A Review of the Recent Developments in Integrating Machine Learning Models with Sensor Devices in the Smart Buildings Sector with a View to Attaining Enhanced Sensing, Energy Efficiency, and Optimal Building Management

Dana-Mihaela Petroșanu [1,*] , **George Căruțașu** [2,3] , **Nicoleta Luminița Căruțașu** [4] and **Alexandru Pîrjan** [2]

[1] Department of Mathematics-Informatics, Faculty of Applied Sciences, University Politehnica of Bucharest, Splaiul Independenței 313, 060042 Bucharest, Romania

[2] Department of Informatics, Statistics and Mathematics, Romanian-American University, Expoziției 1B, 012101 Bucharest, Romania; carutasu.george@profesor.rau.ro (G.C.); alex@pirjan.com (A.P.)

[3] Doctoral School, University Politehnica of Timișoara, Piața Victoriei 2, 300006 Timișoara, Romania

[4] Department of Robotics and Production Systems, Faculty of Industrial Engineering and Robotics, University Politehnica of Bucharest, Splaiul Independenței 313, 060042 Bucharest, Romania; nicoleta.carutasu@upb.ro

* Correspondence: danap@mathem.pub.ro; Tel.: +40-761-086-656

Abstract: Lately, many scientists have focused their research on subjects like smart buildings, sensor devices, virtual sensing, buildings management, Internet of Things (IoT), artificial intelligence in the smart buildings sector, improving life quality within smart homes, assessing the occupancy status information, detecting human behavior with a view to assisted living, maintaining environmental health, and preserving natural resources. The main purpose of our review consists of surveying the current state of the art regarding the recent developments in integrating supervised and unsupervised machine learning models with sensor devices in the smart building sector with a view to attaining enhanced sensing, energy efficiency and optimal building management. We have devised the research methodology with a view to identifying, filtering, categorizing, and analyzing the most important and relevant scientific articles regarding the targeted topic. To this end, we have used reliable sources of scientific information, namely the Elsevier Scopus and the Clarivate Analytics Web of Science international databases, in order to assess the interest regarding the above-mentioned topic within the scientific literature. After processing the obtained papers, we finally obtained, on the basis of our devised methodology, a reliable, eloquent and representative pool of 146 papers scientific works that would be useful for developing our survey. Our approach provides a useful up-to-date overview for researchers from different fields, which can be helpful when submitting project proposals or when studying complex topics such those reviewed in this paper. Meanwhile, the current study offers scientists the possibility of identifying future research directions that have not yet been addressed in the scientific literature or improving the existing approaches based on the body of knowledge. Moreover, the conducted review creates the premises for identifying in the scientific literature the main purposes for integrating Machine Learning techniques with sensing devices in smart environments, as well as purposes that have not been investigated yet.

Keywords: internet of things; sensor networks; machine learning models; sensor devices; smart buildings; energy efficiency; optimal building management

1. Introduction

Globally nowadays, all types of buildings affect the environment to an overwhelming extent, by means ranging from the associated electricity consumption, through generated waste and pollution, up to natural habitat degradations, causing irreparable damages to the environment. Therefore, all over the world, concerted action is being carried out in order to limit these negative impacts. In addition to this, modern society faces issues regarding building safety along with comfort, and consequently, major efforts are being carried out all over the world in the direction of monitoring, identifying occupants' presence and activities in order to achieve enhanced sensing, energy efficiency and optimal building management, while at the same time minimizing or even eliminating the negative consequences imposed on the environment.

There is an increasing interest in the scientific literature in studies related to these topics; for example, there have been papers focusing on smart buildings [1–5], smart homes [6–10], smart hospitals [11], smart commercial buildings [12,13], sensor devices [9,14–17], supervised machine learning models for classification purposes [1,11,16–18] or for regression purposes [19–23], unsupervised machine learning models for clustering purposes [24–26], deep learning techniques [18,27,28], human activity recognition and classification with a view to assisted living [15,29–35], Internet of Things (IoT) [21,36–39], energy efficiency and an optimal building management [1,21,23,24,40–46], and the comfort and safety of the inhabitants [39,40,47–56].

In this context, a subject of utmost importance, which could lead to a wide range of advantages for the inhabitants of buildings, for constructors, for providers of different services, and even for society as a whole, is the analysis of recent developments in integrating machine learning models with sensor devices in the smart buildings sector with a view to attaining enhanced sensing, energy efficiency and optimal building management.

Therefore, this study aims to review the latest scientific articles that fuse emerging topics such as machine learning techniques, enhanced sensing, and smart buildings; hence attaining a proper categorization of a high number of scientific works in accordance with a well-defined encompassing taxonomy. In addition to providing a useful up-to-date overview to the researchers from different scientific fields who might be interested in devising project proposals or studying emerging complex topics like the analyzed ones, this review article sets its sights on providing scientists with valuable insights on enhancing existing methods from the current state of the art and on future research directions that have not yet been addressed by reviewing the recent advances that have been made with regard to integrating machine learning models with sensor devices in the smart buildings sector. Consequently, this review article aims to indicate the main purposes within the scientific literature for the integration of machine learning techniques with sensing devices in the smart buildings sector, thereby helping researchers identify possible novel purposes that have not been pursued up until now.

The review paper is structured as follows: the next section, namely "Research Methodology", presents the devised approach, developed with a view to identifying, filtering, classifying and analyzing the most important and relevant scientific articles related to the topic. The section also includes a flowchart of the developed survey, containing details regarding the steps of the devised research methodology. The Third Section, "Enhanced Sensing by Integrating Machine Learning Models with Sensor Devices in the Smart Buildings Sector" presents a review of the papers that were selected by applying the devised methodology, identifying through summarization tables and their analysis the machine learning models that are most suitable for integration with sensor devices in the smart buildings sector. The section also contains a review of the most highly cited scientific papers approaching the reviewed topics, as reported by the Elsevier Scopus and the Clarivate Analytics Web of Science International Databases. Afterwards, the Fourth Section, namely the "Discussion and Conclusions" Section, highlights the most important findings of the paper, presents an analysis of the conducted review research in perspective of previous surveys, highlighting a series of advantages offered by the devised approach, along with a few limitations of this study and future research directions targeted by the authors.

2. Research Methodology

The main purpose of our review is to survey the current state of the art with respect to recent developments in the integration of supervised and unsupervised machine learning models with sensor devices in the smart building sector with a view to attaining enhanced sensing, energy efficiency and optimal building management. We devised the research methodology with a view to identifying, filtering, classifying and analyzing the most important and relevant scientific articles related to the targeted topic.

We devised our review methodology in accordance to the SALSA (Search, AppraisaL, Synthesis and Analysis) framework, which was developed by Grant, M. J. and Booth, A. in their renowned paper [57], which had itself registered—at the time at which we devised our review methodology—a total of 1257 citations in the Clarivate Analytics Web of Science database and 1364 in Elsevier Scopus. Of the 14 review types and their associated methodologies, as depicted by Grant et al., we conducted our review in compliance with the "Literature Review" type. When developing the review methodology, we took into account the specifications corresponding to the "Literature Review" type provided by Grant et al. namely: the descriptive component characterizes "published materials that provide examination of recent or current literature; can cover wide range of subjects at various levels of completeness and comprehensiveness; may include research findings"; the search component of the SALSA framework for this type of review "may or may not include comprehensive searching"; the appraisal component "may or may not include quality assessment"; the synthesis component is "typically narrative"; the analysis component "may be chronological, conceptual, thematic, etc.".

To this end, we used reliable sources of scientific information, namely the Elsevier Scopus and Clarivate Analytics Web of Science international databases, in order to assess the interest in this topic within the scientific literature and to obtain a starting point for building a reliable, eloquent and representative database of scientific works that would be useful for developing our survey. We chose these two databases as we wanted to make sure that we were using globally accepted sources of information that distinctively select and index their contents in a uniformly consistent manner, backed up by decades of reliable, precise and comprehensive indexing. Furthermore, we took into account the fact that prestigious publishing groups categorize and promote their journals by highlighting the quality metrics of their journals as provided by the Web of Science Core Collection or the Elsevier Scopus databases. Therefore, we devised, based on the taxonomy of supervised and unsupervised machine learning techniques [58], custom search queries in order to assess the broad implementation and to identify which of the machine learning methods from the taxonomy represented in Figure 1 are most suitable for implementation with sensor devices in smart buildings with a view to achieving enhanced sensing, energy efficiency and optimal building management.
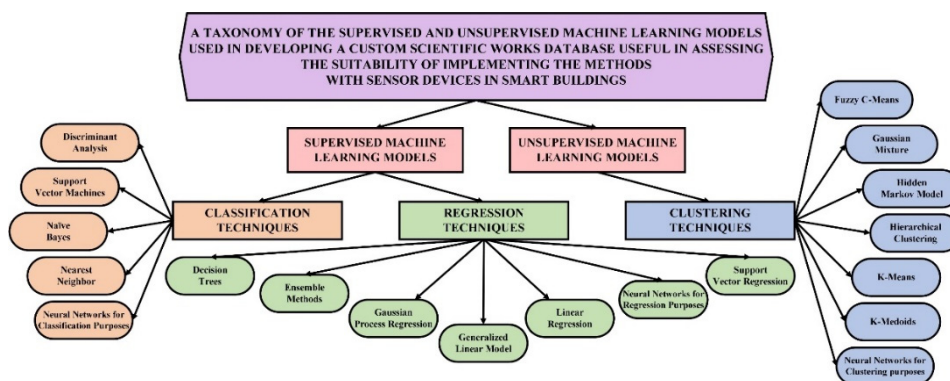

Figure 1. A taxonomy of the supervised and unsupervised machine learning models used in developing a custom scientific works database useful in conducting the survey.

After having tried several search patterns and criteria, we obtained custom search queries, with the terms smart, sensor, and at least one of the terms machine learning, artificial intelligence, supervised learning, and unsupervised learning along with their associated subcategories from the taxonomy depicted in Figure 1 being contained within the title, abstract or keywords. Consequently, according to the specific syntax of each scientific database, the search queries used for interrogating the databases are as follows:

- In the case of the Elsevier Scopus database: TITLE-ABS-KEY(Smart AND Sensor) AND TITLE-ABS-KEY("Machine Learning" OR "Artificial Intelligence" OR "Supervised Learning" OR "Classification" OR "Support Vector Machines" OR "SVM" OR "Discriminant Analysis" OR "DA" OR "Bayes" OR "NB" OR "Nearest Neighbor" OR "NNS" OR "Neural Networks" OR "ANN" OR "Regression" OR "Linear Regression" OR "LR" OR "Generalized Linear Model" OR "GLM" OR "Support Vector Regression" OR "SVR" OR "Gaussian Process Regression" OR "GPR" OR "Ensemble Methods" OR "EM" OR "Decision Tree" OR "DT" OR "Unsupervised Learning" OR "Clustering" OR "Fuzzy" OR "C-Means" OR "Gaussian Mixture" OR "Hidden Markov" OR "Hierarchical Clustering" OR "K-Means" OR "K-Medoids").

- In the case of the Clarivate Analytics Web of Science database: TS = (Smart AND Sensor) AND TS = (Machine Learning OR Artificial Intelligence OR Supervised Learning OR Classification OR Support Vector Machines OR SVM OR Discriminant Analysis OR DA OR Bayes OR NB OR Nearest Neighbor OR NNS OR Neural Networks OR ANN OR Regression OR Linear Regression OR LR OR Generalized Linear Model OR GLM OR Support Vector Regression OR SVR OR Gaussian Process Regression OR GPR OR Ensemble Methods OR EM OR Decision Tree OR DT OR Unsupervised Learning OR Clustering OR Fuzzy OR C-Means OR Gaussian Mixture OR Hidden Markov OR Hierarchical Clustering OR K-Means or K-Medoids).

The search queries were run, and two initial pools of scientific works were retrieved on the 14th of June 2019. Afterwards, the retrieved papers were filtered according to our devised methodology and synthesized into the following flowchart (Figure 2).
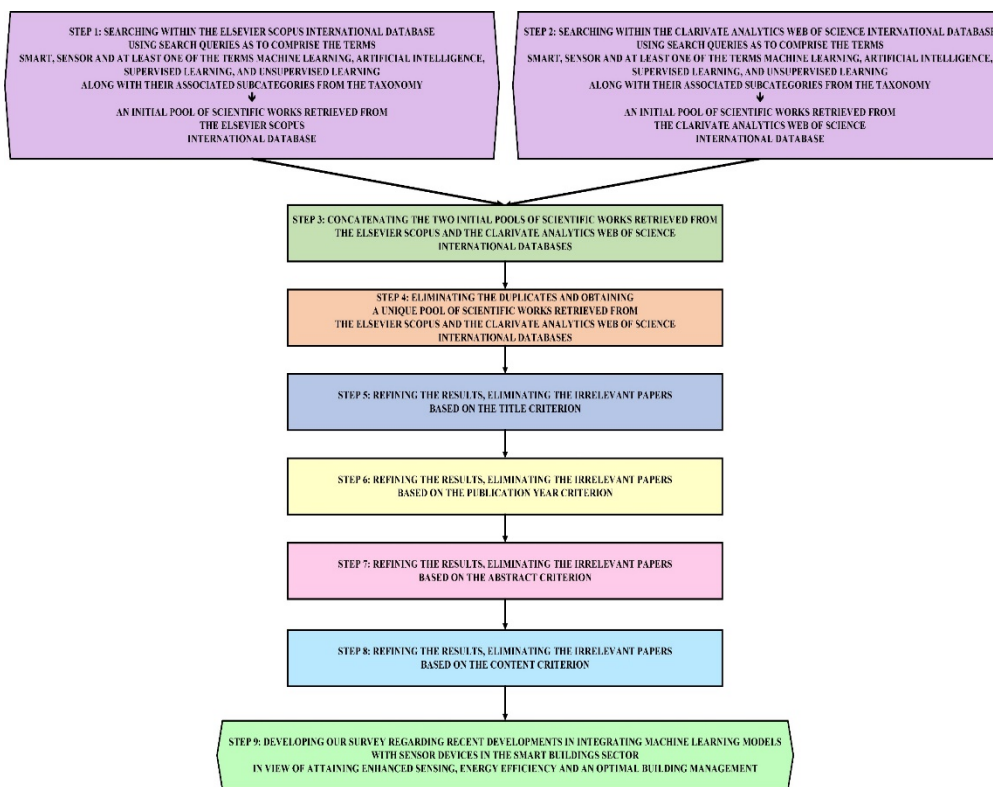

Figure 2. The flowchart of the developed survey.

Therefore, the first two steps of our methodology consist of searching the two international databases using the above-mentioned search queries, consequently obtaining two initial pools of scientific works useful for conducting the survey, consisting of 1255 papers retrieved from the Elsevier Scopus database and 381 papers from the Clarivate Analytics Web of Science database, that is, a total number of 1636 papers (with some papers being included in both databases).

The official data retrieved from the Web of Science and Scopus databases are unique to each database, meaning that the Web of Science database contains no duplicate items, and also that the Scopus database contains only unique entries. When concatenating the scientific articles retrieved from the two international databases, we took into account the fact that some scientific articles might be indexed in both the Web of Science and Scopus databases, thus resulting in duplicate entries, while other scientific works may only be indexed in one of the databases. Consequently, in Step 4 of the review methodology, after having concatenated the works retrieved from the two scientific databases, we eliminated any duplicate entries, retaining only a single instance of each scientific paper.

The particular reason for distinguishing between the two databases is the sheer fact that the two internationally renowned databases have different contents with regard not only to the indexed scientific works, but also with regard to the categories of classification by domain of interest of the papers, and this is why we had to represent the charts depicting the data corresponding to each particular indexing database in different graphics. One can therefore observe that there has been an increasing interest in the literature over the years in the topic targeted by this review, as is clearly depicted by the official data retrieved from the individual databases and distinctly graphically represented for each of database, in accordance with the official records for each database in the absence of duplicate entries.

In order to obtain an initial image regarding the number and content of the scientific papers retrieved from the two databases, we computed, for both the Elsevier Scopus and Clarivate Analytics Web of Science international databases, a series of plots highlighting the number of publications per year (Figure 3), the number of publications by type (Figure 4) and the number of publications per subject area (Figure 5).
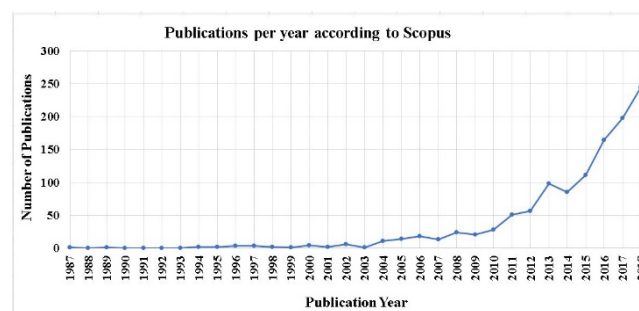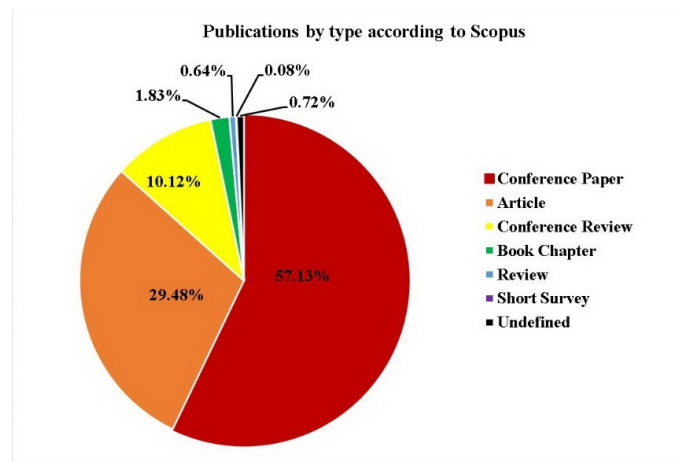

(**a**) Publications per year according to Elsevier Scopus international database
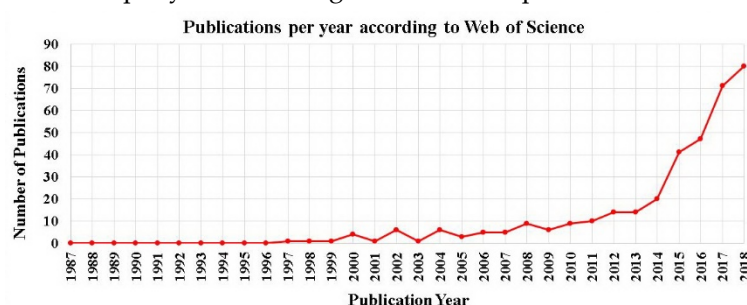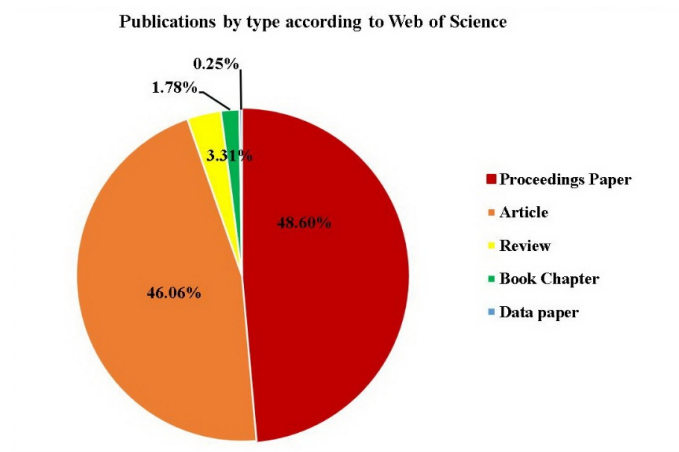

(**b**) Publications per year according to Clarivate Analytics Web of Science international database

Figure 3. The number of publications per year according to the two used databases.

By analyzing Figure 3, we noticed that during the last 5 years, the targeted subjects have been the focus of the research activity of an exponential growing number of papers indexed in both of the used databases, reflecting not only the interest of the authors of these papers but also the development of machine learning models and their integration with sensor devices in smart buildings during the analyzed period of time.

Publications by type according to Scopus

(**a**) Publications by type according to Elsevier Scopus international database

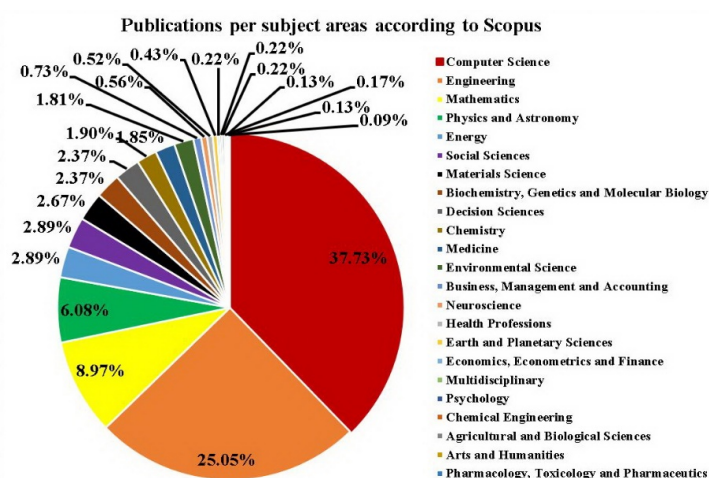**Publications by type according to Web of Science**

(**b**) Publications by type according to Clarivate Analytics Web of Science international database

Figure 4. The number of publications by type according to the two used databases.

Analyzing Figure 4, it can be remarked that the searches performed across the two databases returned a wide range of publication types. Therefore, even if the two consulted databases had returned different search results, the statistics regarding the number of publications by type according to the two databases would be similar, to a large extent, with respect to the hierarchy of the types, if not the percentages. Even though the two databases structure their searches into slightly different categories, the order of the categories of publications returned (in descending order by number of papers) by the searches performed within the two databases are highly similar. With respect to the percentages of different types of publications within the returned results, by analyzing Figure 4, it can be observed that in the case of the Elsevier Scopus international database, the "Article" type of paper represents a percentage of 29.48, while in the case of the Clarivate Analytics Web of Science international database, this type of paper represents a percentage of 46.06 of the total number of published scientific works. With respect to the papers of the "Review" type, they represent a percentage of 0.64 in the case of the Elsevier Scopus international database and a percentage of 3.31 in the case of the Clarivate Analytics Web of Science international database. With respect to "Book Chapters", the search within the Elsevier

Scopus database returned a percentage of 1.83 of the total number of retrieved scientific works, while the Clarivate Analytics Web of Science database returned a percentage of 1.78.

(**a**) Publications per subject area according to Elsevier Scopus international database

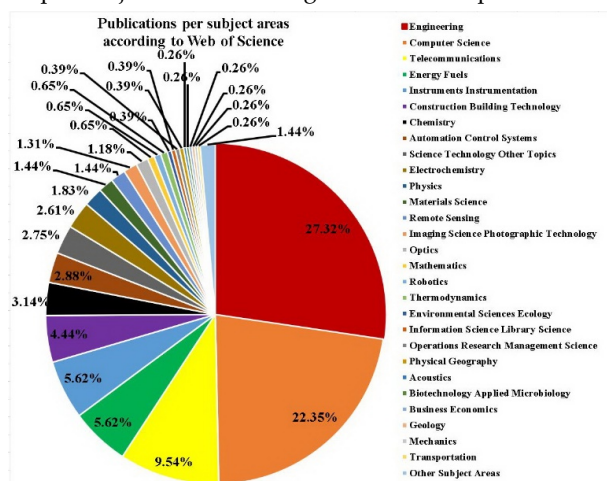(**b**) Publications per subject area according to Clarivate Analytics Web of Science international database

Figure 5. The number of publications per subject area according to the two used databases.

Examining Figure 5, it can be observed that the searches returned an extensive assortment of subject areas based on the search terms of the queries. One interesting aspect of the results depicted in Figure 5 is the fact that, in the cases of both Elsevier Scopus and Clarivate Analytics Web of Science international databases, some papers are considered to belong to more than one subject area.

Even if the results returned are structured by the two databases into slightly different types and subject areas, it is still possible to observe a series of similarities regarding the statistics of the returned results. Therefore, in the case of the Elsevier Scopus database, the most frequently approached subject areas are: Computer Science, Engineering and Mathematics (representing percentages of 37.73, 25.05 and 8.97, respectively, of the returned results), while in the case of the Clarivate Analytics Web of Science database, the hierarchy of the three most frequently approached subject areas is: Engineering, Computer Science, and Telecommunications (with percentages of 27.32, 22.35 and 9.54, respectively).

In the third step of the devised approach, by concatenating the two initial pools of scientific works retrieved from the Elsevier Scopus and Clarivate Analytics Web of Science international databases, we obtained a raw custom scientific works database. However, the raw set of scientific papers obtained still required further refinement, due to the fact that at the end of the third step, the constructed set

contained duplicate copies of some papers. Therefore, during the fourth step, we eliminated the duplicates from the set of scientific papers.

Afterwards, in order to make further improvements to the obtained set of scientific papers, in the fifth, sixth, seventh, and eighth steps, we successively refined the obtained set of scientific works by taking into account the following criteria: title, year of publication, abstract, and content of the paper. Regarding the year of publication, we decided not to plot papers published in 2019 in Figure 3 (as only half of the year had passed at the point at which we retrieved the papers used for our survey), or those papers scheduled to be published in the following year, 2020, because in these two cases, there would be further papers still to be published, and therefore the actual numbers of published papers from these two years would not be able to be taken into account when computing statistics regarding the number of publications per year according to the two used databases. However, in the subsequent analyses, in Figures 4 and 5 and throughout the whole developed survey, for reasons of consistency, we took into account papers whose publication year is (or is scheduled to be) up to 2020. Regarding the earliest year of publication taken into consideration when devising our survey, as we were targeting recent developments in integrating machine learning models with sensor devices in the smart buildings sector with a view to attaining enhanced sensing, energy efficiency, and optimal building management, in our review article we focused mainly on scientific papers published after the year 2012. Moreover, the topic that we are addressing in our survey actually began to soar after this year, as can be seen from Figure 3a,b.

Regarding the filtering performed in the eighth step, when refining the results based on the content criterion, we also eliminated documents published in conference proceedings from the custom database, on account of the fact that the most prominent proceedings papers have also been published in extenso in prestigious journals as scientific articles or reviews, while the remainder, being proceedings, do not contain comprehensive details regarding the developed methodologies and their implementations. Therefore, at this point our database contained a total number of 146 papers.

In the last step of the devised methodology, based on the final form of the custom tailored database of scientific papers, we developed our survey regarding recent developments in the integration of machine learning models with sensor devices in the smart buildings sector with a view to attaining enhanced sensing, energy efficiency, and optimal building management.

In the following, we present a review of the papers that were identified by applying the devised methodology, identifying on the basis of summarization tables and their analysis the machine learning models that are most suitable for integration with sensor devices in the smart buildings sector.

3. Enhanced Sensing by Integrating Machine Learning Models with Sensor Devices in the Smart Buildings Sector

In the following, we conducted a review of the most recent scientific articles, on the basis of the devised research methodology. For each of the identified supervised or unsupervised machine learning models, we summarize, according to the search criteria and methodology, the papers addressing those respective models. A selection of the most recent papers (sorted in descending order of publication year) is presented in the following sections, while comprehensive summarization tables are presented in the Supplementary Materials (Tables S1–S16).

3.1. Supervised Learning

3.1.1. Classification

Based on the devised methodology, we selected and summarized scientific papers that implement the Support Vector Machines (SVM) method integrated with sensor devices in smart buildings. A summary of 25 articles from the scientific papers pool that address Support Vector Machine approaches integrated with sensor devices in smart buildings can be found in Table S1 in the Supplementary Materials file, while a selection of five of the most recent papers is presented in Table 1.

Table 1. Five of the most recent scientific articles addressing the Support Vector Machines method integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the SVM Method with Sensor Devices	SVM Only or Hybrid	Performance Metrics
[1]	2019	smart building	indoor environment sensors: thermocouple TX-FF-0.32-1P (FUKUDEN) for the temperature; photosensor HD2021T AA-SP (Deltaohm) for the illuminance; OPUS20 TCO (Lufft) sensor for the relative humidity and CO_2 concentration; occupancy information sensor: PN1500 (Botem); electricity meters: PR300 (Yokogawa) for the lighting power; Enertalk Plug (Encored Technologies) for the PC electricity consumption and EHP electricity meter	assessing occupancy status information in order to improve the energy prediction performance of a building energy model	Support Vector Machine compared with Decision Tree and Artificial Neural Networks	Overall Accuracy and Standard Deviation
[6]	2018	smart home	motion sensors, item sensors (kitchen items), door sensor, temperature sensor, electricity usage, burner, cold water, hot water sensors	human activity recognition in order to help disabled persons	Support Vector Machines with a polynomial kernel of degree 3 (P-SVM); a comparison with other four classifiers: Radial Basis Function kernel—Support Vector Machine (RBF-SVM), Naïve-Bayes, Logistic Recognition, Recurrent Neural Network (RNN)	True Positives, False Positives, Precision, Recall, the F-Measure, the Receiver-Operating-Characteristic (ROC) Curve
[7]	2018	smart home	smart phones' built-in three-axis acceleration sensors and Kinect motion sensors	human fall detection	Support Vector Machine (SVM)	True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Sensitivity or True Positive Rate (TPR), Specificity (SPC) or True Negative Rate (TNR), Accuracy (ACC)
[8]	2018	smart home	passive radar-based sensor to achieve multiple level activities detection by adjusting Doppler resolution	human activity recognition and classification	Support Vector Machine (SVM) in order to classify the feature vectors into corresponding activity groups	Confusion Matrices, Classification Accuracy
[2]	2018	smart building	thermal sensor	human behavior recognition	Support Vector Regression (SVR) and Recurrent Neural Network (RNN)	Average Error, Error Rate

Examining the 25 papers selected and summarized in Table S1, presented in the Supplementary Materials file, it can be observed that 32% of them take into consideration smart buildings in general (including smart care houses, smart hospitals, smart offices), while the remaining percentage of scientific papers refer solely to smart homes. With respect to the publication year, 60% of the identified articles were published during the last 5 years.

In their research, the authors of these papers implement various types of sensors, according to their purposes, namely: indoor sensors [1], occupancy information sensors [1], electricity meters [1,6,44], motion sensors [6,7,30,59,60], item kitchen sensors [6], door sensors [6,59,61,62], temperature sensors [1,2,6,59,63], photosensors [1,3,63], status of water and burner sensors [6,59], acceleration sensors [4,7], Kinect motion sensors [7], modern smartphone sensors [4,7,60], passive radar-based sensors [8], unobtrusive sensors [9,14], infrared sensors [15,30], wireless sensor networks [61,62], accelerometers [5,63], altimeters [63], gyroscopes [63], barometers [63], heart rate monitor [63], embedded sensors [4,10,32,60,63], binary sensors [29,31,59,61], sensors installed in everyday objects [62], ubiquitous sensors [29], building management systems [44], weather stations [44], video systems [52], multi-appliance recognition systems [64], sensors for the Heating, Ventilation, and Air Conditioning (HVAC) technology [65].

With respect to the reasons for using the SVM method with sensor equipment in smart buildings, it can be observed that the recognition of human activity is at the forefront, as this is addressed in most of the papers [3,4,6,8–10,14,15,29–32,59,60,62,63]. Assisted living was a strong motivation for using the SVM method with sensor devices in the smart buildings sector; seven of the identified papers focusing on the recognition of human activity did so in order to provide appropriate assisted living [6,14,15,30–32,63], while other papers aimed to achieve assisted living by focusing on human fall detection [7], human behavior recognition [2], assessment of occupancy status information, and identification of human behavior [61]. Other reasons for applying SVM with sensors in smart buildings include measuring the occupancy status of a building's inhabitants in order to improve the energy prediction performance of the building's energy model [1], classifying the gender of occupants [5], forecasting electricity consumption [44], detecting and classifying human behavior with a view to maximizing comfort with optimized energy consumption [52], recognizing household appliances in order to assess their usage and develop habits of power preservation [64], and selecting optimal sensors for use in complex system monitoring problems such as HVAC chillers [65].

With respect to the devised methods, in [1], the authors made use of the Support Vector Machine technique and compared the obtained results with those obtained using Decision Tree and Artificial Neural Networks. In [6], the Support Vector Machine approach was implemented with a polynomial kernel of degree 3 (P-SVM), and afterwards, a comparison was conducted with other four classifiers: Radial Basis Function kernel–Support Vector Machine (RBF-SVM), Naïve Bayes, logistic recognition, and Recurrent Neural Network (RNN). The authors of [7,8,32,52,59,60] developed their research based solely on the Support Vector Machine technique. In [2], the Support Vector Regression (SVR) and Recurrent Neural Network (RNN) approaches were used. In [9] the Support Vector Machine technique was implemented for classification purposes, along with two different feature extraction methods: a manually defined method, and a Convolutional Neural Network (CNN). The authors of [3] implemented the Support Vector Machine (SVM), Convolutional Neural Network-Hidden Markov Model (CNN-HMM) and Long Short-Term Memory networks (LSTM) learning algorithms. In [10], the authors developed a hybrid approach combining the Beta Process Hidden Markov Model (BP-HMM) and the Support Vector Machine (SVM). In [4], the authors developed a Coordinate Transformation and Principal Component Analysis (CT-PCA) scheme and compared the results obtained using the K-Nearest Neighbor (KNN), Decision Tree (DT), Artificial Neural Network (ANN), Support Vector Machine (SVM) techniques. The authors of [14] used a hybrid approach, combining the Neural Network, C4.5 Decision Tree, Bayesian Network and Support Vector Machine techniques. Also based on a hybrid approach, the authors of [15] made use of SVM, Linear Kernel, Multinomial Kernel, and Radial Basis Function (RBF) kernel, and compared their results with those obtained using the K-Nearest Neighbor,

Gaussian Mixture Hidden Markov Model (GM-HMM), and Naïve Bayes approaches. The hybrid approach developed in [61] combines resampling methods such as oversampling and undersampling with Support Vector Machines and Linear Discriminant Analysis (LDA). In [5], the authors combined Bagged Decision Tree, Boosted Decision Tree, Support Vector Machines (SVMs) and Neural Networks in order to carry out gender classification. In [30], the authors used a series of learning classification algorithms, namely Naïve Bayesian (NB), Support Vector Machine (SVM), and Random Forest (RF). The authors of [63] developed their research using the Multilayer Perceptron Neural Network (MLP), Radial Basis Function Neural Network (RBF), and Support Vector Machine (SVM) techniques. In [31], the authors made use of the Support Vector Machine (SVM), Evidence-Theoretic K-Nearest Neighbor (ET-KNN), Probabilistic Neural Network (PNN), K-Nearest Neighbor (KNN), and Naïve Bayes (NB) techniques. The authors of [62] conducted their research using various methods of feature extraction, including Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA); afterwards, the new features selected by each method were used as inputs for a Weighted Support Vector Machines (WSVM) classifier. In [29], a hybrid method was developed by combining the Synthetic Minority Oversampling Technique (SMOTE) with Cost-Sensitive Support Vector Machines (CS-SVM). The authors of [44] developed a model based on Support Vector Regression (SVR). In [64], the authors developed a hybrid method by combining the Support Vector Machine with the Gaussian Mixture Model (SVM/GMM) classification model with a view to classifying electric appliances. In [65], the authors compared the Support Vector Machines (SVMs), Principal Component Analysis (PCA), and Partial Least Squares (PLS) methods.

The performance metrics considered in the scientific papers that used Support Vector Machines integrated with sensor devices in smart buildings include: Accuracy [1,3,5,7–10,29,31,59,61,62,64]; Standard Deviation [1,63]; True Positive Rate [6,7,59]; False Positive Rate [6,7,59]; Precision [6,29, 30,59,61,62]; Recall [6,29,59,61,62]; F-measure [6,29,30,59,61,62]; True Negative rate [7,59]; False Negative Rate [7,59]; Sensitivity [7,14,30]; Specificity [7,14,30,59]; Recognition Rate [10,60,64]; Receiver-Operating-Characteristic (ROC) Curve [6,14]; Confusion Matrix [8,15]; Average Error and Error Rate [2]; Root Mean Square Error (RMSE) [9] and Mean Squared Error (MSE) [3]; Classification Rate [15,52]; Absolute Mean, Variance, Median Absolute Deviation, Maximum, Minimum, Signal Magnitude range, Power, Interquartile range for computing the time and the Maximum, Mean, Skewness, Kurtosis, and Power of the frequency [4]; Matthews Correlation Coefficient [59]; Similarity Degree [32]; Mean, Standard Deviation (STD), Maximum, Minimum, Median, Mode, Kurtosis, Skewness, Intensity, Difference, Root-Mean-Square (RMS), Energy, Entropy, and Key Coefficient [63]; Coefficient of Variation (CV) and Standard Error [44]; and Success Rate [64].

With regard to the five most recent scientific articles making use of Support Vector Machines with sensor devices in smart buildings (Table 1), it can be seen that in [1], Kim et al. aimed to enhance the accuracy of energy forecasting for buildings that were not under construction, by means of assessing occupancy status information using a machine learning approach consisting of applying Support Vector Machines, Decision Tree and Artificial Neural Networks to process the data recorded by different types of sensors. The authors gathered the necessary data using indoor environmental sensors like the thermocouple TX-FF-0.32-1P manufactured by Fukuden with a view to measuring the temperature, a Deltaohm HD2021T AA-SP photosensor for measuring the illuminance level, a Lufft OPUS20 TCO sensor for measuring the relative humidity and CO_2 concentration, a PN1500 occupancy status sensor built by Botem, a Yokogawa PR300 electricity meter along with an Enertalk Plug produced by Encored Technologies for measuring the electricity consumption of the Personal Computer (PC), and an Electric Heat Pump (EHP). After carrying out the training and validation processes, the authors noticed that all of the tested machine learning algorithms provided their best results during the summer and their worst results during the spring, whereas the Support Vector Machine approach provided an increased level of accuracy compared with the other two approaches. In light of the promise of the obtained results, the authors aimed to extend their research by addressing open office spaces, which are frequently encountered in office buildings, overcoming the limitation of using only a single private office.

In [6], Machot et al. proposed a method making use of Support Vector Machines with a Polynomial Kernel of Degree 3 (P-SVM) for the recognition of human activity in order to help persons with disabilities in smart homes. The authors put forward a windowing technique relying on data recorded by different types of sensors used for motion, kitchen items, doors, temperature measurements, electricity metering, burner state determination, and cold and hot water usage. In addition to the data recorded from smart homes, available from the Center for Advanced Studies in Adaptive Systems (CASAS) dataset, Machot et al. performed experimental tests on data simulated by the Human Behavior Monitoring and Support (HBMS) software tool, identifying a set of temporal and spatial characteristics that were then used in order to compute, assess and build a conclusive feature vector. The authors compared their proposed method with the Radial Basis Function kernel–Support Vector Machine (RBF-SVM), Naïve Bayes, Logistic Recognition, and Recurrent Neural Network (RNN) approaches, obtaining improved results, as highlighted by the applied performance metrics, which included True Positives, False Positives, Precision, Recall, F-Measure, and the Receiver-Operating-Characteristic Curve.

Acknowledging the importance of accurate human fall detection and the numerous challenges arising due to the plethora of possible activities carried out by a person within a residential environment, in [7], Li et al. propounded a collaborative platform for detecting human falls. The platform comprises two sub-systems: one that uses a smart phone's built-in three-axis acceleration sensors and another that processes, using an SVM approach, the recorded data from a Kinect's motion sensors. The developed platform identifies a fall by combining the data provided by the two sub-systems based on two approaches: a logical rules process and a Dempster–Shafer theory-based method. In terms of performance, Li et al. computed and analyzed the True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Sensitivity/True Positive Rate (TPR), Specificity (SPC)/True Negative Rate (TNR) and Accuracy (ACC) metrics, concluding that the proposed approach was promising when taking into account the rapid development, diversification and integration of sensors.

In [8], Li et al. proposed a passive radar-based human activity recognition and classification method that was able to distinguish the particular body movements, physical activity patterns, and respiration of a person. A wireless energy transmitter device, such as a WiFi access point, was used to provide the signals necessary to identify the residents' activity in the smart home. The method devised by the authors comprises two stages: the Doppler data is obtained and subsequently processed by means of SVM classification in order to recognize human physical activity, while in order to detect the respiration process, a micro Doppler extraction is performed upon a Doppler spectrogram followed by the application of a Savitzky–Golay noise removal filter. The analysis of the performance metrics, which included Confusion Matrices and Classification Accuracy, confirmed that the proposed method offered satisfactory performance levels for the two analyzed situations, namely, physical activity recognition and breathing detection. The authors concluded by stating that the obtained results were promising in the healthcare field, with one advantage being the fact that no wearables or intrusive sensors were needed, meaning that the proposed system could therefore prove useful when the monitoring is being carried out over longer periods of time. The authors remarked that the developed system targets single user scenarios, and that implementing it in real-world working environments would necessitate the development of enhanced methods for separating multiple signals and behavior patterns.

Simulated sensor data related to temperature and heat were used by Zhao et al. in [2] with the aim of recognizing human behavior in smart buildings. Using the EnergyPlus software, the authors simulated different time-series of building-related data samples on which they subsequently applied two methods, one based on Support Vector Regression (SVR) and the other based on Recurrent Neural Networks (RNNs). The results obtained after conducting the experimental tests indicated that the two approaches provided similar levels of performance, as shown by the registered performance metrics, namely the Average Error and the Error Rate. This study confirmed that the Support Vector Regression approach was more flexible, and made it possible to add or remove features from the model without significantly affecting the model's accuracy; meanwhile, the Recurrent Neural Network approach provides a higher level of accuracy when the model's features do not change much over the course of time.

Then, from the obtained pool of scientific articles resulting from applying the devised review methodology, we identified, analyzed and summarized those that make use of the Discriminant Analysis technique integrated with sensor devices in smart buildings for classification purposes. A complete summarization table (Table S2) is provided in the Supplementary Materials file, while Table 2 presents five of the most recent papers that address this subject.

Analyzing the papers in Table S2 in the Supplementary Materials file, it can be observed that 83% of them refer to smart homes, while the remainder deal with any type of smart buildings (like smart offices, smart hospitals, smart foster care houses, smart retirement homes).

In these papers, the authors make use of a variety of different types of sensors. In [17], Brennan et al. considered a scalable wireless sensor network with CO_2-based estimation. In [61], Abidine et al. used a wireless sensor network comprising binary sensors like reed switches to determine the open-closed state of the doors and cabinets, pressure mats to determine whether the subject was lying down in the bed or on the couch, and float sensors to determine whether the toilet had been flushed. In [62], Abidine et al. analyzed sensor networks in a pervasive environment, with sensors installed in everyday objects such as doors, cupboards, the refrigerator, and the toilet flush to record activation/deactivation events (opening/closing events). Liao et al. based their study in [66] on sensors for motion detection. In [16], Tian et al. used a wearable accelerometer, which provided inertial information of human activity. In [33], Alam et al. considered four kinds of biosensors: Electro-Dermal Activity sensors (EDA), Electrocardiogram sensors (ECG), Blood Volume Pulse sensors (BVP) and surface Electromyography sensors (EMG).

In the identified papers, the reasons for using the Discriminant Analysis method with sensor devices in smart buildings were equally distributed between human activity recognition/classification [16,17,62] and the detection of human behavior in the context of assisted living [33,61,66].

With respect to the devised methods, in [16], the authors used the Kernel Fisher Discriminant Analysis (KFDA) technique and the Extreme Learning Machine (ELM) and performed a comparison between Best Base ELM, SVM, Bagging, AdaBoost and the proposed method. In [17], the authors compared Gradient Boosting, K-Nearest Neighbor (KNN), Linear Discriminant Analysis, and Random Forest. In [61], the authors used a hybrid method, combining resampling methods like Oversampling and Undersampling with Support Vector Machines and Linear Discriminant Analysis (LDA). The authors of [66] implemented the Discriminant Analysis technique. In [33], the authors implemented a Hidden Markov Model (HMM), Viterbi path counting, and a scalable Stochastic Variational Inference (SVI)-based training algorithm, along with Generalized Discriminant Analysis. In [62], the authors made use of various methods of feature extraction (Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA)) and the new features selected by each method were subsequently used as the inputs for a Weighted Support Vector Machines (WSVM) classifier.

The performance metrics considered in the scientific papers that use the Discriminant Analysis technique integrated with sensor devices in Smart Buildings include: Accuracy [16,17,33,61,62,66]; Precision [61,62]; Recall [16,61,62] and F-measure [33,61,62]; Root-Mean-Square Error (RMSE) [17]; Coefficient of Variance (CV) [17]; Normalized Root-Mean-Square Error (NRMSE) [17]; Coefficient of Variation of the RMSD (CV) [17]; Sensitivity (Sen) [33], Specificity (Spe) [33]; and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) [33].

Regarding five of the most recent scientific articles that make use of the Discriminant Analysis technique with sensor devices in smart buildings (Table 2), it can be observed that in [16], Tian et al. put forward a method for human activity recognition in a smart home. The proposed approach makes use of a wearable tri-axial accelerometer that provides inertial data related to the resident's activity. The collected data from the sensors are further processed using the Kernel Fisher Discriminant Analysis (KFDA) technique in order to refine and improve the feature vectors that were to be used in the subsequent processing step, which consisted of applying the Extreme Learning Machine classifier trained using the bootstrap method. After comparing the proposed method with the Best Base ELM, SVM, Bagging and AdaBoost approaches, the authors stated that their obtained results were superior, as confirmed by the Accuracy and Recall performance metrics.

Human activity recognition in smart buildings was also addressed in another recent paper [17], in which Brennan et al. studied the performance of several machine learning models, namely, Linear Discriminant Analysis, Gradient Boosting, K-Nearest Neighbor and Random Forest, with data gathered from a scalable wireless sensor network with CO_2-based estimation, with a view to accurately recognizing human activity without having to make use of expensive and privacy intrusive equipment such as computer vision and smart video cameras. In order to compare the results obtained using each of the models, the authors computed performance metrics which included Accuracy, Root-Mean-Square Error (RMSE), Normalized Root-Mean-Square Error (NRMSE) and Coefficient of Variance (CV), thereby concluding that all of the models were able to provide increased levels of performance when the training dataset comprised information regarding the sensor data in terms of structure and magnitude.

In [61], Abidine et al. aimed to assess the occupancy status information and detect human behavior within a smart home with a view to providing assisted living health care. The authors recorded the data using a wireless sensor network comprising binary sensors like reed switches to determine the open-closed state of the doors and cabinets, pressure mats to determine whether someone was lying down in the bed or on the couch, and float sensors to identify whether the toilet had been flushed. The collected data were processed using a hybrid approach, obtained by combining resampling methods like Oversampling and Undersampling with Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). The authors compared the obtained results in terms of accuracy, precision, recall and F-measure with other methods from the scientific literature that rely on the Hidden Markov Model (HMM) and the Conditional Random Field (CRF) statistical modeling technique, concluding that Oversampling with Linear Discriminant Analysis offers the best performance level.

Another scientific work that uses the Discriminant Analysis technique with sensing equipment in a smart home is that of Liao et al. [66], in which the authors aimed to overcome the limitations of existing human fall detection methods in terms of both accuracy detection and privacy intrusion issues. To this end, the authors collected data using motion detection sensors and made use of the Discriminant Analysis method to extract certain features corresponding to a resident's behavior, and to build an associated feature vector, which was then compared with features representing the state of having fallen down. After performing the experimental tests with respect to the robustness of the proposed approach, the authors stated that the results obtained confirmed the performance of the devised method.

Acknowledging the numerous benefits that assisted living brings to a patient's health and wellbeing, in [33], Alam et al. proposed a framework for Ambient Assisted Living (AAL) with a view to predicting emergencies concerning the psychiatric states of patients in a smart home environment. In order to record the different symptoms of psychiatric patients, the authors made use of four types of biosensors, namely Electro-Dermal Activity (EDA) sensors, Electrocardiogram (ECG) sensors, Blood Volume Pulse (BVP) sensors, and surface Electromyography (EMG) sensors. The recorded data were processed using a method that made use of several machine learning techniques, specifically the Hidden Markov Model (HMM) for modeling the psychiatric states, the Viterbi algorithm and the Stochastic Variational Inference (SVI) scalable algorithm for approximating the model's parameters, and Generalized Discriminant Analysis (GDA) in order to focus better on the characteristics belonging to the same psychiatric state class. After conducting an experimental study and analyzing the results in terms of prediction Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), F-Measure (FM) and Area Under the ROC Curve (AUC), the authors concluded that their proposed approach was able to supplement existing psychiatric care in residential spaces.

Subsequently, taking into consideration the devised methodology, we identified and summarized scientific papers that implemented the Naïve Bayes method integrated with sensor devices in smart buildings. The research articles that address Naïve Bayes approaches integrated with sensor devices in smart buildings are summarized in Table S3 in the Supplementary Materials file, while a selection of five of the most recent papers is presented in Table 3.

Table 2. Five of the most recent scientific articles addressing Discriminant Analysis integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Types of Sensor	Reason for Using the Discriminant Analysis Method with Sensor Devices	Discriminant Analysis Only or Hybrid	Performance Metrics
[16]	2019	smart home	wearable sensor, accelerometer providing inertial information of human activity	human activity recognition	Kernel Fisher Discriminant Analysis (KFDA) technique, Extreme Learning Machine (ELM); comparison among Best Base ELM, SVM, Bagging, AdaBoost and the proposed method	Accuracy, Recall
[17]	2018	smart buildings	a scalable wireless sensor network with CO_2-based estimation	human activity recognition	comparison of Gradient Boosting, K-Nearest Neighbor (KNN), Linear Discriminant Analysis, and Random Forest	Accuracy, Root-Mean-Square Error (RMSE), Normalized Root-Mean-Square Error (NRMSE), Coefficient of Variance (CV)
[61]	2016	smart home	wireless sensor network comprising binary sensors like reed switches to determine the open-closed state of the doors and cabinets; pressure mats to determine if one is staying laid down in the bed or on the couch; float sensors to determine if the toilet has been flushed	assessing the occupancy status information and detecting the human behavior with a view to assisted living	hybrid, combining resampling methods like Oversampling and Undersampling with Support Vector Machines and Linear Discriminant Analysis (LDA)	Accuracy, Precision, Recall and F-measure
[66]	2016	smart home	sensors for motion detection	human fall detection	Discriminant Analysis	Accuracy
[33]	2016	smart home	four kinds of biosensors: Electro-Dermal Activity sensor (EDA), Electrocardiogram sensor (ECG), Blood Volume Pulse sensor (BVP) and surface Electromyography sensor (EMG)	ambient assisted living framework for emergency psychiatric state prediction	Hidden Markov Model (HMM), Viterbi path counting, scalable Stochastic Variational Inference (SVI)-based training algorithm Generalized Discriminant Analysis	Prediction Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), F-Measure (FM) and Area Under the ROC Curve (AUC)

Table 3. Five of the most recent scientific articles addressing the Naïve Bayes integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Naïve Bayes Method with Sensor Devices	Bayes Only or Hybrid	Performance Metrics
[11]	2019	smart hospital	biomedical sensors, providing medical data (based on physiological signals), behavioral patterns (e.g., smoking, drinking alcoholics, taking medications, etc.), ambient data (e.g., humidity, temperature, noise, etc.), contextual information (e.g., location, activity, etc.)	achieving remote monitoring of patients outside the hospital in real time	a hybrid algorithm of Naïve Bayes (NB) and Whale Optimization Algorithm (WOA); a comparison between six classifiers: Decision tree (J48), Random Forest (RF), Ripper (JRip), Naïve Bayes (NB), Nearest Neighbor (IBK), Support Vector Machine (SVM)	Accuracy, Recall, Precision, F-Measure
[67]	2018	smart home	acoustic sensor network	accurate knowledge of the positions of surrounding objects useful for autonomous systems and smart devices	Bayesian filter	Mean Value and Standard Deviation
[68]	2018	smart home	carbon dioxide, total volatile organic compounds, air temperature, and air relative humidity sensors	occupancy detection in smart homes	comparison of the supervised learning models: Naïve Bayes (NB), C4.5 Decision Tree, Logistic Regression, K-Nearest Neighbor, Random Forest	For occupancy: Accuracy, True Positive Rate, True Negative Rate; For the number of occupants: Mean Absolute Error, Root Mean Square Error
[36]	2018	smart home	WiFi-enabled sensors for food nutrition quantification, and a smart phone application that collects nutritional facts of the food ingredients	Internet of Things (IoT)-based fully automated nutrition monitoring system	Bayesian algorithms and 5-layer Perceptron Neural Network method for diet monitoring	Accuracy of classification of food items and meal prediction
[34]	2016	smart home	Passive Infrared Sensor (PIR) and environmental sensors to measure pressure, temperature, humidity, and the light intensity in a particular area of the home	human presence identification and location with sub room accuracy in the context of home-based assisted living	Bayes filter algorithm	Error Rate

Analyzing the papers in Table S3, it can be observed that, according to the authors of these papers, all of the studies focused on smart homes. The authors of these scientific articles made use in their analyses of different types of sensors, including: biomedical sensors [11]; ambient data sensors [11,34,68]; acoustic sensor networks [67]; WiFi-enabled sensors [36]; Passive Infrared (PIR) sensors [30,34]; binary sensors [31,69]; and motion sensors [30,70].

With respect to the reasons for using the Naïve Bayes method with sensor equipment in smart buildings, one can observe that the recognition of human activity was the main subject of the identified papers summarized in Table S3, being addressed in papers [11,30,31,34,68–70]. Meanwhile, several of the above-mentioned scientific papers that use the Naïve Bayes integrated with sensor devices in Smart Buildings also addressed issues regarding assisted living [11,30,31,34,36]. Other reasons for applying the Naïve Bayes method with sensors in smart buildings include obtaining accurate information regarding the positions of surrounding objects, an aspect especially useful for autonomous systems and smart devices [67] or in developing an Internet of Things (IoT)-based fully automated nutrition monitoring system [36].

With respect to the devised methods, in [11], the authors made use of a hybrid approach based on the Naïve Bayes (NB) Algorithm and the Whale Optimization Algorithm (WOA), subsequently presenting a comparison among six classifiers: Decision tree (J48), Random Forest (RF), Ripper (JRip), Naïve Bayes (NB), Nearest Neighbor (IBK), Support Vector Machine (SVM). In [67], the authors implemented the Bayesian filter in order to estimate the trajectories of source positions using an acoustic sensor network. In [68], a comparison of the supervised learning models was presented: Naïve Bayes (NB), C4.5 Decision Tree, Logistic Regression, K-Nearest Neighbor, and Random Forest were used in order to detect and estimate occupancy in smart homes. In [36], the authors developed a hybrid approach by combining Bayesian algorithms and a 5-layer Perceptron Neural Network method for diet monitoring purposes; the authors of [34] used the Bayes filter algorithm to locate people. In [30], the authors made use of learning classification algorithms, including Naïve Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF). The authors of [31] made use of the Naïve Bayes (NB), Support Vector Machine (SVM), Evidence-Theoretic K-Nearest Neighbor (ET-KNN), Probabilistic Neural Network (PNN), and K-Nearest Neighbor (KNN) methods. In [69], the Dempster–Shafer theory was implemented, and was subsequently compared with the Naïve Bayes classifier and J48 Decision Tree. In [70], the authors applied a hybrid approach based on the Naïve Bayes classifier, Hidden Markov Model and Viterbi algorithm.

The performance metrics that were chosen by the authors of the scientific papers that use the Naïve Bayes method integrated with sensor devices in smart buildings include: Accuracy [11,31,34,36,68,70]; Precision [11,30,69]; Recall [11,69]; F–measure [11,30,69]; Mean Value and Standard Deviation [67]; Accuracy, True Positive Rate, True Negative Rate with a view to assessing the performance in detecting the occupancy, along with the Mean Absolute Error and the Root Mean Square Error, for establishing the number of occupants [68]; and Error Rate [34].

Regarding five of the most recent scientific articles that make use of the Naïve Bayes machine learning classifiers with sensor devices in smart buildings (Table 3), it can be observed that in [11], Hassan et al. proposed a hybrid approach, consisting of a hybrid algorithm combining Naïve Bayes (NB) and Whale Optimization Algorithm (WOA) in order to achieve real-time remote monitoring in a smart hospital of patients affected by chronic illnesses who reside outside of a hospital, thereby increasing the number and quality of monitored patients while reducing the associated hospitalization costs. The datasets were recorded by means of biomedical sensors for acquiring medical data based on physiological signals, behavioral patterns (e.g., smoking, drinking alcoholic beverages, taking medications), ambient data (e.g., humidity, temperature, noise), and contextual information (e.g., location, activity). After comparing the obtained results of their proposed hybrid approach with those recorded by using six machine learning classifiers, namely, Decision tree (J48), Random Forest (RF), Ripper (JRip), Naïve Bayes (NB), Nearest Neighbor (IBK) and Support Vector Machine (SVM), the

authors concluded that the performance metrics Accuracy, Recall, Precision and F-Measure confirmed the superiority of their proposed approach.

With a view to acquiring accurate knowledge of the positions of surrounding objects in a smart home, an aspect that is useful for both autonomous systems and smart devices, in [67], Evers et al. used a Bayesian filter in order to approximate the position trajectories of sources by acquiring data using a network of acoustic sensors. The authors aimed to overcome the challenges implied by approximating the direction of arrival for the source positions, directions that become more difficult to approximate due to the sound field becoming more diffuse as the distance from the sensor increases, causing an increase in reverberations and noises. The authors proposed using a coherent to diffuse ratio to measure the reliability of a direction of arrival in the case of localizing a single source, and showed that it is possible to triangulate the positions of a source by probabilistic means, taking advantage of the spatial diversity of network nodes.

In [68], Zimmerman et al. made use of environmental sensors that record data related to carbon dioxide, total volatile organic compounds, air temperature, and relative air humidity in order to determine the occupancy level within smart homes. The datasets retrieved from sensors were categorized using a correlation method, and the authors subsequently compared several supervised learning models: Naïve Bayes (NB), C4.5 Decision Tree, Logistic Regression, K-Nearest Neighbor, and Random Forest. These were used to detect and estimate the occupancy level. On the basis of the Accuracy, True Positive Rate and True Negative Rate for assessing the occupancy, along with the Mean Absolute Error and Root Mean Square Error for evaluating the number of occupants, the authors evaluated the performance of various classifiers (ZeroR, JRip, Naïve Bayes, J48, Logistic, K-Nearest Neighbor, Random Forest), concluding that the best performance metrics were registered when using the NB machine learning technique.

Taking into account how important the correct nutritional intake is for people, especially for infants, in [36], Sundaravadivel et al. put forward an automated nutrition monitoring system based on the Internet of Things (IoT) concept, aiming to achieve smart nutritional healthcare in smart homes. The authors' proposed system comprises WiFi-enabled sensors for food nutrition quantification, a smart phone application that collects nutritional facts regarding food ingredients, a five-layer perceptron ANN, and an algorithm based on a Bayesian Artificial Neural Network for predicting and monitoring meals. After performing the experimental tests, the authors concluded, on the basis of the Accuracy for the classification of food items and meal prediction, that their proposed system was a reliable tool for monitoring one's diet, having the potential to become an indispensable tool for childcare and for household residents.

In order to accurately identify human presence and to locate residents with sub-room accuracy in a smart home for assisted living purposes, in [34], Ballardini et al. proposed a probabilistic method that relied on the Bayes filter algorithm. In order to collect the necessary data, the authors made use of a Passive Infrared Sensor (PIR) and environmental sensors to measure pressure, temperature, humidity, and light intensity in a particular area of the home. After having analyzed the obtained results and the obtained Error Rate, the authors concluded that their developed system provided a high level of performance, with its only limitation being the fact that the system was only suitable for situations in which the smart home is inhabited by only a single resident.

Afterwards, using the devised methodology, we selected and summarized scientific papers that implement the Nearest Neighbor method integrated with sensor devices in smart buildings. A summary of the papers that address the Nearest Neighbor approaches integrated with sensor devices in smart buildings is presented in Table S4 in the Supplementary Materials file, while a selection containing five of the most recent papers is presented in Table 4.

Table 4. Five of the most recent scientific articles addressing the Nearest Neighbor method integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Nearest Neighbor Method with Sensor Devices	Nearest Neighbor Only or Hybrid	Performance Metrics
[17]	2018	smart buildings	a scalable wireless sensor network with CO_2-based estimation	human activity recognition	comparison of Gradient Boosting, K-Nearest Neighbor (KNN), Linear Discriminant Analysis, and Random Forest	Accuracy, Root-Mean-Square Error (RMSE), Normalized Root-Mean-Square Error (NRMSE), Coefficient of Variance (CV)
[68]	2018	smart home	carbon dioxide, total volatile organic compounds, air temperature, and air relative humidity sensors	occupancy detection in smart homes	comparison of the supervised learning models: Naïve Bayes (NB), C4.5 Decision Tree, Logistic Regression, K-Nearest Neighbor, Random Forest	For occupancy: Accuracy, True Positive Rate, True Negative Rate; For the number of occupants: Mean Absolute Error, Root Mean Square Error
[71]	2018	smart home	a single point Electromagnetic Interference (EMI) smart sensor	detect and track the operation of the information technology (IT) appliances (such as desktops and printers), operating in non-working hours in office buildings	Nearest Neighbor only	Precision and Recall
[72]	2015	smart home	an accelerometer in order to indicate a potential fall and the Kinect sensor in order to authenticate the eventual fall alert	human activity recognition and fall detection	the k-Nearest Neighbor (k-NN) classifier and comparison with the results obtained using linear SVM	Sensitivity, Specificity, Precision, Classification Accuracy
[31]	2015	smart home	binary sensors	human activity recognition and classification in home-based assisted living	Support Vector Machine (SVM), Evidence-Theoretic K-Nearest Neighbor (ET-KNN), Probabilistic Neural Network (PNN), K-Nearest Neighbor (KNN), Naïve Bayes (NB)	The Classification Accuracy (the Classification Error Results)

80% of the scientific papers selected and summarized in Table S4, presented in the Supplementary Materials file, present research exclusively focused on smart homes, while the remaining 20% take into consideration smart buildings in general. In these papers, the authors make use of different types of sensors. In [17], a scalable wireless sensor network with CO_2-based estimation was used. In [68], carbon dioxide, total volatile organic compounds, air temperature, and air relative humidity sensors were employed. In [71], a single-point Electromagnetic Interference (EMI) smart sensor was used. In [72], an accelerometer was used. In [31], binary sensors were used.

In these papers, the reasons for using the Nearest Neighbor integrated with sensor devices in smart buildings were mainly related to human activity recognition/classification [17,31,68,72], the detection of human behavior in the context of assisted living [31,72], and the detection and tracking of the operation of information technology (IT) appliances (such as desktops and printers) operating during non-working hours in office buildings [71].

With regard to the devised research methods, in [17], Brennan et al. compared the Gradient Boosting, K-Nearest Neighbor (KNN), Linear Discriminant Analysis, and Random Forest methods. In [68], Zimmermann et al. compared a series of supervised learning models, including Naïve Bayes (NB), C4.5 Decision Tree, Logistic Regression, K-Nearest Neighbor, Random Forest. In [71], Gulati et al. developed a Nearest Neighbor-based classification algorithm for the statistical features extracted from histograms of the measured common mode electromagnetic emissions. In [72], Kwolek et al. made use of the K-Nearest Neighbor (K-NN) classifier and compared the results with those obtained using linear SVM. In [31], Fahad et al. used the Support Vector Machine (SVM), Evidence-Theoretic K-Nearest Neighbor (ET-KNN), Probabilistic Neural Network (PNN), K-Nearest Neighbor (KNN) and Naïve Bayes (NB) techniques.

The performance metrics considered in the scientific papers that use the Nearest Neighbor method integrated with sensor devices in smart buildings include: Accuracy [17,68,72]; Root-Mean-Square Error (RMSE), Normalized Root-Mean-Square Error (NRMSE) and Coefficient of Variance (CV) [17]; Precision [71,72]; True Positive Rate, True Negative Rate, Mean Absolute Error, and Root Mean Square Error [68]; Recall [71]; Classification Accuracy [31,72]; and Sensitivity and Specificity [72].

With respect to the five most recent scientific articles addressing the Nearest Neighbor method integrated with sensor devices in smart buildings (Table 4), it can be observed that in [17], Brennan et al. developed a Wireless Sensor Network (WSNs) prototype based on CO_2 measurements in order to estimate the occupancy estimation in a smart building. With a view to improving the developed method, the authors compared the performance provided by four learning models, namely Gradient Boosting, K-Nearest Neighbor (KNN), Linear Discriminant Analysis and Random Forest, using as performance metrics the Accuracy, Root-Mean-Square Error (RMSE), Normalized Root-Mean-Square Error (NRMSE), and Coefficient of Variance (CV), finally concluding that the KNN model had produced the best results.

In [68], Zimmerman et al. made use of environmental sensors (carbon dioxide, total volatile organic compounds, air temperature, and air relative humidity sensors) in order to assess the occupancy detection in smart homes. Data retrieved from sensors were classified using a correlation method, and the authors subsequently compared a few supervised learning models: Naïve Bayes (NB), C4.5 Decision Tree, Logistic Regression, K-Nearest Neighbor, and Random Forest. These were used in order to detect and estimate occupancy. Based on the Accuracy, True Positive Rate and True Negative Rate for assessing the occupancy, along with the Mean Absolute Error and Root Mean Square Error for evaluating the number of occupants, the authors evaluated the performance of different classifiers (ZeroR, JRip, Naïve Bayes, J48, Logistic, k-Nearest Neighbor, Random Forest) and concluded that the best performance metrics were registered when using the NB technique.

In paper [71], the authors analyzed the case in which a single-point Electromagnetic Interference (EMI) smart sensor is used in order to detect and track the operation of the information technology (IT) devices, operating during non-working hours in office buildings. To this end, Gulati et al. developed a Nearest Neighbor-based classification algorithm for the statistical features extracted from histograms

of the measured common mode electromagnetic emissions. Based on the developed experiments, and computing in each case the Precision and Recall performance metrics, the authors concluded that their proposed approach was extremely useful in practice.

In paper [72], Kwolek et al. aimed to improve fall detection using an accelerometer (in order to indicate a potential fall) and a Kinect sensor (in order to authenticate the eventual fall alert) as sensors. The authors used the K-Nearest Neighbor (K-NN) classifier, and subsequently compared the results obtained with those obtained using the linear SVM approach by computing and comparing the Sensitivity, Specificity, Precision, and Classification Accuracy performance metrics. The authors concluded that in the case of their dataset, the K-NN approach outperformed the linear SVM one from a classification performance point of view.

In [31], Fahad et al. made use of binary sensors in order to analyze human activity recognition and classification in home-based assisted living. The authors carried out a comparative analysis by taking into consideration five different learning models, namely the Support Vector Machine (SVM), Evidence-Theoretic K-Nearest Neighbor (ET-KNN), Probabilistic Neural Network (PNN), K-Nearest Neighbor (KNN) and Naïve Bayes (NB) models. Based on Classification Accuracy, the authors noted that the SVM and ET-KNN registered an improved performance when compared to the other three analyzed learning models (PNN, KNN and NB).

Afterwards, of the obtained pool of scientific articles obtained based on the devised review methodology, we identified, analyzed and summarized those that made use of Neural Networks for classification purposes integrated with sensor devices in smart buildings. A complete summarization table (Table S5) is provided in the Supplementary Materials file, while Table 5 presents five of the most recent papers addressing this subject.

Table 5. Five of the most recent scientific articles addressing Neural Networks for classification purposes integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Neural Networks for Classification Method with Sensor Devices	Neural Networks for Classification Only or Hybrid	Performance Metrics
[18]	2019	smart home	wearable hybrid sensor system comprising motion sensors and cameras	human activity recognition in medical care, smart homes, and security monitoring	hybrid approach, combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) methods	Confusion Matrices, F1 Accuracy
[27]	2019	smart home	a two-dimensional acoustic array	human activity recognition	Convolutional Neural Networks compared with traditional recognition approaches such as K-Nearest Neighbor and Support Vector Machines	Overall Accuracy
[23]	2019	smart building	Wireless Sensor Network (WSN)	energy consumption forecasting	Multilayer Perceptron (MLP) compared with: Linear Regression (LR), Support Vector Machine (SVM), Gradient Boosting Machine (GBM) and Random Forest (RF)	Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE)

Table 5. *Cont.*

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Neural Networks for Classification Method with Sensor Devices	Neural Networks for Classification Only or Hybrid	Performance Metrics
[1]	2019	smart building	indoor environment sensors: thermocouple TX-FF-0.32-1P (FUKUDEN) for the temperature; photosensor HD2021T AA-SP (Deltaohm) for the illuminance; OPUS20 TCO (Lufft) sensor for the relative humidity and CO_2 concentration; occupancy information sensor: PN1500 (Botem); electricity meters: PR300 (Yokogawa) for the lighting power; Enertalk Plug (Encored Technologies) for the PC electricity consumption and EHP electricity meter	assessing the occupancy status information in order to improve the energy prediction performance of a building energy model	Support Vector Machine compared with Decision Tree and Artificial Neural Networks	Overall Accuracy and Standard Deviation
[73]	2018	smart home	Environmental sensors: Passive Infrared (PIR) and temperature sensors	human activity recognition	Deep Convolutional Neural Network (DCNN) compared with Naïve Bayes (NB), Back-Propagation (BP) algorithms	Precision, Specificity, Recall, F1 Score, Accuracy, Total Accuracy, Confusion Matrix

Analyzing the papers from the Table S5, it can be observed that 79% of them refer to smart homes, while the remainder take into consideration the more general case of smart buildings. The authors of these scientific articles make use of different types of sensors in their analyses. These include wearable sensors [18,74]; environmental sensors [73,74]; motion sensors [18,75]; a two-dimensional acoustic array [27]; a Wireless Sensor Network (WSN) [23] and sensor networks [76]; temperature sensors [1,63,73,77]; photosensors [1,63]; Passive Infra-Red Sensors (PIR) [73,75]; sensors for humidity and for evaluating the carbon dioxide concentration [1,77]; microphones [77]; cameras [18]; occupancy information sensors [1]; electricity meters [1,75]; accelerometers [5,63]; sensors of IoT devices [38]; an altimeter, a gyroscope and a barometer [63]; sensors mounted on different objects [75]; an unobtrusive sensing module [14]; and binary and ubiquitous sensors [29].

With respect to the reasons for implementing Neural Networks for classification integrated with sensor devices in smart buildings, these are mainly related to the recognition/classification of human activity in the papers [1,5,14,18,23,27,29,63,73–77]. In some of these papers, human activity recognition has as a final purpose the detection and prediction of abnormal behavior [75], monitoring the activities of elderly who are living alone [14,63], classification of the gender of occupants in a building [5], and monitoring the activities of elderly who are living in smart homes care [18,77]. In addition to these purposes, in other papers, the authors target the study of energy consumption forecasting [1,23] or achieving advanced connectivity between devices, systems, and services that continuously record enormous amounts of data from the sensors of IoT devices [38].

With respect to the devised methods, in the paper [18], the authors made use of a hybrid approach, combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) methods. In [27], the authors implemented Convolutional Neural Networks, comparing them with traditional recognition approaches such as K-Nearest Neighbor and Support Vector Machines. In [23], the authors used the Multilayer Perceptron (MLP) method and compared it with Linear Regression (LR),

Support Vector Machine (SVM), Gradient Boosting Machine (GBM) and Random Forest (RF). The authors of [1] made use of the Support Vector Machine technique and compared it with the Decision Tree and the Artificial Neural Networks techniques. In [73], a Deep Convolutional Neural Network (DCNN) approach was implemented, and this was compared with the Naïve Bayes (NB) and the Back-propagation (BP) algorithms. In [38], the authors made use of a Bayesian Network approach that was subsequently compared with the Decision Tree and Monolithic Bayesian Network methods. In [77], the authors developed an Artificial Neural Network based on the Levenberg–Marquardt algorithm (LMA). In [14], an approach was used combining Neural Network, C4.5 Decision Tree, Bayesian Network and Support Vector Machine techniques. The authors of [5] implemented the Bagged Decision Tree, Boosted Decision Tree, Support Vector Machines (SVMs), and Neural Networks methods in order to classify gender. In [74], Recurrent Neural Networks (RNNs) were used for the activity recognition process. In [63], the authors used the Multilayer Perceptron Neural Network (MLP), Radial Basis Function (RBF) Neural Network and Support Vector Machine (SVM) methods. The authors of [29] used a hybrid method, combining Synthetic Minority Oversampling Technique (SMOTE) with Cost-Sensitive Support Vector Machines (CS-SVM). In [76], the authors developed a Bayesian Belief Network (BBN), which was improved using an Edge-Encode Genetic Algorithm (EEGA) approach and afterwards; they compared the developed approach with the Naïve Bayesian Network (NBN) and Multiclass Naïve Bayes Classifier (MNBC). In [75], the authors made use of the Echo State Network (ESN), Back Propagation Through Time (BPTT) and Real Time Recurrent Learning (RTRL) methods.

The performance metrics considered in the scientific papers that use Neural Networks for classification purposes integrated with sensor devices in smart buildings include: Confusion Matrix [18,38,73]; F1 Score [18,73,74,76]; Accuracy [1,5,18,27,29,38,73,74,76]; Root Mean Square Error (RMSE) [23,75,77]; Precision [29,38,73,74,76]; Recall [29,38,73,74,76]; Standard Deviation (STD) [1,63]; Mean Absolute Percentage Error (MAPE) [23,77]; Mean Squared Error (MSE) [77]; Coefficient of Determination (R^2) and Mean Absolute Error (MAE) [23]; Specificity [14,73]; Sensitivity (SN), Area Under the Receiver Operating Characteristic Curve (AUC) [14]; and Maximum, Minimum, Median, Mode, Kurtosis, Skewness, Intensity, Difference, Root-Mean-Square (RMS), Energy, Entropy and Key Coefficient [63].

With regard to the five most recent scientific articles that make use of neural networks for classification purposes with sensor devices in smart buildings (Table 5), it can be observed that in [18], Yu et al. aimed to enhance human activity recognition in medical care and smart homes and to ensure secure monitoring by means of a hybrid approach, combining the Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) methods. The authors recorded the necessary data using a wearable hybrid sensor system comprising motion sensors for identifying and categorizing the different states of the performed activities, along with cameras that recorded photo streams to finalize the human activity recognition within the different groups of identified states. After carrying out the experimental tests and computing the performance metrics, which included Confusion Matrices and F1-Accuracy, the authors concluded that their devised approach had managed to optimally fuse the data from the motion sensors with those from the cameras' photo streams, thereby increasing the performance when compared with a direct fusing approach.

In [27], Guo et al. proposed a method that made use of Convolutional Neural Networks for human activity recognition in smart homes in reliance on the data recorded by a two-dimensional sensor array. The authors aimed to overcome the limitations of traditional methods that make use of ultrasonic sensors with respect to the numerous operations needed for extracting features from a recorded data stream by using a single feature for recognizing human activity. The authors compared their proposed method with traditional recognition approaches such as K-Nearest Neighbor and Support Vector Machines, obtaining improved results, as highlighted by the Overall Accuracy performance metric.

Considering the numerous benefits and the importance attached to accurate electricity consumption forecasting in smart buildings and the numerous prediction methods arising from the literature due

to the evolution of wireless sensing devices and IoT equipment, in [23], Chammas et al. proposed a Multilayer Perceptron (MLP) approach for forecasting the electricity consumption in a building. The authors recorded the necessary data using a Wireless Sensor Network (WSN) comprising sensors for measuring temperature, humidity, and ambient light, along with the information regarding the weather and timestamp data. Chammas et al. compared their proposed approach with the Linear Regression (LR), Support Vector Machine (SVM), Gradient Boosting Machine (GBM) and Random Forest (RF) machine learning methods with respect to the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) performance metrics, concluding that the developed approach was efficient.

Paper [1] was reviewed previously, when analyzing the most recent scientific articles that integrate Support Vector Machine approaches with sensor devices in smart buildings (Table 1).

Passive Infrared (PIR) and temperature environmental sensors were used by Tan et al. [73] with a view to recognizing and classifying, in an unobtrusive manner, the activity of multiple inhabitants within the same smart home. The authors proposed a method based on analyzing the sensor-acquired Red-Green-Blue (RGB) images by means of a Deep Convolutional Neural Network (DCNN), which was trained and tested using the Cairo open dataset. The results obtained after conducting the experimental tests indicated a higher level of performance than those achieved using the Naïve Bayes (NB) and the Back-Propagation (BP) algorithms, as confirmed by the Precision, Specificity, Recall, Confusion Matrix, F1 Score, Accuracy, and Total Accuracy performance metrics. The authors concluded that the devised method could be used for practical purposes in cases of smart homes inhabited by two or three residents, and that the enhancement of the Deep Convolutional Neural Network for the classification of more intricated human activities would be worth investigating in a future study.

3.1.2. Regression

Subsequently, from the obtained pool of scientific articles obtained based on the devised review methodology, we identified, analyzed and summarized those making use of Decision Tree integrated with sensor devices in smart buildings. A complete summarization table (Table S6) is presented in the Supplementary Materials file, while Table 6 presents five of the most recent papers addressing this subject.

Table 6. Five of the most recent scientific articles addressing the Decision Tree integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the DT Method with Sensor Devices	DT Only or Hybrid	Performance Metrics
[19]	2019	smart building	smartphone sensors and Bluetooth beacons data	group activity detection and recognition	a framework for indoor Group Activity Detection and Recognition (GADAR) and Hierarchical Clustering, along with Decision Tree classifier, K-Neighbors classifier, Deep Neural Network, Gaussian Process classifier, Logistic regression, Support Vector Machine, Linear Discriminant Analysis, Gaussian Naïve Bayes (comparison)	Confusion Matrix, Accuracy (Mean), Accuracy (Variation), Precision, Recall, F1 Score

Table 6. *Cont.*

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the DT Method with Sensor Devices	DT Only or Hybrid	Performance Metrics
[1]	2019	smart building	indoor environment sensors: thermocouple TX-FF-0.32-1P (FUKUDEN) for the temperature; photosensor HD2021T AA-SP (Deltaohm) for the illuminance; OPUS20 TCO (Lufft) sensor for the relative humidity and CO_2 concentration; occupancy information sensor: PN1500 (Botem); electricity meters: PR300 (Yokogawa) for the lighting power; Enertalk Plug (Encored Technologies) for the PC electricity consumption and EHP electricity meter	assessing the occupancy status information in order to improve the energy prediction performance of a building energy model	Support Vector Machine compared with Decision Tree and Artificial Neural Networks	Overall Accuracy and Standard Deviation
[50]	2019	smart office buildings	air temperature, relative humidity, air speed, CO_2	personal thermal comfort	comparison between Decision Tree, Random Forest, Boosted Trees	the Overall Prediction Accuracy, the On-State Accuracy, the Present State Accuracy, Confusion Matrix, the Mean Squared Error (MSE), the Root-Mean-Squared Error (RMSE) and the Average Test Accuracy
[21]	2019	smart building	wireless sensor networks	forecasting Packet Delivery Ratio (PDR) and Energy Consumption (EC) in Internet of Things (IoT)	comparison between Linear Regression, Gradient Boosting, Random Forest, Baseline and Deep Learning Neural Networks	Root Mean Square Error (RMSE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE)
[78]	2019	smart office building	common sensors: motion detection, power consumption, CO_2 concentration	estimating the number of occupants	Decision Tree C4.5, parameterized rule-based classifier	Average Error of Occupancy Estimation

It can be seen that 32% of the scientific papers selected and summarized in Table S6, presented in the Supplementary Materials file, analyze smart buildings in general, while 53% target exclusively smart homes, 11% take into consideration smart office buildings, and the remaining 4% analyze smart spaces. The authors of these papers make use of different types of sensors, including wireless sensor networks [17,21,53,79]; sensors for detecting carbon dioxide concentration [1,17,50,53,68,78]; sensors for detecting total volatile organic compounds [68]; air temperature and humidity sensors [1,50,53,68,80]; pressure sensors [5,80]; wind speed sensors [50,80]; motion sensors [30,78,81]; Passive Infrared (PIR) sensors [30,82]; electricity meters [1,78,81]; smartphone sensors and Bluetooth beacon data [19]; indoor environment sensors [1]; occupancy information sensors [1]; sensors measuring the visibility outside the building [80]; sensors embedded in the environment [81]; wearable and environmental sensors [53,74]; binary infrared sensors [83]; unobtrusive sensing modules, including a gateway and a set of passive

sensors [14]; simple non-intrusive sensors, door sensors and occupancy sensors [82]; high-sensitivity underfloor mounted accelerometers [5]; binary sensors installed in doors, cupboards, and toilet flushes [69]; and cameras, microphones, accelerometers, multisensor board and PC monitoring, and external sensors integrated in the user's home automation system [84].

In these papers, the reasons for using the Decision Tree integrated with sensor devices in smart buildings were mainly related to human activity recognition [1,5,14,17,19,21,30,50,53,68,69,74,78–84]. In some of these papers, human activity recognition was just a first step, subsequently focusing on: analyzing and improving the energy prediction performance [1,80]; analyzing and ensuring the thermal comfort of the occupants [50,53]; forecasting energy consumption [21]; estimating the number of occupants [78]; identifying behavioral patterns [79]; detecting deviating human behavior [82]; monitoring the activities of elderly people living alone [14]; classifying the gender of occupants [5]; and improving home-based assisted living [30].

With respect to the devised research methods, in [19], Chen et al. made use of a hybrid approach, combining a framework for indoor group activity detection/recognition and hierarchical clustering, along with the Decision Tree classifier, the K-Neighbors classifier, Deep Neural Network, the Gaussian Process classifier, Logistic Regression, Support Vector Machine, Linear Discriminant Analysis, and Gaussian Naïve Bayes, drawing a comparison among these techniques. In [79], Zamil et al. used the ordered Decision Tree and compared their results with those obtained using the ClaSP and CMCla methods. In [81], Malazi et al. made use of the Emerging Patterns and Random Forest (CARER) method, comparing it with the Hidden Markov Model, Bayesian Network, Naïve Bayes, SVM, Decision Tree, and Random Forest. In [84], Bjelica et al. implemented the Decision-Tree technique only. In [69], Sebbak et al. made use of the Dempster–Shafer theory, comparing it with the Naïve Bayes classifier and J48 Decision Tree. In [17], Brennan et al. compared Gradient Boosting, K-Nearest Neighbor (KNN), Linear Discriminant Analysis, and Random Forest. In [82], Lundström et al. made use of a hybrid approach, combining Random Forest and the third-order Markov chain. In [1], Kim et al. used the Support Vector Machine and compared their results with those obtained using Decision Tree and Artificial Neural Networks. In [78], Amayri et al. made use of Decision Tree C4.5, a parameterized rule-based classifier. In [30], Nef et al. used a series of learning classification algorithms, namely Naïve Bayesian (NB), Support Vector Machine (SVM), and Random Forest (RF). In [68], Zimmermann et al. presented a comparison of the following supervised learning models: Naïve Bayes (NB), C4.5 Decision Tree, Logistic Regression, K-Nearest Neighbor, and Random Forest. These were used to detect and estimate occupancy. In [5], Bales et al. combined Bagged Decision Tree, Boosted Decision Tree, Support Vector Machines (SVMs) and Neural Networks in order to classify gender. In [74], Palumbo et al. made use of Recurrent Neural Networks (RNNs) for the activity recognition process. In [50], Shetty et al. compared the Decision Tree, Random Forest and Boosted Trees methods. In [21], Ateeq et al. compared the Linear Regression, Gradient Boosting, Random Forest, Baseline and Deep Learning Neural Networks. In [53], Li et al. compared Logistic Regression, K-Nearest Neighbor, Support Vector Machine, and Random Forest. In [80], Fong et al. made use of an improved version of the Very Fast Decision Tree (VFDT) classification algorithms and compared their results with those obtained with CART Decision Tree version 4.8, the Active Learning classifier for evolving data streams, Fast Incremental Model Trees with Drift Detection (FIMT-DD), Hoeffding Tree or VFDT, the K-Nearest Neighbor algorithm, Naïve Bayes, Online Regression Tree with Options, and Stochastic Gradient Descent. In [83], Zhao et al. implemented the Fuzzy Decision Tree method. In [14], Kim et al. used a hybrid approach, combining the Neural Network, C4.5 Decision Tree, Bayesian Network and Support Vector Machine techniques.

The performance metrics chosen by the authors of the scientific papers that used the Decision Tree method integrated with sensor devices in smart buildings included Accuracy [1,5,17,19,50,53,68,74,80]; Confusion Matrix [19]; Precision [19,30,69,74,80]; Recall [19,69,74,80]; F1 Score [19,74]; Standard Deviation [1]; Root Mean Square Error (RMSE) [17,21,68,80]; Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE) [21]; Average Error of Occupancy Estimation [78]; Normalized

Root-Mean-Square Error and Coefficient of Variation of the RMSD (CV) [17]; True Positive Rate and True Negative Rate [68,80]; Mean Absolute Error (MAE) [68,80,84]; Runtime [79]; the Receiver Operating Characteristic (ROC) curve [80]; the F-Measure [30,69,80,81]; Recognition Success Rate [83,84]; Sensitivity (SN), Specificity (SP) and Area Under the Receiver Operating Characteristic Curve (AUC) [14]; Local Outlier Factor (LOF), Z-Score values, and Cluster Transition Probability [82]; Average Specificity, Sensitivity [30].

With respect to five of the most recent scientific articles making use of Decision Tree along with sensor devices in smart buildings (Table 6), it can be observed that in [19], Chen et al. put forward a framework for indoor group activity detection and recognition (GADAR), achieving hierarchical clustering in smart buildings by using a Decision Tree classifier and data collected from smartphone sensors and Bluetooth beacons. The developed framework was designed to contain four layers: one for the user, one for the data package, one for processing, and one for output. The selection of the Decision Tree classifier was based on the experimental results obtained after comparing several machine learning approaches, namely Decision Tree, the K-Neighbors classifier, Deep Neural Network, the Gaussian Process classifier, Logistic Regression, Support Vector Machine, Linear Discriminant Analysis, and Gaussian Naïve Bayes. A group activity recognition system was developed based on the devised framework and tasked with distinguishing different types of educational group activities. The best results were obtained when using the DT classifier, as confirmed by the Confusion Matrix, Accuracy (Mean), Accuracy (Variation), Precision, Recall and F1 Score performance metrics. The most important result was the Accuracy of 89% in the cases of both group activity detection and group activity recognition.

The Decision Tree classifier was employed and compared with Support Vector Machines and artificial neural networks in paper [1], which was previously analyzed when reviewing the most recent scientific articles that integrate SVM approaches with sensor devices in smart buildings (Table 1).

Ensuring the wellbeing of inhabitants in smart office buildings in terms of personal thermal comfort is a topic that has been approached in a recent paper [50], in which Shetty et al. analyzed and compared the performance of several machine learning approaches, namely Decision Tree, Random Forest, and Boosted Trees with data recorded from sensors measuring the air temperature, relative humidity, air speed and CO_2, with to the aim of classifying a desk fan's state and forecasting its speed in accordance with individual preferences regarding desk fan usage. In order to compare the results obtained for each of the machine learning approaches, the authors computed the Overall Prediction Accuracy, the On State Accuracy, the Present State Accuracy, the Confusion Matrix, the Mean Squared Error (MSE), the Root-Mean-Squared Error (RMSE), and the Average Test Accuracy performance metrics, concluding that the Random Forest approach registered the highest performance level.

In article [21], Ateeq et al. aimed to forecast the Packet Delivery Ratio (PDR) and Energy Consumption (EC) of wireless sensor networks, given their paramount importance for Internet of Things (IoT) devices, which are increasingly being employed in small- to medium-sized smart buildings. The authors compared the results obtained after applying the Linear Regression, Gradient Boosting, Random Forest, Single Hidden Layer, and Deep Learning Neural Networks approaches to predict the PDR and EC, using an open dataset regarding the IEEE 802.15.4 technical standard. After conducting the experimental study and analyzing the results in terms of Root Mean Square Error (RMSE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE), the authors concluded that the Deep Learning Neural Networks registered the best level of performance, followed closely by the Random Forest approach.

Estimating the number of people within a smart office environment with a minimum number of interactions through video stream acquisition, so as not to disturb the occupants and avoid invading their privacy, was the topic of interest in [78], where Amayri et al. studied Decision Tree C4.5 and a Parameterized Rule-Based Classifier using data recorded from commonly available sensors for motion detection, power consumption, and CO_2 concentration. Analyzing the obtained results, the authors concluded that the C4.5 DT algorithm provided the highest level of performance after approximately

14 interaction spaces, while the Parameterized Rule-Based approach performed better at the beginning but, due to having only two parameters, in the end the C4.5 DT assessed the number of people within the smart office environment with a higher degree of accuracy, as determined on the basis of the Average Error of Occupancy Estimation performance metric.

Subsequently, from the obtained pool of scientific articles resulting from the application of the devised review methodology, we identified, analyzed and summarized those making use of Ensemble Methods integrated with sensor devices in smart buildings for classification purposes. A complete summarization table (Table S7) is presented in the Supplementary Materials file, while Table 7 presents five of the most recent papers addressing this subject.

Table 7. Five of the most recent scientific articles addressing Ensemble Methods integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Ensemble Methods with Sensor Devices	Ensemble Methods Only or Hybrid	Performance Metrics
[20]	2019	smart building	smartphone sensors (acceleration, gyroscope)	human activity recognition	Extreme Learning Machine (ELM) for ensemble learning, compared with Artificial Neural Networks (ANN), Extreme Learning Machine (ELM), Support Vector Machine (SVM), Random Forest (RF), and deep Long Short-Term Memory (LSTM) approaches	Accuracy
[16]	2019	smart home	wearable sensor, accelerometer providing inertial information of human activity	human activity recognition	Kernel Fisher Discriminant Analysis (KFDA) technique, Extreme Learning Machine (ELM); comparison among Best Base ELM, SVM, Bagging, AdaBoost and the proposed method	Accuracy, Recall
[3]	2018	smart building	Light-Emitting Diode (LED) luminaires used as light sensors	human activity recognition	Support Vector Machine (SVM), Convolutional Neural Network-Hidden Markov Model (CNN-HMM), Long Short-Term Memory networks (LSTM) learning algorithms	Accuracy and Mean Square Error (MSE)
[85]	2014	smart home	wireless sensors associated with different objects, monitoring the activities	human activity recognition	Cluster-Based Classifier Ensemble (ensemble method)	Confusion Matrix presenting number of True Positives, True Negatives, False Positives and False Negatives, Precision, Recall and F-Measure
[86]	2013	smart home	embedded sensors: stove-sensor, refrigerator-sensor, door-sensor	Activity recognition	ensemble method, combining one of the methods: Artificial Neural Networks (ANN), Hidden Markov Model (HMM), Conditional Random Fields (CRF) with the Genetic Algorithm (GA) approach	Precision, Recall, F-measure and Accuracy

Analyzing the scientific articles summarized in Table S7, presented in the Supplementary Materials file, it can be observed that 40% of them analyze smart buildings in general, while the remaining 60% take smart homes into consideration. The authors of these scientific articles make use of different types of sensors in their analyses, including smartphone sensors [16,20]; accelerometers providing inertial information of human activity [16]; Light-Emitting Diode (LED) luminaires used as light sensors [3]; and sensors associated with different objects [85,86]. In all of the papers selected and summarized in Table S7, the reason for using the Ensemble Methods integrated with the sensor devices in smart buildings was the recognition of human activity.

Regarding the devised research methods, in [20], Chen et al. made use of the Extreme Learning Machine (ELM) for ensemble learning, and compared it with the Artificial Neural Network (ANN), Extreme Learning Machine (ELM), Support Vector Machine (SVM), Random Forest (RF), and deep Long Short-Term Memory (LSTM) approaches. In [16], Tian et al. implemented the Kernel Fisher Discriminant Analysis (KFDA) technique, along with the Extreme Learning Machine (ELM), and compared their proposed method with Best Base Extreme Learning Machine (ELM), Support Vector Machine (SVM), Bagging, and AdaBoost. In [3], Hao et al. made use of the Support Vector Machine (SVM), Convolutional Neural Network–Hidden Markov Model (CNN-HMM), and Long Short-Term Memory (LSTM) networks learning algorithms. In [85], Jurek et al. implemented the Cluster-Based Classifier Ensemble as an ensemble method. In [86], Fatima et al. developed an ensemble approach, combining each of Artificial Neural Networks (ANN), Hidden Markov Model (HMM), and Conditional Random Fields (CRF) with the Genetic Algorithm (GA) approach.

The performance metrics chosen by the authors of the scientific papers that use Ensemble Methods integrated with sensor devices in smart buildings include Accuracy [3,16,20,86]; Recall [16,85,86]; Precision and F-measure [85,86]; Mean Squared Error (MSE) [3]; and Confusion Matrix presenting a number of true Positives, True Negatives, False Positives and False Negatives [85].

With respect to the scientific articles making use of Ensemble Methods along with sensor devices in smart buildings (Table 7), after applying the devised review methodology, five recent scientific works were identified. In [20], Chen et al. proposed an ensemble Extreme Learning Machine (ELM) approach using Gaussian Random Projection to initialize the input weights with a view to achieving accurate recognition of a diversity of human activities in smart buildings using non-intrusively recorded data by means of smartphone sensors, namely accelerometers and gyroscopes. The authors compared the results provided by their approach with those obtained by using the Artificial Neural Networks (ANNs), Extreme Learning Machine (ELM) that didn't use Gaussian Random Projection to initialize the input weights, Support Vector Machine (SVM), Random Forest (RF), and deep Long Short-Term Memory (LSTM) approaches. They concluded that their proposed approach was superior in terms of recognition accuracy when compared to other existing methods.

An ensemble Extreme Learning Machine method was devised by Tian et al. in [16] and compared with Best Base ELM, SVM, Bagging and AdaBoost. This paper was previously analyzed when reviewing the most recent scientific articles that use Discriminant Analysis approaches with sensor devices in smart buildings (Table 2).

Human activity recognition while the persons are moving in smart buildings is a topic addressed in a recent paper [3], in which Hao et al. proposed an ensemble learning approach consisting of the Support Vector Machine (SVM), Convolutional Neural Network-Hidden Markov Model (CNN-HMM) and Long Short-Term Memory (LSTM) networks learning algorithms. The authors used light-emitting diode luminaires as light sensors and applied a forward sequential pruning technique to improve the performance of their proposed ensemble method. The results obtained from the experimental tests were analyzed in terms of the Accuracy and Mean Squared Error (MSE) performance metrics, with results of 88% and 0.13 MSE, respectively, for the dynamical occupancy dataset.

In article [85], Jurek et al. aimed to recognize human activity in smart homes by proposing a cluster-based classifier ensemble method, using numeric and binary data collected by means of wireless sensors attached to different objects. After conducting the experimental tests and analyzing the results in terms of the Confusion Matrix presenting the number of True Positives, True Negatives, False Positives and False Negatives, Precision, Recall and F-Measure, the authors concluded that their proposed approach offered a higher level of performance than a range of state-of-the-art single clustering algorithms.

Achieving reliable human activity recognition in the context of the many distinctive features that different smart homes may exhibit is a topic addressed in [86], where Fatima et al. studied an ensemble method developed by combining one of the Artificial Neural Networks (ANN), Hidden Markov Model (HMM) or Conditional Random Fields (CRF) approaches with the Genetic Algorithm (GA) approach,

using data recorded from embedded sensors mounted on refrigerators, stoves and doors. Analyzing the obtained results, the authors concluded that their proposed approach offered a higher level of performance than single classifiers and classical multi-class models, as reflected in the Precision, Recall, F-Measure and Accuracy performance metrics.

Subsequently, from the pool of scientific articles obtained based on the devised review methodology, we identified, analyzed and summarized those making use of the Gaussian Process Regression (GPR) integrated with sensor devices in smart buildings. A complete summarization table (Table S8) is presented in the Supplementary Materials file, while Table 8 presents five of the most recent papers addressing this subject.

Table 8. Five of the most recent scientific articles addressing the Gaussian Process Regression (GPR) integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Gaussian Process Regression with Sensor Devices	Gaussian Process Regression Only or Hybrid	Performance Metrics
[20]	2019	smart building	smartphone sensors (acceleration, gyroscope)	human activity recognition	Extreme Learning Machine (ELM) for ensemble learning, compared with Artificial Neural Networks (ANN), Extreme Learning Machine (ELM), Support Vector Machine (SVM), Random Forest (RF), and deep Long Short-Term Memory (LSTM) approaches	Accuracy
[87]	2017	smart home	smart meter	human activity monitoring	Non-Intrusive Load Monitoring (NILM) algorithm, Dempster—Shafer theory compared with the Gaussian Mixture model	Score for test events
[88]	2017	smart home	smart phones as sensors to capturing voice signals, Electroglottography (EGG) electrodes as sensors to capture EGG signals	voice pathology assessment	Gaussian Mixture model-based classifier, using different numbers of Gaussian mixtures	Accuracy
[89]	2017	smart home	wearable sensors providing inertial data, environment sensors and data processed video streams that anonymize the individual	machine monitoring of human health	linear-Gaussian transition model with hard boundaries, nonlinear-Gaussian observation model, post-regularized particle filter (C-ERPF), compared to other methods: Extended Kalman Filter (EKF), constrained-EKF, and Extended Regularized Particle Filtering (ERPF) without transition constraints	Average Error
[35]	2017	smart home	The smart meter or another third-party device	ambient assisted living	the developed PQD-PCA Classifier along with the Gaussian Mixture Mode (GMM) and the Dempster—Shafer Theory (DST) compared with other classifiers (K-Nearest-Neighbors KNN, Gaussian Naïve Bayes GNB, Logistic Regression Classifier LGC, Decision Tree DTree and Random Forest Rforest)	True Positive Percentage (TPP), False Positive Percentage (FPP), Precision, Recall, F1 Score, F2 Score

A total of 83% of the scientific papers selected and summarized in Table S8, presented in the Supplementary Materials file, focus their research exclusively on smart homes, while the remaining 17% analyze both smart homes and smart buildings in general. In these papers, the authors make use of different types of sensors, including smartphone sensors [88]; electroglottography (EGG) electrodes [88]; smart meters [35,87]; wearable sensors providing inertial data, environment sensors and data processed video streams [89]; electricity, water and natural gas consumption sensors [90]; and multi-appliance

recognition systems, designing a single smart meter using a current sensor and a voltage sensor in combination with a microprocessor to meter multi-appliances [64].

With respect to the reasons for implementing the GPR integrated with sensor devices in smart buildings, these are mainly related to human activity recognition/monitoring [35,87–89]; voice pathology assessment [88]; monitoring of human health [89]; ambient assisted living [35]; recognizing household appliances in order to assess their usage and develop habits of power preservation [64]; and developing a framework for automatic leakage detection in smart water and gas grids [90].

With respect to the devised research methods, in [87], Alcalá et al. implemented the Non-Intrusive Load Monitoring (NILM) algorithm and the Dempster–Shafer theory and compared them with the Gaussian Mixture model. In [88], Muhammad et al. used the Gaussian Mixture model-based classifier, using different numbers of Gaussian Mixtures. In [89], Villeneuve et al. made use of the linear-Gaussian transition model with hard boundaries, the nonlinear-Gaussian observation model, and post-regularized particle filter (C-ERPF), and compared these to other methods, including Extended Kalman Filter (EKF), constrained-EKF, and Extended Regularized Particle Filtering (ERPF) without transition constraints. In [35], Alcalá et al. implemented a PQD-PCA Classifier along with the Gaussian Mixture Mode (GMM) and the Dempster–Shafer Theory (DST) and compared their approach with other classifiers (K-Nearest-Neighbor (KNN), Gaussian Naïve Bayes (GNB), Logistic Regression Classifier (LGC), Decision Tree (DTree) and Random Forest (Rforest)). In [90], Fagiani et al. compared Gaussian Mixture Model (GMM), Hidden Markov Model (HMM) and One-Class Support Vector Machine (OC-SVM). In [64], Lai et al. developed a hybrid approach, combining Support Vector Machine with Gaussian Mixture Model (SVM/GMM) with a view to classifying electric appliances.

The performance metrics chosen by the authors of papers using Gaussian Process Regression (GPR) integrated with sensor devices in smart buildings included Score for test events [87]; Accuracy [88]; Average Error [89]; True Positive Percentage (TPP), False Positive Percentage (FPP), Precision, Recall, F1 Score, and F2 Score [35]; the probability of correctly detecting an anomaly, the probability of erroneously detecting an anomaly, the Receiver Operating Characteristic (ROC) curve, and Area Under the ROC Curve (AUC) [90]; and Accuracy, the Success Rate and the Recognition Rate [64].

Regarding the five most recent scientific articles retrieved according to the review methodology (Table 8), in [20], Chen et al. put forward an ensemble Extreme Learning Machine (ELM) approach using Gaussian Random Projection to initialize the input weights. This paper was reviewed previously when analyzing the most recent scientific works using Ensemble Methods approaches with sensor devices in smart buildings (Table 7).

Acknowledging the importance of human activity monitoring in ensuring a certain level of independence for the elderly without sacrificing their wellbeing, in [87], Alcalá et al. aimed to overcome the challenges arising from the rejection of intrusive monitoring techniques due to privacy issues by the residents of smart homes. To this end, the authors proposed a Non-Intrusive Load Monitoring (NILM) algorithm developed based on the Dempster–Shafer theory using only the data retrieved from a smart metering device, and compared this with the Gaussian Mixture model using the Score for Test Events as a performance metric. Based on the obtained results, the authors stated that their proposed method offered a higher level of performance than the model based on the Gaussian Mixture approach.

Considering the numerous disabilities that affect people's overall quality of life by limiting their movements, senses, or activities, in [88], Muhammad et al. put forward a system for assessing voice pathological features within smart homes by means of processing the data, which consisted of voice signals recorded using smartphone sensors and electroglottography (EGG) electrodes for capturing EGG signals, through different numbers of Gaussian mixtures. The authors performed the experimental tests on the open Saarbrucken public database, which consists of a variety of voice samples, concluding the viability of the proposed system on the basis of the Accuracy performance metric, as well as the processing speed. Muhammad et al. remarked that in the case of acute pathological voice features, the information obtained after processing only the electroglottography data was insufficient; for moderate

cases, the use of either the EGG or voice recorded signals offered similar levels of performance, while the highest accuracy level was obtained through a fusion of both sources.

Machine monitoring of human health in smart homes is the topic of another recent scientific article [89], in which Villeneuve et al. devised a system based on the Linear-Gaussian transition model with hard boundaries, the Nonlinear-Gaussian observation model, and the Post-Regularized Particle Filter (C-ERPF). This system was designed to process data, recorded by wearable inertial sensors, environmental sensing devices and video streams, that had been anonymized with respect to the residents' identity. The authors compared the results obtained with their proposed approach with those obtained when using the extended Kalman Filter (EKF), the constrained-EKF, and the Extended Regularized Particle Filtering (ERPF) without transition constraints in terms of Average Error as a performance metric, concluding that two wearable wrist accelerometer sensors were sufficient to predict the kinematics of the arm.

In the scientific article [35], Alcalá et al. aimed to achieve ambient assisted living for the elderly in smart homes by proposing a Power Quality Disturbances (PQD)–Principal Component Analysis (PCA) classifier along with the Gaussian Mixture Mode (GMM) and the Dempster–Shafer Theory (DST) using data recorded by means of a smart meter or another single third-party sensing device. After conducting the experimental tests and analyzing the results with respect to True Positive Percentage (TPP), False Positive Percentage (FPP), Precision, Recall, F1 Score, F2 Score, the authors concluded that their devised method was a viable option for the elderly population who live alone.

Subsequently, from the obtained pool of scientific articles resulting from the application of the devised review methodology, we identified, analyzed and summarized those making use of the Linear Regression integrated with sensor devices in smart buildings. A complete summarization table (Table S9) is presented in the Supplementary Materials file, while Table 9 presents five of the most recent papers addressing this subject.

Table 9. Five of the most recent scientific articles addressing Linear Regression integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Linear Regression Method with Sensor Devices	Linear Regression Only or Hybrid	Performance Metrics
[21]	2019	smart building	wireless sensor networks	forecasting Packet Delivery Ratio (PDR) and Energy Consumption (EC) in Internet of Things (IoT)	comparison between Linear Regression, Gradient Boosting, Random Forest, Baseline and Deep Learning Neural Networks	Root Mean Square Error (RMSE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE)
[91]	2018	smart building	three virtual sensors: temperature, airflow, and fan speed	improving electricity consumption by correctly identifying faults within a smart building's ventilation system	Linear Regression compared with Autoregressive Moving Average with Exogenous Variables (ARMAX) models, Support Vector Machine (SVM), Artificial Neural Network (ANN).	the Coefficient of Determination (for linear models) and Acceptable Ranges (for non-linear ones)
[92]	2018	smart home	wireless sensor networks	adaptive interference suppression	Linear Regression only	range of power savings, ratio of received packet

Table 9. *Cont.*

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Linear Regression Method with Sensor Devices	Linear Regression Only or Hybrid	Performance Metrics
[41]	2017	smart home	temperature and humidity sensors from a Wireless Sensor Network	forecasting the energy use of appliances	comparing: Multiple Linear Regression, Support Vector Machine with Radial Kernel, Random Forest, Gradient Boosting Machines (GBM)	Root Mean Square Error (RMSE), Coefficient of Determination, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE)
[93]	2014	smart home	passive radio-frequency identification antennas various sensors: ultrasonic, infrared, load cells	gesture recognition	Linear regression only	Accuracy

Analyzing the scientific articles summarized in Table S9, presented in the Supplementary Materials file, it can be observed that 50% of these scientific papers analyze smart buildings in general, while the remaining 50% take smart homes into consideration. The authors of these scientific articles make use of different types of sensors in their analyses, including wireless sensor networks [21,41,92,94]; temperature, airflow, and fan virtual sensors [91]; temperature and humidity sensors [41]; and Passive Radio-frequency identification antennas along with various sensors such as ultrasonic, infrared, load cells [93].

With respect to the reasons for implementing the Linear Regression integrated with sensor devices in smart buildings, these were related to the analysis of forecasting Packet Delivery Ratio (PDR) and Energy Consumption (EC) in the Internet of Things (IoT) [21]; improving electricity consumption by correctly identifying faults within a smart building's ventilation system [91]; analyzing Adaptive Interference Suppression [92]; forecasting the energy use of appliances [41]; gesture recognition [93]; and controlling smart lighting [94].

Regarding the devised research methods, in [21], Ateeq et al. compared Linear Regression, Gradient Boosting, Random Forest, Baseline and Deep Learning Neural Networks. In [91], Mattera et al. made use of Linear Regression compared with Autoregressive Moving Average With Exogenous Variables (ARMAX), Support Vector Machine (SVM) and Artificial Neural Network (ANN) methods. In [92], Lynggaard implemented Linear Regression only. In [41], Candanedo et al. compared the Multiple Linear Regression, Support Vector Machine with Radial Kernel, Random Forest, and Gradient Boosting Machines (GBM) methods. In [93], Bouchard et al. made use of Linear Regression only. In [94], Basu et al. made use of the Linear Regression and Support Vector Regression (SVR) models.

The authors of the scientific papers using Linear Regression integrated with sensor devices in smart buildings chose various performance metrics, including the Root Mean Squared Error (RMSE) Mean Absolute Percentage Error (MAPE) [21,41,94]; Mean Percentage Error (MPE) [21]; Coefficient of Determination [41]; Mean Absolute Error (MAE) [41]; range of power savings, ratio of received packet [92]; Accuracy [93]; Normalized Mean Square Error (NMSE) [94]; and Coefficient of Determination (for linear models) and Acceptable Ranges (for non-linear ones) [91].

Concerning the five most recent scientific articles retrieved according to the review methodology (Table 9), in [21], Ateeq et al. proposed a method for predicting Packet Delivery Ratio and energy consumption, and compared the results obtained using the Linear Regression, Gradient Boosting, Random Forest, Baseline and Deep Learning neural networks approaches. This paper was reviewed previously when analyzing the most recent scientific works that use Decision Tree approaches with sensor devices in smart buildings (Table 6).

Considering the major negative impacts that faulty ventilation units can have on the electricity consumption of a building, in [91], Mattera et al. proposed a method for correctly identifying faults that might occur within a smart building's ventilation system by means of developing temperature, airflow and fan speed virtual sensors based on the data provided by existing physical sensors, thereby overcoming the expense and space conditions needed to install supplementary hardware sensing devices. To identify the moments in which virtual sensors were operating outside the correct parameters of a hardware sensor, the authors used and compared Linear Regression, Autoregressive Moving Average with Exogenous Variables (ARMAX), Support Vector Machine (SVM), and Artificial Neural Network (ANN) approaches in terms of the Coefficient of Determination (for linear models) and Acceptable Ranges (for nonlinear ones). Analyzing the obtained results, the authors concluded that their proposed approach yielded satisfactory results, thereby offering the possibility of reducing costs and equipment expenditure while ensuring an appropriate reliability level.

Acknowledging the problems that will arise due to limited radio spectrum availability in the context of IoT devices, which are increasingly present in smart homes, in [92], Lynggaard put forward an adaptive interference suppression system based on the Linear Regression method in order to correctly forecast in wireless sensor networks, using the information related to the radio channels' states, the power needed to successfully transmit a data package. The author performed comprehensive experimental tests using data retrieved from wireless sensor networks in smart homes, and concluded that the savings in terms of power ranged from 42% to 82%, while the receive ratio of a data packet was greater than or equal to 92%.

In the scientific article [41], Candanedo et al. aimed to forecast the electricity usage of appliances in smart homes by comparing the results obtained after applying Multiple Linear Regression, Support Vector Machine with Radial Kernel, Random Forest and Gradient Boosting Machines (GBM) approaches on data recorded by means of temperature and humidity sensors in a wireless sensor network. After conducting the experimental tests and analyzing the results in terms of the Root Mean Square Error (RMSE), Coefficient of Determination, Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), the authors concluded that for all of the machine learning approaches, the timestamps were the most significant information for accurately forecasting the electricity consumption of appliances.

Gesture recognition of the elderly in smart home environments was studied in [93], in which Bouchard et al. devised an algorithm based on the Linear Regression in order to distinguish movement direction and segment the datasets in order to identify a gesture's starting and ending points with a view to recognizing gestures in situations that exhibit a high degree of uncertainty by processing data recorded through means of a Passive Radio-frequency identification antennas system, along with load cells and ultrasonic and infrared sensors. The authors analyzed the results obtained using their proposed approach in terms of the Accuracy performance metric and concluded that even though the accuracy level was low, the passive radio-frequency identification system was a promising tool for the recognition of human activity. The authors intended to enhance the system in the future by means of fuzzy inference methods.

Subsequently, from the pool of scientific articles obtained based on the devised review methodology, we identified, analyzed and summarized those that make use of the Neural Networks for Regression Purposes integrated with sensor devices in smart buildings. A complete summarization table (Table S10) is presented in the Supplementary Materials file, while Table 10 presents five of the most recent papers addressing this subject.

Table 10. Five of the most recent scientific articles addressing the Neural Networks for regression purposes integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the ANN Regression Method with Sensor Devices	ANN Regression Only or Hybrid	Performance Metrics
[22]	2019	smart buildings	sensors for registering the electricity consumption	forecasting the electricity consumption	ANN compared with Linear Regression (LR), Auto-Regressive Integrated Moving Average (ARIMA), Evolutionary Algorithms (EAs) for Regression Trees (EVTree), Generalized Boosted Regression Models (GBM), Random Forest (RF), Ensemble, Recursive Partitioning and Regression Trees (Rpart), Extreme Gradient Boosting (XGBoost)	Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE)
[23]	2019	smart building	Wireless Sensor Network (WSN)	energy consumption forecasting	Multilayer Perceptron (MLP) compared with: Linear Regression (LR), Support Vector Machine (SVM), Gradient Boosting Machine (GBM) and Random Forest (RF)	Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE)
[95]	2018	smart home	smart metering system and sensors installed at a residential consumer, corresponding to 15 individual appliances (water heater, refrigerator, microwave, furnace, master bedroom, front bedroom, kitchen stove wall, dishwasher disposal, kitchen sink wall, family room, kitchen half-bath foyer, washing machine, guest bedroom, dryer, basement)	forecasting the electricity consumption	mixed Artificial Neural Network (ANN) approach using both Non-Linear Autoregressive with Exogenous Input (NARX) ANNs and Function Fitting Neural Networks (FITNETs)	Mean Squared Error (MSE), Correlation Coefficient (R), the differences between the real consumption and the forecasted ones
[12]	2018	smart commercial and residential buildings	weather sensors	forecasting the electricity consumption	deep Recurrent Neural Network (RNN) models	Root Mean Square Error relative to Root Mean Squared (RMS) average of electricity consumption in test data, Root Mean Square Error relative to Root Mean Squared (RMS) average of electricity consumption in training data, Pearson Coefficient
[43]	2018	smart home	flowmeter sensor	identifying the occurrence of a specific pattern in a Water Management System (WMS)	three types of ANN for Multi-Step-Ahead (MSA) forecasting: "Multi-Input Multi-Output (MIMO), Multi-Input Single-Output (MISO), and Recurrent Neural Network (RNN)"	Accuracy, Precision, Recall, and F-Measure

A total of 45.5% of the scientific articles summarized in Table S10, presented in the Supplementary Materials file, analyzed smart buildings in general; the same percentage of papers considered smart homes, while the remaining 9% analyzed both smart homes and smart buildings. The authors of these scientific papers make use of different types of sensors in their analyses, including sensors for registering the electricity consumption [22]; Wireless Sensor Networks (WSNs) [23,45,96]; Passive Infrared (PIR) sensors or motion detectors [75,97]; smart metering systems and sensors installed by the residential consumer, corresponding to 15 individual appliances [95]; weather sensors [12]; flowmeter sensors [43]; temperature sensors, external humidity sensors, solar radiation sensors [98]; thermal sensors [2]; and door/window entry point sensors, electricity power usage sensors, bed/sofa pressure sensors, and flood sensors [75].

With respect to the reasons for implementing the Neural Networks for regression purposes integrated with sensor devices in smart buildings, these were mainly related to forecasting electricity consumption [12,22,23,45,95]; identifying the occurrence of a specific pattern in a Water Management System (WMS) [43]; indoor temperature monitoring and forecasting [96,98]; human behavior recognition [2,75]; and short-term prediction of occupancy [97].

With respect to the devised research methods, in [22], Divina et al. made use of an Artificial Neural Network (ANN) approach, and compared this with Linear Regression (LR), Auto-Regressive Integrated Moving Average (ARIMA), Evolutionary Algorithms (EAs) for Regression Trees (EVTree), Generalized Boosted Regression Models (GBM), Random Forest (RF), Ensemble, Recursive Partitioning and Regression Trees (Rpart), and Extreme Gradient Boosting (XGBoost). In [23], Chammas et al. developed a Multilayer Perceptron (MLP) Neural Network approach and compared it with Linear Regression (LR), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Random Forest (RF). In [95], Oprea et al. made use of a mixed Artificial Neural Network (ANN) approach using both Nonlinear Autoregressive with Exogenous Input (NARX) ANNs and Function Fitting Neural Networks (FITNETs). In [12], Rahman et al. implemented deep Recurrent Neural Network (RNN) models. In [43], Khan et al. used three types of ANN for Multi-Step-Ahead (MSA) forecasting methods: Multi-Input Multi-Output (MIMO), Multi-Input Single-Output (MISO), and Recurrent Neural Network (RNN). In [98], Attoue et al. made use of an Artificial Neural Network (ANN) with Multilayer Perceptron (MLP) structure. In [2], Zhao et al. implemented the Support Vector Regression (SVR) and Recurrent Neural Network (RNN) methods. In [97], Li et al. used an ANN approach and compared the obtained results with the Traditional inhomogeneous Markov chain model, the New Markov chain model, the Probability Sampling model, and Support Vector Regression (SVR). In [45], Collotta et al. developed a hybrid method, combining the Bluetooth Low-Energy Home Energy Management System (BluHEMS) and an Artificial Neural Network (ANN) approach. In [96], Pardo et al. developed two ANNs: a linear model and a Multilayer Perceptron (MLP) model with one hidden layer, comparing the results with the Bayesian standard model. In [75], Lotfi et al. made use of different types of recurrent Neural Networks, such as Echo State Network (ESN), Back Propagation Through Time (BPTT), and Real-Time Recurrent Learning (RTRL).

The performance metrics considered in the scientific papers using the Neural Networks for Regression Purposes integrated with sensor devices in smart buildings included Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) [12,22,23,45]; Coefficient of Determination (R^2) [23]; Mean Absolute Percentage Error (MAPE) [23,45]; Mean Squared Error (MSE) [45,95,98]; Correlation Coefficient (R) [95,98]; the differences between the real consumption and the forecasted ones [95]; Pearson Coefficient [12]; Accuracy [43,97]; Precision, Recall, and F-Measure [43]; Average Error and Error Rate [2]; Mean Absolute Error (MAE) [96]; and Root Mean Squared Error (RMSE) [75].

With respect to the five most recent scientific articles retrieved according to the review methodology (Table 10), in [22], Divina et al. addressed issues regarding the prediction of smart buildings' electricity consumption, using data retrieved from sensors that registered electricity consumption. To this end, the authors analyzed a series of prediction methods, comparing the ANN approach with Linear Regression (LR), Auto-Regressive Integrated Moving Average (ARIMA), Evolutionary Algorithms

(EAs) for Regression Trees (EVTree), Generalized Boosted Regression Models (GBM), Random Forest (RF), Ensemble, Recursive Partitioning and Regression Trees (Rpart), Extreme Gradient Boosting (XGBoost). Based on this comparison, the authors observed that the methods based on machine learning models were the most suitable for task under consideration.

Article [23] was previously detailed when analyzing the most recent scientific articles that integrate Neural Networks for Classification Purposes with sensor devices in smart buildings (Table 5).

In [95], Oprea et al. presented a forecasting method for providing accurate predictions of electricity consumption at the residential level, refined to the electrical devices level. The authors considered smart home complexes that were capable of partially sustaining their electricity consumption based on renewable energy resources. The authors stated that, in contrast to other existing studies, their approach did not require supplementary meteorological datasets. The devised method was based on an ANN approach that combined the Nonlinear Autoregressive with Exogenous Input (NARX) model and Function Fitting Neural Networks (FITNETs). The input dataset was retrieved from a smart metering system and from sensors installed in the residence, corresponding to a selection of the electrical devices. In the case of the NARX model, they also used a timestamp dataset as exogenous variables. In order to validate the developed prediction method, the authors computed the Mean Squared Error (MSE), the Correlation Coefficient (R), and the differences between the real consumption and the forecasted ones and used these as performance metrics. Subsequently, they compared the obtained results with those found in the scientific literature. The authors concluded that the developed approach was a practical and efficient alternative to the existing approaches in the literature.

To obtain medium-to-long term predictions of aggregated hourly electricity consumption in both commercial and residential buildings, in [12], Rahman et al. presented a Recurrent Neural Network approach. Using the Root Mean Square Error relative to Root Mean Squared (RMS) average of electricity consumption in test data, Root Mean Square Error relative to Root Mean Squared (RMS) average of electricity consumption in training data, and the Pearson Coefficient as performance metrics, the authors evaluated the performance of their developed approach and compared it with that provided by the multilayered perceptron model. The authors compared their results to those obtained in the case of the Multilayered Perceptron Model, and the authors concluded that in the case of commercial buildings, their approach registered a lower relative error, while in the case of residential buildings, the results registered by the two methods were comparable.

In [43], Khan et al. addressed issues regarding real-time analysis of data retrieved from sensors in order to develop a process for making decisions by automated means, without any human involvement, in smart homes based on Internet of Things. To identify the patterns in a Water Management System (WMS), the authors made use of three types of ANNs: Multi-Input Multi-Output (MIMO), Multi-Input Single-Output (MISO), and Recurrent Neural Network (RNN). These were compared in order to achieve multi-step-ahead forecasting based on flowmeter sensors. Conducting a series of experiments, using Accuracy, Precision, Recall, and F-Measure as performance metrics, the authors remarked that the Recurrent Neural Network approach provided the best performance, and using its prediction, the implementation of an automated decision-making system provided an accuracy of 86%.

Subsequently, from the pool of scientific articles resulting from the application of the devised review methodology, we identified, analyzed and summarized those making use of the Support Vector Regression (SVR) integrated with sensor devices in smart buildings. A complete summarization table (Table S11) is presented in the Supplementary Materials file, while Table 11 presents five of the most recent papers approaching this subject.

Most of the scientific articles summarized in Table S11, presented in the Supplementary Materials file, analyze smart buildings in general (75%), while 12.5% consider smart homes, and the remaining 12.5% consists of studies regarding commercial buildings. The authors of these scientific articles make use of different types of sensors in their analyses, including wireless sensor networks [23,41,51,94]; thermal sensors [2]; passive infrared motion detecting sensors [97]; temperature and humidity

sensors [41]; occupancy and light sensors [13]; and energy smart meters, building management systems, and weather stations [44].

The reasons for implementing the Support Vector Regression (SVR) integrated with sensor networks in smart buildings were mainly related to forecasting electricity consumption [13,23,41,44]; controlling smart lighting [94]; human behavior recognition [2]; thermal comfort optimization [51]; and short-term prediction of occupancy [97].

Regarding the devised research methods, in [23], Chammas et al. developed a Multilayer Perceptron (MLP) Neural Network approach and compared it with Linear Regression (LR), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Random Forest (RF). In [2], Zhao et al. implemented the Support Vector Regression (SVR) and Recurrent Neural Network (RNN) methods. In [51], Viani et al. implemented the Support Vector Regression method. In [97], Li et al. used an ANN approach and compared the obtained results with the traditional inhomogeneous Markov chain model, the New Markov chain model, Probability Sampling model, and Support Vector Regression (SVR). In [41], Candanedo et al. compared the Multiple Linear Regression, Support Vector Machine with Radial Kernel, Random Forest, and Gradient Boosting Machines (GBM) methods. In [13], Caicedo et al. implemented the Support Vector Regression method. In [44], Jain et al. developed a model based on the Support Vector Regression (SVR) method. In [94], Basu et al. made use of the Linear Regression and Support Vector Regression (SVR) models.

The authors of the scientific papers using the Support Vector Regression (SVR) method integrated with sensor devices in smart buildings chose various performance metrics, including Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) [23,41]; Average Error and Error Rate [2]; Prediction Error [51]; Accuracy [97]; comparison between the actual energy consumption per day and predicted energy consumption per day [13]; Coefficient of Variation (CV) and Standard Error [44]; and Root Mean Square Error (RMSE) along with Normalized Mean Square Error (NMSE) [94].

With respect to the five most recent scientific articles addressing the Support Vector Regression (SVR) method integrated with sensor devices in smart buildings (Table 11) it can be observed that paper [23] was previously reviewed when analyzing the most recent scientific articles that integrate Neural Networks for classification purposes with sensor devices in smart buildings (Table 5); paper [2] was reviewed previously when analyzing the most recent scientific articles that integrate Support Vector Machines with sensor devices in smart buildings (Table 1); article [41] was reviewed previously when analyzing the most recent scientific articles that integrate Linear Regression with sensor devices in smart buildings (Table 9).

Table 11. Five of the most recent scientific articles addressing the Support Vector Regression (SVR) integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Support Vector Regression Method with Sensor Devices	Support Vector Regression Only or Hybrid	Performance Metrics
[23]	2019	smart building	Wireless Sensor Network (WSN)	energy consumption forecasting	Multilayer Perceptron (MLP) compared with: Linear Regression (LR), Support Vector Machine (SVM), Gradient Boosting Machine (GBM) and Random Forest (RF)	Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE)
[2]	2018	smart building	thermal sensor	human behavior recognition	Support Vector Regression (SVR) and Recurrent Neural Network (RNN)	Average Error, Error Rate
[51]	2017	smart building	Wireless Sensor Networks	thermal comfort optimization	Support Vector Regression	Prediction Error
[97]	2017	smart building	passive infrared motion detecting sensors	short-term prediction of occupancy	ANN compared with traditional inhomogeneous Markov Chain model, New Markov Chain model, Probability Sampling model, Support Vector Regression (SVR)	Accuracy (Correctness)
[41]	2017	smart home	temperature and humidity sensors from a Wireless Sensor Network	forecasting the energy use of appliances	comparing: Multiple Linear Regression, Support Vector Machine with Radial Kernel, Random Forest, Gradient Boosting Machines (GBM)	Root Mean Square Error (RMSE), Coefficient of Determination, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE)

In [51], Viani et al. addressed issues regarding the thermal comfort forecasting in smart buildings in order to improve the management of the Heating, Ventilation, Air Conditioning (HVAC) systems, to fulfill the users' requirements and to obtain reduced energy costs. Using a Wireless Sensor Network in order to evaluate the indoor conditions, the authors developed a customized SVR technique in order to determine the indoor temperature necessary to ensure the comfort of the inhabitants. Subsequently, the authors conducted a series of experiments in order to evaluate the performance of their prediction and concluded that the forecasting error was lower than 1 degree Celsius, and that their approach was therefore proved to be useful for ensuring the thermal comfort of the smart building's inhabitants.

In paper [97], Li et al. made use of passive infrared motion detection sensors in order to provide a short-term prediction of occupancy based on an inhomogeneous Markov model. The proposed approach was subsequently compared to existing models such as Probability Sampling, Artificial Neural Network, and Support Vector Regression. With the aim of evaluating the prediction accuracy of their method, the authors took into account various forecasting time intervals, including a quarter of hour, half an hour, one hour, and 24 h. In order to assess the precision of the devised approach at the spatial level, the authors evaluated the forecasting accuracy at both room and house level. The authors observed that their approach outperformed the existing models analyzed, especially when considering the quarter of an hour prediction timeframe, while for the day—ahead prediction, the differences were insignificant.

3.2. Unsupervised Learning

Clustering

Subsequently, from the obtained pool of scientific articles obtained based on the devised review methodology, we identified, analyzed and summarized those that make use of the Fuzzy C-Means method integrated with sensor devices in smart buildings. A complete summarization table (Table S12) is provided in the Supplementary Materials file, while Table 12 presents five of the most recent papers addressing this subject.

Examining the papers selected and summarized in Table S12, presented in the Supplementary Materials file, it can be observed that 53% of them focus on smart homes and smart houses, 37% refer to smart buildings in general, and the remaining 10% are equally divided among smart structures, residential buildings and smart spaces. With respect to the publication year, 63% of the identified articles were published during the last 5 years. The authors of these scientific articles made use in their analyses of different types of sensors, including sensors and actuators related to the primary heating circuits and power generation systems [24]; telecare medicine information systems (TMIS) comprising specialized sensors that provide key health data parameters [99]; distributed sensors [100]; temperature, humidity and flame sensors [101]; string-type strain gauges [49]; temperature and occupancy sensors [54]; wireless sensors [47,102]; environment sensors for measuring indoor illuminance, temperature-humidity, carbon dioxide concentration and outdoor rain and wind direction [103]; sensors for measuring the indoor and outdoor temperature and the humidity [39]; vision sensors [55]; sensor networks [56,104]; binary infrared sensors [83]; motion detectors, light sensors, meteorological sensors for the wind and solar radiation data [105]; light and motion sensors [106]; environmental sensors [107]; in-house and city sensors [108]; meteorological stations [46]; smart home sensors, remote monitoring systems, and data and video review systems [102]; temperature and infrared sensors [109]; temperature sensors [110]; inside and outside home sensors [111]; different sensors and effectors [112]; smart systems for controlling the vibration of building structures by means of smart dampers [113]; virtual sensor based on a fisheye video camera [48]; and indoor and outdoor light sensors [114].

Table 12. Five of the most recent scientific articles addressing the Fuzzy C-Means integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Fuzzy C-Means Method with Sensor Devices	Fuzzy C-Means Only or Hybrid	Performance Metrics
[24]	2019	smart building	more than 450 sensors and actuators related to the primary heating circuits and power generation system, managed by a Supervisory Control and Data Acquisition (SCADA) system	appropriate energy management	a state-of-the-art scalable distributed genetic fuzzy system (GFS) based on scalable fuzzy rule learning through evolution for regression (S-FRULER)	Root Mean Square Error (RMSE), Rules, Time
[99]	2019	smart home	Telecare Medicine Information System (TMIS) comprising specialized sensors that provide key health data parameters	identifying the patients based on their biometric data using a fuzzy extractor within a proposed security protocol	Fuzzy Extractor	the performance is assessed at the level of the whole developed protocol, taking into account the computational costs, user anonymity, mutual authentication, off-line password guessing attacks, impersonation attacks, replay attacks, and the assurance of formal security
[100]	2018	smart home	distributed sensors	object localization	Fuzzy logic techniques compared with similar approaches from other papers: Wireless Network, Radio-Frequency Identification (RFID), Visional Approach	Inaccuracy Rate, Experiment Environment Dimension and Root-Mean-Square Error (RMSE), the dependency of the localization approach to the number of wireless nodes (topology), which are employed to localize the objects
[101]	2018	smart buildings	temperature, humidity and flame sensors	fire monitoring and warning	Fuzzy Logic	Accuracy
[49]	2018	smart building	string-type strain gauge	integrity of the building, assuring public safety	Fuzzy Theory	Coefficient of Determination (R^2)

In these papers, the reasons for using the Fuzzy C-Means with the sensor devices in smart buildings were mainly related to monitoring and controlling energy management processes [24,39,46,47,54,55,106, 109]; monitoring building integrity, thus ensuring public safety [49,101,111]; human activity recognition in the context of assisted living [83,99,102,114]; improving indoor environments [48,56,103,105,108,110]; object localization [100]; identifying user location within the smart home [107]; assessing the behavior of a smart home sensor network's nodes [104]; passive Radio-Frequency Identification (RFID) localization in smart homes [112]; and identifying and isolating sensors faults [113].

With respect to the devised research methods, in [24], Rodriguez-Mier et al. developed a state-of-the-art scalable distributed Genetic Fuzzy System (GFS) based on scalable Fuzzy rule learning through evolution for regression (S-FRULER). In [100], Amirjavid et al. made use of Fuzzy Logic techniques and compared them with similar approaches from other papers, including the wireless network, Radio-Frequency Identification (RFID), and Visional approaches. In [83], Zhao et al. implemented the Fuzzy Decision Tree method. In [48], Anthierens et al. made use of the Fuzzy

Logic algorithm. In [112], Fortin-Simard et al. implemented a hybrid approach, using the elliptical trilateration and the Fuzzy Logic method. In [102], Yuan et al. made use of the pervasive healthcare system, the Context-Aware Real-time Assistant (CARA), which combines a case-based reasoning engine and the Fuzzy Logic method. In [101], Sarwar et al. implemented a Fuzzy Logic approach. In [103], Wang et al. made use of a Fuzzy microcontroller implemented by Arduino UNO. In [114], Chen et al. used Fuzzy Logic and Neuro-Fuzzy systems. In [109], Panna et al. implemented a fuzzy temperature controller. In [110], Wang et al. used a Fuzzy Cognitive Map (FCM) in order to develop a genetic algorithm with a view to identifying the connection matrix of the FCM. In [106], Liu et al. implemented a Fuzzy Logic controller. In [49], Chang et al. made use of the Fuzzy theory. In [39], Meana-Llorián et al. used the Fuzzy Logic approach. In [54], Ain et al. implemented a Fuzzy inference system. In [104], Usman et al. used Fuzzy Logic, with the same method being used by Motamed et al. in [55] and by Ulpiani in [56]. In [99], Khatoon et al. made use of the Fuzzy Extractor. In [113], Sharifi et al. developed a semi-active nonlinear Fuzzy Control System. In [107], Ahvar et al. made use of the Fuzzy Set theory. In [111], Sang-Hyun et al. developed an Adaptive Network Fuzzy Inference System (ANFIS). In [105], Kıyak et al. developed a Fuzzy Expert System for testing the light. In [47], Keshtkar et al. developed a Fuzzy Logic Decision-Making algorithm. In [46], Jabłonski made use of a Fuzzy Controller that generates the output settings for the building actuators according to a general Fuzzy Set processing scheme. In [108], a set of concepts and their Fuzzy Semantic relations were defined, extracted and used by Vlachostergiou et al.

The performance metrics considered in the scientific papers that use the Fuzzy C-Means integrated with sensor devices in smart buildings were evaluated based on experiments and simulations [46,47,103,107–109,111,114]; Root Mean Square Error (RMSE) [24]; computational cost, user anonymity, mutual authentication, off-line password guessing attacks, impersonation attacks, replay attacks, and the assurance of formal security [99]; Inaccuracy Rate, experiment environment dimension and Root-Mean-Square Error (RMSE), and the dependency of the localization approach on the number of wireless nodes (topology) employed to locate the objects [100]; Accuracy [101,110]; Coefficient of Determination (R^2) [49]; energy consumption, Electricity Cost, Peak-to-Average Ratio (PAR) [54]; energy saving percentage in different working scenarios [39]; Standard Error of Mean (SEM), Horizontal Illuminance, Daylight Glare Probability, paper-based Landolt test, Freiburg Visual Acuity Test (FrACT), Electric Lighting Energy Consumption, total number of shading and lighting commands [55]; turbulence intensity, draught rates, operative temperature, Predicted Mean Vote (PMV) and Percentage of People Dissatisfied (PPD) [56]; Identification Rate [83]; Energy Consumption and illumination level [105]; energy savings [106]; Detection Accuracy, Energy Consumption, Memory Consumption, Processing Time Estimation [104]; True Positive, False Positive, True Negative, False Negative, and Accuracy [102]; Accuracy and a comparison with the results presented in related works (based on Ultrasonic, Ultrasonic/RFID, ZigBee, Active RFID, Passive RFID) [112]; Fault Detection Index values for certain fault magnitudes, residual values for individual sensors corresponding to different fault magnitudes [113]; and comfort level [48].

With respect to the five most recent scientific articles addressing the Fuzzy C-Means method integrated with sensor devices in smart buildings (Table 12), it can be observed that in [24], Rodriguez-Mier et al. developed a Genetic Fuzzy system designed to build a scalable information database, useful in forecasting smart buildings' energy consumption. To this end, the authors developed a state-of-the-art scalable distributed Genetic Fuzzy System (GFS) based on Scalable Fuzzy Rule Learning through Evolution for Regression (S-FRULER). The authors subsequently carried out experiments based on real data and concluded that the developed approach provided a high level of accuracy.

In [99], Khatoon et al. proposed a secure and efficient authentication method, along with a key agreement protocol for the Telecare Medicine Information System (TMIS), offering healthcare services to patients, particularly to those who were elderly and vulnerable, and were unable to go to hospitals. The developed protocol was based on a Fuzzy Method in order to identify the patients, making use of

their biometric data. To ensure the security of the proposed approach and the privacy of the users, the authors made use of the elliptic curves' theory. Subsequently, the authors stated that "the performance is assessed at the level of the whole developed protocol, taking into account the computational costs, user anonymity, mutual authentication, off-line password guessing attacks, impersonation attacks, replay attacks, and the assurance of formal security".

In [100], Amirjavid et al. addressed issues regarding the tracking of objects within smart homes, proposing a method that did not require the attachment of sensors to the targeted objects, making use only of distributed sensors (among which were included visual sensors). The authors developed a series of simulations and, comparing the obtained results with those provided by other state-of-art methods, they concluded that their approach offered an improved performance, as highlighted by the following performance metrics: Inaccuracy Rate, the experiment environment dimension and Root-Mean-Square Error (RMSE), and the dependency of the localization approach on the number of wireless nodes (topology) employed to locate the objects.

In their paper [101], Sarwar et al. presented a Fire Monitoring and Warning System (FMWS), developed based on a Fuzzy Logic approach, that was designed to detect the actual existence of fire and to send alarms to a system providing a complete infrastructure for fire safety management, namely, the Fire Management System (FMS), using the Global System for Mobile (GSM) Communication technology. The authors made use of temperature, humidity and flame sensors in their study. The performance of the developed method was assessed by computing the Accuracy as a performance metric, then it was compared with similar existing methods, with the authors ultimately concluding that their approach had the potential to reduce the rate of false alarms, providing an increased potential to save lives and reduce material damage.

In [49], Chang et al. approached a subject related to both the civil engineering and automatic control fields, analyzing issues regarding the detection in real time of the falling of the tiles that cover building exteriors in Taiwan, endangering public safety. The authors combined the micro-resistance approach and the Fuzzy Theory, implementing string-type strain gauges as sensors, the Coefficient of Determination as a performance metric. They concluded that their developed method represented a feasible approach that could be further utilized with a view to assessing the status of the tiles in real time.

Subsequently, from the obtained pool of scientific articles resulting from the application of the devised review methodology, we identified, analyzed and summarized those making use of the Hidden Markov Model integrated with sensor devices in smart buildings for classification purposes. A complete summarization table (Table S13) is presented in the Supplementary Materials file, while Table 13 presents five of the most recent papers addressing this subject.

Table 13. Five of the most recent scientific articles addressing the Hidden Markov Model integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Hidden Markov Model with Sensor Devices	Hidden Markov Model Only or Hybrid	Performance Metrics
[25]	2019	smart home	34 sensors (3 door and 31 motion sensors)	sensor-based activity recognition and abnormal behavior detection	Convolutional Neural Networks (CNNs) for detecting abnormal behavior related to dementia, the results being compared with methods such as Naïve Bayes (NB), Hidden Markov Models (HMMs), Hidden Semi-Markov Models (HSMM), Conditional Random Fields (CRFs)	Precision, Recall, F-Measure and Accuracy, Sensitivity, Specificity
[115]	2019	smart building	Wireless Sensor Network	presence detection in a building	Hidden Markov Model (DS-HMM)	Accuracy
[116]	2019	smart home	unobtrusive sensing infrastructures, environmental sensors monitoring the interaction of the inhabitant with home artifacts, context conditions (e.g., temperature) and presence in certain locations	human activity recognition	the developed newNECTAR framework, based on Markov Logic Network compared with state-of-the-art techniques such as Multilayer Perceptron, Random Forest, Support Vector Machine, Naïve Bayes	Average F1 Score, Confusion Matrix
[117]	2019	smart home	passive infrared motion sensors and door sensors	human activity recognition	Hidden Markov Models and Regression Models	Average Accuracy using real data, synthetic data and randomly generated data; Accuracy first using only the real data and then Accuracy using the real data enlarged with a month of synthetically generated data
[118]	2018	smart home	motion sensors, beacons, switches, thermometers	determining the risk of an anomaly related to the healthcare of a resident happening and provide adequate actions to be taken so that a real anomaly does not occur	Markov Logic Network	Precision, Recall, and Correctness

Analyzing the papers selected and summarized in Table S13, it can be observed that 78% of them exclusively analyze smart homes, 16% take into consideration smart buildings in general, 3% analyze both smart homes and buildings, while the remaining 3% of the selected papers refer to smart workplace environments. The authors of these scientific articles make use of different types of sensors in their analyses, including wireless sensor networks [70,115,119–124]; passive infrared motion sensors [82,97, 117,118,122,125,126]; motion sensors [25,70,81,118,120,127,128]; environmental sensors [10,25,81,82,116–118,123,127–132]; temperature sensors [116,118,120,123,125,131–133]; humidity sensors [123,131–133];

pressure sensors [128,130,131,133]; light sensors [3,123,132]; unobtrusive sensing infrastructures [116]; real and virtual sensors [134]; radar sensors [135]; accelerometers [127,136]; light-emitting diodes (LED) [3]; electricity and electrical sensors [81,131,132]; smartphone sensors [127,131]; microphones [125, 129]; distributed sensor networks [137]; simple non-intrusive sensors [82]; infrared sensors [124,129–131]; actuators and home automation equipment [125]; shelf binary sensors [128]; biosensors [33]; smart meters [138]; acoustics and CO_2 sensors [133]; non-wearable ambient sensors [131].

With respect to the reasons for using the Hidden Markov Model with sensor equipment in smart buildings, it can be observed that the recognition of human activity is the main subject of the identified papers summarized in Table S13, and is addressed in papers [3,10,25,70,81,82,116,117,120–123,125–127,130–132,135,136]. Additional applications include abnormal behavior detection [25,82,118,126]; presence detection in a building [115]; fault-tolerant maintenance of a networked environment in the domain of the Internet of Things [134]; providing proximity services in smart home and building automation [119]; forecasting the presence of residents at the room and house level [97]; modeling the decision process in the context of a voice-controlled smart home [129]; event recognition in cyber-physical systems [137]; the detection of visits in the home of older adults living alone [128]; emergency psychiatric state prediction [33]; load disaggregation [138]; occupancy detection with a view to energy saving [133]; state estimation for a special class of flag Hidden Markov Models [124].

With respect to the devised methods, the authors of papers [115,122,124,126,133,136,138] implemented solely the Hidden Markov Model, while in other papers, a hybrid approach was used, based on: hidden Markov models and regression models [117]; continuous-time Markov chains, together with a cooperative control algorithm [134]; two layers of classifiers: a first-level Bayesian classifier whose inferential results are used as inputs for the second level Hidden Markov Model (HMM) [135]; Support Vector Machine (SVM), Convolutional Neural Network-Hidden Markov Model (CNN-HMM), and Long Short-Term Memory (LSTM) networks learning algorithms [3]; Beta Process Hidden Markov Model (BP-HMM) and Support Vector Machine (SVM) [10]; Hidden Markov Model and Conditional Random Field model [120]; Random Forest and third-order Markov chain [82]; Hidden Markov Model (HMM), Conditional Random Fields (CRF) and a sequential Markov Logic Network (MLN), the obtained results of which were compared to those of three non-sequential models: a Support Vector Machine (SVM), a Random Forest (RF) and a non-sequential MLN [125]; Hidden Markov Model (HMM), Viterbi path counting, scalable Stochastic Variational Inference (SVI)-based training algorithm, and Generalized Discriminant Analysis [33]; Naïve Bayes classifier, Hidden Markov Model and Viterbi algorithm [70]; Coupled Hidden Markov Model (CHMM) and Factorial Conditional Random Field (FCRF) [123]. Other methods implemented by the authors of the papers selected and summarized in Table S13 include Convolutional Neural Networks (CNNs) for detecting abnormal behavior related to dementia, with the results being compared to methods such as Naïve Bayes (NB), Hidden Markov Models (HMMs), Hidden Semi-Markov Models (HSMM), and Conditional Random Fields (CRFs) [25]; the developed newNECTAR framework, based on Markov Logic Network compared with state-of-the-art techniques such as Multilayer Perceptron, Random Forest, Support Vector Machine, and Naïve Bayes [116]; the Markov Logic Network [118]; the Markov chain model [119]; the Inhomogeneous Markov model compared with the Probability Sampling (PS), Artificial Neural Network (ANN) and Support Vector Regression approaches [97]; the Complex Activity Recognition using Emerging patterns and Random Forest (CARER) compared with Hidden Markov Model, Bayesian Network, Naïve Bayes, SVM, Decision Tree, and Random Forest [81]; the Markov Logic Network [129]; an original proposed model, compared with the results obtained when using the Hidden Markov Model and the Conditional Random Field Model [131]; semi-supervised learning algorithms and Markov-based models [132]; the Markov modulated multidimensional non-homogeneous Poisson process (M3P2) compared with the classical Markov modulated Poisson process (MMPP) [128]; a coupled Hidden Markov Model [127]; semantical Markov Logic Network [137]; Markov Logic Network (MLN) compared with Artificial Neural Network (ANN), Support Vector Machine, Bayesian Network (BN) and Hidden Markov Model [121]; two different approaches: a factorial Hidden Markov

model for modeling two separate chains corresponding to two residents, and nonlinear Bayesian tracking for decomposing the observation space into the number of residents.

The performance metrics that chosen by the authors of the scientific papers using the Hidden Markov Model integrated with sensor devices in smart buildings included: Accuracy [3,10,25,33,70, 115,117,120,122,123,125,127,131,133,136,138]; Precision [25,118,128,133,135,137]; Recall [25,118,128,135]; F-Measure [25,81,121,130,133]; Sensitivity and Specificity [25,33,133]; F1 Score [116,133]; Confusion Matrix [116,127,129]; and Correctness [97,118]. In addition to the above-mentioned performance metrics, other methods that were used to assess the performance of the developed methods by the authors of the scientific papers selected and summarized in Table S13 included: a numerical case study highlighting the efficiency of the developed model [134]; thread latency [119]; evaluation of energy savings [135]; memory and response time requirements [136]; Mean Squared Error (MSE) [3]; Receiver Operating Characteristic (ROC) scores computed based on the True Positive Rates against the False Positive ones [97]; Mean Recognition Rate [10]; Leave-One-Subject-Out-Cross-Validation (LOSOCV) [129]; execution speed [127]; Local Outlier Factor (LOF), the Z-Score values, cluster transition probability [82]; the APL: Average Path Length, LTA: Location and Time Accuracy, PRDOS: Pressure of Receiving Data On Sink Node, and APRDOS: average PRDOS of sink node [122]; the probability of error [124]; a series of experiments along with the F-Value [128]; simulation tests in order to compare the Generalized Version Space (GVS) algorithm with a simple method using an epsilon greedy mechanism [132]; the Area Under the ROC Curve (AUC) [33]; Correlation Factors depicting the similarities between simulated and real displacement activities [126]; and the heuristic merit of a sensor feature subset S containing k features [123].

With respect to the five most recent scientific articles addressing the Nearest Neighbor method integrated with sensor devices in smart buildings (Table 13), it can be observed that in [25], Arifoglu et al. analyzed the possibility of detecting abnormal behavior in elderly people in order to identify early indicators and symptoms associated with a decline in memory, indicating dementia or brain disease, by making use of Convolutional Neural Networks. After identifying patterns within the daily activity and abnormal activities within them, the authors compared the performance of their approach with those obtained when using other methods, such as Naïve Bayes, Hidden Markov Models (HMMs), Hidden Semi-Markov Models, and Conditional Random Fields (computing the Precision, Recall, F-measure and Accuracy, Sensitivity, Specificity), and concluded that the developed approach was comparable with the state-of-art methods.

In [115], Papatsimpa et al. addressed issues regarding the human presence in a smart building equipped with a Wireless Sensor Network, making use of various Hidden Markov Models (HMMs). The authors proposed a method based on an efficient transmission strategy along with a blending algorithm that was designed to combine data from various Hidden Markov Models perceiving the same Markovian process. To evaluate their approach, the authors analyzed a series of experimental results and stated that these results confirmed the functionality and benefits of their developed method. Taking into account the accuracy of their scheme, along with the reduction in terms of communication requirements, the authors concluded that their method was suitable and applicable for many situations requiring information merging in wireless sensor devices.

In [116], Civitarese et al. focus on human activity recognition with a view to developing an affordable ambient assisted living approach, ensuring the individual's data privacy. To this end, the authors developed a hybrid approach, combining collaborative active learning with probabilistic and knowledge-based reasoning. The authors developed the newNECTAR framework, which was based on the Markov Logic Network, and compared it with state-of-the-art techniques (such as Multilayer Perceptron, Random Forest, Support Vector Machine, Naïve Bayes). The authors concluded that their developed learning solution improved recognition rates, generated a reduced number of feedback requests, and was comparable and sometimes even better than other existing activity recognition methods based on the performance metrics used (the Average F1 Score and Confusion Matrix).

In [117], Dahmen et al. analyzed methods for "testing machine learning techniques for healthcare applications", aiming to overcome the limitations related to the complexity and lack of applicability of many actual approaches. To this end, the authors developed a synthetic data generation method based on Machine Learning techniques, SynSys. The authors made use of Hidden Markov Models and regression models, and afterwards, they tested the generated set of synthetic data on a dataset recorded from a real smart home. To evaluate the developed approach, the authors made use of the following performance metrics: the Average Accuracy using real data, synthetic data and randomly generated data; the Accuracy first using only the real data, and then the Accuracy using the real data enlarged by a month of synthetically generated data. The authors concluded that their data generation method had the ability to provide a higher human activity recognition accuracy than that obtained when solely using real data.

In paper [118], Sfar et al. developed an approach for early detection of abnormal behavior in elderly people living in smart homes, in order to prevent risks related to their health, based on identifying and extracting anomalous causes from datasets, making use of causal association rules mining. These causes were subsequently used in order to detect the risks of anomalies occurring by using the Markov Logic Network Machine Learning method. The authors evaluated their approach by using real datasets, concluding that the devised method proved to be efficient in terms of the computed performance metrics (Precision, Recall, Recognition Rate and Correctness).

Subsequently, from obtained pool of scientific articles obtained based on the devised review methodology, we identified, analyzed and summarized those making use of Hierarchical Clustering integrated with sensor devices in smart buildings. A complete summarization table (Table S14) is presented in the Supplementary Materials file, while Table 14 presents the most recent papers targeting this subject.

Table 14. The most recent scientific articles addressing the Hierarchical Clustering integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Hierarchical Clustering Approach with Sensor Devices	Hierarchical Clustering Only or Hybrid	Performance Metrics
[19]	2019	smart building	smartphone sensors and Bluetooth beacons data	group activity detection and recognition	a framework for indoor Group Activity Detection and Recognition (GADAR) and Hierarchical Clustering, along with Decision Tree classifier, K-Neighbors classifier, Deep Neural Network, Gaussian Process classifier, Logistic regression, Support Vector Machine, Linear Discriminant Analysis, Gaussian Naïve Bayes (comparison)	Confusion Matrix, Accuracy (Mean), Accuracy (Variation), Precision, Recall, F1 Score
[37]	2019	smart building	WiFi-enabled IoT device-user	Personalized location-based service	hybrid: Hierarchical Clustering and Location Similarity Matching	Accuracy
[139]	2014	smart building	smart meters organized into clusters	data collection in hierarchical smart building networks	hybrid hierarchical clustering containing a two-layer transmission process	simulations scenarios, comparison of the proposed scheme's performance with the performance of the Uniform Algorithm

Analyzing the papers selected and summarized in Table S14, presented in the Supplementary Materials file, it can be observed that all of them analyze smart buildings in general. In these papers, the authors make use of different types of sensors, for example: smartphone sensors and Bluetooth beacons data [19]; WiFi-Enabled IoT Device-User [37]; smart meters organized into clusters [139]. In these

papers, the reasons for using the Hierarchical Clustering approach with the sensor devices in smart buildings are related to group activity detection and recognition [19]; Personalized Location-Based Services [37]; and data collection in hierarchical smart building networks [139].

Regarding the devised research methods, in [19], Chen et al. made use of a hybrid approach, combining a framework for indoor group activity detection/recognition and hierarchical clustering, along with the Decision Tree classifier, K-Neighbors classifier, Deep Neural Network, Gaussian Process Classifier, Logistic Regression, Support Vector Machine, Linear Discriminant Analysis, Gaussian Naïve Bayes, making a comparison between these techniques. In [37], Zou et al. developed a hybrid approach, combining Hierarchical Clustering and location similarity matching. In [139], Luan et al. made use of a hybrid Hierarchical Clustering containing a two-layer transmission process.

The performance metrics considered in the scientific papers that use the Hierarchical Clustering method integrated with sensor devices in smart buildings include: the Confusion Matrix, Accuracy (Mean), Accuracy (Variation), Precision, Recall, F1 Score [19]; Accuracy [37]; and the development of simulated scenarios and a comparison of the proposed scheme's performance with that of the uniform algorithm, in which the cluster heads are uniformly distributed and the resources are uniformly allocated [139].

With respect to the most recent scientific articles addressing the Hierarchical Clustering method integrated with sensor devices in smart buildings (Table 14), it can be observed that in [37], Zou et al. addressed personalized location-based services in smart buildings. To this end, the authors developed a method that used a non-intrusive device, based on WiFi technology, and an association scheme based on an unsupervised learning algorithm. The authors developed a hybrid approach, combining Hierarchical Clustering and location similarity matching. To test the performance of the developed approach, the authors conducted a series of experiments and, using Accuracy as a performance metric, concluded that their method had the potential to be implemented in real-world situations, "for practical personalized context-aware and location-based services in the era of IoT".

The scientific paper [19] was reviewed previously when analyzing the most recent scientific articles that integrate Decision Tree approaches with sensor devices in smart buildings (Table 6).

In [139], Luan et al. proposed a hybrid cooperation scheme useful in collecting data in hierarchical smart buildings networks, making use of machine-to-machine communication. In this study, the authors used smart meters organized into clusters as sensors, sending information to the cluster-heads. The authors developed hybrid Hierarchical Clustering, containing a two-layer transmission process. In the first-layer transmission, the distributed smart meters send the data to their respective cluster heads. In the second-layer transmission, the cluster-heads forward all of the data to the base station. With a view to highlighting the advantages and properties of their developed scheme, the authors developed a series of simulated scenarios and compared the proposed scheme's performance with that of the uniform algorithm, whereby the cluster heads were uniformly distributed and the resources were uniformly allocated.

Subsequently, from the obtained pool of scientific articles resulting from the application of the devised review methodology, we identified, analyzed and summarized those making use of the K-Means integrated with sensor devices in smart buildings for classification purposes. A complete summarization table (Table S15) is presented in the Supplementary Materials file, while Table 15 presents the most recent papers addressing this subject.

Examining the papers selected and summarized in Table S15, presented in the Supplementary Materials file, it can be observed that 67% of them take into consideration smart buildings in general, while the remaining 33% refer to smart homes. The authors of these scientific articles made use of different types of sensors in their analyses, including binary sensors [26]; sensor networks [140]; smart meters, Personal Weather Stations (PWS), and sensors providing data useful in computing the mean values of: hourly indoor temperature, hourly outdoor temperature, hourly value of precipitation, hourly value of wind direction, hourly value of solar radiation, hourly value of ultraviolet index, hourly value of humidity, hourly value of pressure [42].

Table 15. The most recent scientific articles addressing the K-Means integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the K-Means Method with Sensor Devices	K-Means Only or Hybrid	Performance Metrics
[26]	2018	smart home	binary sensors	extraction of behavioral patterns	hybrid: K-Means Algorithm combined with Nominal Matrix Factorization method	comparison with existing methods based on both synthetic and publicly available real smart home datasets
[140]	2018	smart buildings	sensor network	discovering electricity consumption patterns	Cluster Validation Indices (CVIs) for establishing the optimal number of clusters k for the dataset, combined with the Parallelized Version of K-Means Clustering Algorithm for discovering patterns from the dataset	Cluster Analysis, Centroids of the electricity consumption clusters, Centroids of the clusters with lower consumptions, Computing times
[42]	2018	smart building	Smart meters, Personal Weather Stations (PWS), sensors providing data useful in computing the mean values of: hourly indoor temperature, hourly outdoor temperature, hourly value of precipitation, hourly value of wind direction, hourly value of solar radiation, hourly value of ultraviolet index, hourly value of humidity, hourly value of pressure	managing energy consumption	Data Mining Engine, METATECH (METeorological data Analysis for Thermal Energy CHaracterization)	Support, Confidence and Lift

In these papers, the reasons for using the K-Means method with the sensor devices in smart buildings were related to extraction of behavioral patterns [26]; determining electricity consumption patterns [140]; and managing energy consumption [42]. With respect to the devised research methods, in [26], Li et al. made use of a hybrid approach, combining the K-Means algorithm with Nominal Matrix Factorization method. In [140], Pérez-Chacón et al. used the Cluster Validation Indices (CVIs) method to establish the optimal number of clusters for the dataset, combined with the parallelized version of K-Means clustering algorithm for discovering patterns from the dataset. In [42], Di Corso et al. implemented the data mining engine, METATECH (METeorological data Analysis for Thermal Energy CHaracterization), which computes the similarity between two objects by using the Euclidean distance, and integrates a partitional algorithm, the K-Means algorithm.

The performance metrics considered in the scientific papers using the K-Means integrated with sensor devices in smart buildings were evaluated based on a comparison with existing methods based on both synthetic and publicly available real smart home datasets [26]; cluster analysis, centroids of the electricity consumption clusters, centroids of the clusters with lower consumptions, and computing times [140]; and support, confidence and lift [42].

Regarding the most recent scientific articles that make use of the K-Means method along with sensor devices in smart buildings (Table 15), it can be observed that in [26], Li et al. aimed to devise a methodology for the automatic detection of the behavioral patterns of elderly people living in smart

homes. The authors made use of binary sensors and devised a hybrid approach, combining the K-Means algorithm with Nominal Matrix Factorization method in order to obtain the daily routines. To assess the performance and suitability of their method, the authors compared their developed approach with existing methods based on both synthetic and publicly available real smart home datasets and considered their obtained results to be promising.

In [140], Pérez-Chacón et al. proposed a method for identifying patterns in big data time series with respect to energy consumption in smart buildings, making use of sensor networks. The authors based their approach on Cluster Validation Indices (CVIs) for establishing the optimal number of clusters for the dataset, combined with the parallelized version of K-Means clustering algorithm (from the Apache Spark's Machine Learning Library) in order to discover patterns from the dataset. The devised method was tested using a large dataset, representing the energy consumption of eight smart buildings over a seven-year period (2011–2017). As performance metrics, the authors used cluster analysis, centroids of the electricity consumption cluster, and centroids of the clusters with lower consumptions, along with computing times, and concluded that their devised approach represented a valuable tool for the optimization of energy usage.

In paper [42], Di Corso et al. proposed a data mining engine, METeorological Data Analysis for Thermal Energy CHaracterization (METATECH), which computes the similarity between two objects by using the Euclidean distance, and integrates a partitional algorithm, the K-Means algorithm. The authors made use of various types of sensors, including Smart meters, Personal Weather Stations (PWS), and sensors providing data useful in computing the mean values of: hourly indoor temperature, hourly outdoor temperature, hourly value of precipitation, hourly value of wind direction, hourly value of solar radiation, hourly value of ultraviolet index, hourly value of humidity, and hourly value of pressure [42]. The devised method aimed to develop models for correlating meteorological conditions and the energy consumption in smart buildings at various levels of granularity. To validate the devised approach, the authors performed a series of experimental tests using real datasets and concluded that these tests highlighted the effectiveness of their method in the process of data mining.

3.3. Deep Learning Techniques

Taking into account recent increases in the computational power of hardware processing architectures (especially parallel processing ones), which have led to the widespread application of Deep Learning techniques, in addition to the above-mentioned categories, we also identified, analyzed and summarized, with respect to the obtained pool of scientific papers, those that make use of Deep Learning techniques with sensor devices in the smart building sector. A selection of the most recent papers (sorted in descending order of publication year) is presented in Table 16, while a comprehensive summarization table can be found in the Supplementary Materials file (Table S16).

It can be observed that 78% of the scientific papers selected and summarized in Table S16, presented in the Supplementary Materials file, focused their research exclusively on smart homes, while 17% focused on smart buildings in general, and the remaining 5% focused on smart commercial and residential buildings.

In these papers, the authors made use of different types of sensors, including motion sensors [18, 25,28,141–145]; temperature sensors [28,40,73,143,144]; wireless sensor networks [21,40,141,145]; door sensors [25,143]; smartphone inertial sensors [146] and a smartphone application [36]; cameras [18]; a two-dimensional acoustic array [27]; daily activity recognition sensors [28]; actuators [143]; tactile sensors, power meters, and microphones in the ceiling [144]; non-wearable sensors [147]; unobtrusive sensors [9]; environmental sensors [73,142]; weather sensors [12]; WiFi-enabled sensors for food nutrition quantification [36]; and binary sensors [148].

Table 16. Five of the most recent scientific articles addressing Deep Learning techniques integrated with sensor devices in smart buildings.

Reference	Publication Year	Type of Smart Building	Type of Sensors	Reason for Using the Deep Learning Method with Sensor Devices	Deep Learning Only or Hybrid	Performance Metrics
[18]	2019	smart home	wearable hybrid sensor system comprising motion sensors and cameras	human activity recognition in medical care, smart homes, and security monitoring	hybrid approach, combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) methods	Confusion Matrices, F1 Accuracy
[27]	2019	smart home	a two-dimensional acoustic array	human activity recognition	Convolutional Neural Networks compared with traditional recognition approaches such as K-Nearest Neighbor and Support Vector Machines	Overall Accuracy
[28]	2019	smart home	daily activities recognition sensors, infrared motion and temperature sensors	human activity recognition	hybrid, using Term Frequency-Inverse Document Frequency (TF-IDF), along with each of the Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), and Random Forest (RF), Long Short-Term Memory (LSTM) methods and comparison between them	Accuracy, Precision, and F-Measure
[25]	2019	smart home	34 sensors (3 door and 31 motion sensors)	sensor-based activity recognition and abnormal behavior detection	Convolutional Neural Networks (CNNs) for detecting abnormal behavior related to dementia, the results being compared with methods such as Naïve Bayes (NB), Hidden Markov Models (HMMs), Hidden Semi-Markov Models (HSMM), Conditional Random Fields (CRFs)	Precision, Recall, F-Measure and Accuracy, Sensitivity, Specificity
[21]	2019	smart building	wireless sensor networks	forecasting Packet Delivery Ratio (PDR) and Energy Consumption (EC) in Internet of Things (IoT)	comparison between Linear Regression, Gradient Boosting, Random Forest, Baseline and Deep Learning Neural Networks	Root Mean Square Error (RMSE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE)

In the scientific papers selected and summarized in Table S16, the reasons for using Deep Learning techniques integrated with sensor devices in smart buildings were mainly related to human activity recognition [9,18,25,27,28,73,142,143,145–148]; ensuring health care [18,25,142]; forecasting Packet Delivery Ratio (PDR) and Energy Consumption (EC) in Internet of Things (IoT) [21]; realizing small and big data management [141]; adaptive decision-making in smart homes [144]; thermal comfort modeling [40]; forecasting the electricity consumption [12]; and Internet of Things (IoT)-based fully automated nutrition monitoring systems [36].

With respect to the devised methods, in paper [18], the authors made use of a hybrid approach, combining the Long Short-Term Memory (LSTM) networks with the Convolutional Neural Network (CNN) approach. In [27], the authors implemented Convolutional Neural Networks, and compared them with traditional recognition approaches such as K-Nearest Neighbor and Support Vector Machines. The authors of [28] developed a hybrid approach, using Term Frequency-Inverse Document Frequency (TF-IDF), along with the Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), and Random Forest (RF), Long Short-Term Memory (LSTM) methods and compared them. In [25], the authors made use of Convolutional Neural Networks (CNNs) for detecting abnormal behavior related to dementia, the results were compared with methods such as Naïve Bayes (NB), Hidden Markov Models (HMMs), Hidden Semi-Markov Models (HSMM), and Conditional Random Fields (CRFs). In [21], Ateeq et al. compared the Linear Regression, Gradient Boosting, Random Forest, Baseline and Deep Learning Neural Networks. The authors of [141] used Deep Neural Networks for system monitoring and optimization. In [146], the authors implemented a Deep Belief Network (DBN), comparing it with Support Vector Machine (SVM) and Artificial Neural Network (ANN) approaches. In [142], the authors developed a hybrid Deep Learning-based gesture/locomotion recognition model, integrating CNN and RNN. In [143], the authors made use of different Deep Learning (DL) models based on Long Short-Term Memory (LSTM), comparing their approach with the Hidden Markov Model (HMM), Conditional Random Field (CRF), and Naïve Bayes (NB) approaches. In [144], the authors developed a hybrid method, namely the Adaptive Reinforced Context-Aware Deep Decision System (ARCADES), combining Deep Neural Networks and Reinforcement Learning (RL). In [145], the authors compared Recurrent Neural Networks (Long Short-Term Memory, Gated Recurrent Units), Convolutional Neural Network, Behavior Explanatory Models, and Sensor Profiles. In [147], the authors developed a Deep Learning technique, namely the Recurrent Neural Network (RNN), using the Long Short-Term Memory (LSTM) architecture. In [9], the authors made use of the SVM classifier along with two different feature extraction methods: a manually defined method and a Convolutional Neural Network (CNN). The authors of [40] developed an intelligent Thermal Comfort Management (iTCM) system black-box neural network (ITCNN), whose performance was compared with the Fanger's Predicted Mean Vote (PMV) model and six classical machine learning approaches: three traditional white-box machine learning approaches and three classical black-box machine learning methods. In [73], the authors made use of a Deep Convolutional Neural Network (DCNN), comparing it with the Naïve Bayes (NB) and Back-Propagation (BP) algorithms. In [12], the authors used deep Recurrent Neural Network (RNN) models. In [36], the authors implemented Bayesian algorithms and the 5-layer Perceptron Neural Network method for diet monitoring. In [148], the authors developed an Activity Recognition (AR) model based on Deep Learning for two cases: one-layer Denoising Autoencoder (DAE) and two-layer Stacked Denoising Autoencoder (SDAE). The results obtained were compared with those obtained by five commonly used baselines: Naïve Bayes (NB), Hidden Markov Model (HMM), Hidden Semi-Markov Model (HSMM), K-Nearest-Neighbor (KNN), and Support Vector Machine (SVM) with linear kernel.

The performance metrics chosen by the authors of the papers focusing on Deep Learning techniques integrated with sensor devices in smart buildings included Confusion Matrices and F1 Accuracy [18]; Overall Accuracy [27]; Accuracy, Precision and F-Measure [28]; Precision, Recall, F-Measure, Accuracy, Sensitivity, and Specificity [25]; Root Mean Squared Error (RMSE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE) [21]; Overall Accuracy and Mean Recognition Rate [146]; Accuracy, Precision, Recall and F1 Score [142,143]; reward per episode, Precision, Recall, F1 Score [144]; methods discussed and evaluated on the basis of real-life data and the Confusion Matrix [145]; Accuracy and Root Mean Square Error (RMSE) [9]; energy cost savings [40]; Precision, Specificity, Recall, F1 Score, Accuracy, Total Accuracy, and Confusion Matrix [73]; Root Mean Squared Error relative to Root Mean Squared (RMS) average of electricity consumption in test data, Root Mean Squared Error relative to Root Mean Squared (RMS), average of electricity consumption in training data, and Pearson

Coefficient [12]; Accuracy of classification of food items and meal prediction [36]; and time-slice accuracy and class accuracy [148].

With respect to the most recent scientific articles that make use of Deep Learning techniques along with sensor devices in smart buildings (Table 16), it can be observed that papers [18] and [27] were reviewed previously when analyzing the most recent scientific articles integrating Neural Networks for classification purposes with sensor devices in smart buildings (Table 5). Paper [25] was reviewed when analyzing the most recent scientific articles integrating the Hidden Markov Model with sensor devices in smart buildings (Table 13). Article [21] was detailed when analyzing the most recent scientific articles integrating Decision Tree with sensor devices in smart buildings (Table 6).

In paper [28], Guo et al. aimed to achieve human activity recognition based on a non-invasive method in order to improve residents' lives. In their research, the authors made use of daily activity recognition sensors, and infrared motion and temperature sensors, and developed a hybrid approach using Term Frequency–Inverse Document Frequency (TF-IDF), along with the Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), Random Forest (RF), and Long Short-Term Memory (LSTM) methods, carrying out a comparison between them. By computing the Accuracy, Precision and F-Measure performance metrics, the authors evaluate the Machine Learning methods and Deep Learning technique, thereby concluding that their strategy, based on the Term Frequency-Inverse Document Frequency (TF-IDF) approach, has the potential to improve the performance of human activity recognition systems.

In the following, we review the most frequently cited articles from the scientific papers pool addressing the reviewed topics, as reported by the two considered international databases.

3.4. Frequently Cited Scientific Papers Addressing the Reviewed Topics, as Reported by the Elsevier Scopus and the Clarivate Analytics Web of Science International Databases

We devised our research methodology and conducted our review with a view to identifying, filtering, categorizing, and analyzing the most important and relevant scientific articles with respect to recent developments in the integration of machine learning models with sensor devices in the smart buildings sector with a view to attaining enhanced sensing, energy efficiency, and optimal building management. Therefore, we focused our attention on the most recent scientific papers, meanwhile being aware of the fact that these topics represent an important subject, and that new research is disseminated day by day throughout the scientific literature. In addition to this, the choice to review the most recent scientific works addressing developments concerning the integration of machine learning models with sensor devices in the smart buildings sector offers the possibility of grasping the recent advancements in technology and sensing equipment.

Another criterion that can be addressed when devising a review paper is based on the visibility of the papers in the scientific literature, evaluated on the basis of their number of citations. Nevertheless, this approach has its disadvantages, due to the fact that in this way, the most recent papers may not be taken into account, as they have not had the chance to be cited as frequently as those published at an earlier date, as sufficient time has not yet elapsed since their publication. However, in order to highlight the most visible papers in the scientific literature that address the reviewed topics, in addition to the above-mentioned analysis, we also identified, analyzed and summarized from the obtained scientific papers pool the most frequently cited scientific papers, as reported by the Clarivate Analytics Web of Science (WoS) and the Elsevier Scopus (ES) international databases. These papers are summarized in Table 17, sorted into descending order of number of citations.

Table 17. The most frequently cited scientific articles addressing Machine Learning Models integrated with sensor devices in smart buildings as reported by the WoS and the ES international databases.

Reference	Publication Year	Number of Citations According to		Type of Smart Building	Type of Sensors	Reason for Using the Machine Learning Models with Sensor Devices	Machine Learning Models Only or Hybrid	Performance Metrics
		WoS	ES					
[44]	2012	170	197	smart building	energy smart meters, building management systems, and weather stations	energy consumption forecasting	a model based on Support Vector Regression (SVR) using the Scikit-learn module, which provides a Python front-end to LIBSVM, a widely cited Support Vector Machine library	Coefficient of Variation (CV) and Standard Error In %
[75]	2011	79	118	smart home	Passive Infra-Red (PIR) sensors or motion detectors; door/window entry point sensors; electricity power usage sensors; bed/sofa pressure sensors; flood sensors	human activity recognition for detecting and predicting abnormal behavior	Echo State Network (ESN), Back Propagation Through Time (BPTT) and Real Time Recurrent Learning (RTRL)	Root Mean Square Error (RMSE)
[123]	2011	60	76	smart home	wireless sensor network highlighting the user movement (i.e., both hands), user location, human-object interaction (i.e., objects touched and sound), human-to-human interaction (i.e., voice), environmental information (i.e., temperature, humidity and light)	human activity recognition	Coupled Hidden Markov Model (CHMM) and Factorial Conditional Random Field (FCRF)	Accuracy, the heuristic merit of a sensor feature subset S containing k features
[65]	2007	54	66	smart home	sensors for HVAC Chillers	optimal sensor selection in complex system monitoring problems	comparison of: Support Vector Machines (SVMs), Principal Component Analysis (PCA), and Partial Least Squares (PLS)	Recognition Rate
[138]	2007	53	64	smart home	smart meter	load disaggregation	Nonintrusive Load Monitoring (NILM), Hidden Markov Models	Accuracy

Analyzing the papers selected and summarized in Table 17, it can be observed that 80% of them focus exclusively on smart homes, while the remaining 20% take into consideration smart buildings in general. The authors of these scientific articles make use of different types of sensors in their analyses, including energy smart meters, building management systems, and weather stations [44]; Passive Infra-Red (PIR) sensors or motion detectors; door/window entry point sensors; electricity power usage sensors; bed/sofa pressure sensors; flood sensors [75]; wireless sensor network highlighting user movement, user location, human-object interaction, human-to-human interaction, environmental information [123]; sensors for HVAC chillers [65]; and smart meters [138].

In these papers, the reasons for using Machine Learning Models with sensor equipment in the smart buildings are mainly related to the recognition of human activity [75,123]; forecasting of energy consumption [44]; optimal sensor selection in complex system monitoring problems [65]; and load disaggregation [138].

With respect to the devised methods, in [44], Jain et al. developed a model based on Support Vector Regression (SVR). In [75], Lotfi et al. made use of the Echo State Network (ESN), Back Propagation Through Time (BPTT) and Real-Time Recurrent Learning (RTRL) methods. In [123], Wang et al. made use of the Coupled Hidden Markov Model (CHMM) and Factorial Conditional Random Field (FCRF) methods. In [65], Namburu et al. compared Support Vector Machines (SVMs), Principal Component Analysis (PCA), and Partial Least Squares (PLS) methods. In [138] Egarter et al. solely implemented the Hidden Markov Model.

The performance metrics chosen by the authors of the most frequently cited scientific articles addressing Machine Learning Models integrated with sensor devices in smart buildings reported by the WoS and the ES International Databases included the Coefficient of Variation (CV) and Standard Error [44]; Root Mean Square Error (RMSE) [75]; Accuracy, the heuristic merit of a sensor feature subset containing a certain number of features [123]; Recognition Rate [65]; and Accuracy [138].

By analyzing the most frequently cited scientific articles addressing Machine Learning Models integrated with sensor devices in smart buildings reported by the Clarivate Analytics Web of Science and the Elsevier Scopus international databases (Table 17), it can be observed that in [44], Jain et al. started their study by highlighting the importance of the accurate forecasting of a building's energy consumption in order to achieve appropriate, efficient urban energy management. To this end, the authors developed a forecasting model based on the Support Vector Regression method, and applied it to a residential building in New York City, endowed with various types of sensors such as weather stations, smart meters and building management systems. The authors analyzed the impact of spatial and temporal granularity on forecasting accuracy by taking into consideration several parts of the building and a variety of time intervals. By comparing the obtained results, using the Coefficient of Variation (CV) and the Standard Error as performance metrics, the authors concluded that the best results were those registered when forecasting the energy consumption at the floor level, with an hourly timeframe.

In [75], Lotfi et al. proposed a method for monitoring the activities of elderly people living alone in homes equipped with sensor networks (comprising motion and door sensors) by detecting and predicting any abnormal behavior. The authors presented methods for analyzing the large datasets retrieved from the sensors, representing them in formats that were suitable for grouping the abnormalities. Subsequently, they used recurrent neural networks in order to predict potential upcoming values of the activities monitored by each implemented sensor. Thereby, if an abnormal behavior were forecasted to take place, health professionals could be informed. The authors compare their Echo State Network (ESN) approach with those based on other recurrent neural network techniques such as the Back Propagation Through Time (BPTT) and Real-Time Recurrent Learning (RTRL), using the Root Mean Square Error (RMSE) and the training time as performance metrics, concluding that the forecasting results provided by the ESN approach were better than those of the other two approaches with respect to training time. The developed forecasting method was evaluated by implementing it in a smart home inhabited by elderly people suffering from brain diseases.

A wireless sensor network highlighting environmental information, user location, user movement, human-to-human interactions, and human-object interactions was used by Wang et al. in [123] with the aim of multi-user activity recognition in smart homes. The authors made use of a wearable sensor platform in order to retrieve data from multiple users, modeling the interaction processes by the means of two models, namely, the Coupled Hidden Markov Model (CHMM) and the Factorial Conditional Random Field (FCRF). The authors conducted a series of experiments in order to assess the performance of the two developed probabilistic models, concluding that the CHMM model provided an accuracy of 96.41%, while the FCRF model registered an accuracy of 87.93% with respect to multi-user activity recognition.

Acknowledging the importance of the Chillers as components in Heating, Ventilating and Air-Conditioning (HVAC) systems, and the fact that they involve significant energy consumption, in [65], Namburu et al. proposed a generic Fault Detection and Diagnosis (FDD) scheme for centrifugal chillers and "a nominal data-driven model of the chiller" that could be useful in forecasting the system response under changing loading conditions. The authors made use of sensors for HVAC Chillers in order to achieve "an optimal sensor selection in complex system monitoring problems", and compared the Support Vector Machines (SVMs), Principal Component Analysis (PCA), and Partial Least Squares (PLS) classification techniques using the Recognition Rate as a performance metric. Using an approach based on a genetic algorithm, the authors selected the sensor suite that was most suitable for forecasting system response in the context of new loading conditions and also assessed the performance provided by the above-mentioned classification techniques when using the identified sensor suite. Using the loading conditions obtained through the nominal model, the authors forecast the responses of the sensor suite. Afterwards, the authors used real HVAC equipment in order to obtain a benchmark dataset for use in validating the developed approach.

In [138], Egarter et al. addressed issues regarding Particle Filter-Based Load Disaggregation (PALDi) in smart homes. The authors commenced their study by highlighting the fact that smart meters provide information that can be used in order to disaggregate appliance consumption by means of Nonintrusive Load Monitoring (NILM), a method that analyzes the consumption provided by the smart meter device within the smart home and identifies the appliances that are being used in the house, along with their individual associated consumption. The authors made use of the NILM method and estimated the appliance states using the particle filtering approach. Using Hidden Markov Models for modeling the appliances and their combinations, the authors obtained a description of the household power demand. Afterwards, in order to evaluate the developed approach, the authors made use of generated and real datasets and concluded that their method registered an accuracy of 90% when detecting the appliance states in the real dataset case.

4. Discussion and Conclusions

The conducted review focused on recent developments in the scientific literature with respect to the integration of Machine Learning models with sensor devices in the smart buildings sector with a view to attaining enhanced sensing, energy efficiency, and optimal building management. To ensure the quality and reliability of the reviewed works, prominent scientific databases (the Elsevier Scopus and the Clarivate Analytics Web of Science) were used as a means to devise custom tailored queries.

In contrast to other, previously existing review papers, our approach was focused on recent scientific articles, highlighting and comparing, for these papers, the details regarding publication year, type of smart building, types of sensor device implemented, reason for using the respective method with sensor devices, developed approach, and the performance metrics implemented in the study. We first conducted an overall comparative analysis of the pool of scientific papers identified according to the devised review methodology with respect to a previously identified and constructed taxonomy. Subsequently, for each taxonomy branch, the most recent scientific articles were analyzed separately, emphasizing the details of the implementation, along with the specific aspects pertaining to the respective papers.

A review of the most recent scientific articles that deal with emergent topics like machine learning, sensor devices and smart buildings offers a series of undeniable advantages in terms of categorizing a high number of scientific articles according to a clear, comprehensive taxonomy. This review article offers a useful up-to-date overview for researchers from different fields who may wish to submit a project proposal or study complex topics like those reviewed.

At the same time, by reviewing recent advancements in the integration of Machine Learning models with sensor devices in the smart buildings sector, the current study offers scientists the possibility of identifying future research directions that have not yet been addressed in the scientific literature or of improving the approaches that already exist within the body of knowledge. The conducted review provides the possibility of identifying the main applications for which approaches have been developed in the literature integrating Machine Learning techniques with sensing devices in smart environments, as well as those applications that have not yet been pursued.

An important challenge that still remains after decades of evolving research in the semiconductors field is the need to develop novel low-power sensing equipment, considering that the vast majority of sensing devices rely for their operation on different power sources, thereby incurring power consumption costs for the acquisition, processing and transmission of the data streams in addition to the physical wiring installation and maintenance costs when using them at the level of an entire smart building. As can be seen from the results of the performed survey, several methods process the data locally, while others adopt a cloud-based approach. Both of these proposed approaches raise important challenges with regard to data processing, power consumption and data transmission power consumption costs. While local processing of the acquired data consumes computational and power resources on the long run, uploading the data into the cloud raises several security-related challenges, including confidentiality, authenticity, integrity, non-repudiation, and accountability. In addition, there is a need for future studies to focus on developing optimized compression algorithms and uploading schemes for the acquired data into the cloud systems, considering that this process is consumes resources from an energy requirements point of view. It is the authors' opinion that the integration of machine learning techniques in sensing equipment benefits not only enhanced sensing, but the development of optimized processing and uploading strategies, in the end leading to a reduction in the overall energy consumption.

When analyzing the pool of scientific works obtained after applying the devised review methodology, we noticed an important aspect that had not been taken into consideration by the scientific papers focusing on human-centric society and on the improvement of the life quality, namely, the perceived notion of "comfort". According to the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) 25010 specifications [149], comfort is defined as the "degree to which the user is satisfied with physical comfort", and this physical comfort can often be a matter of individual perception, being dependent to some extent to a human being's acoustic, visual, thermal and sensorial traits, while also being influenced by gender, age, and overall health status.

An important aspect that should be further studied by researchers and implemented in practice is improving the data security and privacy of IoT systems, due to the fact that most of the data that resulting from the processes highlighted by our review paper, in which machine learning models are integrated with sensor devices in the smart buildings sector, contain sensitive, personal information related to the inhabitants of the respective buildings. These data must therefore be protected. In addition to this, the entire ecosystem of hardware and software components is also vulnerable, and threat protection must therefore evolve accordingly. The above-mentioned vulnerabilities could be overcome by means of appropriate technologies designed to protect data, networks, systems, and devices from malicious attacks, implementing cryptography, securing both the hardware and software components, and ensuring communication protection in order to prevent unauthorized access to private information, avoid the interruption of communications, and guarantee the accuracy of information managed by the respective system.

Even if the developed review covers the most relevant and important actual scientific articles dealing with the above-mentioned research topics, we are aware of the fact that, as with any other review paper, this is affected by the rapid development of the body of knowledge with regard to the reviewed topics, which is strongly correlated with the extremely rapid evolution of the technology, of sensor devices, and of machine learning approaches.

With respect to future work, we will aim to conduct a review of the most relevant patents awarded, along with those that are pending, that propose methods and devices related to the fusion of machine learning techniques with sensor devices in the smart buildings sector. In our opinion, this is an aspect worth being studied and reviewed, considering the numerous existing patents that have not been disseminated yet as scientific articles in the literature.

Supplementary Materials: The following are available online at http://www.mdpi.com/1996-1073/12/24/4745/s1. Table S1: Scientific articles addressing the Support Vector Machines integrated with sensor devices in smart buildings; Table S2: Scientific articles addressing the Discriminant Analysis integrated with sensor devices in smart buildings; Table S3: Scientific articles addressing the Naïve Bayes integrated with sensor devices in smart buildings; Table S4: Scientific articles addressing the Nearest Neighbor integrated with sensor devices in smart buildings; Table S5: Scientific articles addressing the Neural Networks for Classification Purposes integrated with sensor devices in smart buildings; Table S6: Scientific articles addressing the Decision Tree integrated with sensor devices in smart buildings; Table S7: Scientific articles addressing the Ensemble Methods integrated with sensor devices in smart buildings; Table S8: Scientific articles addressing the Gaussian Process Regression (GPR) integrated with sensor devices in smart buildings; Table S9: Scientific articles addressing the Linear Regression integrated with sensor devices in smart buildings; Table S10: Scientific articles addressing the Neural Networks for Regression Purposes integrated with sensor devices in smart buildings; Table S11: Scientific articles addressing the Support Vector Regression (SVR) integrated with sensor devices in smart buildings; Table S12: Scientific articles addressing the Fuzzy C-Means integrated with sensor devices in smart buildings; Table S13: Scientific articles addressing the Hidden Markov Model integrated with sensor devices in smart buildings; Table S14: Scientific articles addressing the Hierarchical Clustering integrated with sensor devices in smart buildings; Table S15: Scientific articles addressing the K-Means integrated with sensor devices in smart buildings; Table S16: Scientific articles addressing the Deep Learning techniques integrated with sensor devices in smart buildings.

Author Contributions: Conceptualization, D.-M.P., G.C., A.P., and N.L.C.; Methodology, D.-M.P., G.C., A.P., and N.L.C.; Investigation, D.-M.P., G.C., A.P., and N.L.C.; Resources, D.-M.P., G.C., A.P., and N.L.C.; Data Curation, D.-M.P., G.C., A.P., and N.L.C.; Writing—Original Draft Preparation, D.-M.P., G.C., A.P., and N.L.C.; Writing—Review & Editing, D.-M.P., G.C., A.P., and N.L.C.; Visualization, D.-M.P., G.C., A.P., and N.L.C.; Supervision, A.P.; Funding Acquisition, A.P.

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

References

1. Kim, S.; Song, Y.; Sung, Y.; Seo, D. Development of a consecutive occupancy estimation framework for improving the energy demand prediction performance of building energy modeling tools. *Energies* **2019**, *12*, 433. [CrossRef]

2. Zhao, H.; Hua, Q.; Chen, H.-B.; Ye, Y.; Wang, H.; Tan, S.X.-D.; Tlelo-Cuautle, E. Thermal-Sensor-Based Occupancy Detection for Smart Buildings Using Machine-Learning Methods. *ACM Trans. Des. Autom. Electron. Syst.* **2018**, *23*, 54. [CrossRef]

3. Hao, J.; Yuan, X.; Yang, Y.; Wang, R.; Zhuang, Y.; Luo, J. Visible Light Based Occupancy Inference Using Ensemble Learning. *IEEE Access* **2018**, *6*, 16377–16385. [CrossRef]

4. Chen, Z.; Zhu, Q.; Soh, Y.C.; Zhang, L. Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM. *IEEE Trans. Ind. Inform.* **2017**, *13*, 3070–3080. [CrossRef]

5. Bales, D.; Tarazaga, P.A.; Kasarda, M.; Batra, D.; Woolard, A.G.; Poston, J.D.; Malladi, V.V.N.S. Gender Classification of Walkers via Underfloor Accelerometer Measurements. *IEEE Internet Things J.* **2016**, *3*, 1259–1266. [CrossRef]

6. Al MacHot, F.; Mosa, A.H.; Ali, M.; Kyamakya, K. Activity Recognition in Sensor Data Streams for Active and Assisted Living Environments. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 2933–2945. [CrossRef]

7. Li, X.; Nie, L.; Xu, H.; Wang, X. Collaborative Fall Detection Using Smart Phone and Kinect. *Mob. Netw. Appl.* **2018**, *23*, 775–788. [CrossRef]

8.   Li, W.; Tan, B.; Piechocki, R. Passive Radar for Opportunistic Monitoring in E-Health Applications. *IEEE J. Transl. Eng. Heal. Med.* **2018**, *6*, 2800210. [CrossRef]

9.   Chen, Z.; Wang, Y.; Liu, H. Unobtrusive sensor-based occupancy facing direction detection and tracking using advanced machine learning algorithms. *IEEE Sens. J.* **2018**, *18*, 6360–6368. [CrossRef]

10.   Lu, L.; Qing-ling, C.; Yi-Ju, Z. Activity Recognition in Smart Homes. *Multimed. Tools Appl.* **2017**, *76*, 24203–24220. [CrossRef]

11.   Hassan, M.K.; El Desouky, A.I.; Elghamrawy, S.M.; Sarhan, A.M. A Hybrid Real-time remote monitoring framework with NB-WOA algorithm for patients with chronic diseases. *Futur. Gener. Comput. Syst.* **2019**, *93*, 77–95. [CrossRef]

12.   Rahman, A.; Srikumar, V.; Smith, A.D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Appl. Energy* **2018**, *212*, 372–385. [CrossRef]

13.   Caicedo, D.; Pandharipande, A. Sensor Data-Driven Lighting Energy Performance Prediction. *IEEE Sens. J.* **2016**, *16*, 6397–6405. [CrossRef]

14.   Kim, J.Y.; Liu, N.; Tan, H.X.; Chu, C.H. Unobtrusive Monitoring to Detect Depression for Elderly with Chronic Illnesses. *IEEE Sens. J.* **2017**, *17*, 5694–5704. [CrossRef]

15.   Guan, Q.; Yin, X.; Guo, X.; Wang, G. A novel infrared motion sensing system for compressive classification of physical activity. *IEEE Sens. J.* **2016**, *16*, 2251–2259. [CrossRef]

16.   Tian, Y.; Wang, X.; Chen, L.; Liu, Z. Wearable sensor-based human activity recognition via two-layer diversity-enhanced multiclassifier recognition method. *Sensors* **2019**, *19*, 2039. [CrossRef]

17.   Brennan, C.; Taylor, G.W.; Spachos, P. Designing learned CO_2-based occupancy estimation in smart buildings. *IET Wirel. Sens. Syst.* **2018**, *8*, 249–255. [CrossRef]

18.   Yu, H.; Pan, G.; Pan, M.; Li, C.; Jia, W.; Zhang, L.; Sun, M. A hierarchical deep fusion framework for egocentric activity recognition using a wearable hybrid sensor system. *Sensors* **2019**, *19*, 546. [CrossRef]

19.   Chen, H.; Cha, S.H.; Kim, T.W. A framework for group activity detection and recognition using smartphone sensors and beacons. *Build. Environ.* **2019**, *158*, 205–216. [CrossRef]

20.   Chen, Z.; Jiang, C.; Xie, L. A Novel Ensemble ELM for Human Activity Recognition Using Smartphone Sensors. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2691–2699. [CrossRef]

21.   Ateeq, M.; Ishmanov, F.; Afzal, M.K.; Naeem, M. Multi-parametric analysis of reliability and energy consumption in IoT: A deep learning approach. *Sensors* **2019**, *19*, 309. [CrossRef] [PubMed]

22.   Divina, F.; Torres, M.G.; Vela, F.A.G.; Noguera, J.L.V. A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings. *Energies* **2019**, *12*, 1934. [CrossRef]

23.   Chammas, M.; Makhoul, A.; Demerjian, J. An efficient data model for energy prediction using wireless sensors. *Comput. Electr. Eng.* **2019**, *76*, 249–257. [CrossRef]

24.   Rodriguez-Mier, P.; Mucientes, M.; Bugarín, A. Feature Selection and Evolutionary Rule Learning for Big Data in Smart Building Energy Management. *Cognit. Comput.* **2019**, *11*, 418–433. [CrossRef]

25.   Arifoglu, D.; Bouchachia, A. Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks. *Artif. Intell. Med.* **2019**, *94*, 88–95. [CrossRef]

26.   Li, C.; Cheung, W.K.; Liu, J.; Ng, J.K. Automatic extraction of behavioral paterns for elderly mobility and daily routine analysis. *ACM Trans. Intell. Syst. Technol.* **2018**, *9*, 54. [CrossRef]

27.   Guo, X.; Su, R.; Hu, C.; Ye, X.; Wu, H.; Nakamura, K. A single feature for human activity recognition using two-dimensional acoustic array. *Appl. Phys. Lett.* **2019**, *114*, 214101. [CrossRef]

28.   Guo, J.; Mu, Y.; Xiong, M.; Liu, Y.; Gu, J.; Garcia-Rodriguez, J. Activity Feature Solving Based on TF-IDF for Activity Recognition in Smart Homes. *Complexity* **2019**, *2019*, 5245373. [CrossRef]

29.   Abidine, B.M.; Fergani, B.; Oussalah, M.; Fergani, L. A new classification strategy for human activity recognition using cost sensitive support vector machines for imbalanced data. *Kybernetes* **2014**, *43*, 1150–1164. [CrossRef]

30.   Nef, T.; Urwyler, P.; Büchler, M.; Tarnanas, I.; Stucki, R.; Cazzoli, D.; Müri, R.; Mosimann, U. Evaluation of three state-of-the-art classifiers for recognition of activities of daily living from smart home ambient data. *Sensors* **2015**, *15*, 11725–11740. [CrossRef]

31.   Fahad, L.G.; Ali, A.; Rajarajan, M. Learning models for activity recognition in smart homes. *Inf. Sci. Appl.* **2015**, *339*, 819–826.

32.   Amirjavid, F.; Bouzouane, A.; Bouchard, B. Data driven modeling of the simultaneous activities in ambient environments. *J. Ambient Intell. Humaniz. Comput.* **2014**, *5*, 717–740. [CrossRef]

33. Alam, M.G.R.; Abedin, S.F.; Al Ameen, M.; Hong, C.S. Web of objects based ambient assisted living framework for emergency psychiatric state prediction. *Sensors* **2016**, *16*, 1431. [CrossRef] [PubMed]

34. Ballardini, A.L.; Ferretti, L.; Fontana, S.; Furlan, A.; Sorrenti, D.G. An Indoor Localization System for Telehomecare Applications. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *46*, 1445–1455. [CrossRef]

35. Alcala, J.M.; Urena, J.; Hernandez, A.; Gualda, D. Sustainable Homecare Monitoring System by Sensing Electricity Data. *IEEE Sens. J.* **2017**, *17*, 7741–7749. [CrossRef]

36. Sundaravadivel, P.; Kesavan, K.; Kesavan, L.; Mohanty, S.P.; Kougianos, E. Smart-Log: A Deep-Learning Based Automated Nutrition Monitoring System in the IoT. *IEEE Trans. Consum. Electron.* **2018**, *64*, 390–398. [CrossRef]

37. Zou, H.; Zhou, Y.; Yang, J.; Spanos, C.J. Unsupervised WiFi-Enabled IoT Device-User Association for Personalized Location-Based Service. *IEEE Internet Things J.* **2019**, *6*, 1238–1245. [CrossRef]

38. Yang, K.; Cho, S.B. A context-aware system in Internet of Things using modular Bayesian networks. *Int. J. Distrib. Sens. Netw.* **2017**, *13*, 155014771770898. [CrossRef]

39. Meana-Llorián, D.; González García, C.; Pelayo G-Bustelo, B.C.; Cueva Lovelle, J.M.; Garcia-Fernandez, N. IoFClime: The fuzzy logic and the Internet of Things to control indoor temperature regarding the outdoor ambient conditions. *Futur. Gener. Comput. Syst.* **2017**, *76*, 275–284. [CrossRef]

40. Hu, W.; Wen, Y.; Guan, K.; Jin, G.; Tseng, K.J. ITCM: Toward Learning-Based Thermal Comfort Modeling via Pervasive Sensing for Smart Buildings. *IEEE Internet Things J.* **2018**, *5*, 4164–4177. [CrossRef]

41. Candanedo, L.M.; Feldheim, V.; Deramaix, D. Data driven prediction models of energy use of appliances in a low-energy house. *Energy Build.* **2017**, *140*, 81–97. [CrossRef]

42. Di Corso, E.; Cerquitelli, T.; Apiletti, D. METATECH: METeorological data analysis for thermal energy characterization by means of self-learning transparent models. *Energies* **2018**, *11*, 1336. [CrossRef]

43. Khan, N.S.; Ghani, S.; Haider, S. Real-time analysis of a sensor's data for automated decision making in an IoT-based smart home. *Sensors* **2018**, *18*, 1711. [CrossRef]

44. Jain, R.K.; Smith, K.M.; Culligan, P.J.; Taylor, J.E. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl. Energy* **2014**, *123*, 168–178. [CrossRef]

45. Collotta, M.; Pau, G. An Innovative Approach for Forecasting of Energy Requirements to Improve a Smart Home Management System Based on BLE. *IEEE Trans. Green Commun. Netw.* **2017**, *1*, 112–120. [CrossRef]

46. Jabłoński, I. Smart transducer interface—From networked on-site optimization of energy balance in research-demonstrative office building to smart city conception. *IEEE Sens. J.* **2015**, *15*, 2468–2478. [CrossRef]

47. Keshtkar, A.; Arzanpour, S.; Keshtkar, F.; Ahmadi, P. Smart residential load reduction via fuzzy logic, wireless sensors, and smart grid incentives. *Energy Build.* **2015**, *104*, 165–180. [CrossRef]

48. Anthierens, C.; Leclercq, M.; Bideaux, E.; Flambard, L. A smart sensor to evaluate visual comfort of daylight into buildings. *Int. J. Optomech.* **2008**, *2*, 413–434. [CrossRef]

49. Chang, C.Y.; Hung, S.S.; Liu, L.H.; Lin, C.P. Innovative strain sensing for detection of exterior wall tile lesion: Smart skin sensory system. *Materials* **2018**, *11*, 2432. [CrossRef]

50. Shetty, S.S.; Hoang, D.C.; Gupta, M.; Panda, S.K. Learning desk fan usage preferences for personalised thermal comfort in shared offices using tree-based methods. *Build. Environ.* **2019**, *149*, 546–560. [CrossRef]

51. Viani, F.; Polo, A. A forecasting strategy based on wireless sensing for thermal comfort optimization in smart buildings. *Microw. Opt. Technol. Lett.* **2017**, *59*, 2913–2917. [CrossRef]

52. Ahmed, H.S.; Faouzi, B.M.; Caelen, J. Detection and classification of the behavior of people in an intelligent building by camera. *Int. J. Smart Sens. Intell. Syst.* **2013**, *6*, 1317–1342. [CrossRef]

53. Li, D.; Menassa, C.C.; Kamat, V.R. Personalized human comfort in indoor building environments under diverse conditioning modes. *Build. Environ.* **2017**, *126*, 304–317. [CrossRef]

54. Ain, Q.-u.; Iqbal, S.; Khan, S.A.; Malik, A.W.; Ahmad, I.; Javaid, N. IoT operating system based fuzzy inference system for home energy management system in smart buildings. *Sensors* **2018**, *18*, 2802. [CrossRef] [PubMed]

55. Motamed, A.; Deschamps, L.; Scartezzini, J.L. On-site monitoring and subjective comfort assessment of a sun shadings and electric lighting controller based on novel High Dynamic Range vision sensors. *Energy Build.* **2017**, *149*, 58–72. [CrossRef]

56. Ulpiani, G. Overheating phenomena induced by fully-glazed facades: Investigation of a sick building in Italy and assessment of the benefits achieved via fuzzy control of the AC system. *Sol. Energy* **2017**, *158*, 572–594. [CrossRef]

57. Grant, M.J.; Booth, A. A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Info. Libr. J.* **2009**, *26*, 91–108. [CrossRef]

58. Shobha, G.; Rangaswamy, S. Machine Learning. In *Handbook of Statistics*; Rao, C.R., Ed.; Elsevier: Amsterdam, The Netherlands, 2018; Volume 38, pp. 197–228, ISBN 9780444640420.

59. Gillani Fahad, L.; Khan, A.; Rajarajan, M. Activity recognition in smart homes with self verification of assignments. *Neurocomputing* **2015**, *149*, 1286–1298. [CrossRef]

60. Kim, T.S.; Cho, J.H.; Kim, J.T. Mobile motion sensor-based human activity recognition and energy expenditure estimation in building environments. *Smart Innov. Syst. Technol.* **2013**, *22*, 987–993. [CrossRef]

61. Abidine, M.B.; Fergani, B.; Ordóñez, F.J. Effect of over-sampling versus under-sampling for SVM and LDA classifiers for activity recognition. *Int. J. Des. Nat. Ecodyn.* **2016**, *11*, 306–316. [CrossRef]

62. Abidine, M.B.; Fergani, B. News schemes for activity recognition systems using PCA-WSVM, ICA-WSVM, and LDA-WSVM. *Information* **2015**, *6*, 505–521. [CrossRef]

63. Chernbumroong, S.; Cang, S.; Yu, H. Genetic algorithm-based classifiers fusion for multisensor activity recognition of elderly people. *IEEE J. Biomed. Heal. Inform.* **2015**, *19*, 282–289. [CrossRef] [PubMed]

64. Lai, Y.X.; Lai, C.F.; Huang, Y.M.; Chao, H.C. Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home. *Inf. Sci.* **2013**, *230*, 39–55. [CrossRef]

65. Namburu, S.M.; Azam, M.S.; Luo, J.; Choi, K.; Pattipati, K.R. Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers. *IEEE Trans. Autom. Sci. Eng.* **2007**, *4*, 469–473. [CrossRef]

66. Liao, R.; Changqing, S. Smart Home Design Based on Cloud Computing and Internet of Things. *J. Comput. Theor. Nanosci.* **2016**, *13*, 8075–8080. [CrossRef]

67. Evers, C.; Habets, E.A.P.; Gannot, S.; Naylor, P.A. DoA reliability for distributed acoustic tracking. *IEEE Signal Process. Lett.* **2018**, *25*, 1320–1324. [CrossRef]

68. Zimmermann, L.; Weigel, R.; Fischer, G. Fusion of nonintrusive environmental sensors for occupancy detection in smart homes. *IEEE Internet Things J.* **2018**, *5*, 2343–2352. [CrossRef]

69. Sebbak, F.; Benhammadi, F.; Chibani, A.; Amirat, Y.; Mokhtari, A. Dempster–Shafer theory-based human activity recognition in smart home environments. *Ann. Des Telecommun. Telecommun.* **2014**, *69*, 171–184. [CrossRef]

70. Fang, H.; Srinivasan, R.; Cook, D.J. Feature selections for human activity recognition in smart home environments. *Int. J. Innov. Comput. Inf. Control* **2012**, *8*, 3525–3535.

71. Gulati, M.; Ram, S.S.; Majumdar, A.; Singh, A. Single Point Conducted EMI Sensor with Intelligent Inference for Detecting IT Appliances. *IEEE Trans. Smart Grid* **2018**, *9*, 3716–3726. [CrossRef]

72. Kwolek, B.; Kepski, M. Improving fall detection by the use of depth sensor and accelerometer. *Neurocomputing* **2015**, *168*, 637–645. [CrossRef]

73. Tan, T.H.; Gochoo, M.; Huang, S.C.; Liu, Y.H.; Liu, S.H.; Huang, Y.F. Multi-resident activity recognition in a smart home using RGB activity image and DCNN. *IEEE Sens. J.* **2018**, *18*, 9718–9727. [CrossRef]

74. Palumbo, F.; Gallicchio, C.; Pucci, R.; Micheli, A. Human activity recognition using multisensor data fusion based on Reservoir Computing. *J. Ambient Intell. Smart Environ.* **2016**, *8*, 87–107. [CrossRef]

75. Lotfi, A.; Langensiepen, C.; Mahmoud, S.M.; Akhlaghinia, M.J. Smart homes for the elderly dementia sufferers: Identification and prediction of abnormal behaviour. *J. Ambient Intell. Humaniz. Comput.* **2012**, *3*, 205–218. [CrossRef]

76. Zeng, X.H.; Chen, X.T.; Ye, C.Y. An EEGA-based bayesian belief network model for recognition of human activity in smart home. *J. Donghua Univ.* **2012**, *29*, 497–500.

77. Vanus, J.; Belesova, J.; Martinek, R.; Nedoma, J.; Fajkus, M.; Bilik, P.; Zidek, J. Monitoring of the daily living activities in smart home care. *Hum. Cent. Comput. Inf. Sci.* **2017**, *7*, 30. [CrossRef]

78. Amayri, M.; Ploix, S.; Bouguila, N.; Wurtz, F. Estimating occupancy using interactive learning with a sensor environment: Real-time experiments. *IEEE Access* **2019**, *7*, 53932–53944. [CrossRef]

79. Al Zamil, M.G.H.; Samarah, S.M.J.; Rawashdeh, M.; Hossain, M.A. An ODT-based abstraction for mining closed sequential temporal patterns in IoT-cloud smart homes. *Clust. Comput.* **2017**, *20*, 1815–1829. [CrossRef]

80. Fong, S.; Li, J.; Song, W.; Tian, Y.; Wong, R.K.; Dey, N. Predicting unusual energy consumption events from smart home sensor network by data stream mining with misclassified recall. *J. Ambient Intell. Humaniz. Comput.* **2018**, *9*, 1197–1221. [CrossRef]

81. Tabatabaee Malazi, H.; Davari, M. Combining emerging patterns with random forest for complex activity recognition in smart homes. *Appl. Intell.* **2018**, *48*, 315–330. [CrossRef]

82. Lundström, J.; Järpe, E.; Verikas, A. Detecting and exploring deviating behaviour of smart home residents. *Expert Syst. Appl.* **2016**, *55*, 429–440. [CrossRef]

83. Zhao, Q.; Tsai, C.M.; Chen, R.C.; Huang, C.Y. Resident activity recognition based on binary infrared sensors and soft computing. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 291–299. [CrossRef]

84. Bjelica, M.Z.; Mrazovac, B.; Papp, I.; Teslic, N. Context-aware platform with user availability estimation and light-based announcements. *IEEE Trans. Syst. ManCybern. Part A Syst. Hum.* **2013**, *43*, 1228–1239. [CrossRef]

85. Jurek, A.; Nugent, C.; Bi, Y.; Wu, S. Clustering-based ensemble learning for activity recognition in smart homes. *Sensors* **2014**, *14*, 12285–12304. [CrossRef] [PubMed]

86. Fatima, I.; Fahim, M.; Lee, Y.K.; Lee, S. A genetic algorithm-based classifier ensemble optimization for activity recognition in smart homes. *KSII Trans. Internet Inf. Syst.* **2013**, *7*, 2853–2873. [CrossRef]

87. Alcalá, J.M.; Ureña, J.; Hernández, Á.; Gualda, D. Assessing human activity in elderly people using non-intrusive load monitoring. *Sensors* **2017**, *17*, 351. [CrossRef] [PubMed]

88. Muhammad, G.; Alhamid, M.F.; Shamim Hossain, M.; Almogren, A.S.; Vasilakos, A.V. Enhanced living by assessing voice pathology using a co-occurrence matrix. *Sensors* **2017**, *17*, 267. [CrossRef]

89. Villeneuve, E.; Harwin, W.; Holderbaum, W.; Janko, B.; Sherratt, R.S. Reconstruction of angular kinematics from wrist-worn inertial sensor data for smart home healthcare. *IEEE Access* **2017**, *5*, 2169–3536. [CrossRef]

90. Fagiani, M.; Squartini, S.; Gabrielli, L.; Severini, M.; Piazza, F. A statistical framework for automatic leakage detection in smart water and gas grids. *Energies* **2016**, *9*, 665. [CrossRef]

91. Mattera, C.G.; Quevedo, J.; Escobet, T.; Shaker, H.R.; Jradi, M. A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors. *Sensors* **2018**, *18*, 3931. [CrossRef]

92. Lynggaard, P. Using Machine Learning for Adaptive Interference Suppression in Wireless Sensor Networks. *IEEE Sens. J.* **2018**, *18*, 8820–8826. [CrossRef]

93. Bouchard, K.; Giroux, S.; Bouchard, B.; Bouzouane, A. Regression analysis for gesture recognition using passive RFID technology in smart home environments. *Int. J. Smart Home* **2014**, *8*, 245–260. [CrossRef]

94. Basu, C.; Caubel, J.J.; Kim, K.; Cheng, E.; Dhinakaran, A.; Agogino, A.M.; Martin, R.A. Sensor-based predictive modeling for smart lighting in grid-integrated buildings. *IEEE Sens. J.* **2014**, *14*, 4216–4229. [CrossRef]

95. Oprea, S.-V.; Pîrjan, A.; Căruțașu, G.; Petroșanu, D.-M.; Bâra, A.; Stănică, J.-L.; Coculescu, C. Developing a Mixed Neural Network Approach to Forecast the Residential Electricity Consumption Based on Sensor Recorded Data. *Sensors* **2018**, *18*, 1443. [CrossRef] [PubMed]

96. Pardo, J.; Zamora-Martínez, F.; Botella-Rocamora, P. Online learning algorithm for time series forecasting suitable for low cost wireless sensor networks nodes. *Sensors* **2015**, *15*, 9277–9304. [CrossRef] [PubMed]

97. Li, Z.; Dong, B. A new modeling approach for short-term prediction of occupancy in residential buildings. *Build. Environ.* **2017**, *121*, 277–290. [CrossRef]

98. Attoue, N.; Shahrour, I.; Younes, R. Smart building: Use of the artificial neural network approach for indoor temperature forecasting. *Energies* **2018**, *11*, 395. [CrossRef]

99. Khatoon, S.; Rahman, S.M.M.; Alrubaian, M.; Alamri, A. Privacy-Preserved, Provable Secure, Mutually Authenticated Key Agreement Protocol for Healthcare in a Smart City Environment. *IEEE Access* **2019**, *7*, 47962–47971. [CrossRef]

100. Amirjavid, F.; Spachos, P.; Plataniotis, K.N. 3-D Object Localization in Smart Homes: A Distributed Sensor and Video Mining Approach. *IEEE Syst. J.* **2018**, *12*, 1307–1316. [CrossRef]

101. Sarwar, B.; Bajwa, I.S.; Ramzan, S.; Ramzan, B.; Kausar, M. Design and application of fuzzy logic based fire monitoring and warning systems for smart buildings. *Symmetry* **2018**, *10*, 615. [CrossRef]

102. Yuan, B.; Herbert, J. Context-aware hybrid reasoning framework for pervasive healthcare. *Pers. Ubiquitous Comput.* **2014**, *18*, 865–881. [CrossRef]

103. Wang, J.M.; Yang, M.T.; Chen, P.L. Design and implementation of an intelligent windowsill system using smart handheld device and fuzzy microcontroller. *Sensors* **2017**, *17*, 830. [CrossRef] [PubMed]

104. Usman, M.; Muthukkumarasamy, V.; Wu, X.W. Mobile agent-based cross-layer anomaly detection in smart home sensor networks using fuzzy logic. *IEEE Trans. Consum. Electron.* **2015**, *61*, 197–205. [CrossRef]

105. Kıyak, İ.; Oral, B.; Topuz, V. Smart indoor LED lighting design powered by hybrid renewable energy systems. *Energy Build.* **2017**, *148*, 342–347. [CrossRef]

106. Liu, J.; Zhang, W.; Chu, X.; Liu, Y. Fuzzy logic controller for energy savings in a smart LED lighting system considering lighting comfort and daylight. *Energy Build.* **2016**, *127*, 95–104. [CrossRef]

107. Ahvar, E.; Lee, G.M.; Han, S.N.; Crespi, N.; Khan, I. Sensor network-based and user-friendly user location discovery for future smart homes. *Sensors* **2016**, *16*, 969. [CrossRef]

108. Vlachostergiou, A.; Stratogiannis, G.; Caridakis, G.; Siolas, G.; Mylonas, P. User Adaptive and Context-Aware Smart Home Using Pervasive and Semantic Technologies. *J. Electr. Comput. Eng.* **2016**, *2016*, 4789803. [CrossRef]

109. Panna, R.; Thesrumluk, R.; Chantrapornchai, C. Development of energy saving smart home prototype. *Int. J. Smart Home* **2013**, *7*, 47–66.

110. Wang, K.J.; Wu, C.Y.; Ning, W.L. Fuzzy cognitive map control on room temperature in a smart house. *J. Chin. Soc. Mech. Eng.* **2013**, *34*, 431–440.

111. Sang-Hyun, L.; Lee, J.G.; Kyung-Il, M. Smart home security system using multiple ANFIS. *Int. J. Smart Home* **2013**, *7*, 121–132.

112. Fortin-Simard, D.; Bouchard, K.; Gaboury, S.; Bouchard, B.; Bouzouane, A. Accurate passive RFID localization system for smart homes. *Netw. Embed. Syst. Enterp. Appl.* **2012**, *1*, 391–399. [CrossRef]

113. Sharifi, R.; Kim, Y.; Langari, R. Sensor fault isolation and detection of smart structures. *Smart Mater. Struct.* **2010**, *9*, 105001. [CrossRef]

114. Chen, S.Y.; Chiu, M.L. Designing Smart Skins for Adaptive Environments: A fuzzy logic approach to smart house design. *Comput. Aided. Des. Appl.* **2007**, *4*, 751–760. [CrossRef]

115. Papatsimpa, C.; Linnartz, J.-P. Distributed Fusion of Sensor Data in a Constrained Wireless Network. *Sensors* **2019**, *19*, 1006. [CrossRef] [PubMed]

116. Civitarese, G.; Bettini, C.; Sztyler, T.; Riboni, D.; Stuckenschmidt, H. newNECTAR: Collaborative active learning for knowledge-based probabilistic activity recognition. *Pervasive Mob. Comput.* **2019**, *56*, 88–105. [CrossRef]

117. Dahmen, J.; Cook, D. SynSys: A synthetic data generation system for healthcare applications. *Sensors* **2019**, *19*, 1181. [CrossRef]

118. Hela, S.; Amel, B.; Badran, R. Early anomaly detection in smart home: A causal association rule-based approach. *Artif. Intell. Med.* **2018**, *91*, 57–71. [CrossRef]

119. Lan, D.; Pang, Z.; Fischione, C.; Liu, Y.; Taherkordi, A.; Eliassen, F. Latency Analysis of Wireless Networks for Proximity Services in Smart Home and Building Automation: The Case of Thread. *IEEE Access* **2019**, *7*, 2169–3536. [CrossRef]

120. Al Zamil, M.G.; Rawashdeh, M.; Samarah, S.; Hossain, M.S.; Alnusair, A.; Rahman, S.M.M. An Annotation Technique for In-Home Smart Monitoring Environments. *IEEE Access* **2017**, *6*, 1471–1479. [CrossRef]

121. Gayathri, K.S.; Easwarakumar, K.S.; Elias, S. Probabilistic ontology based activity recognition in smart homes using Markov Logic Network. *Knowl. Based Syst.* **2017**, *121*, 173–184. [CrossRef]

122. Wang, C.; Peng, Y.; De, D.; Song, W.Z. DPHK: Real-time distributed predicted data collecting based on activity pattern knowledge mined from trajectories in smart environments. *Front. Comput. Sci.* **2016**, *10*, 1000–1011. [CrossRef]

123. Wang, L.; Gu, T.; Tao, X.; Chen, H.; Lu, J. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive Mob. Comput.* **2011**, *7*, 287–298. [CrossRef]

124. Doty, K.; Roy, S.; Fischer, T.R. Explicit State-Estimation Error Calculations for Flag Hidden Markov Models. *IEEE Trans. Signal Process.* **2016**, *64*, 4444–4454. [CrossRef]

125. Chahuara, P.; Fleury, A.; Portet, F.; Vacher, M. On-line human activity recognition from audio and home automation sensors: Comparison of sequential and non-sequential models in realistic Smart Homes. *J. Ambient Intell. Smart Environ.* **2016**, *8*, 399–422. [CrossRef]

126. Noury, N.; Hadidi, T. Computer simulation of the activity of the elderly person living independently in a Health Smart Home. *Comput. Methods Programs Biomed.* **2012**, *108*, 1216–1228. [CrossRef]

127. Roy, N.; Misra, A.; Cook, D. Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments. *J. Ambient Intell. Humaniz. Comput.* **2016**, *7*, 1–19. [CrossRef]

128. Nait Aicha, A.; Englebienne, G.; Kröse, B. Unsupervised visit detection in smart homes. *Pervasive Mob. Comput.* **2017**, *34*, 157–167. [CrossRef]

129. Chahuara, P.; Portet, F.; Vacher, M. Context-aware decision making under uncertainty for voice-based control of smart home. *Expert Syst. Appl.* **2017**, *75*, 63–79. [CrossRef]

130. Alemdar, H.; Ersoy, C. Multi-resident activity tracking and recognition in smart environments. *J. Ambient Intell. Humaniz. Comput.* **2017**, *8*, 513–529. [CrossRef]

131. Chikhaoui, B.; Wang, S.; Pigot, H. ADR-SPLDA: Activity discovery and recognition by combining sequential patterns and latent Dirichlet allocation. *Pervasive Mob. Comput.* **2012**, *8*, 845–862. [CrossRef]

132. Karami, A.B.; Fleury, A.; Boonaert, J.; Lecoeuche, S. User in the loop: Adaptive smart homes exploiting user feedback-State of the art and future directions. *Information* **2016**, *7*, 35. [CrossRef]

133. Von Bomhard, T.; Wörner, D.; Röschlin, M. Towards smart individual-room heating for residential buildings. *Comput. Sci. Res. Dev.* **2016**, *31*, 127–134. [CrossRef]

134. Casado-Vara, R.; Vale, Z.; Prieto, J.; Corchado, J.M. Fault-tolerant temperature control algorithm for IoT networks in smart buildings. *Energies* **2018**, *11*, 3430. [CrossRef]

135. Papatsimpa, C.; Linnartz, J.P.M.G. Propagating sensor uncertainty to better infer office occupancy in smart building control. *Energy Build.* **2018**, *179*, 73–82. [CrossRef]

136. Khan, W.M.; Zualkernan, I.A. SensePods: A ZigBee-Based Tangible Smart Home Interface. *IEEE Trans. Consum. Electron.* **2018**, *64*, 145–152. [CrossRef]

137. Mohammed, A.W.; Xu, Y.; Liu, M.; Hu, H. Semantical Markov Logic Network for Distributed Reasoning in Cyber-Physical Systems. *J. Sens.* **2017**, *2017*, 4259652. [CrossRef]

138. Egarter, D.; Bhuvana, V.P.; Elmenreich, W. PALDi: Online load disaggregation via particle filtering. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 467–477. [CrossRef]

139. Luan, X.; Zheng, Z.; Wang, T.; Wu, J.; Xiang, H. Hybrid cooperation for machine-to-machine data collection in hierarchical smart building networks. *IET Commun.* **2015**, *9*, 421–428. [CrossRef]

140. Pérez-Chacón, R.; Luna-Romera, J.M.; Troncoso, A.; Martínez-Alvarez, F.; Riquelme, J.C. Big data analytics for discovering electricity consumption patterns in smart cities. *Energies* **2018**, *11*, 683. [CrossRef]

141. Zeiler, W.; Labeodan, T. Human-in-the-loop energy flexibility integration on a neighbourhood level: Small and Big Data management. *Build. Serv. Eng. Res. Technol.* **2019**, *40*, 305–318. [CrossRef]

142. Zhu, H.; Chen, H.; Brown, R. A sequence-to-sequence model-based deep learning approach for recognizing activity of daily living for senior care. *J. Biomed. Inform.* **2018**, *84*, 148–158. [CrossRef] [PubMed]

143. Liciotti, D.; Bernardini, M.; Romeo, L.; Frontoni, E. A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing* **2019**, 1–13. [CrossRef]

144. Brenon, A.; Portet, F.; Vacher, M. ARCADES: A deep model for adaptive decision making in voice controlled smart-home. *Pervasive Mob. Comput.* **2018**, *49*, 92–110. [CrossRef]

145. Mora, N.; Matrella, G.; Ciampolini, P. Cloud-based behavioral monitoring in smart homes. *Sensors* **2018**, *18*, 1951. [CrossRef] [PubMed]

146. Hassan, M.M.; Uddin, M.Z.; Mohamed, A.; Almogren, A. A robust human activity recognition system using smartphone sensors and deep learning. *Futur. Gener. Comput. Syst.* **2018**, *81*, 307–313. [CrossRef]

147. Hsiu-Yu, L.; Yu-Ling, H.; Wen-Nung, L. Convolutional recurrent neural networks for posture analysis in fall detection. *J. Inf. Sci. Eng.* **2018**, *34*, 577–591. [CrossRef]

148. Chen, G.; Wang, A.; Zhao, S.; Liu, L.; Chang, C.Y. Latent feature learning for activity recognition using simple sensors in smart homes. *Multimed. Tools Appl.* **2018**, *77*, 15201–15219. [CrossRef]

149. ISO/IEC 25010:2011 Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—System and Software Quality. Available online: https://www.iso.org/standard/35733.html (accessed on 10 October 2019).