned about whether the integration is successful.

4.5 Real-time big data intelligence

Collecting and analyzing data in real-time are of significant importance while designing powerful and efficient BAMSs. The first step towards this is by adopting a real-time sub-metering, which helps in tracking track utility costs by building region (floor, room, etc), by the tenants, by individual facility equipment (e.g., HVAC, lighting), etc. Therefore, granular utility sub-metering data provides the essential tools for monitoring energy costs/performance, water consumption/waste. This will result in accurately identifying usage

anomalies, enabling data-driven portfolio analysis, etc. Although the value of real-time sub-metering is unquestioned, most of the BAMSs are still unable to provide the real-time data monitoring, which delays decision-making measures and hence reduces then the quality of efficiency and optimization operations.

5 Case studies

This literature review established several applications of AI big data analytics for buildings in terms of energy management, load forecasting, water management, FDAD, or IEQ monitoring. In this section, we present their deployment for energy-related applications given the continuous rise in energy consumption worldwide of the buildings sector under the global energy dilemma and the energy optimization potential of BAMSs, given the increasing concerns about energy efficiency in buildings. More specifically, the case studies handle two of the lead causes of energy waste in buildings, which are (i) system faults and equipment malfunctioning, and (ii) poor management and regulation of the buildings' systems (Alsalemi et al. 2022; Elnour et al. 2022, 2020). The first two case studies present strategies for energy anomaly detection that can be due to both or either of the former causes. They demonstrate the deployment of two different methods: unsupervised and supervised learning. The last case study presents the use of AI data analytics to establish reliable and efficient regulation of HVAC systems, given that those are considered major energy consumers in buildings (energy.gov, Elnour et al. 2022; Fadli et al. 2021).

5.1 Unsupervised AI-based energy anomaly detection

This section presents an example of using unsupervised ML algorithms for detecting abnormal energy consumption (Himeur et al. 2022). Therefore, four algorithms are considered, namely (i) OCSVM with linear kernel, (ii) OCSVM with Gaussian kernel, (iii) DBSCAN, and (iv) LOF. They have been applied on the Dutch residential energy dataset (DRED), which incorporates electricity consumption, occupancy patterns and ambient conditions of a typical household (in the Netherlands). Figure 6 portrays the scatter plot of energy footprints in which normal and abnormal patterns are identified using the aforementioned approaches. It has been clearly seen OCSVM (with linear kernel) detects more energy samples that fall outside the inlier region, which refer to consumption anomalies. While by using OCSVM (with a Gaussian kernel), the number of samples that fall inside the inlier region has been reduced because of its separation capability introduced by the hyperplane generated using the Gaussian kernel. From another side, LOF and DBSCAN help detect abnormal patterns with almost the same efficiency as OCSMV (with the Gaussian kernel), and only a slight difference has been registered in classifying a few numbers of samples.

5.2 Supervised AI-based energy anomaly detection

Supervised ML algorithms excelled in detecting abnormal energy usage, although they require labeled energy data. To that end, in Himeur et al. (2021), a micro-moment-based approach is introduced to cluster energy footprints of an office building (at Qatar University) into five classes with reference to the energy consumption, occupancy patterns and appliance operation specifications. These classes are named "class 0: good

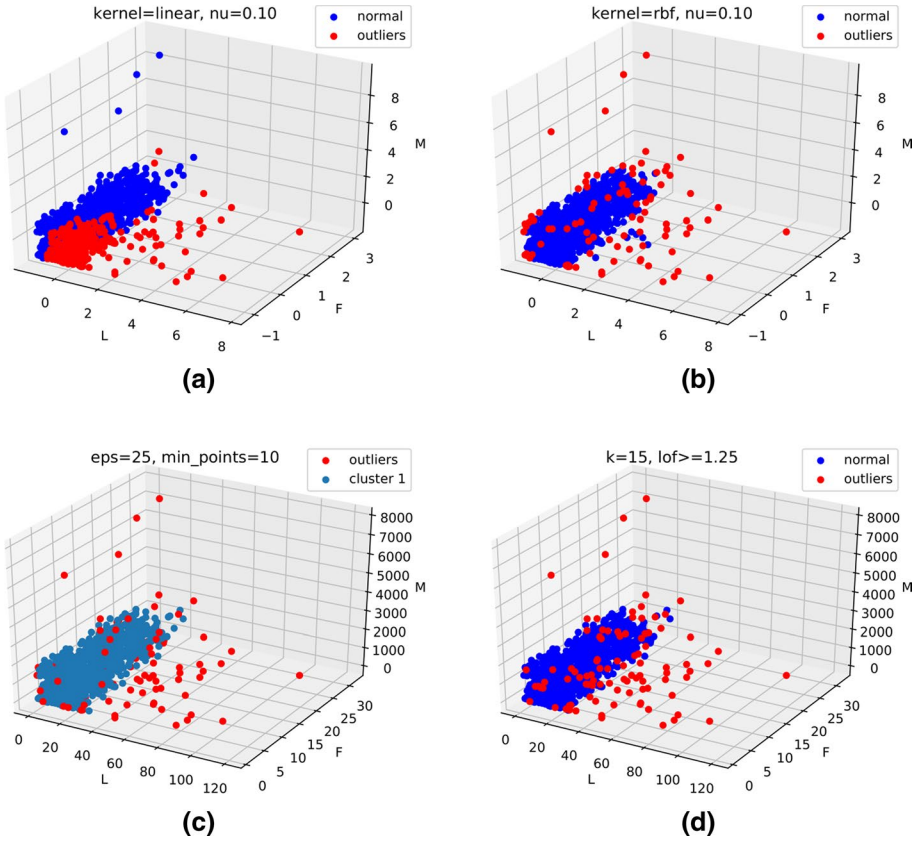

Fig. 6 Energy consumption anomaly detection in residential buildings using a) OCSVM with linear kernel, b) OCSVM with Gaussian kernel, c) DBSCAN and d) LOF

usage", "class 1: turn on appliance", "class 2: turn off appliance", "class 3: excessive consumption" and "class 4: consumption while outside". Following, an improved KNN model is developed and used to learn abnormal energy usage using this annotated data. Fig. 7 illustrates the flowchart of the micro-moment based scheme used to extract and learn intent-driven moments of energy consumption. Typically, energy micro-moment features *MF* are extracted based on analyzing occupancy profiles (O) and power consumption (p) of each device in reference to device active consumption range (*DACR*), device operation time (*DOT*) and device standby power consumption (*DSPC*). Then, the appliance operation parameters are called, including *DACR*, *DOT* and *DSPC*. Table 11 presents an example of different appliance parameter specifications that are used in the rule-based algorithm to extract power consumption micro-moments (Himeur et al. 2022).

To have a clear view of how abnormal energy consumption is distributed over the time, the scatter plot of energy consumption profiles of a television is illustrated in Fig. 8. Accordingly, the corresponding normal and abnormal energy patterns are detected using an IKNN model and micro-moment analysis. Because this approach uses a supervised learning with regard to occupancy data, it has the capability of

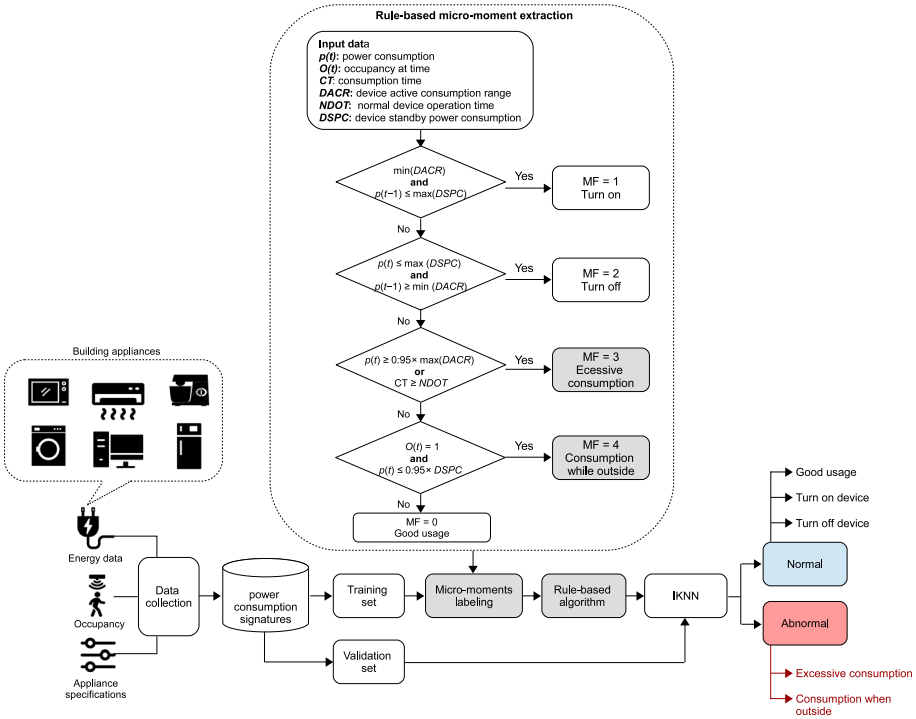

Fig. 7 Block diagram of the supervised ML solution used to detect abnormal energy consumption in office buildings

identifying new consumption anomalies that correspond to the absence of the end-users when the television is on (this abnormality can be extended to other devices that require the presence of the user during their operation, e.g. the air conditioner, heater, fan, etc.). Detecting such abnormalities was not doable if an unsupervised ML model was deployed, in which only energy patterns were analyzed.

5.3 Energy and performance optimization for sports facilities

In light of the increased global energy demand and its associated environmental impacts, the management and optimization of sports facilities are becoming imperative as they are characterized by high energy demand and occupancy profiles. This case study demonstrates the application of the model predictive control (MPC) theory and NNs for energy and performance management of sports facilities. Figure 9 presents the proposed NN-based MPC framework. The work is carried out using the building information model of a sports hall in the sports complex of Qatar University using EnergyPlus and practical data for model calibration. MPC systems are robust as they allow integrated dynamic optimization that accounts for the future system behavior in the decision-making process. NNs are advantageous for their ability to represent complex functions with high accuracy.

Table 11 Power consumption specifications for different home appliances

Appliance	DOT	DACR (watts)	DSPC (watts)	Appliance	DOT	DACR (watts)	DSPC (watts)
Air conditionner	15 h 30 min	1000	4	Washing machine	1 h	500	6
Microwave	1h	1200	7	Light	8 h	60	0
Oven	3h	2400	6	Television	12 h 42 min	65	6
Dishwasher	1h 45 min	1800	3	Refrigerator	17 h 30 min	180	0
Laptop	12 h 42 min	100	20	Desktop	12 h 42 min	250	12

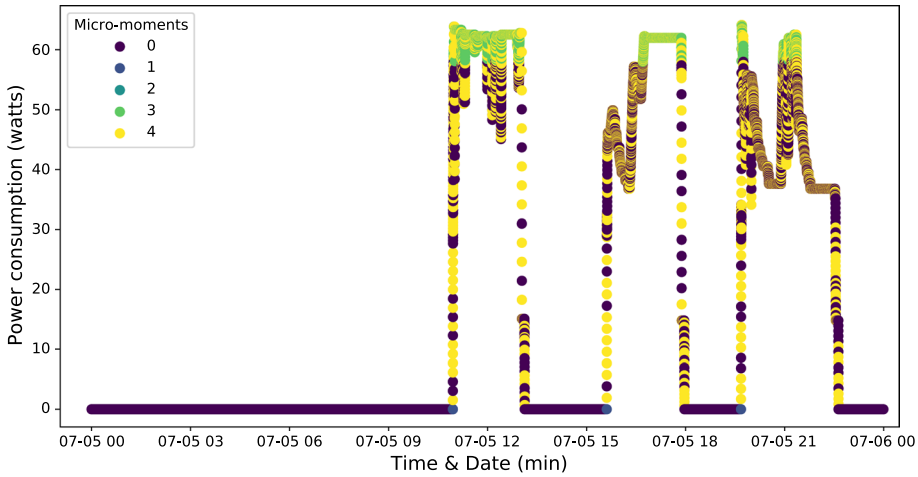

Fig. 8 Scatter plot of energy micro-moments identified using IKNN (Himeur et al. 2021)

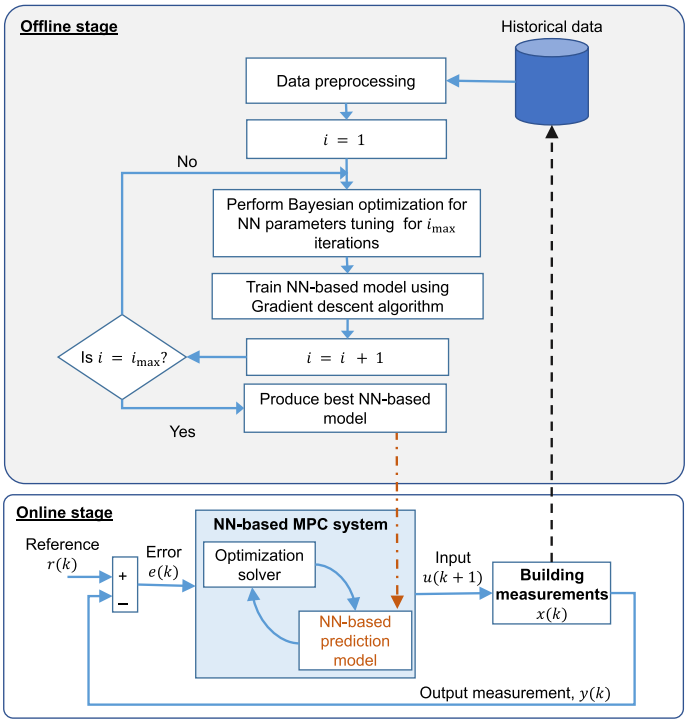

Fig. 9 Block diagram of the NN-based MPC framework for sports facilities energy and performance optimization

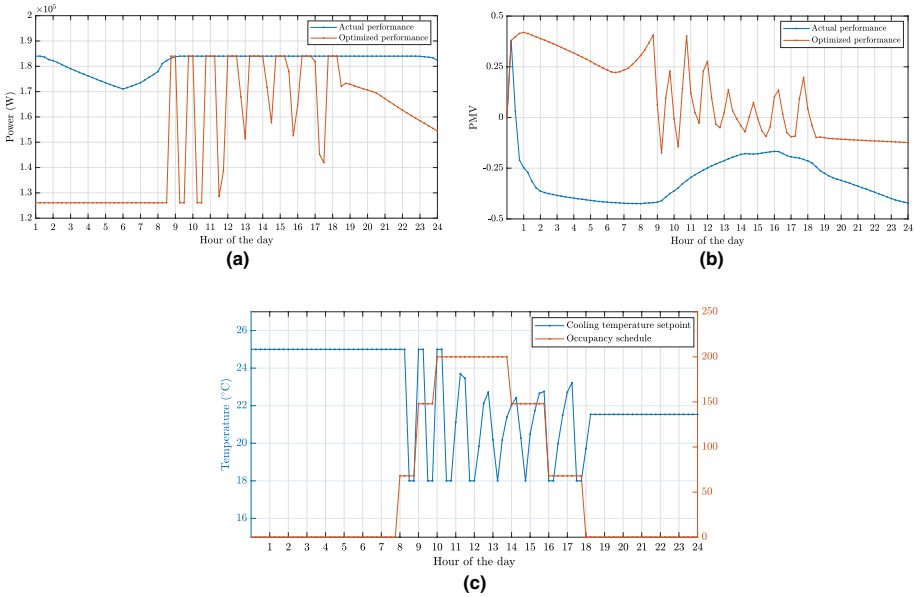

Fig. 10 The performance of the NN-based MPC system for energy and performance optimization in the sports hall in Qatar University sports complex

The NN-based dynamic prediction model aims to express and capture the behavior of the building operation over time given its states $x(k)$ (i.e., power usage, thermal comfort, indoor and outdoor air properties, etc.) and its inputs (i.e., HVAC system settings). The NN-based prediction model is:

$$F = \text{Train\_NN}(x(k+1), [x(k), u(k+1)]), \tag{1}$$

and the prediction is computed by:

$$\hat{x}(k+1) = F(x(k), u(k+1)). \tag{2}$$

The optimization of the NN's hyper-parameters was performed using the Bayesian optimization algorithm which keeps track of past iterations to find better choices for the next set of hyper-parameters to evaluate (Andonie 2019).

The MPC system consists of an optimizer and an NN-based prediction model of the building operation, and based on the system output $y(k) \subset x(k)$ and its reference value $r(k)$ (i.e., power usage and thermal comfort level), the HVAC system settings for temperature setpoints and dampers positions are determined (i.e., $u(k)$) using numerical optimization to achieve tracking. When compared to routine performance, the proposed approach was able to achieve significant energy reduction and adequate thermal comfort levels as demonstrated in Fig. 10. Energy savings of around 15% was observed, which was approximated by evaluating the relative change in the total energy consumption in the two settings for the scenario under study, that is, the relative difference between the areas under the two power curves in Fig. 10a. Considerations about the NN model

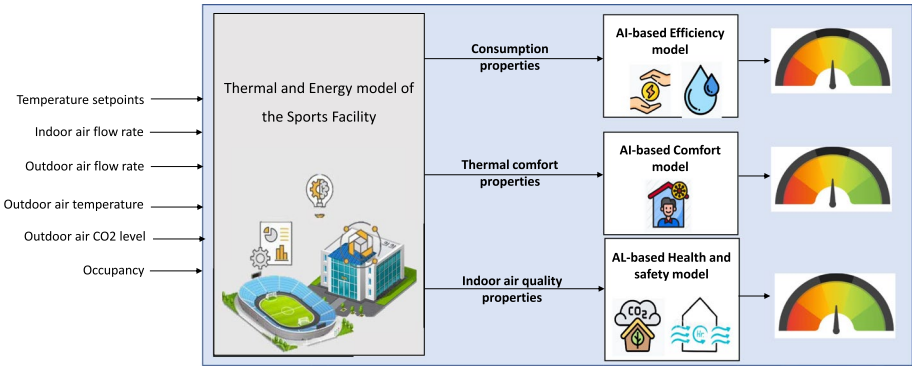

Fig. 11 Block diagram of the improved CUSP model with efficiency-comfort-health model

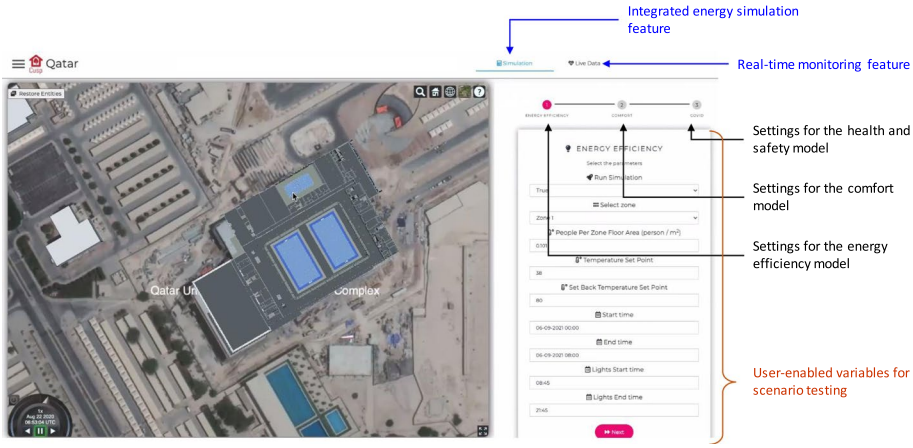

Fig. 12 The web interface of the expanded and improved CUSP platform with the three integrated models for efficiency, comfort, and health & safety

performance, tuning of the MPC settings, and optimization sub-optimality or failure are essential during the design and implementation phases of the proposed framework.

5.3.1 Improved computational sustainability model for sports facility management

The computational urban sustainability platform (CUSP), developed at Cardiff University, is an immersive decision support tool built to deliver a powerful urban analytics and enable interactive monitoring and inform decision making through a web interface. It can be used to promote co-simulation across disciplines, and predict future scenarios towards a sustainable future operation and urban Intelligence [Computational Urban Sustainability Platform (CUSP)]. The CUSP model can be improved to include three integrated models, which are

(1) energy-water efficiency, (2) health, safety, and wellbeing, and (3) comfort as demonstrated in Fig. 11.

The improved CUSP model integrates an energy simulation tool that is used to generate data of the particular scenario under consideration for data analytics for quality monitoring and planning purposes. It contains three AI-based models developed to assess each of the three aspects of the facility operation, which are efficiency and sustainability, health and safety, and users' thermal satisfaction. Through the web interface shown in Fig. 12, the integrated simulation tools will enable facility managers to evaluate the possible scenario in terms of the HVAC system settings, and occupancy and operation schedules towards achieving a reasonable trade-off between those three aspects prior to applying them in the facility.

6 Future directions

6.1 Multimodal data analysis

Due to the advancement of today's sensing and mobile technologies, various modalities of data can be easily and effectively gathered using different and advanced means. Thus, it is now possible to record and process big data about environmental satisfaction levels of buildings' occupants in real-time and non-invasive manners (Plageras et al. 2018). Buildings' end-users naturally react to ambient environmental conditions for minimizing any environmental stress, increasing their comfort based on their autonomic nervous systems and expressed by different poses, which can effectively influence different building operation parameters (Amato et al. 2018). Therefore, it becomes of utmost importance to develop tools for enhancing the interdisciplinary knowledge (i.e. AI, IoT, big data, DL, computer vision) when managing building operations. This helps significantly advance building indoor environmental control and sensing technologies as a function of human bio-signals (i.e. physiological signals) and poses.

Analyzing multi-modal data helps BAMSs in boosting workplace productivity and optimising office spaces, which in turn cutting costs and increasing revenues for companies. Moreover, data generated from these systems could be used to reduce the spread of viruses and other diseases inside buildings, increasingly important since the outbreak of Covid-19 (Sun and Zhai 2020). For instance, in Ding et al. (2020), investigate the collective contagion of the COVID-19 virus inside indoor environments (i.e. healthcare facilities and public vehicles) along with the engineering control against virus spread with ventilation systems.

6.2 In-situ sensor calibration in BAMSs

Sensors are key players in helping BAMSs to attain expected efficiency and automation. However, they are affected by continuous failures and degradation over time. To that end, in-situ sensor calibration plays a crucial role in calibrating different BAMS working sensors (i.e., physical sensors) and avoiding significant errors for reliable results when it is deployed to large-scale building sensor networks (Yu and Li 2015). Most of the studies opted for the conventional periodical calibration as a solution to overcome sensor degradation and failure; however, this is impractical and difficult for various sensors. By contrast, virtual in-situ calibration (VIC) can be a good alternative since

it relies on mathematically extracting the characteristics of essential aspects involved in a calibration, such as the uncertainty quantification, benchmark establishment, and environment assessment (Yoon and Yu 2018). Moreover, because BAMSs need digitally enhanced data-rich environments, virtual sensors offer reliable and informative sensing contexts for operational datasets in BAMSs. More specifically, in-situ virtual sensors help develop the counterparts of target physical sensors in the field. Therefore, they can provide extra data related to residuals between physical and virtual sensors for for deployment in data-driven modeling, diagnostics and analytics (Koo et al. 2022).

6.3 Smart building digital twins

The increasing amounts of data generated by BMAMs, and the need for new methods to leverage it, have motivated scientists to investigate new strategies. One promising solution is using the digital twins (DT) paradigm, which assumes a complete cohesion and integration between the visual and physical worlds. Typically, DT can deliver considerable benefits to the BAMSs and the built environment in general by helping bring together static and dynamic data from various sources (in 2D/3D models) and assisting in making effective and informed decisions. Moreover, it combines the knowledge from the physical and digital worlds by collecting real-time data from the physical environments and provides a real-time understanding of buildings' performance (Delgado and Oyedele 2021). Besides, despite the gradual exploration of digital twinning within the fields of building information modeling (BIM) and cyber-physical systems (CPS), available tools and techniques need to be considered in the next level of integration (technologies and procedures). This is to (i) provide DTs with more adaptability and more cohesion over the managed information and (ii) extract more value from our virtual models (Shahzad et al. 2022).

6.4 Transfer learning

Specifically, transfer learning has recently been proposed as solution that can be investigated for the case of buildings with poor information data (Himeur et al. 2022). Put simply, data and knowledge of already existing buildings (or old buildings) with rich energy usage records, water management data, occupancy patterns, IEQ monitoring footprints and and ambient environmental conditions can be used. Therefore, various frameworks have been introduced for target energy forecasting (Gao et al. 2020; Li et al. 2021), anomaly detection of energy consumption (Liang et al. 2018; Xu et al. 2021), fault diagnosis of energy systems (Liu et al. 2021; Zhu et al. 2021), HVAC fault detection (Dowling et al. 2020), IEQ monitoring (Tariq et al. 2021), indoor occupancy detection (Khalil et al. 2021), etc.

6.5 Blockchain

Due to the security and privacy issues that are still open in BAMSs, blockchain is considered as a promising solution that provides the digital trust. It can function as a permanent,

cloud-based and digital ledger of activities between different users and partners (Nawari and Ravindran 2019; Liu et al. 2021). Also, blockchain can operate as a distributed, single source of shared truth and has the possibility of becoming the top-system for recording all transactions. Therefore, its deployment in BAMSs aims at (i) tracking and validating changes (e.g. security and surveillance, access control, etc.), (ii) monitoring HVAC activities, (iii) recording property transfers, and (iv) detecting occupancy patterns (Siountri et al. 2020). Additionally, it can help manage intelligent buildings and IoT devices with renewable energy, e.g. wind and solar. For example, suppose a facility is in a two-way energy communication with the grid. In that case, blockchain can make it more secure and easier to develop a digital record of energy-in and energy-out transactions (Tiwari and Batra). On another side, as the global market of building automation exceeds $120 billion, smart contracts can be utilized for automating warranties and providing refunds when IoT-connected devices or equipment do not perform as expected (Himeur et al. 2022).

Overall, there are numerous potential applications of blockchain in BAMSs, although the principal advantages are data is easy to access, is secure, and can not be corrupted. Specifically, data stored in the blockchain database can be easily and quickly reviewed, even though it is managed by distinct entities, which results in an accurate and fast data analysis (Nawari and Ravindran 2019). Moving forward, blockchain helps in streamlining processes and lowering costs through reducing and/or eliminating those dreaded manual operations, especially in public buildings, sports facilities, and commercial centers. This could be adapted to almost any process, including preventive maintenance, work orders, environmental health, and safety planning, and space management (Nawari and Ravindran 2019).

Only for energy management in smart buildings, blockchain has found diverse applications. For instance, in Van Cutsem et al. (2020), use a blockchain-based approach to cooperate energy management of multiple end-users in smart-buildings, where smart-contracts have been utilized to allow decentralizing community energy management. In Mukherjee et al. (2021), a smart energy management solution is safeguarded with blockchain and hence ensures judicious generation, uniform distribution and shielded monitoring along with guaranteed security and privacy of the havoc data. In Tiwari and Batra, blockchain is introduced for enabling the reparation of smart buildings-cyber physical systems. Moving on, decentralized and flexible access control using smart contracts is developed for smart and large commercial buildings in Bindra et al. (2021). This solution has been proposed as an alternative to inefficient, unsystematic, and human-intensive access control schemes usually used in these buildings. While the widespread implementation of blockchain is still a long way off, it is also challenging to deploy this technology reliably and widely in BAMSs. This research area needs to be further investigated in the near future. This promising new technology could benefit the other tasks of smart buildings, i.e. water management, IEQ monitoring, occupancy detection, etc.

6.6 Cyber-security standards for BAMSs

While using AI in BAMSs represents a powerful asset, it also presents some data security and privacy concerns and problems with the regulations. Typically, AI-driven BAMSs involve the deployment of lower-cost sensors (both wired and wireless) and the adoption of cloud, fog, edge, and/or hybrid computing architectures, increasing cyber risks. To that end, the need for a sound cybersecurity strategy has become crucial for promoting secure remote BAMSs. Data flows must be planned and monitored, possibly

making it necessary to use one-way data diodes. On the other hand, BAMSs integrate heterogeneous sensing, computation, and control capabilities. They combine cyberspace with the physical world to develop cyber-physical systems. However, the security of BAMSs is significantly threatened by software/hardware failures and/or cyber/physical attacks. For example, sensor failures can engender false detection of abnormal energy/water consumption behaviors and result in actuator misbehavior.

To handle the above issues, privacy and security protection mechanisms should be enforced. This is possible by providing recommendations to the building automation community, e.g., the data protection directive 95/46/EC (Tokarski) suggests recommendations for supporting the security of the implementation of smart metering and smart using data controllers. Addressing these recommendations can enable moving to fully harmonized data protection environments and improving security measures in BAMSs. Moreover, different cyber security standards can be used to secure BAMSs by addressing the cybersecurity for operational technology in automation and control systems, such as ISA/IEC 62443 series (Bicaku et al.) and ASHRAE 135 series (BACnet) (Tang et al. 2020).

6.7 Self-learning for long-term building operation

Self-learning ML models are key to realizing the BAMS in the long-term building operation. The systems built upon these models have recently gained industry recognition and market share as they are based on using a ”user-friendly” technology (Cortiços 2019). Typically, self-learning, also called self-supervision, is an emerging technology that helps develop computationally efficient, low-cost, autonomous, and self-supervised ML algorithms (Kaklauskas et al. 2019). For example, for energy management, a self-learning control scheme assists in assessing the energy flexibility of buildings, in addition to guaranteeing robustness, scalability, and adaptability. Moreover, automated self-learning systems have promising perspectives when they are to integrate demand-response strategies for effective home-energy management systems (Bampoulas et al. 2019).

6.8 Edge analytics for BAMSs

With the advancement of BAMSs and the latest generation of IoT devices, data acquisition from multiple types of equipment has become much easier in today’s buildings. Real-time access to this data helps in better managing facility operations, sustaining efficiency, and lowering costs. However, as most BAMSs are only implemented using cloud computing, real-time data analysis may not be guaranteed. To overcome this issue, open software platforms hosted on edge nodes in close proximity to the building’s IoT devices can enable access to the building data and advanced analytics deployed on these platforms in real-time. In this context, edge computing employs the processing power of IoT devices for filtering, pre-processing, aggregating and storing recorded data, and actions can then be performed in real-time using adequate analytical algorithms. This is because edge computing enables resolving bandwidth and latency problems and reducing response time. Following, the filtered data could be transmitted to the cloudlet platforms for aggregation and enrichment, and running of complex analytics (Sharma et al. 2018).

Thus, various use cases where edge computing and the IoT can efficiently be utilized in BAMSs are emerging, among them fault diagnosis, which helps to (i) find patterns in sensor data representing equipment failures, anomalies, or degraded performance; (ii) detect abnormal energy consumption, e.g. if the lighting or HVAC systems are activated too early or operate too late with regard to the actual occupancy schedules; and (iii) identify correlations across different types of data, which are essential to infer the factors impacting energy consumption (e.g. the patterns related to weather, age of facilities, etc). Overall, open edge software platforms combine multi-protocol connectivity and the ability to aggregate data from multiple sources and facilitate the task of advanced analytics in turning this data into actionable information that can be used to improve the overall operational efficiency buildings (Petri et al.).

7 Conclusion

This paper carried out a comprehensive overview of the application of AI-big data analytics in BAMSs to conduct different tasks, including energy forecasting, fault and anomaly detection, water monitoring, and IEQ monitoring. The pros and cons of AI models within the unsupervised, supervised, semi-supervised, and reinforcement learning categories have been identified. Moreover, it concluded that supervised learning algorithms excelled well in performing the diver BAMS tasks, but their performance always relies on the availability of annotated data and its accuracy. Unsupervised learning models with no prior knowledge can address this issue with less efficiency.

It was demonstrated in this framework that technologies of ML, IoT, and new connectivity capabilities have a critical role in shaping the future of BAMSs. With building owners and facility managers focusing heavily on improving energy efficiency and increasing cost savings, features like advanced fault detection and diagnostics, energy analytics, IEQ monitoring, and water management are becoming critical. The growing interest devoted to developing intelligent analytics in BAMSs has been highlighted by the increasing number of works and studies proposed in the literature to address several challenges. In the coming years, data analytics is expected to expand the capabilities of intelligent building technologies, spurring further advancements in building automation systems and equipment standards while minimizing the environmental impact of commercial buildings.

The AI-big data analytics technology is up-and-coming to BAMSs. However, it faces various challenges for achieving market penetration, including legal, regulatory, security and privacy preservation, interoperability and scalability, and competition barriers. Additional research initiatives, investigations, projects, and collaborations should be considered a primary requirement for showing if the technology can reach its absolute power, prove its commercial viability, and lastly, be adopted in the mainstream.

Acknowledgements This publication was made possible by NPRP Grant No. NPRP12S-0222-190128 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

Funding Open Access funding provided by the Qatar National Library.

References

Aazam M, Zeadally S, Harras KA (2018) Deploying fog computing in industrial internet of things and industry 4.0. IEEE Trans Ind Inform 14(10):4674–4682

Abba S, Pham QB, Usman A, Linh NTT, Aliyu D, Nguyen Q, Bach Q-V (2020) Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. J Water Process Eng 33:101081

Abdulhammed R, Musafer H, Alessa A, Faezipour M, Abuzneid A (2019) Features dimensionality reduction approaches for machine learning based network intrusion detection. Electronics 8(3):322

Aboelwafa MM, Seddik KG, Eldefrawy MH, Gadallah Y, Gidlund M (2020) A machine-learning-based technique for false data injection attacks detection in industrial IoT. IEEE Internet Things J 7(9):8462–8471

Acquaah Y, Steele JB, Gokaraju B, Tesiero R, Monty GH, Occupancy detection for smart HVAC efficiency in building energy: a deep learning neural network framework using thermal imagery. In: 2020 IEEE applied imagery pattern recognition workshop (AIPR). IEEE, pp 1–6

Afaifia M, Djiar KA, Bich-Ngoc N, Teller J (2021) An energy consumption model for the Algerian residential building's stock, based on a triangular approach: geographic information system (gis), regression analysis and hierarchical cluster analysis. Sustain Cities Soc 74:103191

Agha G, Palmskog K (2018) A survey of statistical model checking. ACM Trans Model Comput Simul (TOMACS) 28(1):1–39

Aghemo C, Blaso L, Pellegrino A (2014) Building automation and control systems: a case study to evaluate the energy and environmental performances of a lighting control system in offices. Autom Constr 43:10–22

Ahmad T, Chen H (2018) Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. Energy Build 166:460–476

Ahmad T, Chen H (2018) Utility companies strategy for short-term energy demand forecasting using machine learning based models. Sustain Cities Soc 39:401–417

Ahmad T, Chen H, Huang R, Yabin G, Wang J, Shair J, Akram HMA, Mohsan SAH, Kazim M (2018) Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. Energy 158:17–32

Ahmad T, Huanxin C, Zhang D, Zhang H (2020) Smart energy forecasting strategy with four machine learning models for climate-sensitive and non-climate sensitive conditions. Energy 198:117283

Ahn J, Shin D, Kim K, Yang J (2017) Indoor air quality analysis using deep learning with sensor data. Sensors 17(11):2476

Akil M, Tittelein P, Defer D, Suard F (2019) Statistical indicator for the detection of anomalies in gas, electricity and water consumption: application of smart monitoring for educational buildings. Energy Build 199:512–522

Al Dakheel J, Del Pero C, Aste N, Leonforte F (2020) Smart buildings features and key performance indicators: a review. Sustain Cities Soc 61:102328

Al-Ali A-R, Zualkernan IA, Rashid M, Gupta R, AliKarar M (2017) A smart home energy management system using IoT and big data analytics approach. IEEE Trans Consum Electron 63(4):426–434

Alawadi S, Mera D, Fernández-Delgado M, Alkhabbas F, Olsson CM, Davidsson P (2022) A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. Energy Syst 13:1–17

Alghamdi A, Hu G, Haider H, Hewage K, Sadiq R (2020) Benchmarking of water, energy, and carbon flows in academic buildings: a fuzzy clustering approach. Sustainability 12(11):4422

Alhussein M, Aurangzeb K, Haider SI (2020) Hybrid CNN-LSTM model for short-term individual household load forecasting. IEEE Access 8:180544–180557

Aliberti A, Bottaccioli L, Macii E, Di Cataldo S, Acquaviva A, Patti E (2019) A non-linear autoregressive model for indoor air-temperature predictions in smart buildings. Electronics 8(9):979

Aljabre SH (2002) Hospital generated waste: a plan for its proper management. J Family Commun Med 9(2):61

Al-Kababji A, Alsalemi A, Himeur Y, Fernandez R, Bensaali F, Amira A, Fetais N (2022) Interactive visual study for residential energy consumption data. J Clean Prod 366:132841

Alsalemi A, Himeur Y, Bensaali F, Amira A, Sardianos C, Varlamis I, Dimitrakopoulos G (2020) Achieving domestic energy efficiency using micro-moments and intelligent recommendations. IEEE Access 8:15047–15055

Alsalemi A, Himeur Y, Bensaali F, Amira A, Sardianos C, Chronis C, Varlamis I, Dimitrakopoulos G (2020) A micro-moment system for domestic energy efficiency analysis. IEEE Syst J 15(1):1256–1263

Alsalemi A, Himeur Y, Bensaali F, Amira A (2022) An innovative edge-based internet of energy solution for promoting energy saving in buildings. Sustain Cities Soc 78:103571

Alsalemi A, Al-Kababji A, Himeur Y, Bensaali F, Amira A (2020) Cloud energy micro-moment data classification: a platform study. In: 2020 IEEE/ACM 13th international conference on Utility and Cloud Computing (UCC). IEEE, pp 420–425

Alsalemi A, Himeur Y, Bensaali F, A Amira (2021) Smart sensing and end-user behavioral change in residential buildings: an edge internet of energy perspective. IEEE Sensors J 21(24), 27623–27631

Altunkaynak A, Nigussie TA (2017) Monthly water consumption prediction using season algorithm and wavelet transform-based models. J Water Resour Plan Manag 143(6):04017011

Amato G, Barsocchi P, Falchi F, Ferro E, Gennaro C, Leone GR, Moroni D, Salvetti O, Vairo C (2018) Towards multimodal surveillance for smart building security. In: Multidisciplinary digital publishing institute proceedings, vol 2, p 95

Andonie R (2019) Hyperparameter optimization in learning systems. J Membr Comput 1(4):279–291. https://doi.org/10.1007/s41965-019-00023-0

Antunes A, Andrade-Campos A, Sardinha-Lourenço A, Oliveira M (2018) Short-term water demand forecasting using machine learning techniques. J Hydroinf 20(6):1343–1366

Aparicio-Ruiz P, Barbadilla-Martín E, Guadix J, Cortés P (2021) Knn and adaptive comfort applied in decision making for HVAC systems. Ann Oper Res 303(1):217–231

Aquino I, Nawari NO (2015) Sustainable design strategies for sport stadia. Suburban Sustain 3(1):3

Ashaj SJ, Erçelebi E (2020) Energy saving data aggregation algorithms in building automation for health and security monitoring and privacy in medical internet of things. J Med Imaging Health Inform 10(1):204–210

Aste N, Manfren M, Marenzi G (2017) Building automation and control systems and performance optimization: a framework for analysis. Renew Sustain Energy Rev 75:313–330

Azuatalam D, Lee W-L, de Nijs F, Liebman A (2020) Reinforcement learning for whole-building HVAC control and demand response. Energy AI 2:100020

Bampoulas A, Saffari M, Pallonetto F, Mangina E, Finn DP (2019) Self-learning control algorithms for energy systems integration in the residential building sector, In: (2019) IEEE 5th world forum on internet of things (WF-IoT). IEEE 2019, pp 815–818

Barrett E, Linder S (2015) Autonomous HVAC control, a reinforcement learning approach. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, pp 3–19

Bashari M, Rahimi-Kian A (2020) Forecasting electric load by aggregating meteorological and history-based deep learning modules. In: IEEE power & energy society general meeting (PESGM). IEEE, pp 1–5

Bassamzadeh N, Ghanem R (2017) Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. Appl Energy 193:369–380

Berger MA, Mathew PA, Walter T Big data analytics in the building industry. ASHRAE J 58 (LBNL-1005983)

Bertone E, Sahin O, Stewart RA, Zou PX, Alam M, Hampson K, Blair E (2018) Role of financial mechanisms for accelerating the rate of water and energy efficiency retrofits in Australian public buildings: hybrid Bayesian network and system dynamics modelling approach. Appl Energy 210:409–419

Bessani M, Massignan JA, Santos TM, London JB Jr, Maciel CD (2020) Multiple households very short-term load forecasting using Bayesian networks. Electric Power Syst Res 189:106733

Bicaku A, Zsilak M, Theiler P, Tauber M, Delsing J (2022) Security standard compliance verification in system of systems. IEEE Syst J 16(2) 2195–2205

Bilous I, Deshko V, Sukhodub I (2018) Parametric analysis of external and internal factors influence on building energy performance using non-linear multivariate regression models. J Build Eng 20:327–336

Bindra L, Eng K, Ardakanian O, Stroulia E (2022) Flexible, decentralised access control for smart buildings with smart contracts. Cyber-Phys Syst 8(4): 286–320

Bode G, Baranski M, Schraven M, Kümpel A, Storek T, Nürenberg M, Müller D, Rothe A, Ziegeldorf JH, Fütterer J et al (2019) Cloud, wireless technology, internet of things: the next generation of building automation systems? J Phys 1343:012059

Bode G, Schreiber T, Baranski M, Müller D (2019) A time series clustering approach for building automation and control systems. Appl Energy 238:1337–1345

Bode G, Thul S, Baranski M, Müller D (2020) Real-world application of machine-learning-based fault detection trained with experimental data. Energy 198:117323

Bui V, Le NT, Nguyen VH, Kim J, Jang YM et al (2021) Multi-behavior with bottleneck features LSTM for load forecasting in building energy management system. Electronics 10(9):1026

Bünning F, Huber B, Heer P, Aboudonia A, Lygeros J (2020) Experimental demonstration of data predictive control for energy optimization and thermal comfort in buildings. Energy Build 211:109792

Cao S-J, Ding J, Ren C (2020) Sensor deployment strategy using cluster analysis of fuzzy c-means algorithm: towards online control of indoor environment's safety and health. Sustain Cities Soc 59:102190

Carreira P, Costa AA, Mansur V, Arsénio A (2018) Can hvac really learn from users? A simulation-based study on the effectiveness of voting for comfort and energy use optimization. Sustain Cities Soc 41:275–285

Chemingui Y, Gastli A, Ellabban O (2020) Reinforcement learning-based school energy management system. Energies 13(23):6354

Chen Y, Norford LK, Samuelson HW, Malkawi A (2018) Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. Energy Build 169:195–205

Chen S, Mihara K, Wen J (2018) Time series prediction of CO2, tvoc and hcho based on machine learning at different sampling points. Build Environ 146:238–246

Chen Y, Tong Z, Zheng Y, Samuelson H, Norford L (2020) Transfer learning with deep neural networks for model predictive control of hvac and natural ventilation in smart buildings. J Clean Prod 254:119866

Chen K, Chen H, Zhou C, Huang Y, Qi X, Shen R, Liu F, Zuo M, Zou X, Wang J et al (2020) Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water Res 171:115454

Chen Y, Zhang F, Berardi U (2020) Day-ahead prediction of hourly subentry energy consumption in the building sector using pattern recognition algorithms. Energy 211:118530

Chen Y, Wen J (2017) Whole building system fault detection based on weather pattern matching and PCA method. In: 2017 3rd IEEE international conference on control science and systems engineering (ICCSSE). IEEE, pp 728–732

Choi S, Hur J (2020) An ensemble learner-based bagging model using past output data for photovoltaic forecasting. Energies 13(6):1438

Choi Y, Yoon S (2021) Autoencoder-driven fault detection and diagnosis in building automation systems: residual-based and latent space-based approaches. Build Environ 203, 108066

Chou J-S, Ngo N-T (2016) Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. Appl Energy 177:751–770

Chou J-S, Tran D-S (2018) Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy 165:709–726

Collins AG, Cockburn J (2020) Beyond dichotomies in reinforcement learning. Nat Rev Neurosci 21(10):576–586

Computational Urban Sustainability Platform (CUSP) (n.d.) CUSP—a smart city solution implemented for Cardiff and Luxembourg. https://www.cuspplatform.com/, Accessed 20 Sept 2021

Cortiços ND (2019) Self-learning and self-repairing technologies to establish autonomous building maintenance. In: MATEC Web of conferences, vol 278. EDP Sci, p 04004

Cotrufo N, Zmeureanu R (2016) PCA-based method of soft fault detection and identification for the ongoing commissioning of chillers. Energy Build 130:443–452

Cruz JCD, Amado TM, Hermogino JQ, Andog MLC, Corpuz FT, Ng JRT, Gonazales JCMB , Inacay JND, Redoblado JPD, Manuel MCE (2020) Machine learning-based indoor air quality baseline study of the offices and laboratories of the northwest and southwest building of mapúa university-manila. In: 2020 11th IEEE control and system graduate research colloquium (ICSGRC). IEEE, pp 155–160

Cruz C, Palomar E, Bravo I, Aleixandre M (2021) Behavioural patterns in aggregated demand response developments for communities targeting renewables. Sustain Cities Soc 72:103001

Culaba AB, Del Rosario AJR, Ubando AT, Chang J-S (2020) Machine learning-based energy consumption clustering and forecasting for mixed-use buildings. Int J Energy Res 44(12):9659–9673

Das A, Alonso MJ, Mathisen HM et al (2020) Use of deep learning models to predict indoor air quality in a school case study. In: 16th Conference of the International Society of Indoor Air Quality and Climate:

creative and smart solutions for better built environments, indoor air 2020. International Society of Indoor Air Quality and Climate, pp 894–900
de Assis MV, Carvalho LF, Rodrigues JJ, Lloret J, Proença ML Jr (2020) Near real-time security system applied to SDN environments in IoT networks using convolutional neural network. Comput Electr Eng 86:106738
de Oliveira EM, Oliveira FLC (2018) Forecasting mid-long term electric energy consumption through bagging Arima and exponential smoothing methods. Energy 144:776–788
Dehalwar V, Kalam A, Kolhe ML, Zayegh A (2016) Electricity load forecasting for urban area using weather forecast information. In: 2016 IEEE international conference on power and renewable energy (ICPRE). IEEE, pp 355–359
Delgado JMD, Oyedele L (2021) Digital twins for the built environment: learning from conceptual and process models in manufacturing. Adv Eng Inform 49:101332
Delsing J (2017) Local cloud internet of things automation: technology and business model features of distributed internet of things automation solutions. IEEE Ind Electron Mag 11(4):8–21
Dey M, Rana SP, Dudley S (2020) Smart building creation in large scale HVAC environments through automated fault detection and diagnosis. Futur Gener Comput Syst 108:950–966
Dey M, Rana SP, Dudley S (2020) A case study based approach for remote fault detection using multi-level machine learning in a smart building. Smart Cities 3(2):401–419
Diamantoulakis PD, Kapinas VM, Karagiannidis GK (2015) Big data analytics for dynamic energy management in smart grids. Big Data Res 2(3):94–101
Ding J, Yu CW, Cao S-J (2020) HVAC systems for environmental control to minimize the covid-19 infection. Indoor Built Environ 29(9):1195–1201
Ding X, Du W, Cerpa A (2019) Octopus: deep reinforcement learning for holistic smart building control. In: Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation, pp 326–335
Do H, Cetin KS (2018) Residential building energy consumption: a review of energy data availability, characteristics, and energy performance prediction methods. Curr Sustain Renew Energy Rep 5(1):76–85
Dogruparmak SC, Keskin GA, Yaman S, Alkan A (2014) Using principal component analysis and fuzzy c-means clustering for the assessment of air quality monitoring. Atmos Pollut Res 5(4):656–663
Doorn N (2021) Artificial intelligence in the water domain: opportunities for responsible use. Sci Total Environ 755:142561
Dowling CP, Zhang B (2020) Transfer learning for HVAC system fault detection. In: American control conference (ACC). IEEE 2020:3879–3885
Du B, Zhou Q, Guo J, Guo S, Wang L (2021) Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting. Expert Syst Appl 171:114571
Dun M, Wu L (2020) Forecasting the building energy consumption in china using grey model. Environ Process 7(3):1009–1022
El Motaki S, Yahyaouy A, Gualous H, Sabor J (2021) A new weighted fuzzy c-means clustering for workload monitoring in cloud datacenter platforms. Cluster Comput 24, 3367–3379
Elkhoukhi H, NaitMalek Y, Bakhouya M, Berouine A, Kharbouch A, Lachhab F, Hanifi M, El Ouadghiri D, Essaaidi M (2020) A platform architecture for occupancy detection using stream processing and machine learning approaches. Concurr Comput Pract Exp 32(17):e5651
Ellis F (1963) The control of operating-suite temperatures. Occup Environ Med 20(4):284–287
Elnour M, Meskin N, Al-Naemi M (2020) Sensor data validation and fault diagnosis using auto-associative neural network for HVAC systems. J Build Eng 27:100935
Elnour M, Meskin N, Khan K, Jain R (2021) Application of data-driven attack detection framework for secure operation in smart buildings. Sustain Cities Soc 69:102816
Elnour M, Himeur Y, Fadli F, Mohammedsherif H, Meskin N, Ahmad AM, Petri I, Rezgui Y, Hodorog A (2022) Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities. Appl Energy 318:119153
Elnour M, Fadli F, Himeur Y, Petri I, Rezgui Y, Meskin N, Ahmad AM (2022) Performance and energy optimization of building automation and management systems: towards smart sustainable carbon-neutral sports facilities. Renew Sustain Energy Rev 162:112401
Elnour M, Meskin N (2022) Novel actuator fault diagnosis framework for multizone hvac systems using 2-D convolutional neural networks. IEEE Trans Automat Sci Eng 19(3) 1985-1996
Elsaeidy A, Munasinghe KS, Sharma D, Jamalipour A (2019) Intrusion detection in smart cities using restricted Boltzmann machines. J Netw Comput Appl 135:76–83

El-Sharkawy MF, Javed W (2018) Study of indoor air quality level in various restaurants in Saudi Arabia. Environ Progress Sustain Energy 37(5):1713–1721
energy.gov (2015) An assessment of energy technologies and research opportunities, https://www.energy.gov/sites/prod/files/2017/03/f34/qtr-2015-chapter5.pdf; Accessed 1 Mar 2021
Englund SM (2007) Safety considerations in the chemical process industries. In: Kent and Riegel's handbook of industrial chemistry and biotechnology. Springer, pp 83–146
Estiri H (2014) Building and household x-factors and energy consumption at the residential sector: a structural equation analysis of the effects of household and building characteristics on the annual energy consumption of us residential buildings. Energy Econ 43:178–184
Fadli F, Rezgui Y, Petri I, Meskin N, Ahmad AM, Hodorog A, Elnour M, Mohammedsherif H (2021) Building energy management systems for sports facilities in the gulf region: a focus on impacts and considerations. CIB
Fan C, Xiao F, Li Z, Wang J (2018) Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: a review. Energy Build 159:296–308
Fan C, Liu Y, Liu X, Sun Y, Wang J (2021) A study on semi-supervised learning in enhancing performance of ahu unseen fault detection with limited labeled data. Sustain Cities Soc 70:102874
Fan C, Liu X, Xue P, Wang J (2021) Statistical characterization of semi-supervised neural networks for fault detection and diagnosis of air handling units. Energy Build 234:110733
Fatema N, Malik H (2021) Data-driven occupancy detection hybrid model using particle swarm optimization based artificial neural network. In: Metaheuristic and evolutionary computation: algorithms and applications. Springer, pp 283–297
Fatema N, Malik H, Iqbal A (2020) Big-data analytics based energy analysis and monitoring for multistorey hospital buildings: case study. In: Soft computing in condition monitoring and diagnostics of electrical and mechanical systems. Springer, pp 325–343
Fathi S, Srinivasan R, Fenner A, Fathi S (2020) Machine learning applications in urban building energy performance forecasting: a systematic review. Renew Sustain Energy Rev 133:110287
Fayed N, Abu-Elkheir M, El-Daydamony E, Atwan A (2019) Sensor-based occupancy detection using neutrosophic features fusion. Heliyon 5(9):e02450
Feng C, Mehmani A, Zhang J (2020) Deep learning-based real-time building occupancy detection using ami data. IEEE Trans Smart Grid 11(5):4490–4501
Feng Y, Duan Q, Chen X, Yakkali SS, Wang J (2021) Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods. Appl Energy 291:116814
Ferdoush Z, Mahmud BN, Chakrabarty A, Uddin J A short-term hybrid forecasting model for time series electrical-load data using random forest and bidirectional long short-term memory. Int J Electric Comput Eng (2088-8708) 11(1)
Fernández JMA, Pablo CL (2021) Body temperature and heating temperature in major burns patients care. Enfermería Global 20(1):478–488
Ferrández-Pastor F-J, Mora H, Jimeno-Morenilla A, Volckaert B (2018) Deployment of IoT edge and fog computing technologies to develop smart building services. Sustainability 10(11):3832
Francisco A, Mohammadi N, Taylor JE (2020) Smart city digital twin-enabled energy management: toward real-time urban building energy benchmarking. J Manag Eng 36(2):04019045
Fu F (2020) Fire induced progressive collapse potential assessment of steel framed buildings using machine learning. J Constr Steel Res 166:105918
Gaboalapswe M (2019) Explore and design an artificial intelligent and data analytic software model to address domestic water usage billing crisis in botswana urban areas. Ph.D. thesis, Botho University
Gao G, Li J, Wen Y (2020) Deepcomfort: energy-efficient thermal comfort control in buildings via reinforcement learning. IEEE Internet Things J 7(9):8472–8484
Gao Y, Ruan Y, Fang C, Yin S (2020) Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. Energy Build 223:110156
Gao Y, Fang C, Ruan Y (2019) A novel model for the prediction of long-term building energy demand: Lstm with attention layer. In: IOP conference series: earth and environmental science, vol 294. IOP Publishing, p 012033
Giovanis E (2019) Worthy to lose some money for better air quality: applications of Bayesian networks on the causal effect of income and air pollution on life satisfaction in Switzerland. Empiric Econ 57(5):1579–1611
Gładyszewska-Fiedoruk K, Sulewska MJ (2020) Thermal comfort evaluation using linear discriminant analysis (lda) and artificial neural networks (anns). Energies 13(3):538
Golabi MR, Radmanesh F, Akhoond-Ali AM, Niksokhan MH, Kisi O (2020) Development of an indirect method for modelling the water footprint of electricity using wavelet transform coupled with the random forest model. Hydrol Sci J 65(15):2521–2534

Gong M, Bai Y, Qin J, Wang J, Yang P, Wang S (2020) Gradient boosting machine for predicting return temperature of district heating system: a case study for residential buildings in tianjin. J Build Eng 27:100950

Grillone B, Mor G, Danov S, Cipriano J, Carbonell J, Gabaldón E (2019) Use of generalised additive models to assess energy efficiency savings in buildings using smart metering data. PROCEEDINGS book 27

Grolinger K, L'Heureux A, Capretz MA, Seewald L (2016) Energy forecasting for event venues: big data and prediction accuracy. Energy Build 112:222–233

Guo Y, Tan Z, Chen H, Li G, Wang J, Huang R, Liu J, Ahmad T (2018) Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving. Appl Energy 225:732–745

Hafeez G, Alimgeer KS, Khan I (2020) Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. Appl Energy 269:114915

Haidar N, Tamani N, Nienaber F, Wesseling MT, Bouju A, Ghamri-Doudane Y, (2019) Data collection period and sensor selection method for smart building occupancy prediction. In: IEEE 89th vehicular technology conference (VTC2019-Spring). IEEE, pp. 1–6

Hall S, Cooper WE, Marciani L, McGee JM (2011) Security management for sports and special events: an interagency approach to creating safe facilities. Hum Kinet

Han H, Cui X, Fan Y, Qing H (2019) Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features. Appl Therm Eng 154:540–547

Han H, Zhang Z, Cui X, Meng Q (2020) Ensemble learning with member optimization for fault diagnosis of a building energy system. Energy Build 226:110351

Haq IU, Ullah A, Khan SU, Khan N, Lee MY, Rho S, Baik SW (2021) Sequential learning-based energy consumption prediction model for residential and commercial sectors. Mathematics 9(6):605

Hasanzadeh Nafari R, Ngo T, Mendis P (2016) An assessment of the effectiveness of tree-based models for multi-variate flood damage assessment in Australia. Water 8(7):282

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

Himeur Y, Alsalemi A, Bensaali F, Amira A (2020) A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. Cogn Comput 12(6):1381–1401

Himeur Y, Alsalemi A, Bensaali F, Amira A (2020) Effective non-intrusive load monitoring of buildings based on a novel multi-descriptor fusion with dimensionality reduction. Appl Energy 279:115872

Himeur Y, Alsalemi A, Al-Kababji A, Bensaali F, Amira A (2020) Data fusion strategies for energy efficiency in buildings: overview, challenges and novel orientations. Inform Fus 64:99–120

Himeur Y, Alsalemi A, Bensaali F, Amira A (2020) Building power consumption datasets: survey, taxonomy and future directions. Energy Build 227:110404

Himeur Y, Alsalemi A, Al-Kababji A, Bensaali F, Amira A, Sardianos C, Dimitrakopoulos G, Varlamis I (2021) A survey of recommender systems for energy efficiency in buildings: principles, challenges and prospects. Inform Fus 72:1–21

Himeur Y, Ghanem K, Alsalemi A, Bensaali F, Amira A (2021) Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives. Appl Energy 287:116601

Himeur Y, Alsalemi A, Bensaali F, Amira A (2021) Smart power consumption abnormality detection in buildings using micromoments and improved k-nearest neighbors. Int J Intell Syst 36(6):2865–2894

Himeur Y, Alsalemi A, Bensaali F, Amira A (2021) An intelligent nonintrusive load monitoring scheme based on 2d phase encoding of power signals. Int J Intell Syst 36(1):72–93

Himeur Y, Alsalemi A, Bensaali F, Amira A (2021) Smart non-intrusive appliance identification using a novel local power histogramming descriptor with an improved k-nearest neighbors classifier. Sustain Cities Soc 67:102764

Himeur Y, Alsalemi A, Bensaali F, Amira A, Varlamis I, Bravos G, Sardianos C, Dimitrakopoulos G (2022) Techno-economic assessment of building energy efficiency systems using behavioral change: a case study of an edge-based micro-moments solution. J Clean Prod 331:129786

Himeur Y, Sayed A, Alsalemi A, Bensaali F, Amira A, Varlamis I, Eirinaki M, Sardianos C, Dimitrakopoulos G (2022) Blockchain-based recommender systems: applications, challenges and future opportunities. Comput Sci Rev 43:100439

Himeur Y, Alsalemi A, Bensaali F, Amira A (2021) Appliance identification using a histogram post-processing of 2d local binary patterns for smart grid applications. In: 2020 25th International conference on pattern recognition (ICPR). IEEE, pp 5744–5751

Himeur Y, Alsalemi A, Bensaali F, Amira A (2021) The emergence of hybrid edge-cloud computing for energy efficiency in buildings. In: Proceedings of SAI intelligent systems conference. Springer, pp 70–83

Himeur Y, Alsalemi A, Bensaali F, Amira A (2022) Detection of appliance-level abnormal energy consumption in buildings using autoencoders and micro-moments. In: International conference on big data and internet of things. Springer, pp 179–193

Himeur Y, Alsalemi A, Bensaali F, Amira A, Al-Kababji (2022) A Recent trends of smart nonintrusive load monitoring in buildings: a review, open challenges, and future directions. Int J Intell Syst 37(10): 7124–7179

Himeur Y, Alsalemi A, Bensaali F, Amira A, Varlamis I, Bravos G, Sardianos C, Dimitrakopoulos G Marketability of building energy efficiency systems based on behavioral change: a case study of a novel micro-moments based solution. arXiv:2105.10460

Himeur Y, Elnour M, Fadli F, Meskin N, Petri I, Rezgui Y, Bensaali F, Amira A (2022) Next-generation energy systems for sustainable smart cities: roles of transfer learning. Sustain Cities Soc 104059

Himeur Y, Sohail SS, Bensaali F, Amira A, Alazab M (2022) Latest trends of security and privacy in recommender systems: a comprehensive review and future perspectives. Comput Security 102746

Hu J, Vasilakos AV (2016) Energy big data analytics and security: challenges and opportunities. IEEE Trans Smart Grid 7(5):2423–2436

Hu H, Wang L, Peng L, Zeng Y-R (2020) Effective energy consumption forecasting using enhanced bagged echo state network. Energy 193:116778

Huang Q, Hao K (2020) Development of CNN-based visual recognition air conditioner for smart buildings. J Inf Technol Constr 25:361–373

Huang S, Zuo W, Sohn MD (2018) A Bayesian network model for predicting cooling load of commercial buildings. In: Building simulation, vol 11. Springer, pp 87–101

Huchuk B, Sanner S, O'Brien W (2019) Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. Build Environ 160:106177

Huda AN, Taib S, Jadin MS, Ishak D (2012) A semi-automatic approach for thermographic inspection of electrical installations within buildings. Energy Build 55:585–591

Idowu S, Saguna S, Åhlund C, Schelén O (2016) Applied machine learning: forecasting heat load in district heating system. Energy Build 133:478–488

IoT Security Foundation, Smart cities—the emergence of the CyberSafe building. IoT Security Foundation. https://www.iotsecurityfoundation.org/smart_cities_the_emergence_of_the_cyber_safe_building/. Accessed 28 June 2020

Ippolito M, Riva Sanseverino E, Zizzo G (2014) Impact of building automation control systems and technical building management systems on the energy performance class of residential buildings: an Italian case study. Energy Build 69:33–40

Ishaq M, Kwon S et al (2021) Short-term energy forecasting framework using an ensemble deep learning approach. IEEE Access 9:94262–94271

Janarthanan R, Partheeban P, Somasundaram K, Elamparithi PN (2021) A deep learning approach for prediction of air quality index in a metropolitan city. Sustain Cities Soc 67:102720

Javadzadeh G, Rahmani AM (2020) Fog computing applications in smart cities: a systematic survey. Wireless Netw 26(2):1433–1457

Jenny H, Wang Y, Alonso EG, Minguez R (2020) Using artificial intelligence for smart water management systems. Asian Development Bank. http://hdl.handle.net/11540/12225.

Ji T, Liu L, Wang T, Lin W, Li M, Wu Q (2019) Non-intrusive load monitoring using additive factorial approximate maximum a posteriori based on iterative fuzzy c-means. IEEE Trans Smart Grid 10(6):6667–6677

Jia M, Komeily A, Wang Y, Srinivasan RS (2019) Adopting internet of things for the development of smart buildings: a review of enabling technologies and applications. Autom Constr 101:111–126

Juntarawijit C, Juntarawijit Y (2017) Cooking smoke and respiratory symptoms of restaurant workers in Thailand. BMC Pulm Med 17(1):1–11

Kaklauskas A, Lill I, Amaratunga D, Ubarte I (2019) Model for smart, self-learning and adaptive resilience building. In: 10th Nordic conference on construction economics and organization, Emerald Publishing Limited

Kalajdjieski J, Zdravevski E, Corizzo R, Lameski P, Kalajdziski S, Pires IM, Garcia NM, Trajkovik V (2020) Air pollution prediction with multi-modal data and deep neural networks. Remote Sens 12(24):4142

Kalantzis V, Kollias G, Ubaru S, Nikolakopoulos AN, Horesh L, Clarkson K (2021) Projection techniques to update the truncated svd of evolving matrices with applications. In: International conference on machine learning, PMLR, pp 5236–5246

Kallio J, Tervonen J, Räsänen P, Mäkynen R, Koivusaari J, Peltola J (2021) Forecasting office indoor CO2 concentration using machine learning with a one-year dataset. Build Environ 187:107409
Kang T, Chen P, Quackenbush J, Ding W (2020) A novel deep learning model by stacking conditional restricted boltzmann machine and deep neural network. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp 1316–1324
Kaspersky, Nearly four in ten smart buildings targeted by malicious attacks in H1 2019, https://www.usa.kaspersky.com/about/press-releases/2019_smart-buildings-threat-landscape/, Accessed on 28 June 2020 (2019)
Katipamula S (2019) Building automation: where is it today and where it should be. In: CASE, p 1
Khalil M, McGough S, Pourmirza Z, Pazhoohesh M, Walker S, Transfer learning approach for occupancy prediction in smart buildings. In: (2021) 12th International renewable engineering conference (IREC). IEEE 2021, pp 1–6
Khamma TR, Zhang Y, Guerrier S, Boubekri M (2020) Generalized additive models: an efficient method for short-term energy prediction in office buildings. Energy 213:118834
Khan LU, Yaqoob I, Tran NH, Kazmi SA, Dang TN, Hong CS (2020) Edge-computing-enabled smart cities: a comprehensive survey. IEEE Internet Things J 7(10):10200–10232
Khan ZA, Ullah A, Ullah W, Rho S, Lee M, Baik SW (2020) Electrical energy prediction in residential buildings for short-term horizons using hybrid deep learning strategy. Appl Sci 10(23):8634
Khan AN, Iqbal N, Ahmad R, Kim D-H (2021) Ensemble prediction approach based on learning to statistical model for efficient building energy consumption management. Symmetry 13(3):405
Khattak HA, Farman H, Jan B, Din IU (2019) Toward integrating vehicular clouds with IoT for smart city services. IEEE Network 33(2):65–71
Khwaja A, Naeem M, Anpalagan A, Venetsanopoulos A, Venkatesh B (2015) Improved short-term load forecasting using bagged neural networks. Electric Power Syst Res 125:109–115
Kiliccote S, Piette MA, Hansen D Advanced controls and communications for demand response and energy efficiency in commercial buildings
Kim R, Hong Y, Choi Y, Yoon S (2021) System-level fouling detection of district heating substations using virtual-sensor-assisted building automation system. Energy 227:120515
Koo J, Yoon S, Kim J (2022) Virtual in situ calibration for operational backup virtual sensors in building energy systems. Energies 15(4):1394
Krause B, Lu L, Murray I, Renals S Multiplicative LSTM for sequence modelling. arXiv:1609.07959
Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
Kučera A, Pitner T (2018) Semantic bms: allowing usage of building automation data in facility benchmarking. Adv Eng Inform 35:69–84
Lagesse B, Wang S, Larson TV, Kim AA (2020) Predicting pm2.5 in well-mixed indoor air for a large office building using regression and artificial neural network models. Environ Sci Technol 54(23):15320–15328
Lee S, Karava P (2020) Towards smart buildings with self-tuned indoor thermal environments—a critical review. Energy Build 224:110172
Li M (2020) Optimizing HVAC, systems in buildings with machine learning prediction models: an algorithm based economic analysis. In: 2020 Management science informatization and economic innovation development conference (MSIEID). IEEE, pp 210–217
Li Y, Tong Z (2021) Model predictive control strategy using encoder-decoder recurrent neural networks for smart control of thermal environment. J Build Eng 42:103017
Li S, Wen J (2014) A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. Energy Build 68:63–71
Li W, Logenthiran T, Phan V-T, Woo WL (2018) Implemented iot-based self-learning home management system (shms) for Singapore. IEEE Internet Things J 5(3):2212–2219
Li Z, Friedrich D, Harrison GP (2020) Demand forecasting for a mixed-use building using agent-schedule information with a data-driven model. Energies 13(4):780
Li B, Cheng F, Zhang X, Cui C, Cai W (2021) A novel semi-supervised data-driven method for chiller fault diagnosis with unlabeled data. Appl Energy 285:116459
Li B, Cheng F, Cai H, Zhang X, Cai W (2021) A semi-supervised approach to fault detection and diagnosis for building HVAC systems based on the modified generative adversarial network. Energy Build 246:111044
Liang P, Yang H-D, Chen W-S, Xiao S-Y, Lan Z-Z (2018) Transfer learning for aluminium extrusion electricity consumption anomaly detection via deep neural networks. Int J Comput Integr Manuf 31(4–5):396–405

Lin F, Liu S-J, Chao H-C, Pan J-S (2021) Short-term household load forecasting model based on variational mode decomposition and gated recurrent unit with attention mechanism. J Netw Intell 6(1):143–153
Li L, Ota K, Dong M (2017) Everything is image: Cnn-based short-term electrical load forecasting for smart grid. In: 2017 14th International symposium on pervasive systems, algorithms and networks & 2017 11th International conference on frontier of computer science and technology & 2017 Third international symposium of creative computing (ISPAN-FCST-ISCC). IEEE, pp 344–351
Lissa P, Deane C, Schukat M, Seri F, Keane M, Barrett E (2021) Deep reinforcement learning for home energy management system control. Energy AI 3:100043
Liu G, Yang J, Hao Y, Zhang Y (2018) Big data-informed energy efficiency assessment of china industry sectors based on k-means clustering. J Clean Prod 183:304–314
Liu N, Liu X, Jayaratne R, Morawska L (2020) A study on extending the use of air quality monitor data via deep learning techniques. J Clean Prod 274:122956
Liu J, Wang Y, Zhang Y (2020) A novel isomap-SVR soft sensor model and its application in rotary kiln calcination zone temperature prediction. Symmetry 12(1):167
Liu J, Zhang Q, Li X, Li G, Liu Z, Xie Y, Li K, Liu B (2021) Transfer learning-based strategies for fault diagnosis in building energy systems. Energy Build 250:111256
Liu Z, Chi Z, Osmani M, Demian P (2021) Blockchain and building information management (bim) for sustainable building development within the context of smart cities. Sustainability 13(4):2090
Li A, Xiao F, Fan C, Hu M (2021) Development of an ann-based building energy model for information-poor buildings using transfer learning. In: Building simulation, vol 14. Springer, pp 89–101
Lobdell KW, Stamou S, Sanchez JA (2012) Hospital-acquired infections. Surg Clin 92(1):65–77
Lopes MADS, Neto ADD, Martins ADM (2020) Parallel t-sne applied to data visualization in smart cities. IEEE Access 8:11482–11490
Loy-Benitez J, Heo S, Yoo C (2020) Imputing missing indoor air quality data via variational convolutional autoencoders: implications for ventilation management of subway metro systems. Build Environ 182:107135
Loy-Benitez J, Heo S, Yoo C (2020) Soft sensor validation for monitoring and resilient control of sequential subway indoor air quality through memory-gated recurrent neural networks-based autoencoders. Control Eng Pract 97:104330
Loy-Benitez J, Li Q, Nam K, Yoo C (2020) Sustainable subway indoor air quality monitoring and fault-tolerant ventilation control using a sparse autoencoder-driven sensor self-validation. Sustain Cities Soc 52:101847
Lu TT (2009) Fundamental limitations of semi-supervised learning. Master's thesis, University of Waterloo
Lu H, Cheng F, Ma X, Hu G (2020) Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: a case study of an intake tower. Energy 203:117756
Lv Z, Qiao L, Kumar Singh A, Wang Q (2021) Ai-empowered IoT security for smart cities. ACM Trans Internet Technol 21(4):1–21
Ma J, Cheng JC, Lin C, Tan Y, Zhang J (2019) Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. Atmos Environ 214:116885
Maatoug A, Belalem G, Mahmoudi S (2019) Fog computing framework for location-based energy management in smart buildings. Multiagent Grid Syst 15(1):39–56
Mad Saad S, Andrew AM, Shakaff AY, Mat Dzahir MA, Hussein M, Mohamad M, Ahmad ZA (2017) Pollutant recognition based on supervised machine learning for indoor air quality monitoring systems. Appl Sci 7(8):823
Mahmud MS, Huang JZ, Salloum S, Emara TZ, Sadatdiynov K (2020) A survey of data partitioning and sampling methods to support big data analysis. Big Data Min Anal 3(2):85–101. https://doi.org/10.26599/BDMA.2019.9020015
Mansor AA, Shamsul S, Abdullah S, Dom N, Napi NM, Ahmed A, Ismail M (2021) Identification of indoor air quality (iaq) sources in libraries through principal component analysis (PCA). In: IOP conference series: materials science and engineering, vol 1144. IOP Publishing, p 012055
Mariano-Hernández D, Hernández-Callejo L, Zorita-Lamadrid A, Duque-Pérez O, García FS (2021) A review of strategies for building energy management system: model predictive control, demand side management, optimization, and fault detect & diagnosis. J Build Eng 33:101692
Marino DL, Amarasinghe K, Manic M (2016) Building energy load forecasting using deep neural networks. In: IECON 2016-42nd annual conference of the IEEE Industrial Electronics Society. IEEE, pp 7046–7051
Markoska E, Lazarova-Molnar S (2018) Towards smart buildings performance testing as a service. In: 2018 Third international conference on fog and mobile edge computing (FMEC). IEEE, pp 277–282

Mawson VJ, Hughes BR (2020) Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector. Energy Build 217:109966
Merz H, Hansemann T, Hübner C (2018) Building automation: communication systems with EIB/KNX, Lon and BACnet, 2nd edn. Springer, Cham
Miori V, Russo D, Ferrucci L (2019) Interoperability of home automation systems as a critical challenge for IoT. In: 2019 4th International conference on computing, communications and security (ICCCS). IEEE, pp 1–7
Mo H, Sun H, Liu J, Wei S (2019) Developing window behavior models for residential buildings using xgboost algorithm. Energy Build 205:109564
Mohamed N, Al-Jaroodi J, Lazarova-Molnar S (2018) Energy cloud: services for smart buildings. In: Sustainable cloud and energy services. Springer, pp 117–134
Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. PLoS Med 6(7):e1000097
Molinara M, Ferdinandi M, Cerro G, Ferrigno L, Massera E (2020) An end to end indoor air monitoring system based on machine learning and sensiplus platform. IEEE Access 8:72204–72215
Molina-Solana M, Ros M, Ruiz MD, Gómez-Romero J, Martín-Bautista MJ (2017) Data science for building energy management: a review. Renew Sustain Energy Rev 70:598–609
Moon J, Park J, Hwang E, Jun S (2018) Forecasting power consumption for higher educational institutions based on machine learning. J Supercomput 74(8):3778–3800
Moon J, Park S, Rho S, Hwang E (2019) A comparative analysis of artificial neural network architectures for building energy consumption forecasting. Int J Distrib Sens Netw 15(9):1550147719877616
Moradzadeh A, Mansour-Saatloo A, Mohammadi-Ivatloo B, Anvari-Moghaddam A (2020) Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings. Appl Sci 10(11):3829
Movahedi A, Derrible S Interrelationships between electricity, gas, and water consumption in large-scale buildings. J Ind Ecol
Mtibaa F, Nguyen K-K, Dermardiros V, Cheriet M (2021) Context-aware model predictive control framework for multi-zone buildings. J Build Eng 42:102340
Muhammad S, Sapri M, Sipan I (2014) Academic buildings and their influence on students' wellbeing in higher education institutions. Soc Indic Res 115(3):1159–1178
Muiruri D et al (2021) Modelling indoor air quality using sensor data and machine learning methods
Mukherjee P, Barik R, Pradhan C (2021) echain: Leveraging toward blockchain technology for smart energy utilization. In: Applications of advanced computing in systems. Springer, pp 73–81
Mumtaz R, Zaidi SMH, Shakir MZ, Shafi U, Malik MM, Haque A, Mumtaz S, Zaidi SAR (2021) Internet of things (IoT) based indoor air quality sensing and predictive analytic-a covid-19 perspective. Electronics 10(2):184
Mundt T, Dähn A, Glock H-W (2014) Forensic analysis of home automation systems. In: 7th Workshop on hot topics in privacy enhancing technologies (HotPETs 2014)
Muntean M, Dănăiaţă D, Hurbean L, Jude C (2021) A business intelligence & analytics framework for clean and affordable energy data analysis. Sustainability 13(2):638
Mutis I, Ambekar A, Joshi V (2020) Real-time space occupancy sensing and human motion analysis using deep learning for indoor air quality control. Autom Constr 116:103237
Nasser AA, Rashad MZ, Hussein SE (2020) A two-layer water demand prediction system in urban areas based on micro-services and LSTM neural networks. IEEE Access 8:147647–147661
Nawari NO, Ravindran S (2019) Blockchain and the built environment: potentials and limitations. J Build Eng 25:100832
Nejat P, Hussen HM, Fadli F, Chaudhry HN, Calautit J, Jomehzadeh F (2020) Indoor environmental quality (ieq) analysis of a two-sided windcatcher integrated with anti-short-circuit device for low wind conditions. Processes 8(7):840
Nguyen VK, Zhang WE, Mahmood A (2021) Semi-supervised intrusive appliance load monitoring in smart energy monitoring system. ACM Trans Multimed Comput Commun Appl (TOMM) 17(3):1–20
Nie P, Roccotelli M, Fanti MP, Ming Z, Li Z (2021) Prediction of home energy consumption based on gradient boosting regression tree. Energy Rep 7:1246–1255
O'Grady T, Chong H-Y, Morrison GM (2021) A systematic review and meta-analysis of building automation systems. Build Environ 195:107770
Ouache R, Nahiduzzaman KM, Hewage K, Sadiq R (2021) Performance investigation of fire protection and intervention strategies: artificial neural network-based assessment framework. J Build Eng 42:102439

Ozturk GB (2020) Interoperability in building information modeling for aeco/fm industry. Autom Constr 113:103122
Pal N, Ghosh P, Karsai G (2019) DeepECO: applying deep learning for occupancy detection from energy consumption data. In: 18th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 1938–1943
Park JY, Yang X, Miller C, Arjunan P, Nagy Z (2019) Apples or oranges? Iidentification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset. Appl Energy 236:1280–1295
Park S, Moon J, Jung S, Rho S, Baik SW, Hwang E (2020) A two-stage industrial load forecasting scheme for day-ahead combined cooling, heating and power scheduling. Energies 13(2):443
Park S, Jung S, Jung S, Rho S, Hwang E (2021) Sliding window-based lightgbm model for electric load forecasting using anomaly repair. J Supercomput 1–22
Pathak N, Ba A, Ploennigs J, Roy N (2018) Forecasting gas usage for big buildings using generalized additive models and deep learning. In: 2018 IEEE international conference on smart computing (SMARTCOMP). IEEE, pp 203–210
Pešić S, Tošić M, Iković O, Radovanović M, Ivanović M, Bošković D (2019) BLEMAT: data analytics and machine learning for smart building occupancy detection and prediction. Int J Artif Intell Tools 28(06):1960005
Petri I, Rana O, Rezgui Y, Fadli F Edge HVAC analytics. Energies 14(17)
Pham A-D, Ngo N-T, Truong TTH, Huynh N-T, Truong N-S (2020) Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. J Clean Prod 260:121082
Pinto T, Praça I, Vale Z, Silva J (2021) Ensemble learning for electricity consumption forecasting in office buildings. Neurocomputing 423:747–755
Pinto G, Wang Z, Roy A, Hong T, Capozzoli A (2022) Transfer learning for smart buildings: a critical review of algorithms, applications, and future perspectives. Adv Appl Energy 100084
Plageras AP, Psannis KE, Stergiou C, Wang H, Gupta BB (2018) Efficient IoT-based sensor big data collection-processing and analysis in smart buildings. Futur Gener Comput Syst 82:349–357
Png E, Srinivasan S, Bekiroglu K, Chaoyang J, Su R, Poolla K (2019) An internet of things upgrade for smart and scalable heating, ventilation and air-conditioning control in commercial buildings. Appl Energy 239:408–424
Pratt BW, Erickson JD (2020) Defeat the peak: behavioral insights for electricity demand response program design. Energy Res Soc Sci 61:101352
Pulimeno M, Piscitelli P, Colazzo S, Colao A, Miani A (2020) Indoor air quality at school and students' performance: Recommendations of the unesco chair on health education and sustainable development & the Italian society of environmental medicine (sima). Health Promot Perspect 10(3):169
Qin J, Zhang J (2017) Sampling for building energy consumption with fuzzy theory. Energy Build 156:78–84
Quinn C, Shabestari AZ, Misic T, Gilani S, Litoiu M, McArthur J (2020) Building automation system-bim integration using a linked data structure. Autom Constr 118:103257
Rahim MS, Nguyen KA, Stewart RA, Giurco D, Blumenstein M (2020) Machine learning and data analytic techniques in digital water metering: a review. Water 12(1):294
Ray PP, Dash D, De D (2019) Edge computing for internet of things: a survey, e-healthcare case study and future direction. J Netw Comput Appl 140:1–22
Razavi R, Gharipour A, Fleury M, Akpan IJ (2019) Occupancy detection of residential buildings using smart meter data: a large-scale study. Energy Build 183:195–208
Rehman SU, Javed AR, Khan MU, Nazar Awan M, Farukh A, Hussien A (2020) Personalisedcomfort: a personalised thermal comfort model to predict thermal sensation votes for smart building residents. Enterprise Inform Syst 1–23
Ribeiro M, Grolinger K, ElYamany HF, Higashino WA, Capretz MA (2018) Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. Energy Build 165:352–363
Roccetti M, Delnevo G, Casini L, Cappiello G (2019) Is bigger always better? a controversial journey to the center of machine learning design, with uses and misuses of big data for predicting water meter failures. J Big Data 6(1):1–23
Rocha Filho GP, Mano LY, Valejo ADB, Villas LA, Ueyama J (2018) A low-cost smart home automation to enhance decision-making based on fog computing and computational intelligence. IEEE Lat Am Trans 16(1):186–191
Roelofsen P The impact of office environments on employee performance: The design of the workplace as a strategy for productivity enhancement. J Facilit Manag

Roger Rozario A et al (2021) Forecasting-mining prediction of water consumption for residential sectors. Ann Roman Soc Cell Biol 25(6):2918–2924
Runge J, Zmeureanu R A review of deep learning techniques for forecasting energy use in buildings. Energies 14(3)
Saini J, Dutta M, Marques G (2020) Indoor air quality monitoring systems based on internet of things: a systematic review. Int J Environ Res Public Health 17(14):4942
Saini J, Dutta M, Marques G (2020) A comprehensive review on indoor air quality monitoring systems for enhanced public health. Sustain Environ Res 30(1):1–12
Sajjad M, Khan ZA, Ullah A, Hussain T, Ullah W, Lee MY, Baik SW (2020) A novel cnn-gru-based hybrid approach for short-term residential load forecasting. IEEE Access 8:143759–143768
Salerno VM, Rabbeni G (2018) An extreme learning machine approach to effective energy disaggregation. Electronics 7(10):235
Salonen H, Lahtinen M, Lappalainen S, Nevala N, Knibbs LD, Morawska L, Reijula K (2013) Physical characteristics of the indoor environment that affect health and wellbeing in healthcare facilities: a review. Intell Build Int 5(1):3–25
Sardianos C, Varlamis I, Dimitrakopoulos G, Anagnostopoulos D, Alsalemi A, Bensaali F, Himeur Y, Amira A (2020) Rehab-c: recommendations for energy habits change. Futur Gener Comput Syst 112:394–407
Sardianos C, Varlamis I, Chronis C, Dimitrakopoulos G, Alsalemi A, Himeur Y, Bensaali F, Amira A (2021) The emergence of explainability of intelligent systems: delivering explainable and personalized recommendations for energy efficiency. Int J Intell Syst 36(2):656–680
Sardianos C, Chronis C, Varlamis I, Dimitrakopoulos G, Himeur Y, Alsalemi A, Bensaali F, Amira A (2020) Real-time personalised energy saving recommendations. In International conferences on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData) and IEEE congress on cybermatics (Cybermatics). IEEE, pp 366–371
Sardianos C, Varlamis I, Chronis C, Dimitrakopoulos G, Himeur Y, Alsalemi A, Bensaali F, Amira A (2020) A model for predicting room occupancy based on motion sensor data. In: 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT). IEEE, pp 394–399
Sayed AN, Himeur Y, Bensaali F (2022) Deep and transfer learning for building occupancy detection: a review and comparative analysis. Eng Appl Artif Intell 115:105254
Sayed A, Alsalemi A, Himeur Y, Bensaali F, Amira A (2022) Endorsing energy efficiency through accurate appliance-level power monitoring, automation and data visualization. In: Networking, intelligent systems and security. Springer, pp 603–617
Sayed A, Himeur Y, Alsalemi A, Bensaali F, Amira A Intelligent edge-based recommender system for internet of energy applications. IEEE Syst J
Schachinger D, Fernbach A, Kastner W (2017) Modeling framework for IoT integration of building automation systems. At-Automatisierungstechnik 65(9):630–640
Serale G, Fiorentini M, Capozzoli A, Bernardini D, Bemporad A (2018) Model predictive control (mpc) for enhancing building and hvac system energy efficiency: problem formulation, applications and opportunities. Energies 11(3):631
Seyedzadeh S, Rahimian FP, Glesk I, Roper M (2018) Machine learning for estimation of building energy consumption and performance: a review. Visual Eng 6(1):1–20
Seyedzadeh S, Pour Rahimian F, Rastogi P, Glesk I (2019) Tuning machine learning models for prediction of building energy loads. Sustain Cities Soc 47:101484
Sha H, Xu P, Hu C, Li Z, Chen Y, Chen Z (2019) A simplified HVAC energy prediction method based on degree-day. Sustain Cities Soc 51:101698
Shahnazari H, Mhaskar P, House JM, Salsbury TI (2019) Modeling and fault diagnosis design for HVAC systems using recurrent neural networks. Comput Chem Eng 126:189–203
Shahzad M, Shafiq MT, Douglas D, Kassem M (2022) Digital twins in built environments: an investigation of the characteristics, applications, and challenges. Buildings 12(2):120
Sharma PK, Mondal A, Jaiswal S, Saha M, Nandi S, De T, Saha S (2021) Indoairsense: a framework for indoor air quality estimation and forecasting. Atmos Pollut Res 12(1):10–22
Sharma A, Sabitha AS, Bansal A (2018) Edge analytics for building automation systems: a review. In: 2018 International conference on advances in computing, communication control and networking (ICACCCN). IEEE, pp 585–590
Shine P, Murphy MD, Upton J, Scully T (2018) Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms. Comput Electron Agric 150:74–87
Sideratos G, Ikonomopoulos A, Hatziargyriou ND (2020) A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks. Electric Power Syst Res 178:106025

Singh S, Yassine A (2018) Big data mining of energy time series for behavioral analytics and energy consumption forecasting. Energies 11(2):452
Siountri K, Skondras E, Vergados DD (2020) Developing smart buildings using blockchain, internet of things, and building information modeling. Int J Interdiscipli Telecommun Netw (IJITN) 12(3):1–15
Skomski E, Lee J-Y, Kim W, Chandan V, Katipamula S, Hutchinson B (2020) Sequence-to-sequence neural networks for short-term electrical load forecasting in commercial office buildings. Energy Build 226:110350
Smolak K, Kasieczka B, Fialkiewicz W, Rohm W, Siła-Nowicka K, Kopańczyk K (2020) Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. Urban Water J 17(1):32–42
Somontina JAB, Garcia FCC, Macabebe EQB (2018) Water consumption monitoring with fixture recognition using random forest. In: TENCON 2018-2018 IEEE region 10 conference. IEEE, pp 0663–0667
Somu N, Gauthama Raman MR, Ramamritham K (2020) A hybrid model for building energy consumption forecasting using long short term memory networks. Appl Energy 261:114131
Somu N, Gauthama Raman MR, Ramamritham K (2021) A deep learning framework for building energy consumption forecast. Renew Sustain Energy Rev 137:110591
Stamatescu G, Stamatescu I, Arghira N, Făgărăşan I (2020) Cybersecurity perspectives for smart building automation systems. In: 2020 12th International conference on electronics, computers and artificial intelligence (ECAI). IEEE, pp 1–5
Stergiou C, Psannis KE, Gupta BB, Ishibashi Y (2018) Security, privacy & efficiency of sustainable cloud computing for big data & IoT. Sustain Comput Inform Syst 19:174–184
Su B, Wang S (2020) An agent-based distributed real-time optimal control strategy for building HVAC systems for applications in the context of future IoT-based smart sensor networks. Appl Energy 274:115322
Sultan MM, Islam MS, Rahman MA (2017) Smart fire detection system with early notifications using machine learning. Int J Comput Intell Appl 16(02):1750009
Sun AY, Scanlon BR (2019) How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. Environ Res Lett 14(7):073001
Sun F, Yu J (2021) Improved energy performance evaluating and ranking approach for office buildings using simple-normalization, entropy-based topsis and k-means method. Energy Rep 7:1560–1570
Sun C, Zhai Z (2020) The efficacy of social distance and ventilation effectiveness in preventing covid-19 transmission. Sustain Cities Soc 62:102390
Sun Y, Haghighat F, Fung BC (2020) A review of the-state-of-the-art in data-driven approaches for building energy prediction. Energy Build 221:110022
Sun L, Wei Q, He L, Yin Z (2020) The prediction of building heating and ventilation energy consumption base on adaboost-bp algorithm. In: IOP Conference series: materials science and engineering, vol 782. IOP Publishing, p 032008
Surya L (2017) Risk analysis model that uses machine learning to predict the likelihood of a fire occurring at a given property. Int J Creat Res Thoughts (IJCRT), 2320–2882
Swiercz M, Mroczkowska H (2019) Application of PCA for early leak detection in a pipeline system of a steam boiler. Prz. Elektrotechniczny Electr. Rev 95:190–203
Syed D, Abu-Rub H, Ghrayeb A, Refaat SS (2021) Household-level energy forecasting in smart buildings using a novel hybrid deep learning model. IEEE Access 9:33498–33511
Tagliabue LC, Cecconi FR, Rinaldi S, Ciribini ALC (2021) Data driven indoor air quality prediction in educational facilities based on IoT network. Energy Build 236:110782
Taheri S, Ahmadi A, Mohammadi-Ivatloo B, Asadi S (2021) Fault detection diagnostic for HVAC systems via deep learning algorithms. Energy Build 250:111275
Taheri S, Razban A (2021) Learning-based CO2 concentration prediction: application to indoor air quality control using demand-controlled ventilation. Build Environ 108164
Tanabe S-I, Iwahashi Y, Tsushima S, Nishihara N (2013) Thermal comfort and productivity in offices under mandatory electricity savings after the Great East Japan earthquake. Archit Sci Rev 56(1):4–13
Tang S, Shelden DR, Eastman CM, Pishdad-Bozorgi P, Gao X (2020) Bim assisted building automation system information exchange using bacnet and ifc. Autom Constr 110:103049
Tariq S, Loy-Benitez J, Nam K, Lee G, Kim M, Park D, Yoo C (2021) Transfer learning driven sequential forecasting and ventilation control of pm2.5 associated health risk levels in underground public facilities. J Hazard Mater 406:124753
Taştan M, Gökozan H (2019) Real-time monitoring of indoor air quality with internet of things-based e-nose. Appl Sci 9(16):3435
Tian Z, Si B, Shi X, Fang Z (2019) An application of Bayesian network approach for selecting energy efficient HVAC systems. J Build Eng 25:100796

Tian Y, Yu J, Zhao A (2020) Predictive model of energy consumption for office building by using improved GWO-BP. Energy Rep 6:620–627
Tien PW, Wei S, Calautit JK, 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 Build 226:110386
Tien PW, Wei S, Calautit J (2021) A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand. Energies 14(1):156
Tiwari A, Batra U Blockchain enabled reparations in smart buildings-cyber physical system. Defence Sci J 71(4)
Tokarski M Protection of individuals in the light of eu regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Safety Defense 2
Trianti-Stourna E, Spyropoulou K, Theofylaktos C, Droutsa K, Balaras C, Santamouris M, Asimakopoulos D, Lazaropoulou G, Papanikolaou N (1998) Energy conservation strategies for sports centers: part A. Sports halls. Energy Build 27(2):109–122
Valenzuela VEL, Lucena VF, Parvaresh P, Jazdi N, Göhner P (2013) Voice-activated system to remotely control industrial and building automation systems using cloud computing. In: 2013 IEEE 18th conference on emerging technologies & factory automation (ETFA). IEEE, pp 1–4
Valgaev O, Kupzog F (2017) Schmeck H (2017) Building power demand forecasting using k-nearest neighbours model-practical application in smart city demo aspern project. CIRED-Open Access Proc J 1:1601–1604
Valladares W, Galindo M, Gutiérrez J, Wu W-C, Liao K-K, Liao J-C, Lu K-C, Wang C-C (2019) Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm. Build Environ 155:105–117
Van Cutsem O, Dac DH, Boudou P, Kayal M (2020) Cooperative energy management of a community of smart-buildings: a blockchain approach. Int J Electric Power Energy Syst 117:105643
Van Engelen JE, Hoos HH (2020) A survey on semi-supervised learning. Mach Learn 109(2):373–440
Varlamis I, Sardianos C, Chronis C, Dimitrakopoulos G, Himeur Y, Alsalemi A, Bensaali F, Amira A (2022) Smart fusion of sensor data and human feedback for personalized energy-saving recommendations. Appl Energy 305:117775
Varlamis I, Sardianos C, Chronis C, Dimitrakopoulos G, Himeur Y, Alsalemi A, Bensaali F, Amira A (2022) Using big data and federated learning for generating energy efficiency recommendations. Int J Data Sci Anal, pp 1–17
Verma S, Singh S, Majumdar A (2019) Multi label restricted boltzmann machine for non-intrusive load monitoring. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 8345–8349
Wang S (2020) Wireless network indoor positioning method using nonmetric multidimensional scaling and rssi in the internet of things environment. Math Problems Eng 2020 8830891
Wang S (2009) Intelligent buildings and building automation. Routledge, London
Wang Z, Hong T (2020) Reinforcement learning for building controls: the opportunities and challenges. Appl Energy 269:115036
Wang Y, Velswamy K, Huang B (2017) A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems. Processes 5(3):46
Wang J, Li G, Chen H, Liu J, Guo Y, Sun S, Hu Y (2018) Energy consumption prediction for water-source heat pump system using pattern recognition-based algorithms. Appl Therm Eng 136:755–766
Wang Y, Chen J, Chen X, Zeng X, Kong Y, Sun S, Guo Y, Liu Y (2020) Short-term load forecasting for industrial customers based on tcn-lightgbm. IEEE Trans Power Syst 36(3):1984–1997
Wang R, Lu S, Feng W (2020) A novel improved model for building energy consumption prediction based on model integration. Appl Energy 262:114561
Wang JQ, Du Y, Wang J (2020) LSTM based long-term energy consumption prediction with periodicity. Energy 197:117197
Wang J, Chen Y (2021) Adaboost-based integration framework coupled two-stage feature extraction with deep learning for multivariate exchange rate prediction. Neural Process Lett 1–25
Wang Z, Liu J, Zhang Y, Yuan H, Zhang R, Srinivasan RS (2021) Practical issues in implementing machine-learning models for building energy efficiency: moving beyond obstacles. Renew Sustain Energy Rev 143:110929

Wan B, Xu C, Mahapatra RP, Selvaraj P (2021) Understanding the cyber-physical system in international stadiums for security in the network from cyber-attacks and adversaries using AI. Wireless Personal Communications, pp 1–18
Wei W, Ramalho O, Malingre L, Sivanantham S, Little JC, Mandin C (2019) Machine learning and statistical models for predicting indoor air quality. Indoor Air 29(5):704–726
Wen L, Zhou K, Yang S (2020) Load demand forecasting of residential buildings using a deep learning model. Electric Power Syst Res 179:106073
Wilcox HS (2020) Virtual metering for monitoring building energy consumption. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM
Wu K, Wu J, Feng L, Yang B, Liang R, Yang S, Zhao R (2021) An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system. Int Trans Electric Energy Syst 31(1)
Wu L, Kong C, Hao X, Chen W (2020) A short-term load forecasting method based on gru-cnn hybrid neural network model. Mathematical problems in engineering
Wu L, Wang Y Stationary and moving occupancy detection using the sleepir sensor module and machine learning, IEEE Sens J
Wytock M, Kolter Z (2013) Sparse gaussian conditional random fields: algorithms, theory, and application to energy forecasting. In: International conference on machine learning, PMLR, pp 1265–1273
Xiao Z, Gang W, Yuan J, Zhang Y, Fan C (2021) Cooling load disaggregation using a nilm method based on random forest for smart buildings. Sustain Cities Soc 74:103202
Xiao-wei X (2020) Study on the intelligent system of sports culture centers by combining machine learning with big data. Pers Ubiquit Comput 24(1):151–163
Xu C, Chen H (2020) A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data. Energy Build 215:109864
Xu C, Wang J, Zhang J, Li X (2021) Anomaly detection of power consumption in yarn spinning using transfer learning. Comput Ind Eng 152:107015
Yahyaoui A, Yaakoubi F, Abdellatif T et al (2020) Machine learning based rank attack detection for smart hospital infrastructure. In: International conference on smart homes and health telematics. Springer, pp 28–40
Yaïci W, Krishnamurthy K, Entchev E, Longo M (2021) Recent advances in internet of things (IoT) infrastructures for building energy systems: A review. Sensors 21(6):2152
Yang S, Wan MP (2022) Machine-learning-based model predictive control with instantaneous linearization—a case study on an air-conditioning and mechanical ventilation system. Appl Energy 306:118041
Yang S, Wan MP, Chen W, Ng BF, Dubey S (2020) Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization. Appl Energy 271:115147
Yang S, Wan MP, Chen W, Ng BF, Dubey S (2021) Experiment study of machine-learning-based approximate model predictive control for energy-efficient building control. Appl Energy 288:116648
Yoon S, Yu Y (2018) Strategies for virtual in-situ sensor calibration in building energy systems. Energy Build 172:22–34
Yu Y (2011) Ai chiller: an open IoT cloud based machine learning framework for the energy saving of building HVAC system via big data analytics on the fusion of bms and environmental data. arXiv: 2011.01047
Yu Y, Li H (2015) Virtual in-situ calibration method in building systems. Autom Constr 59:59–67
Yu Z, Haghighat F, Fung BC, Yoshino H (2010) A decision tree method for building energy demand modeling. Energy Build 42(10):1637–1646
Yu J, Jang J, Yoo J, Park JH, Kim S (2017) Bagged auto-associative kernel regression-based fault detection and identification approach for steam boilers in thermal power plants. J Electr Eng Technol 12(4):1406–1416
Yuan Z, Wang W, Wang H, Mizzi S (2020) Combination of cuckoo search and wavelet neural network for midterm building energy forecast. Energy 202:117728
Yucong W, Bo W (2020) Research on ea-xgboost hybrid model for building energy prediction. J Phys 1518:012082
Yun W-S, Hong W-H, Seo H (2021) A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systems considering undefined states. J Build Eng 35:102111
Yu L, Qin S, Zhang M, Shen C, Jiang T, Guan X. A review of deep reinforcement learning for smart building energy management. IEEE Internet Things J 8(1)5 12046–12063

Zakharchenko A, Stepanets O (2019) Edge computing in building automation system-pros and cons. In: Modeling, control and information technologies: proceedings of international scientific and practical conference, pp 130–132
Zekić-Sušac M, Has A, Knežević M (2021) Predicting energy cost of public buildings by artificial neural networks, cart, and random forest. Neurocomputing 439:223–233
Zhan S, Liu Z, Chong A, Yan D (2020) Building categorization revisited: a clustering-based approach to using smart meter data for building energy benchmarking. Appl Energy 269:114920
Zhang Q, Fu F, Tian R (2020) A deep learning and image-based model for air quality estimation. Sci Total Environ 724:138178
Zhang C, Li J, Zhao Y, Li T, Chen Q, Zhang X (2020) A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process. Energy Build 225:110301
Zhang G, Tian C, Li C, Zhang JJ, Zuo W (2020) Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature. Energy 201:117531
Zhang G, Li Y, Deng X (2020) K-means clustering-based electrical equipment identification for smart building application. Information 11(1):27
Zhang L, Leach M, Bae Y, Cui B, Bhattacharya S, Lee S, Im P, Adetola V, Vrabie D, Kuruganti T (2021) Sensor impact evaluation and verification for fault detection and diagnostics in building energy systems: a review. Adv Appl Energy 3(1):100055
Zhang Y, Geng P, Sivaparthipan C, Muthu BA (2021) Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. Sustain Energy Technol Assess 45:100986
Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu Y, Livingood W (2021) A review of machine learning in building load prediction. Appl Energy 285:116452
Zhang X, Zeng Z, Wang P, Song J, Kong Z (2020) A hybrid edge-cloud computing method for short-term electric load forecasting based on smart metering terminal. In: 2020 IEEE 4th conference on energy internet and energy system integration (EI2). IEEE, pp 3101–3105
Zhao H, Hua Q, Chen H-B, Ye Y, Wang H, Tan SX-D, Tlelo-Cuautle E (2018) Thermal-sensor-based occupancy detection for smart buildings using machine-learning methods. ACM Trans Design Autom Electr Syst (TODAES) 23(4):1–21
Zhao Y, Zhang C, Zhang Y, Wang Z, Li J (2020) A review of data mining technologies in building energy systems: load prediction, pattern identification, fault detection and diagnosis. Energy Built Environ 1(2):149–164
Zheng H, Yuan J, Chen L (2017) Short-term load forecasting using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation. Energies 10(8):1168
Zheng P, Wang C, Liu Y, Lin B, Wu H, Huang Y, Zhou X (2022) Thermal adaptive behavior and thermal comfort for occupants in multi-person offices with air-conditioning systems. Build Environ 207:108432
Zhong H, Wang J, Jia H, Mu Y, Lv S (2019) Vector field-based support vector regression for building energy consumption prediction. Appl Energy 242:403–414
Zhong F, Calautit JK, Hughes BR (2020) Analysis of the influence of cooling jets on the wind and thermal environment in football stadiums in hot climates. Build Serv Eng Res Technol 41(5):561–585
Zhou K, Yang S (2016) Understanding household energy consumption behavior: the contribution of energy big data analytics. Renew Sustain Energy Rev 56:810–819
Zhu X, Chen K, Anduv B, Jin X, Du Z (2021) Transfer learning based methodology for migration and application of fault detection and diagnosis between building chillers for improving energy efficiency. Build Environ 200:107957
Zou J, Zhao Q, Yang W, Wang F (2017) Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation. Energy Build 152:385–398
Zubaidi SL, Gharghan SK, Dooley J, Alkhaddar RM, Abdellatif M (2018) Short-term urban water demand prediction considering weather factors. Water Resour Manage 32(14):4527–4542
Zverovich V, Mahdjoubi L, Boguslawski P, Fadli F, Barki H (2016) Emergency response in complex buildings: automated selection of safest and balanced routes. Comput Aided Civil Infrastr Eng 31(8):617–632

Authors and Affiliations

Yassine Himeur[1,2] · Mariam Elnour[1] · Fodil Fadli[1] · Nader Meskin[3] · Ioan Petri[4] · Yacine Rezgui[4] · Faycal Bensaali[3] · Abbes Amira[5,6]

Mariam Elnour
me1003659@qu.edu.qa

Fodil Fadli
f.fadli@qu.edu.qa

Nader Meskin
nader.meskin@qu.edu.qa

Ioan Petri
PetriI@cardiff.ac.uk

Yacine Rezgui
RezguiY@cardiff.ac.uk

Faycal Bensaali
f.bensaali@qu.edu.qa

Abbes Amira
aamira@sharjah.ac.ae; abbes.amira@dmu.ac.uk

1 Department of Architecture & Urban Planning, Qatar University, Doha, Qatar

2 College of Engineering and Information Technology, University of Dubai, Dubai, UAE

3 Department of Electrical Engineering, Qatar University, Doha, Qatar

4 School of Engineering, BRE Institute of Sustainable Engineering, Cardiff University, Wales, UK

5 Department of Computer Science, University of Sharjah, Sharjah, UAE

6 Institute of Artificial Intelligence, De Montfort University, Leicester, UK