energies

MDPI

Review

Data-Driven Tools for Building Energy Consumption Prediction: A Review

Razak Olu-Ajayi [1,*], Hafiz Alaka [1], Hakeem Owolabi [2], Lukman Akanbi [2] and Sikiru Ganiyu [3]

[1] Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield AL10 9AB, UK
[2] Faculty of Business and Law (FBL), University of the West of England, Bristol BS16 1QY, UK
[3] Big-DEAL Laboratory, Teesside University, Middlesbrough TS1 3BX, UK
* Correspondence: r.olu-ajayi@herts.ac.uk

Abstract: The development of data-driven building energy consumption prediction models has gained more attention in research due to its relevance for energy planning and conservation. However, many studies have conducted the inappropriate application of data-driven tools for energy consumption prediction in the wrong conditions. For example, employing a data-driven tool to develop a model using a small sample size, despite the recognition of the tool for producing good results in large data conditions. This study delivers a review of 63 studies with a precise focus on evaluating the performance of data-driven tools based on certain conditions; i.e., data properties, the type of energy considered, and the type of building explored. This review identifies gaps in research and proposes future directions in the field of data-driven building energy consumption prediction. Based on the studies reviewed, the outcome of the evaluation of the data-driven tools performance shows that Support Vector Machine (SVM) produced better performance than other data-driven tools in the majority of the review studies. SVM, Artificial Neural Network (ANN), and Random Forest (RF) produced better performances in more studies than statistical tools such as Linear Regression (LR) and Autoregressive Integrated Moving Average (ARIMA). However, it is deduced that none of the reviewed tools are predominantly better than the other tools in all conditions. It is clear that data-driven tools have their strengths and weaknesses, and tend to elicit distinctive results in different conditions. Hence, this study provides a proposed guideline for the selection tool based on strengths and weaknesses in different conditions.

Keywords: building energy consumption prediction; data driven tools; energy conservation; energy efficiency; energy prediction; machine learning

Citation: Olu-Ajayi, R.; Alaka, H.; Owolabi, H.; Akanbi, L.; Ganiyu, S. Data-Driven Tools for Building Energy Consumption Prediction: A Review. *Energies* **2023**, *16*, 2574. https://doi.org/10.3390/en16062574

Academic Editor: Eul-Bum Lee

Received: 19 January 2023
Revised: 2 March 2023
Accepted: 6 March 2023
Published: 9 March 2023

1. Introduction

Many countries are experiencing challenges of excess energy use at all levels of industry and economy. Although energy conservation is considered the most plausible solution to alleviate this issue, identifying the most effective approach for energy conservation across all sectors remains a challenge [1]. Since buildings constitute the most prevalent share of over 36% of total energy consumption and carbon emissions around the world [2,3], various methods have been explored and applied for improving energy efficiency in buildings, such as building energy modeling [4], the use of prediction tools [5,6], demand response control [7], among others. Of these methods, the importance of advanced prediction tools for energy planning and conservation has been well noted in the literature [6,8,9]. Building energy consumption prediction can serve as a guide for informed decision-making towards the conservation of energy in buildings.

Despite the stated importance and extensive application of various tools for energy prediction in buildings, there is no consensus on the most suitable tool for building energy prediction. In recent years, various researchers have applied different contemporary tools, namely statistical or artificial intelligence (AI) tools [10–13]. These tools have been very prominent in research, due to their relatively good performance in energy prediction. However, the selection of these tools for energy prediction has been mostly arbitrarily performed or based on popularity; without consideration for strengths and weaknesses [14–17]. The ineffective method of tool selection often leads to the production of poor model performance and time-consuming comparative analysis of tools, rather than the utilization of the right tool for the specific condition. In research, this tool selection method can be due to a lack of adequate evaluation reports of data-driven energy prediction tool performance, centered on pertinent conditions. It is eminent that the tool's performance in several models (such as building energy prediction models) is greatly contingent on the tool and features selected, among other factors [18–21].

Data-driven building energy prediction models (i.e., statistical or artificial intelligence (AI)) have been explored and developed by many researchers, using various features and building characteristics; climate, among others [22–24]. There is a lack of adequate studies that have identified the pertinent features for the development of energy prediction models, which could be one reason for the production of inaccurate results in studies [16,25]. However, the utilization of the wrong data-driven tool for a specific condition also leads to poor performance in studies [14–17]. Despite the prominence of these efforts, the need for more review studies that evaluate existing data-driven tools based on their performance in various conditions (including the features selected for model training, and energy type predicted, among others) is imperative. This is because such reviews will help facilitate the selection of the right tool for a specific condition, and reveal the relevant features required for model development.

To address this gap, this study aims to deliver a structured review of the performance of data-driven tools, such as Artificial Neural Network (ANN) [26], Random Forest (RF) [27,28], and Linear Regression (LR) [13], among others, employed for the prediction of building energy consumption to identify the optimal tool in different conditions. This study focuses on evaluating the performance based on the following conditions: type of building used, type of energy predicted, and type of features used, among others, in the various studies; and delivers a discussion of the findings; a guideline for tool selection; and proposes future research directions. This study is structured as follows. Section 2 delivers an abridged overview of the existing review articles in the field of building energy prediction and pinpoints the gaps. Section 3 explains the methodology utilized in this study. Section 4 discusses the selected studies based on the features selected, the type of energy, and the type of building, among others, and it also delivers the proposed framework for tool selection. Lastly, Section 5 conveys the conclusion and future research directions.

2. Overview of Existing Review Studies

There has been increasing research on exploring the performance of various data-driven tools for predicting energy use in buildings [20,23–25]. However, only a few review studies have systematically analyzed these tools based on relevant situations. Hence, there is no consensus on the best tools for certain conditions [29]. Several data-driven tools possess to the capacity produce optimal performance in different conditions based on their related strengths; for example, Artificial Neural Network (ANN) is recognized for its production of optimal performance following the availability of a large dataset to train the model [30], and similarly, Support Vector Machine (SVM) using a small dataset [31]. However, ANN has been applied in small data conditions, and vice versa. Therefore, it is imperative to comprehend the strengths and weaknesses of data-driven tools under certain conditions.

Several studies have reviewed the performance of data-driven tools in relation to certain conditions, and instances of these studies are concisely stated in Table 1 below.

Table 1. Overview of existing review studies and this study.

References	Year	Tools Reviewed	Energy Type	Focus	Highlights
[32]	2018	ANN SVM Gaussian process Clustering		Algorithms comparison based on performance extensive review on ANN	Generally, it is difficult to conclude which data-driven tools are the best. From the academic literature, it was deduced that most models produced reasonable performance accuracy using large datasets. This paper conducted a comparative review of four data-driven tools, namely Support Vector Machine (SVM), Artificial Neural Network (ANN), clustering, and Gaussian-based regressions based on popularity in the field of building energy prediction.
[33]	2017	Hybrid ANN SVM	Electricity	Algorithms Hybrid algorithms	This paper concluded that AI-based tools are suitable for prediction as they often produce better performance. It was also stated that in comparison to single method of prediction, the hybrid two-prediction methods could be employed for more accurate results. This paper conducted a comparative review of Hybrid ANN, SVM, and Stochastic time series tools, primarily centered on the performance of these tools for predicting electricity energy.
[34]	2017	ARIMA ARMA ANN SVM	Electricity	Algorithms comparison based on performance	This paper performed a comparative review of studies that employed time series tools, namely Autoregressive Integrated Moving Average (ARIMA), Autoregressive Moving Average (ARMA), and AI-based tools, namely ANN and SVM, for electricity energy prediction. It was stated that direct tool comparison across studies is pointless.
[35]	2017	ANN ARIMA SVM Fuzzy time series Nearest Neighbor (kNN)-Hybrid	Electricity	AlgorithmsHybrid algorithms Features	This paper delivers a comprehensive review of certain data-driven tools for energy use prediction. This study also presents an analysis of a "hybrid model", which combines two or more prediction tools. It also examines tools applied with other time series variables, such as outdoor climate, as well as indoor environmental conditions.
[36]	2019	kNN SVM ANN DNN	Electricity Natural gas	Building typologies Data properties	This paper delivers a review of commonly used tools in the field of energy consumption prediction. This focused on data properties, building typologies, and assessment of accuracy.
[37]	2020	ARIMA ANN SVM LR		Types of Features Algorithms performance	This paper provided a review of building energy prediction tools with a focus on feature engineering, performance, and types of features.
This study	2022	ARIMA ANN SVM LR RF	Electricity Natural Gas Overall building energy	Algorithms performance Types of Features Types of energy Temporal granularities	Although the existing review studies provided comprehensive reviews of data-driven tools for energy use prediction, the tools were reviewed with a focus on performance/accuracy, feature typologies, and specific types of energy. There is still a shortfall of comprehensive reviews that capture the strengths and performance of data-driven tools in various conditions, such as energy types (e.g., electricity, natural gas, etc.), feature types (e.g., building envelope, meteorological/weather, etc.) and temporal granularity. Comparatively, this paper conducts a comprehensive review of five data-driven tools, namely ARIMA, ANN, SVM, Linear Regression (LR), and Random Forest (RF), for energy use prediction with a focus on various conditions (i.e., energy types, feature types, and temporal granularity). This type of review is imperative for proffering the knowledge to promote more informed decision-making for the appropriate selection of data-driven tools, rather than the arbitrary selection of tools or selection based on popularity.

3. Materials and Methods

This paper conducted a systematic review of data-driven tools and their performance in various conditions. The arbitrary selection of tools for building energy prediction engendered few tools that produced good performance (i.e., ANN [30], SVM) [31]. However, to reduce the time-consuming comparative analysis and achieve optimum performance, developers need to gain a better comprehension of the selection of the appropriate tool for a specific condition (for example, the type of building considered, data properties, required accuracy, and type of energy considered). Figure 1 presents the framework of the key stages of this research.

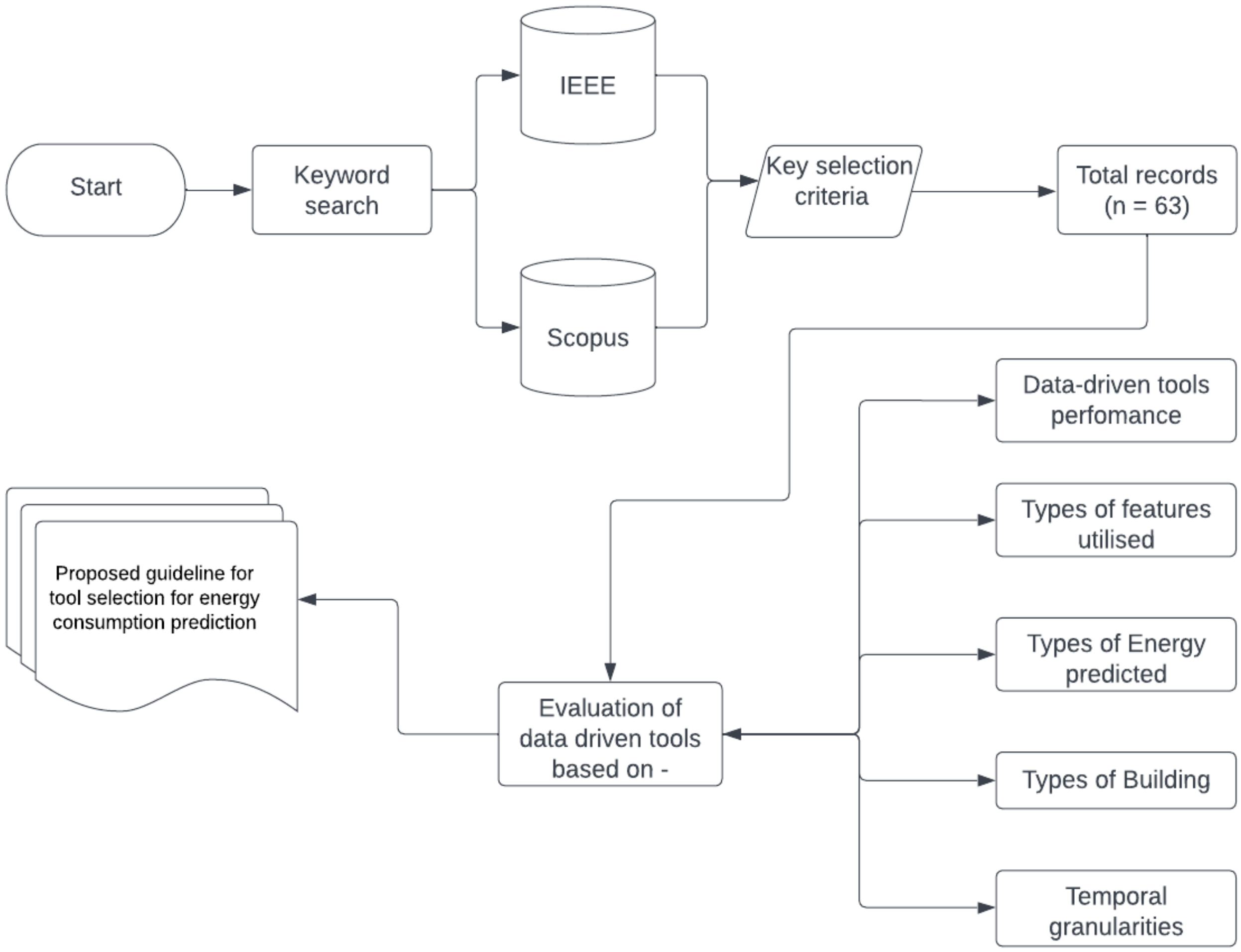

Figure 1. Diagram of the review framework.

In the past decade, several data-driven tools have been applied for energy prediction, to such an extent that it is essentially unviable to comprehensively review all tools in a single study. Hence, five tools were chosen for review, centered on popularity and limited existing reviews. The selected data-driven tools [i.e., Artificial Neural Network (ANN), Support Vector Machine (SVM), Linear Regression (LR), Random Forest (RF), and Autoregressive Integrated Moving Average (ARIMA)] have been noted as being popular and promising in extensive reviews [8,38,39].

Several databases were considered, such as Scopus and the Institute of Electrical and Electronics Engineers (IEEE), Google Scholar, and Web of Science, based on their possession of high-quality articles. However, only Scopus and IEEE were utilized in this study. This is because Google Scholar engendered endless results with wavering precision from the expected result, as also experienced by [40], while the Web of Science was not used due to accessibility limitations. Nevertheless, the utilized databases were measured as sufficient for a systematic review centered on their elevated indexing rate and broad publication coverage [2,41]. Additionally, the two databases consisting of studies from various countries worldwide were utilized to eliminate geographical bias [42].

Firstly, the keywords used for searching the two databases were cautiously selected from existing review studies and a reflection of other energy-related articles [43,44]. Using Scopus and the Institute of Electrical and Electronics Engineers (IEEE), a keyword-based search was conducted. Several review and research articles have used synonyms such as ({"predict", "forecast"}, {"usage", "load"}). The selected keywords were encompassed by utilizing Boolean operators such as "OR" and "AND" to obtain suitable research articles from the two databases (Scopus, IEEE). The keywords utilized were as follows:

"building" AND "energy" OR "electri *" OR "power" OR "load" AND "consumption" OR

"performance" AND "forecast *" OR "Predict *"

The search outcome produced research articles on building energy use prediction. The results showed an increase in research from the year 2017, which is the reason why 2017 was chosen as the start year. The end date for this search was the year 2022. The titles and abstracts of the search results were examined to confirm the suitability of the articles for this study. Regarding inclusion criteria, the study needed to be extensive and have produced satisfactory clarity (i.e., clear explication of methodology and findings). Furthermore, the study needed to have employed one of the selected tools for the development of the building energy prediction model. However, in exceptional cases where the abstract and title were not clear enough to determine their suitability for the study, the full text was examined. Regarding exclusion criteria, only English articles were selected due to constraints in terms of interpretation costs, and research articles that did not utilize or apply the selected tools were eliminated. Additionally, to improve the validity of this study, only journal research articles were chosen because they were considered to be of good quality [42]. After screening the articles, only 51 articles were systematically selected for review. However, to include more studies and avoid bias, the bibliographies of selected articles were explored to identify related articles that utilized at least one of the data-driven tools reviewed. Subsequently, 22 more studies were included, which made up a total of 63 studies reviewed in this study.

4. Results and Discussion

This study conducted a review of the five most applied data-driven tools in model development for predicting energy use in buildings. The five most utilized tools selected included Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), Linear Regression (LR), and Autoregressive Integrated Moving Average (ARIMA). Table 2 shows the performance values of the data-driven tools captured in these studies, including the various energy types, building types, and time granularity explored (indicated with a tick symbol ($\surd$)), while those that were not clearly stated were represented with the (X) symbol.

4.1. Scope of Prediction

The scope of the studies was categorized based on the type of energy load predicted, the building types, and temporal granularity, based on all the different types of predictions conducted in the reviewed studies. Four types of temporal granularities (i.e., yearly, monthly, daily, hourly), three types of energy consumption (i.e., Electricity, Natural Gas, Overall building energy), two types of building (i.e., residential, commercial) and two performance measures (i.e., Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)). The proportion of the data-driven tools based on energy types, temporal granularity, and performance were as follows: 14% of the selected data-driven tools were applied for yearly energy usage prediction, while 6%, 30%, and 50% focused on monthly, daily, and hourly energy usage predictions, respectively. The fraction of the different types of energy load predicted in the selected research articles were 41%, 50%, and 9% for overall building energy (OBE), electricity, and natural gas, respectively, as shown in Figure 2. Numerous data-driven tools were employed in the selected studies, yet the most employed was SVM with 32%, while the other tools accounted for 24%, 16%, 19%, and 9% for ANN, RF, LR, and ARIMA, respectively.

Table 2. Performance of data-driven tools employed in reviewed studies.

S/N	Author and Year	Data Driven Tools					Energy Types			Building Types		Time Granularity			
		ANN	SVM	RF	LR	ARIMA	Natural Gas	Total Electricity	Overall Building Energy	Residential	Commercial	Yearly	Monthly	Daily	Hourly
1	[14]	0.99		0.97	0.98	1.30		√			√			√	
2	[45]				1.10		√			√					√
3	[46]		0.75	0.79	0.80		√			√	√	√			
4	[47]	0.18				0.10			√		√			√	
5	[48]	0.34	0.71		1.36					√		√			
6	[49]		38.70						X						X
7	[50]	29.55	31.36	31.85	51.97				√		√				X
8	[51]	2.42								√					X
9	[15]	1.46	1.57	1.05					√		√				X
10	[52]		16.36		25.75				√		√				X
11	[53]		29.16								√				X
12	[54]	9.52									√				√
13	[55]	8.65	6.24					√			√				X
14	[56]		5.82	6.11				√			√				√
15	[16]	0.06	0.07		0.06	0.09			√	√					√
16	[57]	0.06	0.07		0.06	0.09			√	√				√	
17	[58]				74.26				√	√					X
18	[59]	17.00	24.11					√			√			√	
19	[60]	0.67	0.80		0.68			√		√					√
20	[17]	0.28	0.25	0.29	0.29		√		√		√	√			
21	[29]	24.85							√		√				√
22	[61]		3.13		2.56						√				X
23	[62]			2.78					√						√
24	[63]		0.64							√					√
25	[64]			0.76						√					X
26	[32]	2.82	2.71	2.80						√					X
27	[26]	0.99			0.95						√				X
28	[65]		22.67								√				X
29	[66]		18.10	20.63							√				√
30	[67]	2.70	2.79					√		√					√
31	[68]		0.25	0.20						√					√
32	[69]		2.15	0.97						√					√
33	[70]			1.13						√					X
34	[71]	0.37	0.35	0.41	0.42	0.39		√		√	√			√	
35	[72]					1.58			√		√				X
36	[73]	50.77	64.18							√				√	
37	[74]		0.32								√				√
38	[75]					21.73		√			√			√	
39	[76]	1.69			26.02	1.09		√			√			√	
40	[77]		0.08	0.10					√		√				X
41	[78]		89.49			87.40		√		√					X
42	[79]	0.96							√	√		√			
43	[80]		4.49								√		√		
44	[81]		8.36	6.97	16.72				√		√				√
45	[82]	6.19						√			√			√	

Table 2. *Cont.*

S/N	Author and Year	Data Driven Tools					Energy Types			Building Types		Time Granularity			
		ANN	SVM	RF	LR	ARIMA	Natural Gas	Total Electricity	Overall Building Energy	Residential	Commercial	Yearly	Monthly	Daily	Hourly
46	[83]		0.02						√	√	√				√
47	[84]	11.86							√		√				√
48	[85]		0.69	0.99	0.73				√		√				√
49	[86]	7.84	7.00							√		√			
50	[87]		16.70					√		√			√		
51	[10]			0.50	3.55		√	√		√					X
52	[88]		24.47	34.95	43.90			√			√			√	
53	[89]	27.10	25.80		26.20			√			√				
54	[90]			7.50					√		√	√			
55	[91]				0.06			√		√					
56	[92]							√			√				√
57	[93]		13.68	12.57	17.65			√			√			√	
58	[94]		0.05						√		√			√	
59	[25]	68.31	11.68			4.17		√			√				√
60	[95]	26.88	23.70			26.00		√			√		√	√	
61	[1]	7.04			4.18			√			√				√
62	[96]	64.40	19.87					√		√	√				√
63	[97]	23.22	21.82		30.78			√			√				√

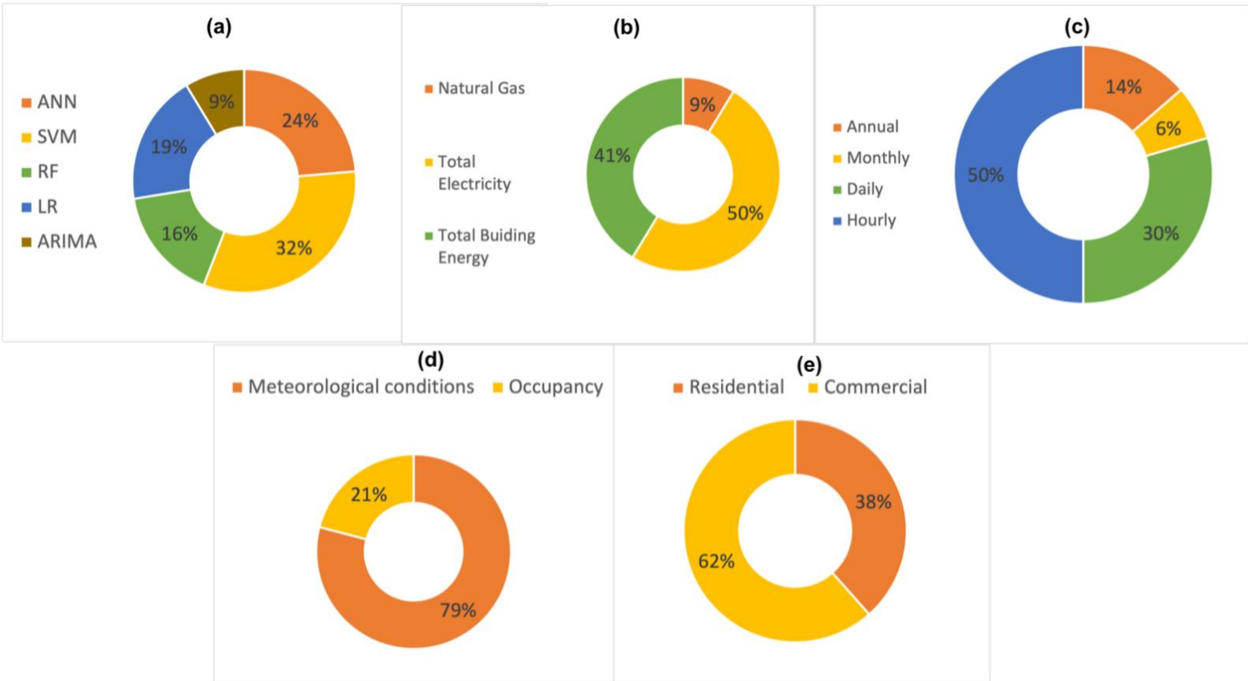

Figure 2. (**a**) Proportion of studies that employed the different data-driven tools (**b**) Proportion of studies that predicted different energy types (**c**) Proportion of studies that predicted different granularity of energy consumption (**d**) Proportion of studies that utilized meteorological and occupancy (**e**) Proportion of studies that employed different types of building.

4.2. Data Properties—Types of Features

The selection of input features is often the first yet most relevant stage in the development of reliable data-driven prediction models [98,99]. Selecting the most significant input feature is imperative in the development of data-driven energy consumption prediction models, as model performance is also predicated on the sets of features or features [11,100]. The relevant input features are often greatly correlated with the target feature(s) or predicted feature(s), with low correlation with each other. However, if features are highly correlated with each other, but have low correlation with the target feature(s), methods such as Principal Component Analysis (PCA) can be applied for dimension reduction [85,101,102]. Researchers often select features based on domain knowledge, the academic literature, or understanding of prediction problems, which sometimes leads to the difference in features applied in studies. Several features were used in the reviewed studies, such as building properties (building wall thickness, wall area, window-wall ratio (WWR), building floor area, building orientation, area of the roof, roof thickness, building height, orientation, area of window glazing, glazing area circulation, mean heat transfer coefficient of the wall, heat transfer coefficient of the roof, solar radiation of external walls, southern/northern/eastern window-wall ratio (WWR), shading coefficient (SC) of the window, SC of the window), occupancy features (building operation schedule, water temperature, and occupants size), and meteorological or weather features (wind speed, direction of wind, dew point temperature, air pressure, solar radiation, dry-bulb temperature, quantity of rainfall, and humidity). Occupancy and meteorological features have received increased attention in recent studies [14,45,46,103] due to their significance in predicting energy consumption in buildings.

Building electricity is mainly influenced by two features, namely occupancy and weather conditions [88,104]. Meteorological features have been more frequently utilized in recent research, having been employed in 79% of the reviewed studies. Only 21% of review studies used occupancy features. The low proportion of studies that utilized occupancy features could be due to the inability to easily obtain these data [22,105]. However, observation and examination of 35 buildings elicited that there is a high correlation between outdoor weather (i.e., temperature, wind speed) and building properties (window, wall) [106]. It is also concluded that the lower the U-values of windows ultimately results in better energy performance of the building [107]. Additionally, the high correlation between the building wall and the outdoor weather of the location of the building is based on the fact that the intensification of wall thickness in cold regions has a significant impact on a building's energy performance [108,109]. Furthermore, it is noted that energy savings through the optimization of wall properties vary in various outdoor weather conditions [109]. For example, in a cold region, [109] indicated that the increase in wall thickness has a sizable impact on the building's heating energy usage, while it displays a relatively minimal impact on the building's cooling energy usage.

Occupant behavior has a considerable effect on the energy consumed in buildings and on prediction, it is considered one of the key reasons for differences or errors between the predicted output and the actual energy consumption values [105]. Although occupant behavior is considered one of the most important features that influence the energy consumed in buildings, it is also considered one of the key reasons for uncertainty in prediction outcomes [11]. However, considering the difficulty in obtaining occupancy data and its relative importance, studies such as [110] have gone further. To obtain occupancy data, [110] attached an infrared thermal sensor to the main entrance to determine the occupancy rate. Nonetheless, this study proves that the occupancy rate could not significantly impact electricity energy consumption.

4.3. Data-Driven Tools

Data-driven tools are required to train a building energy usage prediction model. Prior research in the field of data-driven energy use prediction has employed SVM and ANN, among others. Generally, ANN and SVM were employed for energy use prediction in 24% and 32% of the selected articles, respectively. While ARIMA, LR, and RF were employed in 9%, 19%, and 16%, respectively. In recent years, the most applied tools for developing building energy prediction models have been data-driven tools (mainly statistical and AI-related tools) [57,59]. Based on popularity, ANN is considered the most prevailing for energy use prediction in buildings [111,112]. Aside from its disadvantages, such as high computational costs and deficiencies in terms of transparency [30,33,51], numerous studies have randomly employed the ANN tool due to its acceptance in the field of energy prediction [16,25,60,95,113,114]. ANN is fairly recognized for its production of good outcomes following the availability of large data sizes to train the model [12,30,115]. However, ANN and other fairly common data-driven tools (i.e., RF, LR) have been utilized and compared in various studies using a small data sizes [71,73,116]. More recently, SVM has emerged as one of the most utilized data-driven tools based on its capacity to produce good outcomes regardless of the data size [31,33,117]. However, a drawback of SVM is its large requirements and low computational efficiency [118]. Various comparative analyses of ANN and SVM have been conducted and some studies have concluded that SVM performs better than ANN, while some have concluded otherwise [15,48,50]. The selection of a data-driven tool for energy use prediction or other purposes should not only be based on its strengths and its popularity/acceptance, but also a comprehension of its disadvantages [119].

4.3.1. Performance Evaluation

In the research, after the development of energy prediction models using data-driven tools, the evaluation of these models is often implemented using various measures, such as Mean Absolute Error (MAE), R squared, Root Mean Square Error (RMSE), and Coefficient of Variation (CV), among others. Of these performance measurements, the most utilized measurement was MAE.

1. Mean Absolute Error (MAE) is an evaluating measurement of performance that examines the disparity between the predicted values and the actual values at their re-

spective points in a scatter plot. The score closer to zero represents better performance, while the closer the value is to one indicates a poor performance.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|AE_i - PE_i| \tag{1}$$

2. Root Mean Squared Error (RMSE) is an evaluating measurement of performance employed for calculating the difference between predicted values and actual values. The RMSE score closer to zero represents better performance, while the closer the value is to one indicates a poor performance.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(AE_i - PE_i)^2} \tag{2}$$

Comparative analysis of data-driven tools is mainly implemented to identify the most effective tool based on the prediction performance. In the energy prediction field, overestimation and underestimation can have a detrimental effect on industrial and economic developments [47]. Figure 2 shows a direct comparative analysis of the selected tools in a chart. Some studies utilized MAE, while few studies employed RMSE to evaluate performance. Hence, both performance measures were utilized for this comparison, and this can be undoubtedly measured and compared centered on low error values, representing good model performance. Figure 2 displays the average error values for a set of tools in a direct comparison. The results show that SVM and ARIMA outperformed other tools, such as ANN, LR, and RF.

Figure 3 shows a pairwise comparison of the reviewed tools using average performance; however, this is not enough to deduce an unbiased inference. For a more objective and equitable analysis, Figure 4 also shows the number of studies that concluded that one tool is better than the other, based on different evaluation methods.

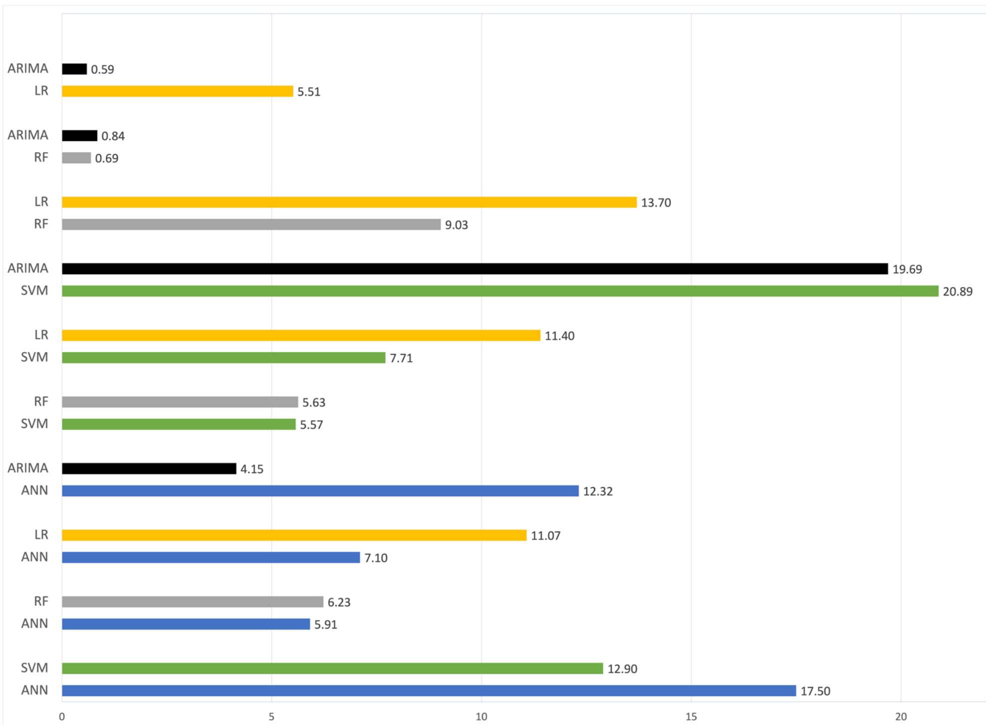

Figure 3. Pair-wise comparison using average performance.

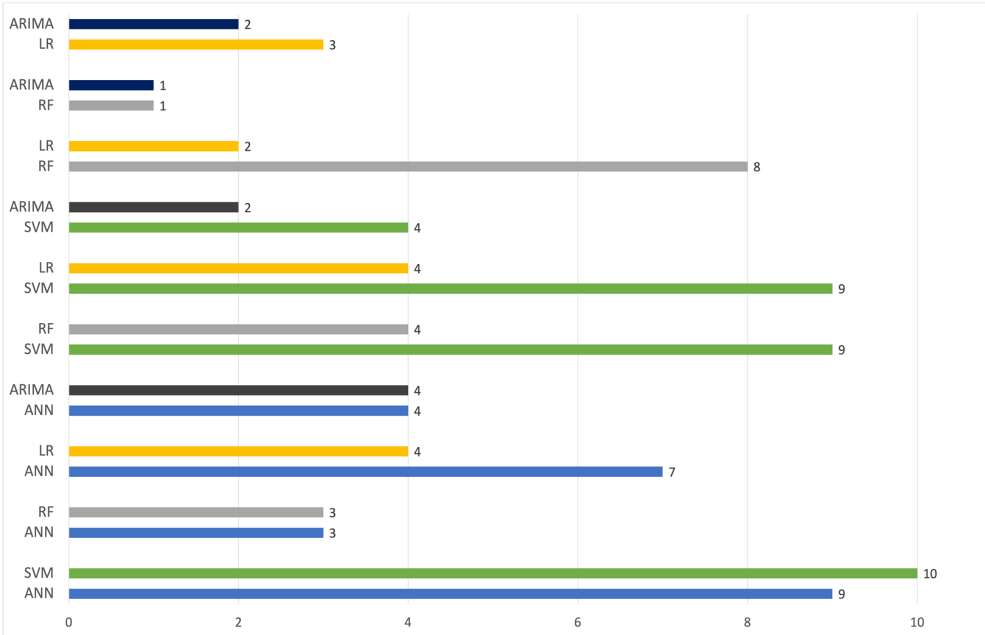

Figure 4. Pair-wise comparison based on the number of studies that noted one tool as being better than the other.

The results from Figures 3 and 4 show that SVM outperformed other tools such as ANN, LR, RF, and ARIMA. Figure 4 shows that AI-related tools such as SVM, ANN, and RF produced better performances in more studies than statistical tools such as LR and ARIMA. However, the good performance of ARIMA in comparison to LR could be due to the capacity of ARIMA to handle temporal dependencies such as weather data [120]. It is noted that meteorological data is employed in 79% of the review studies as it is one of the key factors for energy prediction in buildings [88,104]. Similarly, RF also produced better performance than LR due to its ability to capture nonlinear relationships between predictors [121]. Several factors could justify the reason for the outperformance of SVM, such as data size, data quality, etc. SVM can handle non-linear relationships in the data [122] and produce good performance in small data sizes [31].

4.3.2. Temporal Granularities

There are four major types of building energy consumption prediction that have recently gained more attention: yearly, monthly, daily, and hourly. This is due to the availability of advanced energy consumption meters in buildings, which record energy use at varying intervals [47]. Of all the energy prediction types, hourly energy prediction was the most performed among the selected research articles, constituting a total of 52%. Other energy prediction types were researched in a reasonable portion of the total research articles, such as daily (27%), monthly (6%), and yearly (15%). In research, temporal granularities are separated into classes, namely short term (i.e., daily, hourly) and long term (i.e., monthly, yearly); the short term has been noted as the most prevailing because of their direct relationship with the daily operations of the building [123]. Hence, a low percentage of the selected articles concentrated on long-term (i.e., monthly, yearly) energy consumption predictions. This could also be due to the more pronounced nonlinearity in long-term sample sizes in comparison to short-term sample sizes [124]. However, long-term energy consumption predictions are considered vital to decision-making regarding economic and operational scheduling [114]. In building energy consumption prediction, various data-driven tools have been stated to elicit good outcomes for certain granularities. For example, RF and SVM have been noted for their good outcomes in predicting long-term electricity (heating and cooling load) consumption [48,125]. However, SVM has also shown good performance among other data-driven tools for predicting long-term electricity use [93].

For temporal granularity, the chart visualized in Figure 5 above shows the average performance values (i.e., MAE, RMSE) for predicting the specific granularity of energy consumption in buildings. Figure 5 shows that LR produced the best performance for annual energy prediction, while the AI-related tool also produced a relatively good performance. Considering ARIMA was not employed for annual energy prediction, it cannot be proffered that statistical tools are better at predicting annual energy consumption. Regarding monthly energy use prediction, only the performance of SVM produced a relatively poor performance. In comparison to other tools, ANN produced the best performance for daily energy use prediction and SVM produced the worst performance. Furthermore, RF produced the best performance for hourly energy use prediction and ANN produced the worst performance. To prevent ambiguity, tools employed in only two studies were removed from this comparison, considering that this engendered high average performance values. For example, ARIMA was only employed in two studies for hourly energy use prediction. Additionally, studies with performance values of over 50 were removed. For example, [25] employed ANN for hourly energy prediction, and the MAE was 68.31. Additionally, Figure 5 also shows the number of studies that employed each reviewed tool for different granularities of data. The results show that SVM was the most employed for predicting hourly usage, while ANN was the most employed for daily energy prediction.

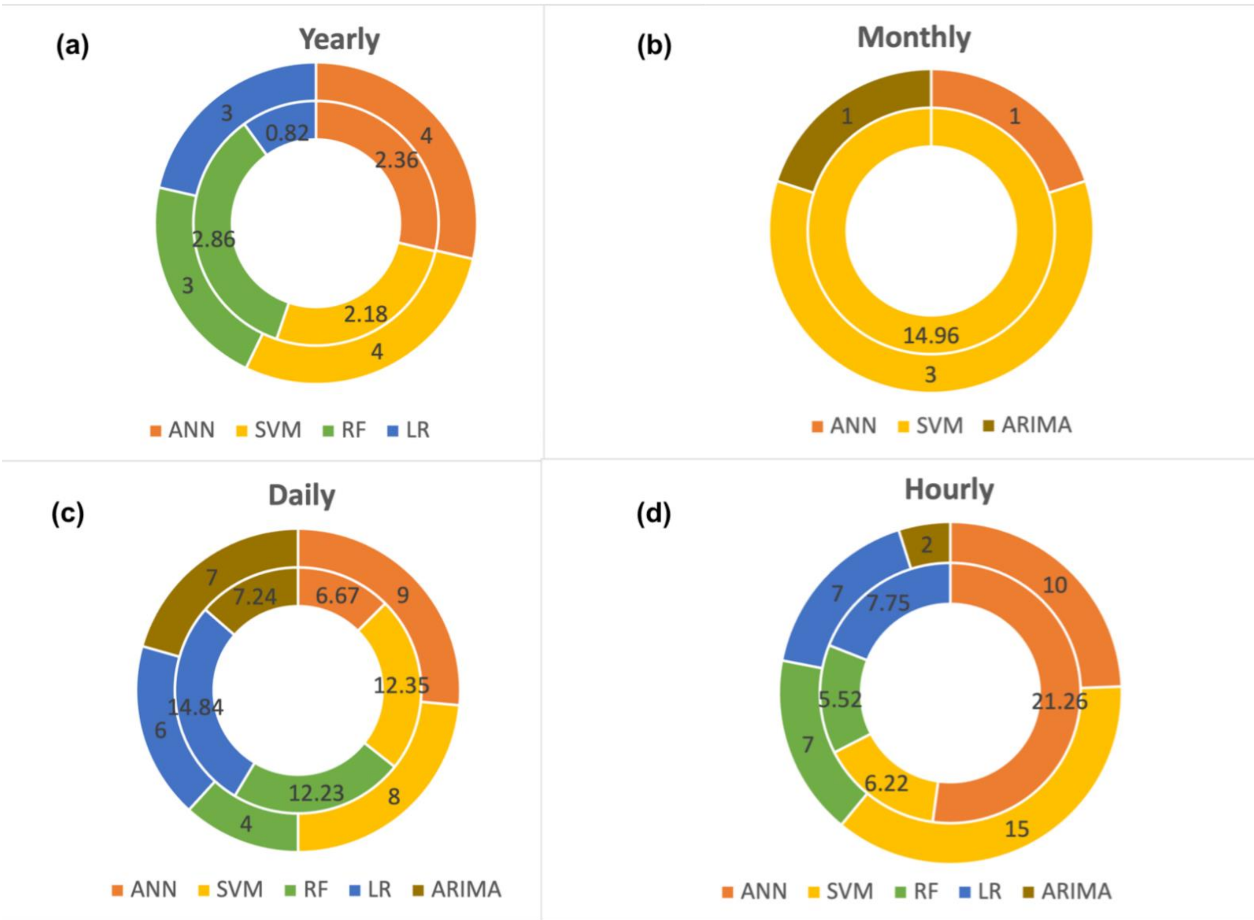

Figure 5. Number of studies and average performance values of each for specific granularity of energy prediction—(**a**) yearly, (**b**) monthly, (**c**) daily, (**d**) hourly.

4.3.3. Building Types

Data of varying sizes have been collected from different building types and these data have been utilized for developing data-driven (statistical or AI tools) energy use prediction models [71,73,83]. In this study, the different buildings types were classified into two sets, namely: residential (i.e., [91,126]), commercial (i.e., hotel buildings [5,25,127], hospital buildings [93,128], educational buildings ([54,59]). The percentage of reviewed studies that utilized different building types were as follows:

A total of 38% of selected research articles explored energy use prediction for residential buildings, which was relatively low compared to the 62% that explored energy use prediction for commercial buildings. The relatively low fraction of articles focused on residential buildings could be due to a shortage of sensor-based data, which is a prerequisite for training the model [129]. Additionally, the difficulty in accessing occupancy data could be another reason for the low concentration of residential buildings. In commercial buildings, researchers have obtained occupancy by attaching data from infrared thermal sensors to the main entrances of buildings to collect occupancy data [110]. Occupancy behavior is measured as the most indeterminate feature in building energy use prediction [22,105]. Nevertheless, the importance of overcoming these setbacks in energy use prediction cannot be overemphasized, as residential electricity constitutes 70% of total electricity in the UK [48]. Furthermore, owing to the elevated degree of total energy use in commercial buildings, there has been a higher number of studies focused on commercial buildings than residential buildings [47]. Commercial buildings consume over 45% more energy than residential buildings in the UK [59].

Based on the reviewed studies, AI tools showed better outcomes than statistical tools. A total of 37% of studies concluded that ANN produced the best performance in comparison to other data-driven models for energy use prediction for both residential and commercial buildings. Following this, SVM was rated as the best prediction model for both residential and commercial buildings in 37% of reviewed studies. ARIMA and LR were concluded as the best models in 31% and 25% of the studies in predicting energy use for both residential and commercial buildings. However, further analysis was conducted using the performance values of all studies. Figure 6 displays the average performance values (i.e., MAE, RMSE) for predicting the energy consumption of two types of buildings (i.e., residential, commercial). Figure 6 shows that RF outperforms the other models for the prediction of energy use for residential buildings, while ARIMA elicited better performance than other models for commercial buildings.

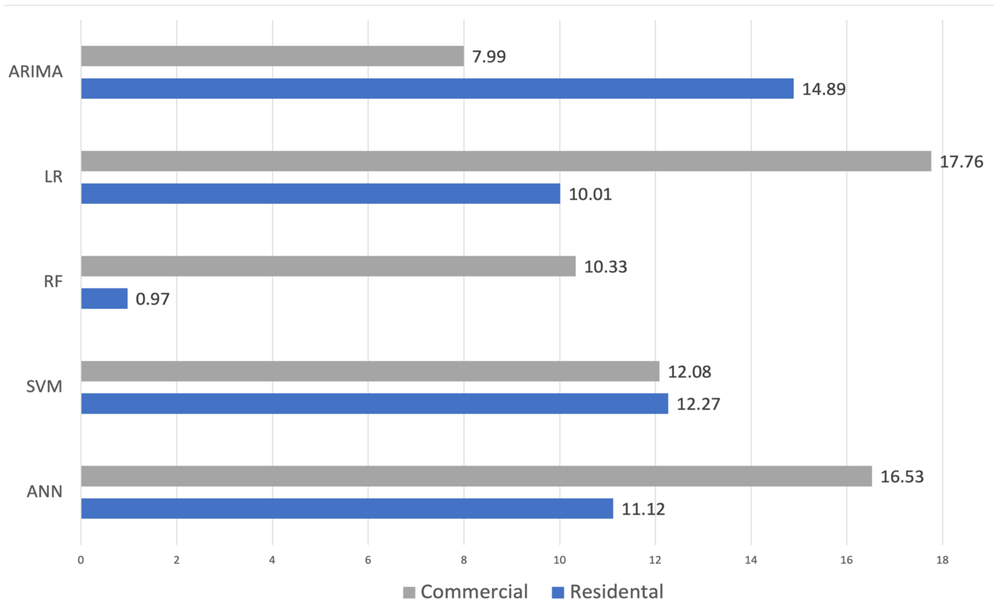

Figure 6. Average performance values of energy prediction for predicting different buildings.

4.3.4. Energy Consumption Types

This study reviewed research articles that developed data-driven models and conducted a comparative analysis of these models for predicting three different energy types:

electricity, natural gas, and overall building energy. Overall building energy denotes the combination of all the energy types consumed in a building. The fraction of studies that explored the different energy types were as follows: overall building energy (50%), electricity (38%), and natural gas (12%). The high percentage of studies focusing on overall building energy could be due to a combination of other types of energy. Following this, electricity has received a good percentage of attention from review studies based on its noted consumption of over 65% of total electricity usage in China, as well as its projection of constituting over one-fifth of electricity consumption in buildings worldwide by 2050 [130]. The low focus on natural gas could be because of the adoption of renewable energy sources for operational buildings and the eradication of non-renewable sources (i.e., natural gas) [1].

4.4. Proposed Framework

Based on the results and deductions of this systematic review, Figure 7 presents a streamlined framework to support or guide the selection of tools by energy prediction researchers and model developers. Essentially, all of the tools reviewed in this study could make predictions. However, some tools are better than others in specific situations. For example, if a low error rate is the target goal for model developers and only a small sample size is available for model training, SVM will be a suitable choice for such a situation.

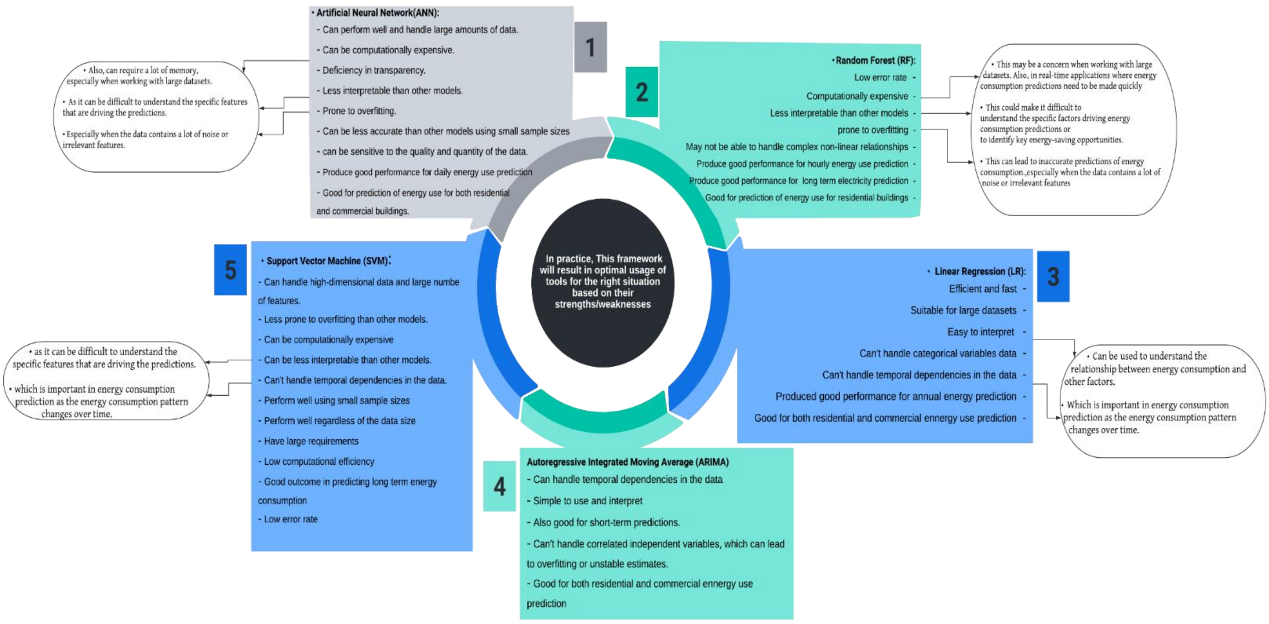

Figure 7. Proposed guideline for tool selection for energy consumption prediction.

In Figure 7, the singular rectangles around the circle show the strengths and weaknesses of each data-driven tool, and their ability to produce good predictions in certain conditions. For example, ANNs are noted to be computationally expensive; however, they produce good predictions for the energy use of residential and commercial buildings.

5. Conclusions

In this review, it is apparent that the application of data-driven tools for building energy use prediction has drawn more research attention. Various models perform well for various purposes, in different conditions, and are trained on various feature sizes and data sizes. However, many tools are applied in the wrong conditions without much consideration of their strengths in dissimilar circumstances or conditions. This paper delivers a systematic literature review of commonly used tools in the field of energy consumption prediction, based on certain relevant conditions. The development of an energy use prediction model requires step-by-step consideration of all the studied aspects in this study. This study delivers a guideline for model developers to facilitate informed decision-making during model development in diverse conditions, and therefore, eradicate the development of models based on popularity.

Results show that AI-related tools such as SVM, ANN, and RF produced better performances in more studies than statistical tools such as LR and ARIMA. However, LR produced optimal performances in specific situations, such as annual energy prediction. ARIMA elicited good performances for energy prediction of commercial buildings, while RF produced good performance for residential buildings.

Based on overall performance, regardless of the different criteria, SVM produced very good results. This could be due to several reasons—SVMs can handle high-dimensional data, which is important in energy consumption prediction as the energy consumption pattern changes over time. Furthermore, it is less prone to overfitting than other tools and it performs well with small data sizes. Although, SVM produced good performances in the majority of the reviewed studies, in general, the finding indicated that no singular data-driven tool is fundamentally better than all other tools in all conditions.

This study has shown that specific areas require further attention: yearly and monthly energy consumption predictions, and natural gas energy predictions. The low focus of attention in these areas could be due to insufficient data; however, this is slowly changing as several buildings have been equipped with smart meters. Therefore, this will elicit more research in these areas. Despite the results, which convey that no singular tool performs best in all conditions, future research should consider the review of hybrid tools performance in several conditions. Furthermore, future research should explore ANN, LR, and RF for yearly, daily, and hourly energy use predictions, as they appear to yield good results in many conditions or circumstances. Future research should also consider focusing on studies that employ deep learning methods in various situations, and developing SVM and other hybrid models for predicting building energy consumption.

Author Contributions: Conceptualization, R.O.-A. and H.A.; methodology, formal analysis, R.O.-A.; investigation, R.O.-A., data curation—R.O.-A. and H.A.; writing—R.O.-A., writing—review and editing, R.O.-A. and H.A. and visualization, R.O.-A.; supervision, H.A., H.O., S.G. and L.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fathi, S.; Srinivasan, R.; Fenner, A.; Fathi, S. Machine learning applications in urban building energy performance forecasting: A systematic review. *Renew. Sustain. Energy Rev.* **2020**, *133*, 110287. [CrossRef]
2. Debrah, C.; Chan, A.P.; Darko, A. Artificial intelligence in green building. *Autom. Constr.* **2022**, *137*, 104192. [CrossRef]
3. Alimohammadisagvand, B.; Jokisalo, J.; Sirén, K. The potential of predictive control in minimizing the electricity cost in a heat-pump heated residential house. In Proceedings of the 3rd IBPSA-England Conference BSO, Newcastle, UK, 19 September 2016.
4. Li, X.; Wen, J. Review of building energy modeling for control and operation. *Renew. Sustain. Energy Rev.* **2014**, *37*, 517–537. [CrossRef]
5. Ahmad, M.W.; Mouraud, A.; Rezgui, Y.; Mourshed, M. Deep Highway Networks and Tree-Based Ensemble for Predicting Short-Term Building Energy Consumption. *Energies* **2018**, *11*, 3408. [CrossRef]
6. Woo, J.; Fenner, A.E.; Asutosh, A.; Kim, D.; Razkenari, M.A.; Kibert, C.J. A review of the state-of-the-art machine learning algorithms for building energy consumption prediction. In *IIE Annual Conference. Proceedings*; Institute of Industrial and Systems Engineers (IISE): Peachtree Corners, GA, USA, 2018; pp. 2169–2174.
7. Alimohammadisagvand, B.; Alam, S.; Ali, M.; Degefa, M.; Jokisalo, J.; Sirén, K. Influence of energy demand response actions on thermal comfort and energy cost in electrically heated residential houses. *Indoor Built Environ.* **2017**, *26*, 298–316. [CrossRef]
8. Ahmad, A.S.; Hassan, M.Y.; Abdullah, M.P.; Rahman, H.A.; Hussin, F.; Abdullah, H.; Saidur, R. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew. Sustain. Energy Rev.* **2014**, *33*, 102–109. [CrossRef]
9. Zhang, L.; Wen, J. A systematic feature selection procedure for short-term data-driven building energy forecasting model development. *Energy Build.* **2019**, *183*, 428–442. [CrossRef]
10. Gassar, A.A.A.; Yun, G.Y.; Kim, S. Data-driven approach to prediction of residential energy consumption at urban scales in London. *Energy* **2019**, *187*, 115973. [CrossRef]

11. Amasyali, K.; El-Gohary, N. Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. *Renew. Sustain. Energy Rev.* **2021**, *142*, 110714. [CrossRef]
12. Bourhnane, S.; Abid, M.R.; Lghoul, R.; Zine-Dine, K.; Elkamoun, N.; Benhaddou, D. Machine learning for energy consumption prediction and scheduling in smart buildings. *SN Appl. Sci.* **2020**, *2*, 297. [CrossRef]
13. Ciulla, G.; D'Amico, A. Building energy performance forecasting: A multiple linear regression approach. *Appl. Energy* **2019**, *253*, 113500. [CrossRef]
14. Divina, F.; Gilson, A.; Goméz-Vela, F.; García Torres, M.; Torres, J.F. Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. *Energies* **2018**, *11*, 949. [CrossRef]
15. Feng, C.; Zhang, J. Assessment of aggregation strategies for machine-learning based short-term load forecasting. *Electr. Power Syst. Res.* **2020**, *184*, 106304. [CrossRef]
16. Izidio, D.; Neto, P.D.M.; Barbosa, L.; de Oliveira, J.; Marinho, M.; Rissi, G. Evolutionary Hybrid System for Energy Consumption Forecasting for Smart Meters. *Energies* **2021**, *14*, 1794. [CrossRef]
17. Robinson, C.; Dilkina, B.; Hubbs, J.; Zhang, W.; Guhathakurta, S.; Brown, M.A.; Pendyala, R.M. Machine learning approaches for estimating commercial building energy consumption. *Appl. Energy* **2017**, *208*, 889–904. [CrossRef]
18. Goyal, K.; Tiwari, N.; Sonekar, J. An Anatomization of Data Classification Based on Machine Learning Techniques. *IJRAR-Int. J. Res. Anal. Rev. (IJRAR)* **2020**, *7*, 713–716.
19. Kabir, M.A. Vehicle Speed Prediction based on Road Status using Machine Learning. *Adv. Res. Energy Eng.* **2020**, *2*, 5.
20. Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Sunmola, F.; Ajayi, S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *J. Build. Eng.* **2022**, *45*, 103406. [CrossRef]
21. Runge, J.; Zmeureanu, R. Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. *Energies* **2019**, *12*, 3254. [CrossRef]
22. Khan, A.-N.; Iqbal, N.; Rizwan, A.; Ahmad, R.; Kim, D.-H. An Ensemble Energy Consumption Forecasting Model Based on Spatial-Temporal Clustering Analysis in Residential Buildings. *Energies* **2021**, *14*, 3020. [CrossRef]
23. Kim, D.D.; Suh, H.S. Heating and cooling energy consumption prediction model for high-rise apartment buildings considering design parameters. *Energy Sustain. Dev.* **2021**, *61*, 1–14. [CrossRef]
24. Mocanu, E.; Nguyen, P.H.; Gibescu, M.; Kling, W.L. Deep learning for estimating building energy consumption. *Sustain. Energy Grids Netw.* **2016**, *6*, 91–99. [CrossRef]
25. Shan, S.; Cao, B.; Wu, Z. Forecasting the Short-Term Electricity Consumption of Building Using a Novel Ensemble Model. *IEEE Access* **2019**, *7*, 88093–88106. [CrossRef]
26. Pino-Mejías, R.; Pérez-Fargallo, A.; Rubio-Bellido, C.; Pulido-Arcas, J.A. Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO_2 emissions. *Energy* **2017**, *118*, 24–36. [CrossRef]
27. Sulaimon, I.A.; Alaka, H.; Olu-Ajayi, R.; Ahmad, M.; Ajayi, S.; Hye, A. Effect of traffic data set on various machine-learning algorithms when forecasting air quality. *J. Eng. Des. Technol.* **2022**, *ahead-of-print*. [CrossRef]
28. Wusu, G.E.; Alaka, H.; Yusuf, W.; Mporas, I.; Toriola-Coker, L.; Oseghale, R. A machine learning approach for predicting critical factors determining adoption of offsite construction in Nigeria. *Smart Sustain. Built Environ.* **2022**, *ahead-of-print*. [CrossRef]
29. Amasyali, K.; El-Gohary, N. Deep Learning for Building Energy Consumption Prediction. In Proceedings of the 6th CSCE-CRC International Construction Specialty Conference, Vancouver, BC, Canada, 31 May–3 June 2017.
30. Alaka, H.A.; Oyedele, L.O.; Owolabi, H.A.; Kumar, V.; Ajayi, S.O.; Akinade, O.O.; Bilal, M. Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Syst. Appl.* **2018**, *94*, 164–184. [CrossRef]
31. Aversa, P.; Donatelli, A.; Piccoli, G.; Luprano, V.A.M. Improved Thermal Transmittance Measurement with HFM Technique on Building Envelopes in the Mediterranean Area. *Sel. Sci. Pap. J. Civ. Eng.* **2016**, *11*, 39–52. [CrossRef]
32. Seyedzadeh, S.; Rahimian, F.; Glesk, I.; Roper, M. Machine learning for estimation of building energy consumption and performance: A review. *Vis. Eng.* **2018**, *6*, 5. [CrossRef]
33. Daut, M.A.M.; Hassan, M.Y.; Abdullah, H.; Rahman, H.A.; Abdullah, P.; Hussin, F. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. *Renew. Sustain. Energy Rev.* **2017**, *70*, 1108–1118. [CrossRef]
34. Kuster, C.; Rezgui, Y.; Mourshed, M. Electrical load forecasting models: A critical systematic review. *Sustain. Cities Soc.* **2017**, *35*, 257–270. [CrossRef]
35. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]
36. Bourdeau, M.; Zhai, X.Q.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [CrossRef]
37. Sun, Y.; Haghighat, F.; Fung, B.C. A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy Build.* **2020**, *221*, 110022. [CrossRef]
38. Zhang, L.; Wen, J.; Li, Y.; Chen, J.; Ye, Y.; Fu, Y.; Livingood, W. A review of machine learning in building load prediction. *Appl. Energy* **2021**, *285*, 116452. [CrossRef]
39. Zhao, H.-X.; Magoulès, F. A review on the prediction of building energy consumption. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3586–3592. [CrossRef]

40. Falagas, M.E.; Pitsouni, E.I.; Malietzis, G.; Pappas, G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB J.* **2008**, *22*, 338–342. [CrossRef]
41. Diirr, B.; Santos, G. Interorganizational Information Systems: Systematic Literature Mapping Protocol. *RelaTe-DIA* **2019**, *12*, 4–5.
42. Schlosser, R.W. Appraising the quality of systematic reviews. *Focus* **2007**, *17*, 1–8.
43. Chiradeja, P.; Ngaopitakkul, A. Energy and Economic Analysis of Tropical Building Envelope Material in Compliance with Thailand's Building Energy Code. *Sustainability* **2019**, *11*, 6872. [CrossRef]
44. Rouleau, J.; Gosselin, L.; Blanchet, P. Understanding energy consumption in high-performance social housing buildings: A case study from Canada. *Energy* **2018**, *145*, 677–690. [CrossRef]
45. Hribar, R.; Potočnik, P.; Šilc, J.; Papa, G. A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy* **2019**, *167*, 511–522. [CrossRef]
46. Kontokosta, C.E.; Tull, C. A data-driven predictive model of city-scale energy use in buildings. *Appl. Energy* **2017**, *197*, 303–317. [CrossRef]
47. Somu, N.; MR, G.R.; Ramamritham, K. A deep learning framework for building energy consumption forecast. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110591. [CrossRef]
48. Li, X.; Yao, R. A machine-learning-based approach to predict residential annual space heating and cooling loads considering occupant behaviour. *Energy* **2020**, *212*, 118676. [CrossRef]
49. Zhang, G.; Tian, C.; Li, C.; Zhang, J.J.; Zuo, W. Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature. *Energy* **2020**, *201*, 117531. [CrossRef]
50. Chammas, M.; Makhoul, A.; Demerjian, J. An efficient data model for energy prediction using wireless sensors. *Comput. Electr. Eng.* **2019**, *76*, 249–257. [CrossRef]
51. Sadeghi, A.; Sinaki, R.Y.; Ii, W.A.Y.; Weckman, G.R. An Intelligent Model to Predict Energy Performances of Residential Buildings Based on Deep Neural Networks. *Energies* **2020**, *13*, 571. [CrossRef]
52. Li, C.; Ding, Z.; Zhao, D.; Yi, J.; Zhang, G. Building Energy Consumption Prediction: An Extreme Deep Learning Approach. *Energies* **2017**, *10*, 1525. [CrossRef]
53. Li, C.; Ding, Z.; Yi, J.; Lv, Y.; Zhang, G. Deep Belief Network Based Hybrid Model for Building Energy Consumption Prediction. *Energies* **2018**, *11*, 242. [CrossRef]
54. Jang, J.; Lee, J.; Son, E.; Park, K.; Kim, G.; Lee, J.H.; Leigh, S.-B. Development of an Improved Model to Predict Building Thermal Energy Consumption by Utilizing Feature Selection. *Energies* **2019**, *12*, 4187. [CrossRef]
55. Shapi, M.K.M.; Ramli, N.A.; Awalin, L.J. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Dev. Built Environ.* **2021**, *5*, 100037. [CrossRef]
56. Pinto, T.; Praça, I.; Vale, Z.; Silva, J. Ensemble learning for electricity consumption forecasting in office buildings. *Neurocomputing* **2021**, *423*, 747–755. [CrossRef]
57. Chou, J.-S.; Tran, D.-S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy* **2018**, *165*, 709–726. [CrossRef]
58. Syed, D.; Abu-Rub, H.; Ghrayeb, A.; Refaat, S.S. Household-Level Energy Forecasting in Smart Buildings Using a Novel Hybrid Deep Learning Model. *IEEE Access* **2021**, *9*, 33498–33511. [CrossRef]
59. Amber, K.; Ahmad, R.; Aslam, M.; Kousar, A.; Usman, M.; Khan, M. Intelligent techniques for forecasting electricity consumption of buildings. *Energy* **2018**, *157*, 886–893. [CrossRef]
60. Koukaras, P.; Bezas, N.; Gkaidatzis, P.; Ioannidis, D.; Tzovaras, D.; Tjortjis, C. Introducing a novel approach in one-step ahead energy load forecasting. *Sustain. Comput. Inform. Syst.* **2021**, *32*, 100616. [CrossRef]
61. Guo, Y.; Wang, J.; Chen, H.; Li, G.; Liu, J.; Xu, C.; Huang, R.; Huang, Y. Machine learning-based thermal response time ahead energy demand prediction for building heating systems. *Appl. Energy* **2018**, *221*, 16–27. [CrossRef]
62. Pham, A.-D.; Ngo, N.-T.; Truong, T.T.H.; Huynh, N.-T.; Truong, N.-S. Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *J. Clean. Prod.* **2020**, *260*, 121082. [CrossRef]
63. Feng, Y.; Duan, Q.; Chen, X.; Yakkali, S.S.; Wang, J. Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods. *Appl. Energy* **2021**, *291*, 116814. [CrossRef]
64. Mohammed, A.; Asteris, P.; Koopialipoor, M.; Alexakis, D.; Lemonis, M.; Armaghani, D. Stacking Ensemble Tree Models to Predict Energy Performance in Residential Buildings. *Sustainability* **2021**, *13*, 8298. [CrossRef]
65. Fu, G. Deep belief network based ensemble approach for cooling load forecasting of air-conditioning system. *Energy* **2018**, *148*, 269–282. [CrossRef]
66. Wang, R.; Lu, S.; Feng, W. A novel improved model for building energy consumption prediction based on model integration. *Appl. Energy* **2020**, *262*, 114561. [CrossRef]
67. Ullah, F.U.M.; Ullah, A.; Haq, I.U.; Rho, S.; Baik, S.W. Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks. *IEEE Access* **2020**, *8*, 123369–123380. [CrossRef]
68. Wang, R.; Lu, S.; Li, Q. Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. *Sustain. Cities Soc.* **2019**, *49*, 101623. [CrossRef]
69. Irfan, M.; Ramlie, F.; Widianto Lestandy, M.; Faruq, A. Prediction of Residential Building Energy Efficiency Performance using Deep Neural Network. *IAENG Int. J. Comput. Sci.* **2021**, *48*, 731–737.

70. Hosseini, S.; Fard, R.H. Machine Learning Algorithms for Predicting Electricity Consumption of Buildings. *Wirel. Pers. Commun.* **2021**, *121*, 3329–3341. [CrossRef]
71. Groß, A.; Lenders, A.; Schwenker, F.; Braun, D.A.; Fischer, D. Comparison of short-term electrical load forecasting methods for different building types. *Energy Inform.* **2021**, *4*, 13. [CrossRef]
72. Goudarzi, S.; Anisi, M.H.; Soleymani, S.A.; Ayob, M.; Zeadally, S. An IoT-Based Prediction Technique for Efficient Energy Consumption in Buildings. *IEEE Trans. Green Commun. Netw.* **2021**, *5*, 2076–2088. [CrossRef]
73. Gao, X.; Qi, C.; Xue, G.; Song, J.; Zhang, Y.; Yu, S.-A. Forecasting the Heat Load of Residential Buildings with Heat Metering Based on CEEMDAN-SVR. *Energies* **2020**, *13*, 6079. [CrossRef]
74. Xie, Y.; Hu, P.; Zhu, N.; Lei, F.; Xing, L.; Xu, L.; Sun, Q. A hybrid short-term load forecasting model and its application in ground source heat pump with cooling storage system. *Renew. Energy* **2020**, *161*, 1244–1259. [CrossRef]
75. Lin, X.; Yu, H.; Wang, M.; Li, C.; Wang, Z.; Tang, Y. Electricity Consumption Forecast of High-Rise Office Buildings Based on the Long Short-Term Memory Method. *Energies* **2021**, *14*, 4785. [CrossRef]
76. Alduailij, M.A.; Petri, I.; Rana, O.; Aldawood, A.S. Forecasting peak energy demand for smart buildings. *J. Supercomput.* **2021**, *77*, 6356–6380. [CrossRef]
77. Guo, N.; Gui, W.; Chen, W.; Tian, X.; Qiu, W.; Tian, Z.; Zhang, X. Using improved support vector regression to predict the transmitted energy consumption data by distributed wireless sensor network. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 1–16. [CrossRef]
78. Nie, P.; Roccotelli, M.; Fanti, M.P.; Ming, Z.; Li, Z. Prediction of home energy consumption based on gradient boosting regression tree. *Energy Rep.* **2021**, *7*, 1246–1255. [CrossRef]
79. Elbeltagi, E.; Wefki, H. Predicting energy consumption for residential buildings using ANN through parametric modeling. *Energy Rep.* **2021**, *7*, 2534–2545. [CrossRef]
80. Chen, Y.; Zhang, F.; Berardi, U. Day-ahead prediction of hourly subentry energy consumption in the building sector using pattern recognition algorithms. *Energy* **2020**, *211*, 118530. [CrossRef]
81. Cho, S.; Lee, J.; Baek, J.; Kim, G.-S.; Leigh, S.-B. Investigating Primary Factors Affecting Electricity Consumption in Non-Residential Buildings Using a Data-Driven Approach. *Energies* **2019**, *12*, 4046. [CrossRef]
82. Jeong, D.; Park, C.; Ko, Y.M. Short-term electric load forecasting for buildings using logistic mixture vector autoregressive model with curve registration. *Appl. Energy* **2021**, *282*, 116249. [CrossRef]
83. Culaba, A.B.; Del Rosario, A.J.R.; Ubando, A.T.; Chang, J. Machine learning—Based energy consumption clustering and forecasting for mixed—Use buildings. *Int. J. Energy Res.* **2020**, *44*, 9659–9673. [CrossRef]
84. Ding, Y.; Su, H.; Kong, X.; Zhang, Z. Ultra-Short-Term Building Cooling Load Prediction Model Based on Feature Set Construction and Ensemble Machine Learning. *IEEE Access* **2020**, *8*, 178733–178745. [CrossRef]
85. Parhizkar, T.; Rafieipour, E.; Parhizkar, A. Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction. *J. Clean. Prod.* **2021**, *279*, 123866. [CrossRef]
86. Szul, T.; Nęcka, K.; Mathia, T.G. Neural Methods Comparison for Prediction of Heating Energy Based on Few Hundreds Enhanced Buildings in Four Season's Climate. *Energies* **2020**, *13*, 5453. [CrossRef]
87. Shen, M.; Lu, Y.; Wei, K.H.; Cui, Q. Prediction of household electricity consumption and effectiveness of concerted intervention strategies based on occupant behaviour and personality traits. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109839. [CrossRef]
88. Fan, C.; Xiao, F.; Zhao, Y. A short-term building cooling load prediction method using deep learning algorithms. *Appl. Energy* **2017**, *195*, 222–233. [CrossRef]
89. Sha, H.; Xu, P.; Hu, C.; Li, Z.; Chen, Y.; Chen, Z. A simplified HVAC energy prediction method based on degree-day. *Sustain. Cities Soc.* **2019**, *51*, 101698. [CrossRef]
90. Pan, Y.; Zhang, L. Data-driven estimation of building energy consumption with multi-source heterogeneous data. *Appl. Energy* **2020**, *268*, 114965. [CrossRef]
91. 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]
92. Shao, X.; Pu, C.; Zhang, Y.; Kim, C.S. Domain Fusion CNN-LSTM for Short-Term Power Consumption Forecasting. *IEEE Access* **2020**, *8*, 188352–188362. [CrossRef]
93. Cao, L.; Li, Y.; Zhang, J.; Jiang, Y.; Han, Y.; Wei, J. Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Rep.* **2020**, *6*, 2751–2767. [CrossRef]
94. Liu, Y.; Chen, H.; Zhang, L.; Wu, X.; Wang, X.-J. Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *J. Clean. Prod.* **2020**, *272*, 122542. [CrossRef]
95. Hwang, J.; Suh, D.; Otto, M.-O. Forecasting Electricity Consumption in Commercial Buildings Using a Machine Learning Approach. *Energies* **2020**, *13*, 5885. [CrossRef]
96. Eseye, A.T.; Lehtonen, M. Short-Term Forecasting of Heat Demand of Buildings for Efficient and Optimal Energy Management Based on Integrated Machine Learning Models. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7743–7755. [CrossRef]
97. Fan, C.; Sun, Y.; Zhao, Y.; Song, M.; Wang, J. Deep learning-based feature engineering methods for improved building energy prediction. *Appl. Energy* **2019**, *240*, 35–45. [CrossRef]
98. Balogun, H.; Alaka, H.; Egwim, C.N.; Ajayi, S. Systematic review of drivers influencing building deconstructability: Towards a construct-based conceptual framework. *Waste Manag. Res.* **2022**, 734242X221124078. [CrossRef]

99. Egwim, C.N.; Alaka, H.; Demir, E.; Balogun, H.; Ajayi, S. Systematic review of critical drivers for delay risk prediction: Towards a conceptual framework for BIM-based construction projects. *Front. Eng. Built Environ.* **2022**, *3*, 16–31. [CrossRef]
100. Egwim, C.N.; Alaka, H.; Toriola-Coker, L.O.; Balogun, H.; Sunmola, F. Applied artificial intelligence for predicting construction projects delay. *Mach. Learn. Appl.* **2021**, *6*, 100166. [CrossRef]
101. Robert, C. Machine Learning, a Probabilistic Perspective. *CHANCE* **2014**, *27*, 62–63. [CrossRef]
102. Balogun, H.; Alaka, H.; Egwim, C.N. Boruta-grid-search least square support vector machine for NO_2 pollution prediction using big data analytics and IoT emission sensors. *Appl. Comput. Inform.* **2021**. *ahead-of-print*. [CrossRef]
103. Somu, N.; MR, G.R.; Ramamritham, K. A hybrid model for building energy consumption forecasting using long short term memory networks. *Appl. Energy* **2020**, *261*, 114131. [CrossRef]
104. Zhang, C.; Cao, L.; Romagnoli, A. On the feature engineering of building energy data mining. *Sustain. Cities Soc.* **2018**, *39*, 508–518. [CrossRef]
105. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]
106. Ren, J.; Zhou, X.; An, J.; Yan, D.; Shi, X.; Jin, X.; Zheng, S. Comparative analysis of window operating behavior in three different open-plan offices in Nanjing. *Energy Built Environ.* **2021**, *2*, 175–187. [CrossRef]
107. Ihara, T.; Gustavsen, A.; Jelle, B.P. Effect of facade components on energy efficiency in office buildings. *Appl. Energy* **2015**, *158*, 422–432. [CrossRef]
108. Huang, Y.; Niu, J.; Chung, T.-M. Study on performance of energy-efficient retrofitting measures on commercial building external walls in cooling-dominant cities. *Appl. Energy* **2013**, *103*, 97–108. [CrossRef]
109. Zhang, L.; Jin, L.; Wang, Z.; Liu, X. Effects of wall configuration on building energy performance subject to different climatic zones of China. *Appl. Energy* **2017**, *185*, 1565–1573. [CrossRef]
110. Kim, M.K.; Kim, Y.-S.; Srebric, J. Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. *Sustain. Cities Soc.* **2020**, *62*, 102385. [CrossRef]
111. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* **2017**, *147*, 77–89. [CrossRef]
112. Kim, J.-H.; Seong, N.-C.; Choi, W. Forecasting the Energy Consumption of an Actual Air Handling Unit and Absorption Chiller Using ANN Models. *Energies* **2020**, *13*, 4361. [CrossRef]
113. Almalaq, A.; Zhang, J.J. Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings. *IEEE Access* **2019**, *7*, 1520–1531. [CrossRef]
114. Lee, E.; Rhee, W. Individualized Short-Term Electric Load Forecasting With Deep Neural Network Based Transfer Learning and Meta Learning. *IEEE Access* **2021**, *9*, 15413–15425. [CrossRef]
115. Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Sunmola, F.; Ajayi, S. Machine learning for energy performance prediction at the design stage of buildings. *Energy Sustain. Dev.* **2022**, *66*, 12–25. [CrossRef]
116. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Multi-Sequence LSTM-RNN Deep Learning and Metaheuristics for Electric Load Forecasting. *Energies* **2020**, *13*, 391. [CrossRef]
117. Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Grishikashvili, K.; Sunmola, F.; Oseghale, R.; Ajayi, S. Ensemble learning for energy performance prediction of residential buildings. In Proceedings of the Environmental Design and Management Conference (EDMIC), Bristol, UK, 8 July 2021.
118. Zhang, J.-P.; Li, Z.-W.; Yang, J. A parallel SVM training algorithm on large-scale classification problems. In *2005 International Conference on Machine Learning and Cybernetics*; IEEE: Piscataway, NJ, USA, 2005; Volume 3. [CrossRef]
119. Chung, K.C.; Tan, S.S.; Holdsworth, D.K. *Insolvency Prediction Model Using Multivariate Discriminant Analysis and Artificial Neural Network for the Finance Industry in New Zealand*; Social Science Research Network: Rochester, NY, USA, 2008.
120. Jagait, R.K.; Fekri, M.N.; Grolinger, K.; Mir, S. Load Forecasting Under Concept Drift: Online Ensemble Learning With Recurrent Neural Network and ARIMA. *IEEE Access* **2021**, *9*, 98992–99008. [CrossRef]
121. Auret, L.; Aldrich, C. Interpretation of nonlinear relationships between process variables by use of random forests. *Miner. Eng.* **2012**, *35*, 27–42. [CrossRef]
122. Ghosh, S. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. *J. Geophys. Res. Atmos.* **2010**, *115*, 18–19. [CrossRef]
123. Fan, C.; Xiao, F.; Wang, S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl. Energy* **2014**, *127*, 1–10. [CrossRef]
124. Li, X.; Wen, J.; Bai, E.-W. Building energy forecasting using system identification based on system characteristics test. In Proceeding of the 2015 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES), Seattle, WA, USA, 13 April 2015; pp. 1–6. [CrossRef]
125. Liao, J.-M.; Chang, M.-J.; Chang, L.-M. Prediction of Air-Conditioning Energy Consumption in R&D Building Using Multiple Machine Learning Techniques. *Energies* **2020**, *13*, 1847. [CrossRef]
126. Al-Rakhami, M.; Gumaei, A.; Alsanad, A.; Alamri, A.; Hassan, M.M. An Ensemble Learning Approach for Accurate Energy Load Prediction in Residential Buildings. *IEEE Access* **2019**, *7*, 48328–48338. [CrossRef]
127. Borowski, M.; Zwolińska, K. Prediction of Cooling Energy Consumption in Hotel Building Using Machine Learning Techniques. *Energies* **2020**, *13*, 6226. [CrossRef]

128. Silvestro, F.; Bagnasco, A.; Lanza, I.; Massucco, S.; Vinci, A. Energy efficient policy and real time energy monitoring in a large hospital facility: A case study. *Int. J. Heat Technol.* **2017**, *35*, S221–S227. [CrossRef]
129. Wang, Z.; Srinivasan, R.S. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renew. Sustain. Energy Rev.* **2017**, *75*, 796–808. [CrossRef]
130. Capuano, D.L. International Energy Outlook 2020. Energy Information Administration: Washington, DC, USA; IEO2020; 14 September 2020. p. 7. Available online: https://www.eia.gov/outlooks/ieo/pdf/ieo2020.pdf (accessed on 5 March 2023).