Article

A Holistic Review of Building Energy Efficiency and Reduction Based on Big Data

Jeeyoung Lim [1], Joseph J. Kim [2] and Sunkuk Kim [3,*]

1   Department of Architectural Engineering, Pusan National University, Busan 46241, Korea;
     jyounglim02@gmail.com
2   Department of Civil Engineering and Construction Engineering Management, California State University
     Long Beach, Long Beach, CA 90840, USA; Joseph.Kim@csulb.edu
3   Department of Architectural Engineering, Kyung Hee University, Yongin-si 17104, Gyeonggi-do, Korea
*   Correspondence: kimskuk@khu.ac.kr; Tel.: +82-31-201-3685

Abstract: The construction industry is recognized as a major cause of environmental pollution, and it is important to quantify and evaluate building energy. As interest in big data has increased over the past 20 years, research using big data is active. However, the links and contents of much literature have not been summarized, and systematic literature studies are insufficient. The objective of this study was a holistic review of building energy efficiency/reduction based on big data. This review study used a holistic analysis approach method framework. As a result of the analysis, China, the Republic of Korea, and the USA had the most published papers, and the simulation and optimization area occupied the highest percentage with 33.33%. Most of the researched literature was papers after 2015, and it was analyzed because many countries introduced environmental policies after the 2015 UN Conference on Climate Change. This study can be helpful in understanding the current research progress to understand the latest trends and to set the direction for further research related to big data.

Keywords: building energy; building energy efficiency; building energy reduction; big data; literature review; holistic review

1. Introduction

Due to climate change [1], problems such as heat waves, drought, and sea-level rise are occurring [2]. One of the major causes of climate change is greenhouse gas [3], and international regulations on greenhouse gas emission are being strengthened [4,5]. In addition, building energy plays a key role in the energy sector, as building energy accounts for more than 40% of global energy use and greenhouse gases account for 40–50% of global energy [6]. Therefore, the construction industry and building energy are recognized as the main causes of environmental pollution [7], and thus it is important to minimize them [5,8].

For research on building energy reduction and environmental load minimization, Trabucco and Wood (2016) created an energy-efficient design with high sustainability and low energy consumption to address the energy problem of tall buildings at each stage of the building life cycle [9]. Hong et al. (2019) built a real-time monitoring system that can reduce consumption and dust emissions by automatically managing energy consumption and dust emissions at construction sites [10]. Park et al. (2016) studied correlation analysis between the information that can be extracted in the road planning stage and the environmental load computed through life cycle assessment using the data of national highway construction cases [11]. Lee et al. (2018) developed and validated an environmental load estimating model for the new Austrian tunneling method tunnel based on the standard quantity of major works in the early design phase [7]. These papers applied life cycle cost or environmental valuation methodology to construct a system to reduce building energy and minimize environmental load.

In 2001, big data were defined by Doug Laney, an analyst of Meta Group (presently Gartner), as "challenges and opportunities brought about by increased data with a 3Vs model, ie, the increase of Volume, Velocity, and Variety," in a research report [12]. Over the past 20 years, data has grown on a massive scale in various fields [13]. The use of big data to provide support for management decisions became an important topic in current management research [14]. Big data analysis was widely adopted [15] as new research methods in several fields such as information system [16], transportation [17], internet and social media [18], medical industry [19], business [20] and finance [21], weather [22], and smart cities [23].

Literature review of building energy efficiency/reduction based on big data (BER-B) is summarized focusing on a few areas such as unsupervised data analytics [24], energy consumption calculation [6], artificial intelligence (AI), and big data for energy-efficient buildings [25] or studied weaknesses in big data or data mining. Su et al. (2020) provided a review of the publications on carbon emissions and environmental management based on big data and streaming data [26], but it is not about research in the field of architecture, engineering, and construction (AEC). Since the contents of the overall literature in BER-B are not summarized yet, and it is difficult to identify the overall flow and trends, it is necessary to review the future direction of BER-B. The objective of this study was a holistic review of building energy efficiency/reduction based on big data.

This study was carried out as follows:

(1) Describe research methodologies.
(2) Analyze the overall status of the literature sample.
(3) Analyze scientific metrics such as journal sources, co-authors, article regions, keywords, and citations.
(4) Conduct qualitative analysis by dividing it into six categories.

2. Research Methodologies

Figure 1 shows the research framework of the holistic review method, and this is adopted by Li et al. (2020a) and consists of quantitative and qualitative reviews [6,27]. Li et al. (2020a) reviewed the life cycle energy of buildings, but this study is an overall review of building energy focusing on big data. Subjective interpretation and biased conclusions can be avoided and can provide an in-depth understanding of research fields and trends [28]. The approach is used to evaluate published research in BER-B as follows.
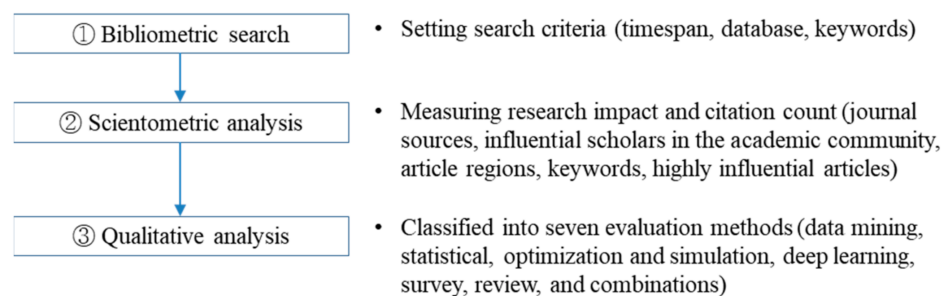

Figure 1. Research framework.

2.1. Bibliometric Search

Bibliographic analysis provides additional insights on topics that have not been fully understood or evaluated before [29]. This analysis is an effective way to gain a comprehensive understanding of the research field that we are searching for [6]. The period was set from 2001 to 2021, when the concept of big data began to be defined. There were no research results until 2004, one publication was published in 2005, 2009, and 2014, and the years 2015–2021 increased steadily. Thesis was selected based on the following criteria. By verifying the language and type of literature, only journal reviews or articles published

in English were retained in the literature samples. Conference proceedings published in journals such as *Energy Procedia* and *Procedia Computer Science* were excluded.

We used the database of the American Society of Civil Engineers Library, Frontiers Database, MDPI Database, ScienceDirect Database, Scopus Database, and Springerlink because these contain many authoritative journals related to AEC and are used by many researchers around the world. Since it is difficult to include all the terms related to the research topic in one study [29], a keyword was selected by combining the abbreviations of the research topics for the search of the paper. The keywords are as follows: ("data mining" OR "big data") AND ("energy" OR "environmental load" OR "life cycle energy" OR "life cycle assessment") AND ("building" OR "construction").

After the initial screening, 312 papers were selected. However, 60 papers were finally selected as literature samples by excluding articles with inconsistent research subjects. For example, Mokhov et al. (2014) conducted a case study based on the premise that energy becomes the dominant factor in the widespread calculation of "big data" companies that cause large-scale electricity bills. However, the paper presented an approach to processor design for a power proportional calculation system, not a big data-based study [30]. In addition, Atitallah et al. (2020) reviewed the literature on the use of the Internet of things (IoT) and deep learning for smart city development, but not in the fields of energy, life cycle energy, or environmental loads. Therefore, these papers were excluded from the literature sample [31].

2.2. Scientometric Analysis

The scientometric approach is the step of measuring research impact and citation count [6]. Scientometric analysis includes co-occurrence analysis of keywords, journal sources, and regions and co-citation analysis of scholars and documents [32,33], and this method aims to investigate the knowledge structure of the research area. In order to provide conclusive information for starting and conducting BER-B-related research, we visualize important categories and trends using tables and graphs with a large amount of bibliographic data.

2.3. Qualitative Analysis

Qualitative analysis is the final step in this three-stage study, which provides an overview of the content of the BER-B literature. It is classified into seven research methodologies, including data mining, statistics, simulation, deep learning, survey, review, and combinations.

3. Results and Findings

3.1. Overall Status of the Literature Samples about BER-B

Figure 2 shows the overall trend of BER-B study results over the past 20 years. More than four papers have been published every year since 2016. From 2014 to 2020 (present), the number of papers published in BER-B has increased. From 2019 to 2020, the number of published papers grew fastest. In particular, because many countries introduced carbon dioxide emission reduction and greenhouse gas control policies after the 2015 United Nations Climate Change Conference [26], it is believed to show a steady increase since 2015.
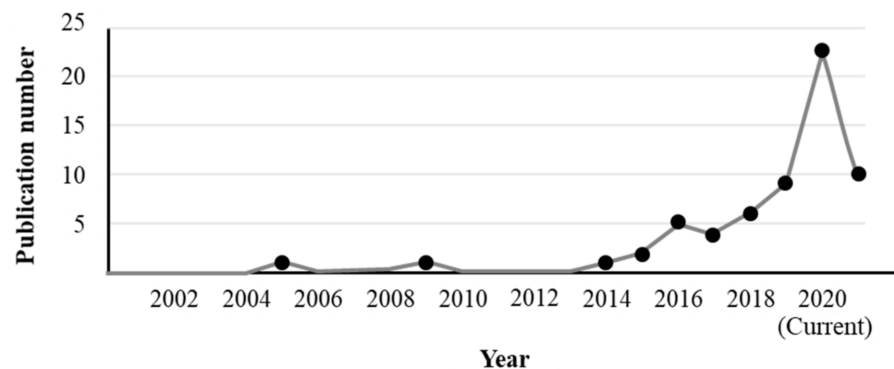
Figure 2. Yearly BER-B-related publication.

This trend shows that interest and research on BER-B have increased in the academic community. In the last six years, scholars have become more active in using big data to identify and solve environmental problems, and it is expected to increase exponentially in the future. In addition, considering that there are currently ten papers in 2021, considerably more papers are expected to be published in one year. In other words, it is expected that the total number of publications will continue to increase, reflecting the academic recognition of the importance of BER-B.

3.2. Results of Scientometric Analysis

3.2.1. Analysis of Journal Sources

Analyzing the influence of journals in a particular field allows researchers to easily obtain useful information and quickly select the most suitable journal for publishing their work [33]. Table 1 shows a summary of the journal sources published in the studies to explore the number of influential journals in BER-B. Quantitative measures of journal impact were summarized into four main measures. Of the four evaluation indicators, total citations are closely related to the number of papers, which indicates the cumulative degree of research results [34,35]. The average number of citations excludes the influence of the number of papers, and the average influence can be observed [6].

Table 1 shows that the total citations of Applied Energy, Frontiers in Mechanical Engineering, Building Simulation, Energy and Buildings, Sustainable Cities and Society, and The International Journal of Advanced Manufacturing Technology were 180, 148, 107, 98, 121, and 205, respectively. However, considering the number of papers, it was determined to be the most influential in BER-B in Applied Energy, Building Simulation, and Energy and Buildings. In addition, despite the fact that Building Simulation has the largest number of papers, it has a low average citation. The reason is that there are many papers published recently.

Artificial Intelligence Review and *Energy Informatics* were expected to have many articles, but the number of citations was as low as zero, indicating the least influential of the journals. The average publication year represents the active year of articles published in the journal. *Renewable and Sustainable Energy Reviews* and *Energy* recently published articles in the BER-B research area, but *Mobile Networks and Applications* has long been inactive in this area. Furthermore, since there are many papers published in 2020 and 2021, the average publication year is close to 2020.

3.2.2. Analysis of Co-Authorship

Scholars always conduct academic collaboration and exchange through knowledge transfer [6]. Scientific cooperation not only promotes the development of expertise but also increases productivity [36–38]. Table 2 shows co-author analysis in BER-B study, and the minimum number of papers and the minimum number of citations were set to 3 and 10 or more, respectively. As a result, 6 out of a total of 227 authors met the inclusion criteria.

As a result of analyzing author-cooperative networks of literature in existing research fields, scholars are divided into two main cooperative groups. For example, it can be seen that Cheng Fan and Jiayuan Wang belong to the same group and have published articles several times. In addition, it can be seen that Shenzhen University, Beijing University of Civil Engineering and Architecture, and Tsinghua University have close cooperation and knowledge communication, and have published several papers in BER-B.

Table 1. Quantitative analysis of journal sources in research.

Journal	Number of Articles	Total Citations	Avg. Citation	Avg. Pub. Year
Applied Energy	7	180	25.71	2018.7
Artificial Intelligence Review	1	0	0.00	2020.0
Building	1	1	1.00	2020.0
Building Simulation	14	107	7.64	2018.5
Cluster Computing	1	2	2.00	2020.0
Energies	2	2	1.00	2021.0
Energy	2	2	1.00	2020.0
Energy and Buildings	6	98	16.33	2018.5
Energy Informatics	1	0	0.00	2020.0
Frontiers in Built Environment	3	5	1.67	2019.7
Frontiers in Energy Research	3	7	2.33	2020.7
Frontiers in Mechanical Engineering	2	148	74.00	2020.0
Frontiers in Sustainable Cities	1	1	1.00	2017.5
Future Generation Computer Systems	1	10	10.00	2020.0
Journal of Ambient Intelligence and Humanized Computing	1	3	3.00	2019.0
Journal of Building Engineering			0.00	2021.0
Journal of Computing in Civil Engineering	1	1	1.00	2019.0
KSCE Journal of Civil Engineering	1	6	6.00	2018.0
Mobile Networks and Applications	2	12	6.00	2011.5
Peer-to-Peer Networking and Applications	1	51	51.00	2018.0
Renewable and Sustainable Energy Reviews	2	3	1.50	2020.0
Soft Computing	1	23	23.00	2016.0
Sustainable Cities and Society	2	121	60.50	2019.0
The International Journal of Advanced Manufacturing Technology	1	205	205.00	2015.0
Visualization in Engineering	1	3	3.00	2016.0
Wireless Personal Communications	1	16	16.00	2016.0

Note: Avg. pub. year represents the average publication year of articles published in the journal [39].

Table 2. Quantitative analysis of influential scholars in the academic community.

Scholar	Affiliation	Number of Articles	Total Citations	Avg. Citation	Avg. Pub. Year
Tianzhen Hong	Lawrence Berkeley National Laboratory	4	10	2.5	2019.0
Cheng Fan	Shenzhen University	5	181	36.2	2019.6
Fu Xiao	The Hong Kong Polytechnic University	5	181	36.2	2019.6
Jiayuan Wang	Shenzhen University	3	120	40	2018.7
Jingjing An	Beijing University of Civil Engineering and Architecture	4	61	15.25	2020.5
Da Yan	Tsinghua University	5	64	12.8	2020.4

Regarding Jingjing An and Da Yan's average publication year, the reason the year exceeds 2020 is that papers that has already been published in 2021 has been published in November 2020. In addition, their thesis is composed only of 2020 and 2021, and therefore they are expected to have continuous research results in the future. It was also found that Cheng Fan, Fu Xiao, and Da Yan were the scholars with the largest number of papers in research in the field of BER-B. However, they only have five publications, and thus it can be seen that no authors do the dominant high research in this field. Moreover, the average publication year is over 2018. Therefore, it shows that this field is in the early stages of development. However, the number of citations is relatively high, which means that interest in research in the field of BER-B is quite high.

3.2.3. Analysis of Article Regions

In order to better understand the global status of BER-B study, the regions of selected literature samples were evaluated and the contributions of different countries were determined. The minimum number of articles and the minimum number of citations were set at 2 and 10 or more, respectively, and 6 out of 15 countries met these criteria. Table 3 shows the results of the countries that are beneficial and influential in the BER-B. According to the number of papers in Table 3, China, the Republic of Korea, and the USA have the most published papers, and their contribution to the research community is clear. Compared to efforts being made to reduce energy such as carbon tax in various European countries, research on BER-B is rather poor.

Table 3. Quantitative analysis of article regions in research.

Country	Number of Articles	Total Citations	Avg. Citations	Avg. Pub. Year
Australia	2	1	0.5	2020.5
China	20	547	27.4	2018.6
India	2	13	6.5	2017.5
Ireland	2	1	0.5	2020.0
Italy	4	7	1.8	2020.0
Norway	2	120	60.0	2019.0
Korea	5	30	6.0	2019.4
Switzerland	2	35	17.5	2018.5
UK	3	151	50.3	2017.3
USA	8	27	3.4	2017.5

3.2.4. Analysis of Keywords

Keywords represent the core content of a research article and summarize topics in the field [40]. The minimum number of occurrences was set to 3. Table 4 shows that 7 out of a total of 204 keywords satisfied the criteria. General keywords such as building and buildings were removed, and keywords with the same semantic meaning such as building energy/energy in buildings were integrated. According to the average citation listed in Table 4, keywords that receive high points mostly include big data, cloud computing, and IoT. In addition, machine learning, data-driven approach, geographic information system (GIS), data mining, data-driven, and artificial intelligence were calculated twice, and these were keywords related to big data. In addition, a number of keywords related to building energy, energy efficiency, and energy consumption were found. For example, the keywords of building energy performance, environmental sustainability, building energy modeling, energy modeling, and energy management were counted as two.

Table 4. Quantitative analysis of keywords in research.

Keywords	Occurrences	Avg. Citation	Avg. Pub. Year
Building energy	3	4.3	2019.3
Big data	8	18.0	2018.1
Cloud computing	3	41.7	2019.0
Energy efficiency	4	6.8	2019.3
Energy consumption	4	1.3	2019.8
IoT	3	53.0	2016.7
Occupant behavior	4	0.8	2020.3

The keywords of urban planning, urban building energy modeling, smart sustainable cities, and simulation were calculated twice. This means that smart city, urban, and urban building can be good research fields in BER-B. In addition, the keywords of the heating, ventilating, and air-conditioning (HVAC) system and building operational performance were counted as two, and therefore many studies were trying to solve systemic problems to save energy. However, in keywords, "occupant behavior" is used more frequently than other keywords. This shows that interest in research related to occupant behavior in the BER-B field is increasing. In other words, it does not just solve the technical problem but tries to save energy by analyzing the behavior pattern of the occupant.

3.2.5. Analysis of Article Citations

For the most cited and influential papers, the minimum number of citations was set at 50. Table 5 shows that seven out of 60 papers met the criteria. Fan, Cheng. had the largest number of articles in Table 2, and the total number of citations for his articles in Table 5 was 161 times. Fan, Cheng. uses big data in BER-B to perform modeling [15] and building energy predictions [41] and review articles in the field [24]. In addition, Fan et al. (2021) is a paper to be published in 2021 (currently November 2020) [15], but it already shows a significant number of citations, and it will continue to increase in the future.

Table 5. Quantitative analysis of highly influential articles in BER-B research.

Article	Title	Citation
Bibri (2018) [42]	The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability	120
Fan et al. (2018) [24]	Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review	51
Fan et al. (2019) [43]	Assessment of deep recurrent neural network-based strategies for short-term building energy predictions	50
Fan et al. (2021) [15]	Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches	60
Li et al. (2015) [44]	Big Data in product lifecycle management	205
Liu et al. (2018) [45]	Big program code dissemination scheme for emergency software-define wireless sensor networks	51
Ma and Cheng (2016) [46]	Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology	78
Van Dronkelaar et al. (2016) [47]	A Review of the Energy Performance Gap and Its Underlying Causes in Non-Domestic Buildings	148

3.3. Qualitative Analysis

Big data technologies include several disciplines including statistics, data mining, machine learning, neural networks, optimization methods, and visualization approaches [48]. There are many specific technologies in this field and some overlap [14]. Collected papers were classified according to the research methodology used to write the papers. Table 6 shows that they were classified into seven research methodologies including four analysis techniques such as data mining, statistics, simulation, and deep learning, review articles, and combinations in BER-B.

3.3.1. Research Methodology Based on Data Mining

Many studies were conducted to save energy by improving the building system using collected data in the building energy field. Catrini et al. (2020) took data from a large Do It Yourself store in northern Italy as a case study to investigate the profitability of a third-generation system for commercial buildings. Compared with the existing energy conversion system, energy consumption and CO_2 emissions have been reduced [49]. Chiosa et al. (2021) proposed an innovative anomaly detection and diagnosis methodology that automatically detects abnormal energy consumption at the level of the entire building and diagnoses the lower level for abnormal patterns. The paper performed a post-mining analysis based on association rule mining and found the major lower level that mainly affects the anomalies detected in the entire building [50]. Conner et al. (2005) studied data collected from several applications deployed throughout the building that monitor room occupancy and environmental statistics and provide access to room reservation status. Using the proposed protocol, it was confirmed that a small number of wall-powered nodes can significantly increase the lifespan of the battery-powered network [51].

Research Methodology	Example Papers	Number of Papers	Percentage of Papers (%)
Data mining	Catrini et al. (2020) [49], Chiosa et al. (2021) [50], Conner et al. (2005) [51], Kang and Choi (2018) [52], Li et al. (2020b) [53], Liu et al. (2018) [45], Ma et al. (2021) [54], Reshma et al. (2019) [55], Shahrokni et al. (2014) [56], Silva et al. (2020) [23], Sohn and Dunn (2019) [57], Suciu et al. (2016) [58], Guo et al. (2020) [59], Tang et al. (2021) [60], Van Dronkelaar et al. (2016) [47]	15	25.00
Statistics	Ahn and Sohn (2019) [61], Buffat et al. (2017) [62], Filippi and Sirombo (2019) [63], Kamel et al. (2020) [64], Liu et al. (2015) [65], Manfren et al. (2020) [66], Pan and Zhang (2020) [67], Piscitelli et al. (2021) [68], Qian et al. (2020) [69]	9	15.00
Simulation	Alsaleemet et al. (2020) [70], Bambara et al. (2021) [71], Dawood et al. (2017) [72], Fan et al. (2019a) [41], Han et al. (2019) [73], Hong et al. (2020) [74], Jin et al. (2020) [75], Ma and Cheng (2016) [46], Mosteiro-Romero et al. (2020) [76], Nematchoua et al. (2020) [77], Chen et al. (2021) [78], Li et al. (2021) [79], Su et al. (2009) [80], Van Dronkelaar et al. (2019) [81], Ye et al. (2016) [82], Ali et al. (2020a) [83], Belafi et al. (2017) [84], Moreno et al. (2016) [85]	20	33.33
Deep learning	Fan et al. (2019b) [43], Niu et al. (2018) [86], Zhao et al. (2021) [87]	3	6.38
Survey and investigation	Arora and Bala (2020) [88], Bibri and Krogstie (2020) [89], Li et al. (2015) [44], Salim et al. (2020) [90]	4	8.33
Review	Bibri (2018) [42], Fan et al. (2018) [24], Fan et al. (2021) [15], Li et al. (2020c) [91], Mehmood et al. (2019) [25], Yan et al. (2020) [92]	6	12.77
Combination	Ali et al. (2020b) [93], Jeong et al. (2020) [94], Risch et al. (2020) [95], Yigit (2021) [96]	3	6.38

Kang and Choi (2018) proposed a building information modeling(BIM)-based data mining method to support data integration and function expansion in consideration of functional variability and scalability. Scenarios related to building energy management were developed, and work efficiency was verified through analysis of the generated results [52]. Li et al. (2020b) developed a new operational approach to improve the energy efficiency of HVAC systems in office spaces. This approach can contribute to a large amount of energy savings when operating an HVAC system [53]. Liu et al. (2018) proposed an adjustable duty cycle-based fast dissemination method for minimum-transmission broadcast in the smart wireless software-defined network. By using the remaining energy of the node, the duty cycle of the node is increased to reduce the transmission time and delay while maintaining the network life [45]. Ma et al. (2021) reduced the transmission time while maintaining the network life by increasing the duty cycle of the node using the remaining energy of the node. Compared to conventional methods, this method was able to provide a more comprehensive analysis of energy-saving potential in a shorter time [54].

Reshma et al. (2019) suggested that the wireless sensor network (WSN) is an energy-aware routing protocol to solve the limited node energy problem. The proposed protocol found energy efficiency in routing vast amounts of data and extending network connectivity [55]. Shahrokni et al. (2014) quantified resource efficiency and greenhouse gas reduction

by studying empirical meter data obtained from district heating use in Stockholm. The study provided a benchmark of excellence in the energy performance monitoring stage [56]. Silva et al. (2020) proposed a big data analysis-embedded smart city architecture, which uses a smart gateway system to integrate the web and smart control system. The system was evaluated based on a certified data set from a variety of reliable data sources [23]. Sohn and Dunn (2019) presented an exploratory analysis of an existing large-scale building energy data set, focusing on two data sets: the commercial building energy consumption survey in the United States and the building performance database [57]. Suciu et al. (2016) studied how to process big data collected from distributed cloud systems based on general systems and remote telemetry units for remote monitoring of renewable energy targets [58]. In these studies, the efficiency of the improved system was verified. Van Dronkelaar et al. (2016) reviewed the discrepancy between predicted and measured energy use based on 62 buildings and presented an outlook to global studies [47].

Meanwhile, Guo et al. (2020) compared data on energy use intensity, energy structure, and carbon emissions in different countries [59]. The paper discussed including support for relevant policies, potential work, and limitations on building energy data to address new trends in related research. Tang et al. (2021) showed that a case study of field measurements at a Hong Kong campus building was conducted to investigate the impact of technical guided occupant behavior on the energy use of a central air conditioning system [60]. The paper developed a new concept, skill-driven occupant behavior, to coordinate between occupants and technology/control strategies for the building's energy system. Unlike most studies, these studies showed a new direction in the field of BER-B. Therefore, it is possible to derive the direction of further research or new research in the BER-B field by referring to these studies.

3.3.2. Research Methodology Based on Statistics

Statistics, the most basic technique for data analysis, exists in almost all fields of study [14]. Kamel et al. (2020) studied feature selection for hot water, heating, cooling, and ventilation loads in residential buildings under the mixed-humid climate zone. Filter method, wrapper recursive feature elimination, wrapper backward elimination, linear regression, Lasso regression, and Extreme Gradient Boosting regression are adopted for heating and cooling days [64]. Filippi and Sirombo (2019) showed the results of two-year residential monitoring of a case project of 152 apartments for large social housing built near Milan, Italy. Data on thermal energy consumption for heating and cooling, domestic hot and cold water use, and control units were evaluated [63]. Liu et al. (2015) studied the correlation between energy consumption and architectural surfaces from the perspective of passive design. Taking Beijing as a case project, a typical high-rise office building floor plan was reviewed in terms of annual dynamic load and energy consumption [65].

Manfren et al. (2020) analyzed the topic of linking the energy performance of the design and operation phases through a regression-based approach of buildings, emphasizing the hierarchical nature of building energy modeling data [66]. Pan and Zhang (2020) is a case study, validated on a multidimensional data set of building energy performance provided by the city government of Seattle. The study was conducted to estimate the building's weather-normalized site energy use intensity and characterize possible nonlinear relationships with 12 other possible sources [67]. Piscitelli et al. (2021) developed and tested a tool to detect and diagnose abnormal daily electrical energy patterns associated with substations on university campuses. The study developed an error-free predictive model using an artificial neural network (ANN) with the following regression tree to find an unusual trend in electrical energy consumption [68]. Qian et al. (2020) proposed a virtual sensor modeling method to determine the actual energy efficiency and power consumption of the variable refrigerant flow system in residential buildings. Statistical and clustering analysis was performed to determine energy efficiency and power consumption to discover the distribution and typical operating load patterns of the variable refrigerant flow system in Chinese residential buildings [69].

There are studies conducted by collecting GIS-based big data as follows. Ahn and Sohn (2019) provided an investigation into the effectiveness of three urban form variables (vertical density, horizontal compactness, and variation of building heights), using the Seattle city government's building energy consumption data and GIS data obtained from the Washington Geospatial Data Archive. Using a regression model, the study confirmed that the correlation between the spatial delay model of building energy use and its surrounding level is spatially dependent on the energy consumption of the building [61]. Buffat et al. (2017) collected Swiss residential building data and modeled the building heat demand of large regions with a high temporal resolution using a large-scale GIS-based method. It was validated against a database containing the measured energy demand of 1845 buildings in the city of St. Gallen and 120 buildings in the town of Alpine in Zernez, Switzerland [62]. Using the research results of these papers, strategies for better energy saving in buildings and urban buildings can be developed. It is also expected to provide useful information to urban designers and planners.

3.3.3. Research Methodology Based on Simulation

Simulation can be divided into general simulation and optimization, and it occupied the highest percentage of research methodologies. First, the following studies were based on general simulation. Alsaleemet et al. (2020) presented a method to access and exploit wearable device data to build a personal thermal comfort model. The paper evaluated various supervised machine-learning algorithms to produce accurate personal thermal comfort models for each building occupant that exhibit superior performance compared to a general model for all occupants [71]. Dawood et al. (2017) studied the potential to improve functionality, the accuracy of energy use calculations, and visualization of carbon emissions using remotely sensed data in commercial and open-source combinations. The study used light detection and ranging data for a mixed-use urban area in the northeastern part of the UK and presented a method to improve the quality of the input data for carbon modeling [72]. Fan et al. (2019a) described and evaluated a data-driven building energy performance model to increase the practical value of advanced machine learning technologies in the building sector [41]. Han et al. (2019) proposed a virtual machine rearrangement method to increase energy efficiency by increasing the density of virtual machines using the Knapsack algorithm. Through the proposed method, the Knapsack algorithm was improved to achieve efficient virtual machine relocation in a short period of time [73].

Hong et al. (2020) described a new methodology for generating synthetic smart meter data for a building's electricity use using detailed building energy modeling, which aims to capture the variability and probability of real energy use in buildings. Their methodology was able to generate customized data sets to represent specific scenarios with a known truth and a controllable amount of synthetic noise [74]. Jin et al. (2021) proposed a novel data-driven predictive control method based on a temporal sequential-based artificial neural network. As a result of the simulation, it was confirmed that the proposed method achieved higher accuracy (97.4%), and fewer false errors (from 79.5 of the existing time delay method to 0.6 times per day) were achieved by the model [75].

Ma and Cheng (2016) proposed a GIS-integrated data mining methodology framework for estimating the energy use intensity of urban-scale buildings, including feature selection, preprocessing, and algorithm optimization. A case study estimating the energy use intensity of 3640 multi-family residential buildings in New York City has been tested and validated [46]. Mosteiro-Romero et al. (2020) presented a computational approach for the quantitative analysis of building energy demand at the district scale, including interdependent factors such as relative humidity and wind speed, local air temperature, diversity in building geometry, and materials. This method, which couples the microclimate model program ENVI-met and the district-scale energy simulation tool City Energy Analyst, was applied to a case study in Zurich, Switzerland [76]. Nematchoua et al. (2020) analyzed and compared the energy consumption and carbon emissions of six building categories

commonly found in Saharan Africa cities that are designed to be deployed in 13 cities unevenly distributed in six climate regions of Madagascar [77].

Among simulation studies, there are studies that have not been conducted solely by data collection. Bambara et al. (2021) compared the energy performance (consumption and generation) of an existing house with that of two energy-efficient homes equipped with building-integrated photovoltaics (BIPV) that are built on the same land lot in Montreal, Canada. An energy model was created for each of the houses, and annual energy simulations were performed to quantify their energy performance [71]. Chen et al. (2021) conducted a study using data generated by calibration and program. The identification method was tested by adopting data generated by EnergyPlus calibrated in real buildings. In order to understand the control strategy of HVAC in a given building using a data mining algorithm, data on the building automation system (BAS) and data on building energy were analyzed [78]. Li et al. (2021) developed a data-based building energy prediction model using a transfer learning method that uses data from other buildings as limited data. To test the proposed method, data from 400 non-residential buildings from the open-source Building Genome Project were used, and the energy prediction accuracy was verified [79].

Su et al. (2009) studied a set of 3600 alternative weather files generated using different parameter weights for Beijing, China. The study evaluated the influence of different parameter weights when generating weather data in the form of a typical year according to a typical weather year methodology [80]. Van Dronkelaar et al. (2019) quantified the impact of underlying causes of this gap by developing building simulation models of four buildings and then calibrating them toward their measured energy use at a high level of data granularity [81]. Ye et al. (2016) confirmed that long-term data can be predicted with short-term data because they do not change over time through simulation. The paper investigated and verified the appropriateness of using empirical models to predict long-term emissions with short-term data, typically less than one month [82].

In addition, the optimization studies through modeling are as follows. Ali et al. (2020b) developed and simulated a methodology for optimizing the urban-scale energy retrofit determination of residential buildings in Ireland using a data-driven approach. The importance of data-based renovation modeling was confirmed by reducing the number of functions in the building inventory database from 203 to 56 with a building grade prediction accuracy of 86% [83]. Belafi et al. (2017) introduced a method of implementing occupant behavior modeling into Hungarian office building projects and a way to support the building audit and renovation analysis process. As a result of optimization simulations, it was found that the building could save 23% of its annual energy use [84]. Moreno et al. (2016) presented a new approach to energy conservation in buildings through the application of soft computing techniques and identification of relevant parameters to generate predictive models of energy consumption in buildings. These models can be used to define strategies for optimizing the building's daily energy consumption [85].

In general, big data-based research is conducted by collecting actual data. However, in the investigated simulation studies, not only the studies were conducted using data collection, but also many studies were conducted using data generated by the program. These studies predicted or optimized the energy using the developed model, and evaluated the model. Collecting data comes with many difficulties without the active cooperation of the building owners. Therefore, research methodology using a program that can generate data is expected to continue to be highly useful in the BER-B field.

3.3.4. Research Methodology Based on Deep Learning

Deep learning is an algorithm based on learning representations of data in machine learning [14]. Fan et al. (2019b) studied the usefulness of advanced recurrent neural network-based strategies for energy prediction. At a high level, three inference approaches were used to generate short-term predictions, including the recursive approach, the direct approach, and the multi-input and multi-output approach [43]. Niu et al. (2018) estimated consumption using five commonly used data-based algorithms: autoregressive with ex-

ternal inputs, subspace state space, state space, discretized variable Bayesian network, and continuous variable Bayesian network [86]. Zhao et al. (2021) proposed the hybrid-model-based deep reinforcement learning method for the HVAC control problem. Using this method, the protection mechanism and adjusting reward methods are used to further reduce the learning cost [87]. These papers conducted using deep learning were conducted for the purpose of energy prediction, and the derived results can be applied to real building models. In addition, it is expected to provide useful data for building experts by providing more insight into the model for building energy prediction.

3.3.5. Research Methodology Based on Survey and Investigation

Arora and Bala (2020) conducted a survey to address the major challenges of big data applications hosted in cloud data centers by reviewing existing energy-efficient technologies. As a result of reviewing various energy efficiency technologies for big data applications, it was confirmed that hybrid energy efficiency technologies are the best way to reduce overall energy consumption [88]. Bibri and Krogstie (2020) investigated and compared Stockholm and Barcelona in terms of smart sustainable cities based on environmental data by adopting a case study as a qualitative research methodology. As a result, it has been confirmed that smart grid, smart building, smart meter, smart environmental monitoring, and smart city ambassador are major data-based smart solutions applied to improve and develop environmental sustainability in eco-city and smart city [89].

Li et al. (2015) introduced the concepts, features, and applications of big data and analyzed various data related to the three main stages (beginning of life, mid-life, and end of life) of product life cycle management. In product life cycle management, the paper summarized existing big data applications, and investigated and analyzed potential applications of big data technology [44]. Salim et al. (2020) presented various sources of resident-centered city data useful for data-driven modeling and classified a range of applications and recent data-driven modeling techniques for urban behavior and energy modeling along with traditional stochastic and simulation-based approaches [90]. In these surveys and investigations in the field of BER-B, extensive data were collected, various problems related to energy efficiency technologies, smart city and building, and resident-oriented cities were raised, and the direction of future research was explained. Therefore, it is expected that these studies will be able to narrow the knowledge gap by providing insights on related topics and confirming detailed information about the research environment and its impact on energy use.

3.3.6. Research Methodology Based on Review

Bibri (2018) reviewed relevant literature for the purpose of identifying and discussing state-of-the-art sensor-based big data applications supported by IoT for environmental sustainability and related data processing platforms and computing [42]. Fan et al. (2018) provided a review of unsupervised data analysis studies in mining building operation data and improving building operation performance [24]. Fan et al. (2021) reviewed articles on data-driven methods for real-world applications for building energy modeling and improving building performance, especially those based on large data sets [15].

Li et al. (2020c) provided a review on the studies on energy consumption calculation of urban buildings and divided the reviewed studies into three categories: database, model, and platform [91]. Mehmood et al. (2019) provided a review of studies on applying AI and big data to energy-efficient buildings, focusing on the use of machine learning and large databases for improved search, optimization speed, and accuracy [25]. Yan et al. (2020) reviewed the AI-based applications in building energy efficiency with a focus on facilitating the implementation of zero energy building, considering the influence of occupant behaviors [92].

3.3.7. Research Methodology Based on Combination

Ali et al. (2020a) predicted the energy performance of over 2 million buildings using data from approximately 650,000 Irish Energy Performance Certificate buildings. The study provided 88% prediction accuracy using a deep learning algorithm and proposed an integrated methodology using building energy modeling [93]. Jeong et al. (2021) applied a data-driven approach to setting a CO_2 emission benchmark using data mining techniques. Data for a total of 1212 multi-family housing complexes were established, and the developed CO_2 emission benchmark was verified using statistical methods such as Mann–Whitney test and Kruskal–Wallis test [94]. Risch et al. (2021) modeled and calibrated three office buildings in Germany based on one-year hourly measurement data. The corrected model was evaluated as a coefficient of variation of root mean squared error and R^2, this study was simulated using the corrected values, and statistical analysis was conducted [95]. Yigit (2021) developed a surrogate model-based integrated optimization system to obtain energy-optimal thermal designs for residential buildings in the most urbanized cities in Turkey under different levels of budget constraints. This system is an integrated system consisting of a genetic algorithm optimization technique and gradient-boosting machine-based surrogate model [96].

4. Discussion

Since the concept of big data emerged recently, there were many papers recently published in the field of BER-B. The papers used for this review were adopted since 2001 when the concept of big data began to be defined. However, there were no research results until 2004, and a total of three publications were published between 2005 and 2014; the number of articles increased steadily from 2015 to 2021. In addition, there were five papers in 2021 (currently November 2020), and one year later, the number of papers in 2021 is expected to increase considerably.

In addition, among the seven research methodologies, 15 articles were published in the field of simulation and optimization, showing the highest points. These methodologies included testing the identification method by adopting the data generated by EnergyPlus [78], developing a predictive model using data obtained from other buildings as limited data [79], evaluating 3600 alternate weather data generated using parameter weights [80], predicting long-term data with short-term data of less than one month [82], development of an optimized model for urban-scale energy in residential buildings [83], optimizing energy consumption through occupant behavior modeling [84], and optimizing the daily energy consumption of buildings [85]. These papers did not consist solely of data collection but created data using programs or implemented an optimal model using existing data.

In the researched papers, data were collected on conference rooms [51], HVAC systems used in office spaces [48], and occupant behavior in terms of energy systems in buildings [60] and hot water, heating, cooling, and ventilation loads in residential buildings [64]. These studies are studies conducted using partial data in buildings, in other words, building energy management systems, building management, and facility management-related fields that have not yet accumulated big data. It is a field that has the potential to contribute to minimizing construction energy through data mining in the future, and therefore more research on this topic is necessary.

5. Conclusions

This study adopted a three-step holistic method to review articles published in BER-B over the past 20 years, after big data was first defined. According to the literature survey, 60 published papers were selected as literature samples to understand the trends of the papers, and the following conclusions are as follows.

First, most of the literature surveyed in this study was from 2015 onward, and this is believed to be due to the introduction of greenhouse gas control and carbon diox-

ide emission reduction policies in many countries after the 2015 UN Conference on Climate Change.

Second, China, the Republic of Korea, and the USA have the most published papers, and as a result of an author-cooperative network analysis, Shenzhen University, Beijing University of Civil Engineering and Architecture, and Tsinghua University have close cooperation and knowledge communication. It can be seen that the thesis in BER-B has been published several times.

Third, in BER-B, the area of simulation was 33.33%, accounting for the highest percentage of the seven research methodologies. In addition, simulation studies are not conducted solely by data collection, but a number of studies have been found using data generated by calibration and programs.

Fourth, there are many cases where the average publication year exceeds 2020 because there are papers already published in 2021 and in November 2020. Papers to be published in 2021 already show a significant number of citations and are expected to increase steadily in the future, and their citations will increase further.

This study helps to understand the recent trends by grasping the current research progress in the BER-B field. This can be helpful in setting the direction for further research related to big data. In addition, research on the selection of construction methods through the collection of big data for various construction methods has not been conducted yet. Therefore, research on environmental load and energy efficiency analysis for new technologies and new methods as well as existing methods is needed.

Author Contributions: Conceptualization, J.L.; methodology, J.L. and S.K.; formal analysis, J.L. and S.K.; investigation, J.L.; writing—original draft preparation, J.L. and J.J.K.; writing—review and editing, J.L. and S.K.; visualization, J.L.; supervision, S.K.; project administration, J.L. and S.K.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Sharma, A.; Reddy, V.; Jain, S.; Bansal, V.; Niranjan, B. Perception of teachers about implementation of oral health education in primary school curriculum. *Int. J. Dent. Med. Spec.* **2016**, *3*, 10–15. [CrossRef]
2. Intergovernmental Panel on Climate Change (IPCC). *Working Group I Contribution to the IPCC Fifth Assessment Report Climate Change 2013*; Cambridge University Press: Cambridge, UK, 2013.
3. Jung, K.O.; Chung, Y.K. The pollution and economic growth based on the multi-country comparative analysis. *J. Ind. Econ. Bus.* **2004**, *17*, 1077–1098.
4. Sartori, I.; Hestnes, A.G. Energy use in the lifecycle of conventional and low energy buildings: A review article. *Energy Build.* **2007**, *39*, 249–257. [CrossRef]
5. Lim, J.; Kim, J.J. Dynamic Optimization Model for Estimating In-Situ Production Quantity of PC Members to Minimize Environmental Loads. *Sustainability* **2020**, *12*, 8202. [CrossRef]
6. Li, C.Z.; Lai, X.; Xiao, B.; Tam, V.W.; Guo, S.; Zhao, Y. A holistic review on life cycle energy of buildings: An analysis from 2009 to 2019. *Renew. Sustain. Energy Rev.* **2020**, *134*, 110372. [CrossRef]
7. Lee, J.H.; Kim, K.; Kim, H. Environmental load estimating model of NATM tunnel based on standard quantity of major works in the early design phase. *KSCE J. Civ. Eng.* **2018**, *22*, 1040–1051. [CrossRef]
8. Lim, J.; Kim, S. Evaluation of CO_2 emission reduction effect using in-situ production of precast concrete components. *J. Asian Archit. Build. Eng.* **2020**, *19*, 176–186. [CrossRef]
9. Trabucco, D.; Wood, A. LCA of tall buildings: Still a long way to go. *J. Build. Eng.* **2016**, *7*, 379–381. [CrossRef]

10. Hong, J.; Hong, T.; Kang, H.; Lee, M. A Framework for Reducing Dust Emissions and Energy Consumption on Construction Sites. *Energy Procedia* **2019**, *158*, 5092–5096. [CrossRef]
11. Park, J.Y.; Lee, D.E.; Kim, B.S. A study on analysis of the environmental load impact factors in the planning stage for highway project. *KSCE J. Civ. Eng.* **2016**, *20*, 2162–2169. [CrossRef]
12. Laney, D. 3-d data management: Controlling data volume, velocity and variety. *META Group Res. Note* **2001**, *6*, 1.
13. Chen, M.; Mao, S.; Liu, Y. Big data: A survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [CrossRef]
14. Yang, F.; Wang, M. A review of systematic evaluation and improvement in the big data environment. *Front. Eng. Manag.* **2020**, 1–20. [CrossRef]
15. Fan, C.; Yan, D.; Xiao, F.; Li, A.; An, J.; Kang, X. Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches. *Build. Simul.* **2021**, *14*, 3–24. [CrossRef]
16. Agarwal, R.; Dhar, V. Big data, data science, and analytics: The opportunity and challenge for IS research. *Inf. Syst. Res.* **2014**, *25*, 443–448. [CrossRef]
17. Liu, J.; Wang, X.; Khattak, A.J.; Hu, J.; Cui, J.; Ma, J. How big data serves for freight safety management at highway-rail grade crossings? A spatial approach fused with path analysis. *Neurocomputing* **2016**, *181*, 38–52.
18. Liu, X.; Singh, P.V.; Srinivasan, K. A structured analysis of unstructured big data by leveraging cloud computing. *Mark. Sci.* **2016**, *35*, 363–388. [CrossRef]
19. Nambisan, P.; Luo, Z.; Kapoor, A.; Patrick, T.B.; Cisler, R.A. Social media, big data, and public health informatics: Ruminating behavior of depression revealed through Twitter. In Proceedings of the 48th Hawaii International Conference on System Sciences, Kauai, HI, USA, 5–8 January 2015; pp. 2906–2913.
20. Akter, S.; Wamba, S.F. Big data analytics in e-commerce: A systematic review and agenda for future research. *Electron. Mark.* **2016**, *26*, 173–194. [CrossRef]
21. Sirignano, J.; Giesecke, K. Risk analysis for large pools of loans. *Manag. Sci.* **2018**, *65*, 107–121. [CrossRef]
22. Biffis, E.; Chavez, E. Satellite data and machine learning for weather risk management and food security. *Risk Anal.* **2017**, *37*, 1508–1521. [CrossRef]
23. Silva, B.N.; Khan, M.; Han, K. Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management. *Future Gener. Comput. Syst.* **2020**, *107*, 975–987. [CrossRef]
24. Fan, C.; Xiao, F.; Li, Z.; Wang, J. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. *Energy Build.* **2018**, *159*, 296–308. [CrossRef]
25. Mehmood, M.U.; Chun, D.; Han, H.; Jeon, G.; Chen, K. A review of the applications of artificial intelligence and big data to buildings for energy-efficiency and a comfortable indoor living environment. *Energy Build.* **2019**, *202*, 109383. [CrossRef]
26. Su, Y.; Yu, Y.; Zhang, N. Carbon emissions and environmental management based on Big Data and Streaming Data: A bibliometric analysis. *Sci. Total Environ.* **2020**, 138984. [CrossRef]
27. Yin, X.; Liu, H.; Chen, Y.; Al-Hussein, M. Building information modelling for off-site construction: Review and future directions. *Autom. Constr.* **2019**, *101*, 72–91. [CrossRef]
28. Zou, P.X.; Sunindijo, R.Y.; Dainty, A.R. A mixed methods research design for bridging the gap between research and practice in construction safety. *Saf. Sci.* **2014**, *70*, 316–326. [CrossRef]
29. Fahimnia, B.; Sarkis, J.; Davarzani, H. Green supply chain management: A review and bibliometric analysis. *Int. J. Prod. Econ.* **2015**, *162*, 101–114. [CrossRef]
30. Mokhov, A.; Rykunov, M.; Sokolov, D.; Yakovlev, A. Design of processors with reconfigurable microarchitecture. *J. Low Power Electron. Appl.* **2014**, *4*, 26–43. [CrossRef]
31. Atitallah, S.B.; Driss, M.; Boulila, W.; Ghézala, H.B. Leveraging deep learning and IoT big data analytics to support the smart cities development: Review and future directions. *Comput. Sci. Rev.* **2020**, *38*, 100303. [CrossRef]
32. Falagas, M.E.; Pitsouni, E.I.; Malietzis, G.A.; Pappas, G. Comparison of PubMed, Scopus, web of science, and Google scholar: Strengths and weaknesses. *FASEB J.* **2008**, *22*, 338–342. [CrossRef]
33. Hoxha, V.; Haugen, T.; Bjorberg, S. Measuring perception about sustainability of building materials in Kosovo. *Facilities* **2017**. [CrossRef]
34. Hammersley, M. On 'systematic' reviews of research literatures: a 'narrative' response to Evans & Benefield. *Br. Educ. Res. J.* **2001**, *27*, 543–554. [CrossRef]
35. Darko, A.; Chan, A.P.; Huo, X.; Owusu-Manu, D.G. A scientometric analysis and visualization of global green building research. *Build. Environ.* **2019**, *149*, 501–511. [CrossRef]
36. Jin, R.; Zuo, J.; Hong, J. Scientometric review of articles published in ASCE's journal of construction engineering and management from 2000 to 2018. *J. Constr. Eng. Manag.* **2019**, *145*, 06019001. [CrossRef]
37. Ding, Y. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *J. Informetr.* **2011**, *5*, 187–203. [CrossRef]
38. Su, H.N.; Lee, P.C. Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight. *Scientometrics* **2010**, *85*, 65–79. [CrossRef]
39. Chen, C. Science mapping: A systematic review of the literature. *J. Data Inf. Sci.* **2017**, *2*, 1–40. [CrossRef]
40. Jin, R.; Zou, P.X.; Piroozfar, P.; Wood, H.; Yang, Y.; Yan, L.; Han, Y. A science mapping approach based review of construction safety research. *Saf. Sci.* **2019**, *113*, 285–297. [CrossRef]

41.   Fan, C.; Xiao, F.; Yan, C.; Liu, C.; Li, Z.; Wang, J. A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Appl. Energy* **2019**, *235*, 1551–1560. [CrossRef]

42.   Bibri, S.E. The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustain. Cities Soc.* **2018**, *38*, 230–253. [CrossRef]

43.   Fan, C.; Wang, J.; Gang, W.; Li, S. Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. *Appl. Energy* **2019**, *236*, 700–710. [CrossRef]

44.   Li, J.; Tao, F.; Cheng, Y.; Zhao, L. Big data in product lifecycle management. *Int. J. Adv. Manuf. Technol.* **2015**, *81*, 667–684. [CrossRef]

45.   Liu, X.; Li, G.; Zhang, S.; Liu, A. Big program code dissemination scheme for emergency software-define wireless sensor networks. *Peer Peer Netw. Appl.* **2018**, *11*, 1038–1059. [CrossRef]

46.   Ma, J.; Cheng, J.C. Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. *Appl. Energy* **2016**, *183*, 182–192. [CrossRef]

47.   Van Dronkelaar, C.; Dowson, M.; Burman, E.; Spataru, C.; Mumovic, D. A review of the energy performance gap and its underlying causes in non-domestic buildings. *Front. Mech. Eng.* **2016**, *1*, 17. [CrossRef]

48.   Chen, C.P.; Zhang, C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf. Sci.* **2014**, *275*, 314–347. [CrossRef]

49.   Catrini, P.; Curto, D.; Franzitta, V.; Cardona, F. Improving energy efficiency of commercial buildings by Combined Heat Cooling and Power plants. *Sustain. Cities Soc.* **2020**, *60*, 102157. [CrossRef]

50.   Chiosa, R.; Piscitelli, M.S.; Capozzoli, A. A Data Analytics-Based Energy Information System (EIS) Tool to Perform Meter-Level Anomaly Detection and Diagnosis in Buildings. *Energies* **2021**, *14*, 237. [CrossRef]

51.   Conner, W.S.; Chhabra, J.; Yarvis, M.; Krishnamurthy, L. Experimental evaluation of topology control and synchronization for in-building sensor network applications. *Mob. Netw. Appl.* **2005**, *10*, 545–562. [CrossRef]

52.   Kang, T.W.; Choi, H.S. BIM-based data mining method considering data integration and function extension. *KSCE J. Civ. Eng.* **2018**, *22*, 1523–1534. [CrossRef]

53.   Li, W.; Koo, C.; Hong, T.; Oh, J.; Cha, S.H.; Wang, S. A novel operation approach for the energy efficiency improvement of the HVAC system in office spaces through real-time big data analytics. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109885. [CrossRef]

54.   Ma, R.; Yang, S.; Wang, X.; Wang, X.C.; Shan, M.; Yu, N.; Yang, X. Systematic Method for the Energy-Saving Potential Calculation of Air-Conditioning Systems via Data Mining. Part I: Methodology. *Energies* **2021**, *14*, 81. [CrossRef]

55.   Reshma, J.; Kumar, T.S.; Vani, B.A.; Sakthivel, S. Big data oriented energy aware routing for wireless sensor networks. *Mob. Netw. Appl.* **2019**, *24*, 298–306. [CrossRef]

56.   Shahrokni, H.; Levihn, F.; Brandt, N. Big meter data analysis of the energy efficiency potential in Stockholm's building stock. *Energy Build.* **2014**, *78*, 153–164. [CrossRef]

57.   Sohn, M.D.; Dunn, L.N. Exploratory Analysis of Energy Use across Building Types and Geographic Regions in the United States. *Front. Built Environ.* **2019**, *5*, 105. [CrossRef]

58.   Suciu, G.; Vulpe, A.; Martian, A.; Halunga, S.; Vizireanu, D.N. Big data processing for renewable energy telemetry using a decen Fan tralized cloud M2M system. *Wirel. Pers. Commun.* **2016**, *87*, 1113–1128. [CrossRef]

59.   Guo, S.; Yan, D.; Hu, S.; An, J. Global comparison of building energy use data within the context of climate change. *Energy Build.* **2020**, *226*, 110362. [CrossRef]

60.   Tang, R.; Wang, S.; Sun, S. Impacts of technology-guided occupant behavior on air-conditioning system control and building energy use. *Build. Simul.* **2021**, *14*, 209–217. [CrossRef]

61.   Ahn, Y.; Sohn, D.W. The effect of neighbourhood-level urban form on residential building energy use: A GIS-based model using building energy benchmarking data in Seattle. *Energy Build.* **2019**, *196*, 124–133. [CrossRef]

62.   Buffat, R.; Froemelt, A.; Heeren, N.; Raubal, M.; Hellweg, S. Big data GIS analysis for novel approaches in building stock modelling. *Appl. Energy* **2017**, *208*, 277–290. [CrossRef]

63.   Filippi, M.; Sirombo, E. Energy and Water Monitoring for a Large Social Housing Intervention in Northern Italy. *Front. Energy Res.* **2019**, *7*, 126. [CrossRef]

64.   Kamel, E.; Sheikh, S.; Huang, X. Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. *Energy* **2020**, *206*, 118045. [CrossRef]

65.   Liu, L.; Lin, B.; Peng, B. Correlation analysis of building plane and energy consumption of high-rise office building in cold zone of China. *Build. Simul.* **2015**, *8*, 487–498. [CrossRef]

66.   Manfren, M.; Nastasi, B.; Tronchin, L. Linking design and operation phase energy performance analysis through regression-based approaches-A review. *Front. Energy Res.* **2020**, *8*, 288. [CrossRef]

67.   Pan, Y.; Zhang, L. Data-driven estimation of building energy consumption with multi-source heterogeneous data. *Appl. Energy* **2020**, *268*, 114965. [CrossRef]

68.   Piscitelli, M.S.; Brandi, S.; Capozzoli, A.; Xiao, F. A data analytics-based tool for the detection and diagnosis of anomalous daily energy patterns in buildings. *Build. Simul.* **2021**, *14*, 131–147. [CrossRef]

69.   Qian, M.; Yan, D.; Liu, H.; Berardi, U.; Liu, Y. Power consumption and energy efficiency of VRF system based on large scale monitoring virtual sensors. *Build. Simul.* **2020**, *13*, 1145–1156. [CrossRef]

70. Alsaleem, F.; Tesfay, M.K.; Rafaie, M.; Sinkar, K.; Besarla, D.; Arunasalam, P. An IoT Framework for Modeling and Controlling Thermal Comfort in Buildings. *Front. Built Environ.* **2020**, *6*, 87. [CrossRef]

71. Bambara, J.; Athienitis, A.K.; Eicker, U. Residential Densification for Positive Energy Districts. *Front. Sustain. Cities* **2021**, *3*, 3. [CrossRef]

72. Dawood, N.; Dawood, H.; Rodriguez-Trejo, S.; Crilly, M. Visualising urban energy use: The use of LiDAR and remote sensing data in urban energy planning. *Vis. Eng.* **2017**, *5*, 1–13. [CrossRef]

73. Han, S.; Min, S.; Lee, H. Energy efficient VM scheduling for big data processing in cloud computing environments. *J. Ambient Intell. Humaniz. Comput.* **2019**, 1–10. [CrossRef]

74. Hong, T.; Macumber, D.; Li, H.; Fleming, K.; Wang, Z. Generation and representation of synthetic smart meter data. *Build. Simul.* **2020**, *13*, 1205–1220. [CrossRef]

75. Jin, Y.; Yan, D.; Zhang, X.; An, J.; Han, M. A data-driven model predictive control for lighting system based on historical occupancy in an office building: Methodology development. *Build. Simul.* **2020**, *14*, 219–235. [CrossRef]

76. Mosteiro-Romero, M.; Maiullari, D.; Pijpers-van Esch, M.; Schlueter, A. An integrated microclimate-energy demand simulation method for the assessment of urban districts. *Front. Built Environ.* **2020**, *6*, 165. [CrossRef]

77. Nematchoua, M.K.; Orosa, J.A.; Buratti, C.; Obonyo, E.; Rim, D.; Ricciardi, P.; Reiter, S. Comparative analysis of bioclimatic zones, energy consumption, CO2 emission and life cycle cost of residential and commercial buildings located in a tropical region: A case study of the big island of Madagascar. *Energy* **2020**, 117754. [CrossRef]

78. Chen, Z.; Xu, P.; Feng, F.; Qiao, Y.; Luo, W. Data mining algorithm and framework for identifying HVAC control strategies in large commercial buildings. *Build. Simul.* **2021**, *14*, 63–74. [CrossRef]

79. Li, A.; Xiao, F.; Fan, C.; Hu, M. Development of an ANN-based building energy model for information-poor buildings using transfer learning. *Build. Simul.* **2021**, *14*, 89–101. [CrossRef]

80. Su, F.; Huang, J.; Xu, T.; Zhang, C. An evaluation of the effects of various parameter weights on typical meteorological years used for building energy simulation. *Build. Simul.* **2009**, *2*, 19–28. [CrossRef]

81. Van Dronkelaar, C.; Dowson, M.; Spataru, C.; Burman, E.; Mumovic, D. Quantifying the underlying causes of a discrepancy between predicted and measured energy use. *Front. Mech. Eng.* **2019**, *5*, 20. [CrossRef]

82. Ye, W.; Won, D.; Zhang, X. Examining the applicability of empirical models using short-term VOC emissions data from building materials to predict long-term emissions. *Build. Simul.* **2016**, *9*, 701–715. [CrossRef]

83. Ali, U.; Shamsi, M.H.; Bohacek, M.; Hoare, C.; Purcell, K.; Mangina, E.; O'Donnell, J. A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings. *Appl. Energy* **2020**, *267*, 114861. [CrossRef]

84. Belafi, Z.; Hong, T.; Reith, A. Smart building management vs. intuitive human control—Lessons learnt from an office building in Hungary. *Build. Simul.* **2017**, *10*, 811–828. [CrossRef]

85. Moreno, M.V.; Dufour, L.; Skarmeta, A.F.; Jara, A.J.; Genoud, D.; Ladevie, B.; Bezian, J.J. Big data: The key to energy efficiency in smart buildings. *Soft Comput.* **2016**, *20*, 1749–1762. [CrossRef]

86. Niu, F.; O'Neill, Z.; O'Neill, C. Data-driven based estimation of HVAC energy consumption using an improved Fourier series decomposition in buildings. *Build. Simul.* **2018**, *11*, 633–645. [CrossRef]

87. Zhao, H.; Zhao, J.; Shu, T.; Pan, Z. Hybrid-Model-Based Deep Reinforcement Learning for Heating, Ventilation, and Air-Conditioning Control. *Front. Energy Res.* **2021**, *8*, 412. [CrossRef]

88. Arora, S.; Bala, A. A survey: ICT enabled energy efficiency techniques for big data applications. *Clust. Comput.* **2019**, 1–22. [CrossRef]

89. Bibri, S.E.; Krogstie, J. Environmentally data-driven smart sustainable cities: Applied innovative solutions for energy efficiency, pollution reduction, and urban metabolism. *Energy Inform.* **2020**, *3*, 1–59. [CrossRef]

90. Salim, F.D.; Dong, B.; Ouf, M.; Wang, Q.; Pigliautile, I.; Kang, X.; Hong, T.; Wu, W.; Liu, Y.; Rumi, S.K.; et al. Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey. *Build. Environ.* **2020**, *183*, 106964. [CrossRef]

91. Li, Z.; Lin, B.; Zheng, S.; Liu, Y.; Wang, Z.; Dai, J. A review of operational energy consumption calculation method for urban buildings. *Build. Simul.* **2020**, *13*, 739–751. [CrossRef]

92. Yan, B.; Hao, F.; Meng, X. When artificial intelligence meets building energy efficiency, a review focusing on zero energy building. *Artif. Intell. Rev.* **2020**, 1–28. [CrossRef]

93. Ali, U.; Shamsi, M.H.; Bohacek, M.; Purcell, K.; Hoare, C.; Mangina, E.; O'Donnell, J. A data-driven approach for multi-scale GIS-based building energy modeling for analysis, planning and support decision making. *Appl. Energy* **2020**, *279*, 115834. [CrossRef]

94. Jeong, K.; Hong, T.; Kim, J.; Lee, J. A data-driven approach for establishing a CO$_2$ emission benchmark for a multi-family housing complex using data mining techniques. *Renew. Sustain. Energy Rev.* **2020**, 110497. [CrossRef]

95. Risch, S.; Remmen, P.; Müller, D. Influence of data acquisition on the Bayesian calibration of urban building energy models. *Energy Build.* **2020**, *230*, 110512. [CrossRef]

96. Yigit, S. A Machine-Learning-Based Method for Thermal Design Optimization of Residential Buildings in Highly Urbanized Areas of Turkey. *J. Build. Eng.* **2021**, 102225. [CrossRef]