energies

MDPI

Review

User Activity Detection and Identification of Energy Habits in Home Energy-Management Systems Using AI and ML: A Comprehensive Review

Filip Durlik [1], Jakub Grela [1,*], Dominik Latoń [1], Andrzej Ożadowicz [1] and Lukasz Wisniewski [2,*]

[1] Department of Power Electronics and Energy Control Systems, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Krakow, 30-059 Krakow, Poland; fdurlik@agh.edu.pl (F.D.); laton@agh.edu.pl (D.L.); ozadow@agh.edu.pl (A.O.)
[2] Institute Industrial IT (inIT), Technische Hochschule OWL, 32657 Lemgo, Germany
* Correspondence: jgrela@agh.edu.pl (J.G.); lukasz.wisniewski@th-owl.de (L.W.)

Abstract

The residential energy sector contributes substantially to global energy-related emissions. Effective energy management requires an understanding occupant behavior through activity detection and habit identification. Recent advances in artificial intelligence (AI) and machine learning (ML) enable the automatic detection of user activities and prediction of energy needs based on historical consumption data. Non-intrusive load monitoring (NILM) facilitates device-level disaggregation without additional sensors, supporting demand forecasting and behavior-aware control in Home Energy Management Systems (HEMSs). This review synthesizes various AI and ML approaches for detecting user activities and energy habits in HEMSs from 2020 to 2025. The analyses revealed that deep learning (DL) models, with their ability to capture complex temporal and nonlinear patterns in multisensor data, achieve superior accuracy in activity detection and load forecasting, with occupancy detection reaching 95–99% accuracy. Hybrid systems combining neural networks and optimization algorithms demonstrate enhanced robustness, but challenges remain in limited cross-building generalization, insufficient interpretability of deep models, and the absence of dataset standardized. Future work should prioritize lightweight, explainable edge-ready models, federated learning, and integration with digital twins and control systems. It should also extend energy optimization toward occupant wellbeing and grid flexibility, using standardized protocols and open datasets for ensuring trustworthy and sustainability.

Keywords: machine learning; deep learning; building automation; activity recognition; user habit detection; home energy management; occupancy prediction; demand response

Check for updates

Academic Editor: Francesco Nocera

Received: 17 December 2025
Revised: 13 January 2026
Accepted: 21 January 2026
Published: 26 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The residential sector represents a primary area of concern for global efforts to reduce greenhouse gas emissions and optimize energy consumption. It is estimated that the buildings energy sector accounts for a significant portion of global energy-related gas emissions and consumes over 40% of total final energy [1]. At the same time, the demand for energy in this sector is increasing in response to climate change. Specifically, the rapid increase in cooling requirements has caused air conditioning to become a growing source of load for power grids and CO_2 emissions. In consideration of these challenges, the implementation of Home Energy Management Systems (HEMSs) has become increasingly

essential. These systems must be capable not only for monitoring energy consumption, but also actively controlling it by loads shifting, etc. [2,3]. Traditional energy-conservation approaches, such as educational campaigns and replacing older devices with more efficient models, have achieved limited success because they often overlook the complexity of user behavior and changing energy market conditions.

The effective management of residential energy environments requires a deep analysis of how building occupants use energy. This includes an examination of their specific activities and various energy habits, as well as the external factors that may influence their behavior [4,5]. Recent advancements in the fields of artificial intelligence (AI) and machine learning (ML) have opened new avenues for addressing this issue. Rather than being dependent on manually programmed control rules, it is now possible for technical systems to learn directly from historical energy-consumption data to automatically discover activity patterns and user habits [6]. Furthermore, learning can be facilitated throughout the entire life cycle of a building to enable control scenarios that may vary over time [7]. Activity detection relies on analyzing energy consumption time series using advanced algorithms. Through disaggregating total energy consumption at the appliance level, often referred as non-intrusive load monitoring (NILM), it is possible to identify individual activities without installing additional sensors in every device [8–10]. These detectable events include daily activities, such as meal preparation, hygiene routines, and laundry, as well as appliance states and potential energy usage anomalies that are indicative for equipment malfunction or irregular usage.

Energy habits are characterized by repetitive, long-term patterns of user behavior beyond immediate activities. Dedicated AI and ML models can learn these patterns from historical data and forecast future behaviors with high accuracy. These habits include occupancy schedules, enabling the automatic adjustment of Heating, Ventilation, and Air Conditioning (HVAC). Furthermore, automatic lighting systems with absence detection accuracy ranging from 95% to 99% can also be applied [11–13]. Additionally, these models can identify seasonal and weekly patterns, thermal comfort preferences, and specific morning or evening routines. The precise detection of activities and habits is the foundation of advanced home energy management and leads to significant benefits, such as demand forecasting, time-of-use optimization, and peak shaving. By identifying peak usage patterns and user habits, HEMSs can anticipate loads and actively participate in demand side management (DSM) and demand response (DR) programs, potentially reducing energy costs and balancing distribution loads. Furthermore, these systems facilitate predictive diagnostics and optimize self-consumption in homes with photovoltaic panels or energy storage systems by adapting charging cycles to detected usage peaks [14].

This review aims to provide a structured synthesis of recent research on applying AI and ML methods to detect user activities and identify energy habits within HEMSs. This article's key contributions are as follows:

- A compilation and comparison of existing approaches and algorithms, as well as their results.
- A comparative analysis of different AI/ML architectures in terms of accuracy and practical application.
- Identification of gaps in current research.
- Delineation of promising research directions and technologies.

In the context of habit identification, the review focuses on predicting building occupancy, thermal comfort based on user preferences, and detecting seasonal user activities, including energy consumption profiling. Regarding activity recognition, the study concentrates on sensor-based activity recognition and energy disaggregation to detect appliances operating within the network.

The remainder of this paper is organized as follows. Section 2 describes the methodology employed for the literature review and article selection. Section 3 presents a comprehensive overview of user habits and activity detection for adaptive HEMSs, covering occupancy prediction, thermal comfort preferences, and seasonal profiling, as well as sensor-based activity recognition and energy-disaggregation techniques. Section 4 discusses existing research gaps and outlines future works and directions for the advancement of intelligent energy-management systems. Finally, Section 5 offers a summary and concludes the article.

2. Methodology of Review

This review is based on a structured literature survey employing a multi-stage search and screening procedure applied to scientific and technical publications published between 2020 and 2025. The objective of the investigation was to identify and categorize methods, recent trends, and research directions related to the use of AI, particularly ML and deep learning (DL), to occupant habits and activity detection for adaptive HEMSs. The search strategy was based on combining thematic keyword groups using Boolean logic. Synonymous terms were connected with the OR operator, while conceptually distinct groups were combined using AND. The keyword groups covered (i) user activities and habits (e.g., activity recognition, occupancy detection, routines), (ii) AI-based methods (machine learning, deep learning), and (iii) the application context (home energy management systems, smart homes, residential energy). Searches were applied to titles, abstracts, and keywords or topic fields, depending on the database.

The literature review process comprised two primary phases. In the first stage, a comprehensive and structured search was conducted across the primary bibliometric databases, namely Scopus, Google Scholar, and Web of Science. The search queries included combinations of keywords such as user habit detection, user activity recognition, ML, and DL. The preliminary set of publications was filtered according to three key criteria: topical relevance, scientific rigor, and the presence of quantitative performance metrics. The results of the bibliometric analysis are presented in Table 1, detailing the number of research and review papers retrieved for each category across the examined databases.

Table 1. The literature review results from bibliometric databases.

Database	Publication Type	User Activity Recognition	User Habits Detection	User Activity Recognition AND Machine Learning *	User Habits Detection AND Machine Learning	User Activity Recognition AND Deep Learning	User Habits Detection AND Deep Learning
Web of Science	Articles	2913	192	736	42	663	33
	Reviews	131	5	33	1	21	1
Scopus	Articles	1972	157	526	35	494	29
	Reviews	110	7	38	1	25	1
Google Scholar	Articles	727,000	41,700	17,900	18,100	17,900	18,100
	Reviews	50,800	13,600	17,300	17,500	16,900	17,600

* AND denotes Boolean conjunction between keyword groups; synonyms were combined using OR.

The second phase of the study involved a focused review of publications from major scientific publishers, including MDPI, Springer, the ACM Digital Library, Elsevier (ScienceDirect), IEEE Xplore, Taylor & Francis, and Wiley Online Library. This stage aimed at identifying high-impact contributions—such as journal articles, review papers, and selected conference proceedings—that demonstrated real-world applicability, methodological robustness, and model validation. The outcomes of this publisher-based search are summarized in Table 2, which outlines the thematic coverage and publication volume across the considered platforms and subcategories.

Table 2. The literature review results from publisher databases.

Database	Publication Type	User Activity Recognition	User Habits Detection	User Activity Recognition AND Machine Learning *	User Habits Detection AND Machine Learning	User Activity Recognition AND Deep Learning	User Habits Detection AND Deep Learning
Springer	Articles Reviews	22,282 3620	4041 565	11,682 2086	2347 343	10,211 1771	1984 298
Science Direct	Articles Reviews	72,768 15,180	9958 2071	31,576 7866	4797 1170	33,178 7480	4470 1086
MDPI	Articles Reviews	327 16	43 1	67 4	2 0	73 1	3 0
IEEE Xplore	Conf. papers Journals	3813 903	193 31	0 0	0 0	0 0	0 0
Taylor & Francis	Articles Reviews	444,431 25,831	56,083 1137	42,499 3706	6400 388	167,536 7355	28,320 932
ACM Digital Library	Research papers Reviews	104,017 28	8986 3	60,426 21	5282 2	64,396 21	6574 3
Wiley Online Library	Journal papers Books	1,033,892 47,293	64,086 6793	84,865 11,601	7484 1805	173,439 20,426	20,194 4056

* AND denotes Boolean conjunction between keyword groups; synonyms were combined using OR.

Tables 1 and 2 illustrate the publication landscape and thematic saturation of research related to the AI-based detection of user activities and habits in residential and smart building contexts. They are intended to provide contextual insight into the scope and distribution of the field rather than to represent the directly screened corpus.

As it has been mentioned, the review was limited to the period from 2020 to 2025 to capture the most recent advances in the AI- and ML-based detection of user activities and habits. Earlier periods have already been widely discussed in prior surveys and reviews; repeating those findings would distract from the focus on current, application-oriented progress. In this context, Table 1 summarizes general keyword-based searches in bibliometric databases, while Table 2 complements these results with targeted queries across publisher platforms and publication types.

Following the initial search across bibliometric databases and publisher platforms, duplicate records were removed and the remaining studies were screened at the title and abstract level. Full-text assessment focused on methodological clarity, the availability of quantitative evaluation, and relevance to AI-based user activity or habit detection in home energy-management systems. This multi-stage screening resulted in a final set of 74 peer-reviewed publications.

For clarity and comparability, the selected publications were further grouped according to their primary application focus, including occupancy prediction, thermal comfort modeling, and seasonal or long-term activity recognition. This thematic classification underpins the structure of Section 3 and the topic-specific tables presented in the subsequent parts of the review.

3. User Habit and Activity Detection for Adaptive HEMSs

Following the preliminary survey and systematic screening process, a final set of 74 publications was selected for in-depth analysis. The selection was guided by several key criteria, including relevance to user behavior and activity detection in real or simulated environments, transparency and reproducibility of the applied methodologies, availability of input and output variables, inclusion of quantitative model performance evaluations, and representation of diverse AI-based modeling paradigms. To improve clarity and reduce potential ambiguity arising from varying terminology used in the original studies, Table 3 was introduced. This table provides definitions and brief descriptions of the most frequently used algorithms and methods discussed in the manuscript. The reviewed studies span a broad range of indoor environments, such as residential dwellings, classrooms, university

laboratories, healthcare facilities, and commercial spaces. As presented in Tables 4–8, a comprehensive compilation of the most relevant publications has been assembled, detailing key parameters, object types, applied methodologies, and performance metrics.

Table 3. List of most frequently used algorithms and methods discussed in the manuscript.

Method	Type	Description
Random Forest (RF)	Machine Learning	Random Forest is an ensemble algorithm that constructs multiple decision trees using randomly selected subsets of data and features, aggregating their outputs through majority voting or averaging. This approach effectively reduces overfitting by decreasing correlation among individual trees, resulting in improved accuracy and robustness in predictive tasks [15].
Support Vector Machines (SVM)	Machine Learning	Support Vector Machine constructs an optimal hyperplane that maximizes the margin between distinct classes in a high-dimensional space, thus enhancing classification accuracy. It handles both linearly separable and non-linear data through kernel functions that map input data into higher-dimensional spaces where linear separation is feasible. The method explicitly controls the trade-off between classifier complexity and misclassification, providing robust generalization capabilities [16].
Long Short-Term Memory (LSTM)	Deep Learning	Long Short-Term Memory (LSTM) is a recurrent neural network architecture designed to overcome the vanishing gradient problem, allowing it to learn long-term dependencies in sequential data. By using memory cells and gating mechanisms, LSTM networks effectively retain and access information over extended sequences, making them highly suitable for tasks like speech recognition, time-series analysis, and natural language processing [17].
Decision Trees (DT)	Machine Learning	Decision Trees are supervised learning models that recursively partition the input space into regions with increasingly homogeneous target values. At each step, the algorithm selects a predictor and a split (or multiway partition) that maximizes the separation of the target variable across the resulting child nodes, producing a rooted, directed tree whose internal nodes represent decision tests and whose leaves correspond to terminal prediction rules. Decision trees natively accommodate nominal, ordinal, and interval predictors, handle missing values through explicit routing strategies, and yield highly interpretable, rule-based models that expose local, context-dependent interaction effects between predictors [18].
k-Nearest Neighbors (kNN)	Machine Learning	Gradient Boosting Machines (GBMs) are an ensemble learning technique that builds predictive models by iteratively combining weak learners, typically decision trees, to form a strong predictive model. GBMs are based on the principle of boosting, where each new model corrects the errors of the previous ensemble. The core idea is to fit new models to the negative gradient of the loss function, which measures the difference between predicted and actual values, thereby minimizing the loss step-by-step [19].
Gated Recurrent Unit (GRU)	Deep Learning	Gated Recurrent Units (GRUs) are a streamlined variant of the Long Short-Term Memory (LSTM) architecture, designed to efficiently model sequential data in recurrent neural networks (RNNs). GRUs employ two gating mechanisms—the update gate and reset gate—to regulate the flow of information between hidden states, enabling improved handling of long-term dependencies compared to traditional RNNs. With fewer parameters than LSTMs, GRUs offer computational efficiency while maintaining comparable performance, making them a widely adopted choice for tasks such as speech recognition, natural language processing, and time-series analysis. Their simplified structure facilitates faster training and reduced memory usage without sacrificing effectiveness in capturing temporal patterns [20].

Table 3. *Cont.*

Method	Type	Description
Multi-Layer Perceptron (MLP)	Deep Learning	A multi-layer perceptron (MLP) is a feed-forward neural network with an input layer, one or more hidden layers, and an output layer, where each neuron applies a weighted sum, bias, and nonlinear activation to its inputs. By stacking these layers and training weights and biases with gradient-based optimization on labeled data, MLPs can model complex nonlinear relationships for classification and regression beyond what a single-layer perceptron can represent [21].
Logistic Regression (LR)	Machine Learning	Logistic regression (LR) is a supervised statistical model used to predict the probability that an observation belongs to one of several categorical outcome classes, most commonly a binary outcome coded as 0 and 1. Instead of modeling the dependent variable directly, logistic regression models the log-odds of category membership as a linear combination of the independent variables and then maps this linear predictor to the [0, 1] range via the logistic (sigmoid) function, yielding interpretable probabilities and odds ratios. Logistic regression does not require multivariate normality or homoscedasticity of predictors and is therefore widely applied as a flexible alternative to linear regression and discriminant analysis when the dependent variable is categorical [22].
Artificial Neural Network (ANN)	Deep Learning	Artificial Neural Networks (ANNs) are computational models inspired by the structure and function of biological neural networks, designed to simulate the way the human brain processes information. ANNs consist of interconnected nodes (neurons) organized in layers—input, hidden, and output—which collectively learn to recognize patterns, classify data, and make predictions through adaptive weight adjustments during training. Their ability to model complex, nonlinear relationships makes them highly effective in diverse applications, including pattern recognition, machine learning, and data analysis [23].
Convolutional Neural Network (CNN)	Deep Learning	A Convolutional Neural Network (CNN) is a deep feed-forward neural architecture that learns hierarchical feature representations by applying trainable convolutional filters, nonlinear activations, and pooling operations to structured data such as images, audio, or time series. Through local connectivity, weight sharing, and down-sampling, CNNs efficiently capture spatial or spatiotemporal patterns while greatly reducing the number of parameters compared with fully connected networks, which enables state-of-the-art performance in tasks like visual recognition, detection, and segmentation without manual feature engineering [24].
Support Vector Regression (SVR)	Machine Learning	Support Vector Regression (SVR) is a supervised learning algorithm that extends the Support Vector Machine framework to continuous prediction by finding a function that approximates the target values within an epsilon-insensitive tube, minimizing structural risk through margin maximization. SVR maps input data into a high-dimensional feature space via kernel functions (e.g., RBF, polynomial, or Pearson VII universal kernel) where a linear regression hyperplane is fitted, with only support vectors—data points outside or on the tube boundary—contributing to the model. This approach yields sparse, robust models with strong generalization for nonlinear relationships, interpretable via kernel matrix visualization, and regression coefficient profiles [25]

3.1. Occupancy Prediction

Research on occupancy prediction is currently focused on two main areas: accurate prediction, which refers to short-term (minutes-scale) classification of occupancy status at the room level, and generalized models using transfer learning, unsupervised methods, and semi-supervised approaches [26–28]. In the context of residential buildings, both

models that directly predict occupancy and those that directly link prediction quality to the operation of HVAC systems are being developed [29–31]. Conversely, in public buildings, on the other hand, there is a strong emphasis on scalability between rooms and buildings, resilience to data anomalies, and minimization of requirements for additional sensor wiring [13,32,33].

In the residential sector, occupancy prediction is most often formulated as a binary classification (presence 0/1) in minute or ten-minute time steps. It is based on environmental data, energy consumption, and simple motion sensors. Seasonal models for a single apartment in the HIKARI building show that separating data by season and using classic ML algorithms (GBDT, RF, ANN) allows F1 scores above 0.85–0.92 to be achieved. This significantly outperforms a generalized year-round model. In the other hand, work that couples occupancy prediction with heating control shows that the way the problem is framed (arrival time vs. 0/1 state) and domain metrics (MissTime, energy savings) have a greater impact on HVAC performance than the choice of the algorithm itself: kNN in occupancy state classification provides the best compromise between comfort and savings, while regressive DTs maximize savings at the expense of comfort. The use of AMI meter data alone (without additional sensors) is possible thanks to DL models combining CNN and BiLSTM. The model built on this combination achieves approximately 89–90% accuracy and clearly outperforms classic classifiers and simpler deep networks. Another approach is to use thermal comfort parameters alone (temperature, RH, light, CO_2). In this setup, simple classifiers (kNN, AO-ANN, DT) achieve accuracies of 99%+ and SHAP analysis exposes the important role of such parameters as lighting and CO_2 in the model's decisions [12,29,30].

In the context of public buildings, the complexity of the sensory environment and the wide variety of space types are of particular significance. A review of methods shows a very wide range of sensors (PIR, cameras, UWB, Wi-Fi/BLE, CO_2, energy meters) and models—from deterministic schedules and Markov chains to deep CNN and LSTM networks—with typical accuracies of 80–95% for a well-chosen combination of sensors and algorithms. In the domain of smaller offices, the use of transfer learning in sequential models (stacked LSTM, deep MLP) allows models to be transferred between rooms: pre-training on a long history of one room and fine tuning on short series from the others, increases accuracy from 62–66% to approx. 69–71%, which is particularly important when labeled data is limited [28,34]. In a multi-sensor building, a scalable CDLSTM architecture with transfer learning based on CO_2 and temperature has been proposed. With only two days of data from the target room, an MCC of approx. 0.81–0.82 can be achieved, which is comparable to a model trained from scratch over five days. The same model transferred to similar rooms achieves an MCC of 0.83–0.85. The idea of scalability is further developed in the Occupancy Sense system, in which the spatiotemporal context (time of day, zone type) is combined with environmental sensors and PIR. Gradient Boosting on trees (CatBoost) achieves 98–99% accuracy here and significantly outperforms RF, XGBoost, SVM, and neural networks [27,32,33].

Other studies explore newer data representations and unsupervised and semi-supervised techniques. The transformation of multidimensional time series into 2D images and the application of transfer learning from AlexNet/GoogLeNet or classic pixel classifiers allows for accuracies of 98–99% to be achieved for various spaces (UCI, student rooms, living rooms), with kNN and RF models performing particularly well when learned on images. Instead of supervised classification, some studies propose estimating occupancy levels using unsupervised methods: ExKMC combines k-means with a small threshold tree, producing easy-to-interpret if–then rules and achieving a weighted F1 score of 82–85%, which is better than supervised DTs, especially for the rare class of "2 people." At the

system level, the use of semi-supervised methods for indirect occupancy estimation from operational signals (energy, simple sensors) allows a similar level of accuracy/F1 to be kept with a significant reduction in the number of labels, making these approaches more practical in existing buildings. On the other hand, in large buildings (grid interactive efficient buildings), the importance of data quality is growing. For example, a hybrid LSTM autoencoder for detecting anomalies in sensor power consumption and LSTM/GCN-LSTM models for occupancy prediction improves prediction quality by 13–14%. Furthermore, DR scenarios based on these forecasts have been shown to facilitate a reduction in HVAC energy of approximately 20% [11,35].

There is also a growing trend towards models that use rich, heterogeneous data and sequential computations. A study comparing DNN, LSTM, Bi-LSTM, GRU, and Bi-GRU models for three types of spaces (office, library, lecture auditorium) shows that bidirectional architectures and GRU best combine accuracy with a minimal feature set. It is worth noting that for several key variables (e.g., cooling energy, number of Wi-Fi devices), they achieve very low MAE and RMSE, which has a direct impact on HVAC control and space-utilization analysis. At the same time, several review papers systematize the results and indicate that combining multiple types of sensors (e.g., CO_2 + Wi-Fi + energy) with intelligent feature selection and DL usually achieves accuracies of $\geq$90%. However, the problem of generalizing models across buildings, seasons, and space types remains one of the main research challenges [27,34,36]. The visual classification of algorithms (types and methods) for occupancy prediction is presented on Figure 1. As demonstrated in Table 4, a comprehensive collection of the most relevant publications on occupancy prediction methods has been compiled. This compendium includes key parameters, object types, methodologies, and performance indicators.

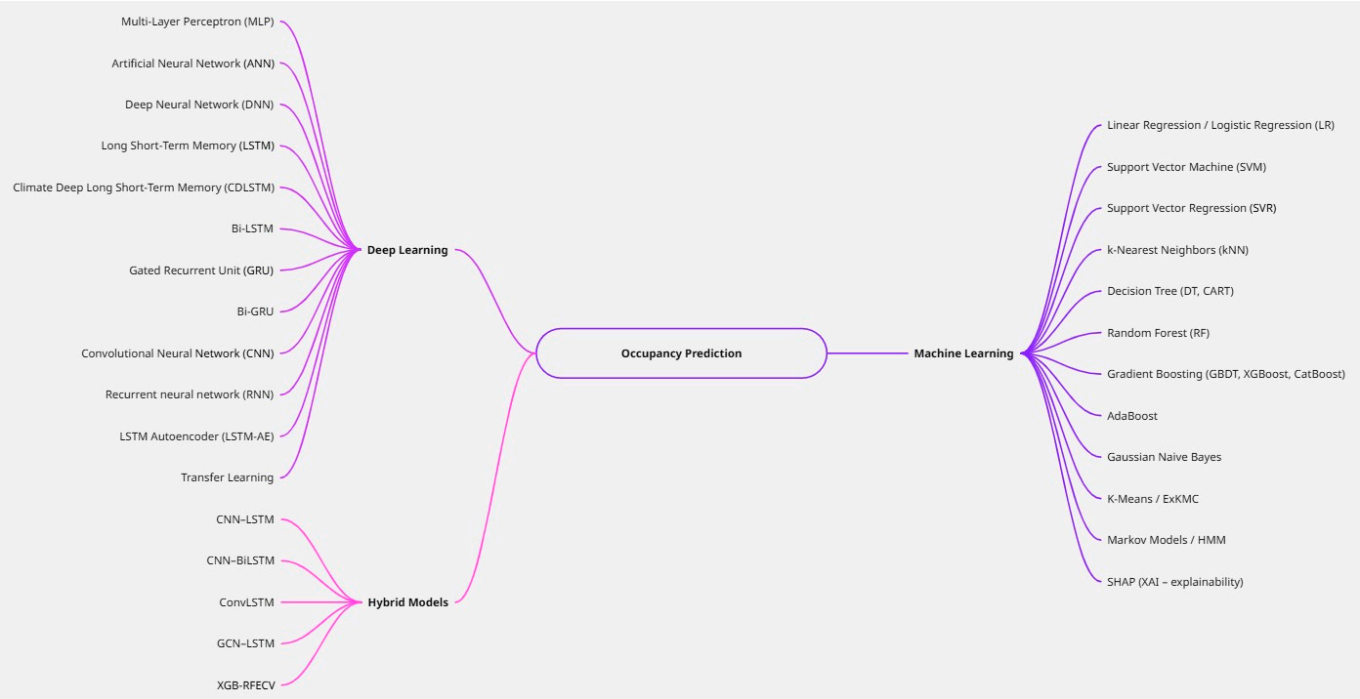


Figure 1. Algorithms classification for occupancy prediction.

Table 4. The list of selected publications: Occupancy Prediction.

	Object Type	Algorithms	Result	Advantages	Limitations
[26]	Residential (apartment—HIKARI, Lyon)	LR, SVM, DT, GBDT, RF, ANN	GBDT, RF, ANN: F1 > 0.85 (in summer and winter up to 0.90–0.92); seasonal models clearly more effective than "universal" ones	High F1 (>0.85–0.92); seasonal models outperform universal ones; GBDT, RF, ANN.	Universal models less adaptive; higher computational cost and tuning needs; limited interpretability and higher maintenance for seasonal training.
[29]	Residential (apartments, test house simulation)	DT, kNN, MLP, GRU	Classification: 74–76% accuracy; best comfort–savings: kNN (TOPSIS R ~0.71, MissTime = 23 min/day, 4–5% savings); regression: MAE arrival 59–67 min; choice of problem formulation more important than choice of ML algorithm	Balanced comfort–energy trade-off; kNN achieved best overall performance (TOPSIS ≈ 0.71, 4–5% savings); consistent classification (74–76%) across model.	Moderate accuracy and prediction delays (MAE ≈ 1 h); limited gain from advanced models; model selection less critical than problem formulation.
[28]	Public (offices—University of Southern Brittany)	Stacked LSTM, deep sequential MLP, SVM, RF; transfer learning	Best: stacked LSTM with transfer learning—71% (room B), 69% (C); without transfer learning ~62–66%	Transfer learning significantly improved accuracy (≈+5–8 pp); stacked LSTM.	Overall accuracy moderate (<75%); models sensitive to spatial variability; high training complexity and data demand.
[27]	Public and residential (overview—various)	Machine Learning (SVM, kNN, DT, RF, K-means), Deep Learning (RNN, LSTM, hybrids)	Most studies: accuracy of detection/prediction of presence 80–95% (sensors + ML); challenge: generalization of models between objects	High detection accuracy (80–95%) across diverse settings; effective integration of sensor data with ML/DL models.	Limited generalization across buildings; dependence on context-specific data and calibration.
[30]	Residential (single-family homes, ECO dataset)	CNN-BiLSTM, CNN-LSTM, ConvLSTM, CNN, LSTM, MLP, kNN, SVM, RF, AdaBoost	CNN-BiLSTM: ~89–90% accuracy, high F1/AUC; alternative DL ~80–89%	CNN-BiLSTM achieved highest accuracy (~90%) with strong F1/AUC; deep models effectively capture temporal–spatial features.	High computational cost; complex architectures; limited interpretability and scalability to unseen homes.
[11]	Residential and public (student room, living room, UCI)	CNN, AlexNet, GoogLeNet (transfer), kNN, DT, RF	Images + kNN/RF: UCI 99.4%, student room 99.1%, living room 98.5%; CNN up to 95–99%	Very high accuracy (≈95–99%); vision-based models with transfer learning deliver robust occupancy detection.	High computational demand; privacy concerns; limited scalability beyond controlled environments.
[36]	Public (office, library, lecture room—university, Singapore)	DNN, LSTM, Bi-LSTM, GRU, Bi-GRU, XGB-RFECV	MAE: office (Bi-GRU) ~0.12; library (GRU) ~0.16; lecture (Bi-GRU) ~0.09; RMSE ~0.33/0.33/0.32 respectively; key features: CO_2, Wi-Fi	Very low prediction error (MAE ≈ 0.1–0.16); Bi-GRU and GRU effectively model temporal dynamics from CO_2 and Wi-Fi data.	Model performance site-dependent; high data and sensor requirements; interpretability remains limited.
[12]	Public (synthetic/lab dataset—Kaggle room occupancy)	kNN, AO-ANN (BP), DT; SHAP	kNN, AO-ANN: 99.5% accuracy; DT: 98.9%; SHAP: dominant features—light, CO_2	Very high accuracy (≈99%); interpretable results via SHAP; key features (light, CO_2) consistent with occupancy behavior.	Synthetic dataset limits real-world validity; potential overfitting; low diversity of environmental conditions.
[34]	Public & residential (overview—various)	Schedules, Markov Models, HMM, SVM, kNN, Trees, RF, Boosting, CNN, LSTM, hybrid, TL, FL	Review: cameras/CNN/UWB ~90–95%+, CO_2/ANN/SVM 80–95%, Wi-Fi/BLE/ML ~90–98%, classic PIR 70–90%	Broad applicability across sensing modalities; modern ML/DL (CNN, LSTM, TL, FL) achieve high accuracy (≥90%) with diverse inputs.	Performance strongly sensor-dependent; privacy and data-sharing issues; heterogeneous setups hinder model generalization.

Table 4. *Cont.*

	Object Type	Algorithms	Result	Advantages	Limitations
[35]	Public (office—Grenoble)	ExKMC (Explainable K-Means Clustering), DT	F1 ~82% (all features), up to 85% without power; better recognition of rare classes than DT	Good overall performance (F1 ≈ 82–85%); explainable clustering enhances interpretability and minority-class detection.	Moderate accuracy; dependence on feature quality; limited scalability to larger or more dynamic datasets.
[32]	Public (meeting rooms, Finland: office + hospital)	CDLSTM (CNN + 3xLSTM + FC), transfer learning	Occupancy detection: MCC ~0.81–0.82 with 2 days of transfer; MCC classes 0.63–0.67 TL	High detection performance (MCC ≈ 0.8); transfer learning effective with limited adaptation data (≈2 days).	Moderate class-wise balance (MCC ≈ 0.63–0.67); complex architecture; increased computational and tuning effort.
[13]	Public (office)	CatBoost, RF, XGBoost, SVM, NN	CatBoost: 98–99% detection/prediction, much better than RF, SVM, NN	Very high accuracy (≈98–99%); CatBoost outperforms other models with efficient handling of categorical and nonlinear features.	Potential overfitting on limited datasets; reduced interpretability; high dependency on feature engineering.
[33]	Public (university, ECS building—US, Long Beach)	LSTM-AE (anomaly det.), GCN-LSTM, LSTM	GCN-LSTM/LSTM on 'cleaned' data: +13–14% quality, HVAC: 20% less energy	Graph-based LSTM improves prediction accuracy (+13–14%) and enables notable HVAC energy savings (~20%).	Requires data pre-processing; high model complexity; limited generalization to unstructured sensor networks.
[31]	Public (office/smart building—real)	DT, RF, SVM, LR; Semi-supervised: self-training, pseudo-labelling	Full labeling: high accuracy/F1; semi-supervised: similar results with significantly fewer labels	High accuracy with full labelling; semi-supervised learning achieves comparable performance using fewer annotated samples.	Dependent on initial label quality; potential error propagation in pseudo-labeling; limited transparency of decision process.

3.2. Thermal Comfort Prediction

The development of ML methods in thermal comfort prediction can be organized around three related trends: (1) prediction of individual preferences and thermal sensations, (2) generalized group models and PMV/SET "control" models, and (3) supporting techniques such as calibration, sequential time modelling, reinforcement learning, and visual detection of adaptive behaviors.

In the initial approach, the majority of studies formulate the task as a multi-class classification of individual preferences ("warmer/no change/cooler") or subjective assessment of TSV/TPV/TCV/TAV in the "here and now" state. A study utilizing the ASHRAE II database shows that the Deep Forest model achieves superior performance in predicting three-class thermal preferences (approx. 82% accuracy, F1-macro ~0.74), clearly outperforming SVM, logistic regression, MLP, RF, and XGBoost [37]. In turn, a comparison of a wide range of algorithms (RF, GBM, AdaBoost, SVM, kNN, ANN, NB, tree, logistics) based on the same database for TSV (3- and 7-point scale) indicates RF as the dominant method (approx. 66% accuracy for three classes, 61% for seven classes), significantly better than the classic PMV model, whose accuracy for three-class TSV does not exceed 42–43% [38]. The concept of personalization is being developed, among others, by models reflecting age-related thermal sensitivity and comfort perception of older people: separate shallow neural networks were built for 28 seniors, achieving an average of ~74% accuracy, Kappa ~0.61, and AUC ~0.83. This represents a significant improvement over the "converted" PMV [39]. In air-conditioned offices in Malaysia, TSV, TPV, TCV, and TAV were predicted simultaneously, showing that adding simple spatial features (distance from the

window and AC unit) to standard Fangerow inputs can increase TSV accuracy in RF to ~71% (a gain of ~38.6 p.p. compared to the baseline model without SP) and improve metrics for TPV/TCV/TAV [40]. In a similar vein, but using a richer spatial context, paper on digital twin and Build2Vec shows that BIM graph embeddings (room topology, location of windows, air vents, furniture) combined with physiological signals and IEQ, feeding RF, allow the accuracy of preference prediction to be increased from 0.58–0.72 to ~0.86 [41]. Another study utilized streaming data from 251 home air conditioners (temperature, humidity, settings, operating modes, time, temperature differences) to predict the user's next action ("raise", "lower", "no change"), showing that RF achieves 55–68% accuracy in the three-class version and up to ~87% in the two-class variant after selecting stable devices and extending time-operational features. In addition, it was determined that after approximately 30–50 adjustment episodes, the user preference model stabilizes [42]. A separate direction in this trend is represented by the sequential R-LSTM model, which, based on a 120 min window of physiological (heart rate, skin temperature) and environmental signals, predicts the next three-class preference, achieving ~78% accuracy and F1~0.74. This clearly outperforms RF trained on static features [43].

The second main trend concerns group comfort models and the integration of prediction with automatic control. In school conditions, a multi-task DeepComfort model was proposed, which simultaneously classifies the TSV, TPV, and TCV of children in naturally ventilated classrooms in winter. Subsequently, the collected data was verified with the ASHRAE II. The achieved classification accuracy reached ~90% (TSV), ~87% (TPV), and ~89% (TCV), exceeding the results obtained using classic SVM, RF, kNN, AdaBoost, and single-task DNN [38]. A related study [44] examines the spatial variability of SVM (one-vs-rest) models trained separately for five schools: F1 for TSV/TPV/TCV ranges from approx. 41–71%, 30–64%, and 37–74%, respectively, with models trained per school generally outperforming the "All Schools" model. The disparities among schools attain several dozen percentage points, confirming the strong dependence of predictor quality on a specific building and its operational and functional organization, including usage schedules and occupancy dynamics. In office buildings, models that predict PMV and/or SET in continuous mode are an important class. In a study of 13 office buildings in Amman, SVR, RF, ANN, and XGBoost were compared, with very high fit for PMV: RF with $R^2 \approx 0.967$, MAE ≈ 0.021, RMSE ≈ 0.073, and XGBoost with $R^2 \approx 0.964$, MAE ≈ 0.010, RMSE ≈ 0.076, making them good candidates for integration with MPC for the simultaneous control of comfort and energy consumption [45]. Even further goes the study in which ANN was developed on real university data (after NCA feature selection and BO tuning) for the simultaneous prediction of PMV, SET, and TS. The aforementioned network achieves extremely small errors (PMV: MSE $\approx 2.5 \times 10^{-6}$, MAE ≈ 0.001, $R^2 \approx 100\%$; SET: MSE ≈ 0.0013, MAE ≈ 0.016 °C, $R^2 \approx 99.98\%$; TS: accuracy $\approx 98.7\%$, F1 ≈ 0.96, AUC of individual classes > 0.97). This allows the air conditioning settings to be controlled to maintain PMV in the range of -0.5 to $+0.5$, with estimated energy savings of 3–6% for each 1 °C setpoint correction [46]. A completely different paradigm is represented by the reinforcement learning approach, where the tabular off-policy SARSA learns preferred temperature ranges and corresponding settings based on synthetic user "overrides" in an office building simulated using EnergyPlus (version 9.6.0). The algorithm can learn comfort ranges within 4–16 working days, reduce the number of manual interactions with the thermostat by ~70–80%, and reduce heating energy consumption by approx. 35–40% in scenarios where preferences allow for lower temperatures [47].

The third trend includes several papers concentrated on data quality, prediction calibration, and alternative (indirect) comfort signals. One study utilized the ASHRAE II database to simultaneously model four perception labels (TSV, TPV, TAV, TCV) using

classical classifiers (SVM, RF, tree, kNN, NB, logistic, MLP) and the M-ARM algorithm to detect biased/inconsistent responses. After filtering out such noises and applying various probabilistic calibration techniques (Platt, isotonic, Beta, BBQ, temperature scaling, GP), it is possible to increase the weighted F1 by 1–5.5 p.p. and reduce the ECE by up to ~80% (e.g., for NB from ~0.20 to ~0.04), thereby markedly enhancing the reliability of probabilistic predictions used later in control [48]. A very original and complementary approach is to build a non-invasive processing chain in which the input signal for further prediction of thermal preference is automatically detected adaptive behavior—based on 654 office surveys, 14 characteristic gestures were selected (putting on/taking off a jacket, wiping sweat, warming hands with breath, crossing arms over chest, etc.). Thereafter, 27,504 video recordings were obtained from eight directions and two heights, utilizing I3D and SlowFast models to recognize them in real time with an accuracy of 95–96% for multi-class classification and ~84–85% for adaptive vs. non-adaptive classification [49]. These types of signals can then feed into the previously described preference prediction models or RL input states.

Overall, the analyzed studies show that there is no universal approach to thermal comfort prediction. The selection of algorithm and features depends on the type of facility (office, school, apartment), time horizon (here and now vs. the near future), purpose (personal recommendations vs. building control), and data quality and structure. Tree models (RF, XGBoost, Deep Forest) have been shown to dominate classification tasks based on large population samples [37,38,42]; deep ANNs with MTL and personalization are particularly effective in individual and multi-label models [39,46,50]. Meanwhile, RL [47], sequential models [43], BIM-IoT hybrids [41], and visual behavior recognition [49] have been shown to open new paths for integrating thermal comfort with intelligent, adaptive building control. The visual classification of algorithms (types and methods) for thermal comfort prediction is presented in Figure 2. As demonstrated in Table 5, a comprehensive collection of the most relevant publications on thermal comfort prediction has been compiled. This compendium includes key parameters, methodologies, object types, and performance indicators.

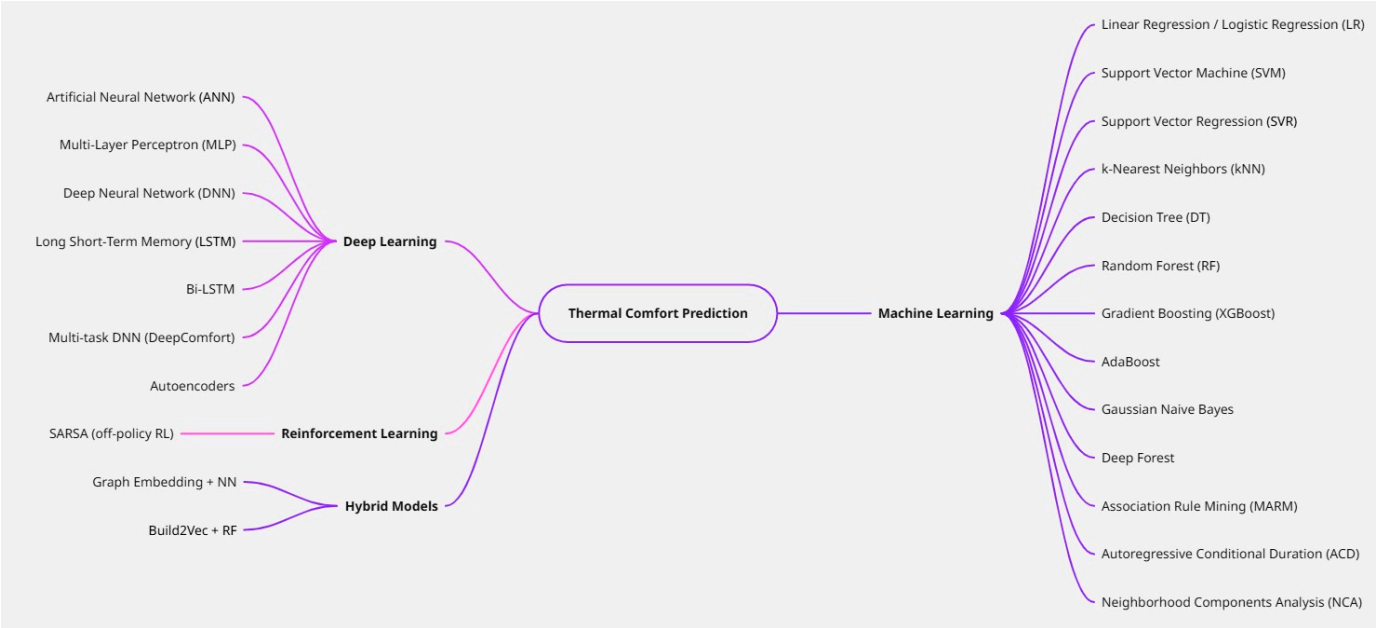


Figure 2. Algorithm classification for thermal comfort prediction.

Table 5. The list of selected publications: Thermal Comfort Prediction.

Title	Object Type	Algorithms	Result	Advantages	Limitations
[37]	Public (mixed buildings from ASHRAE II: offices, residential, others—aggregated)	SVM, Logistic Regression, BP-ANN (MLP), RF, XGBoost, Deep Forest	Best Deep Forest model: approx. 82% accuracy and F1 macro ~0.74 in preference prediction (warmer/no change/cooler); removal of CLO lowers accuracy to ~80%; tree models clearly better than non-tree models, most important features: air temperature and age.	Tree-based models (Deep Forest, RF, XGBoost) outperform others (~82% accuracy); robust feature importance analysis; key variables (temperature, age) well identified.	Moderate class balance (F1 ≈ 0.74); reduced accuracy without clothing data; limited interpretability of aggregated model outputs.
[50]	Public (schools—naturally ventilated primary classrooms)	DeepComfort (multi-task DNN with hard parameter sharing, dropout), SVM, RF, DT, kNN, AdaBoost, single-output DNN	DeepComfort: ~90% accuracy and F1 for TSV, ~87% for TPV, ~89% for TCV on winter student and ASHRAE II data; significantly better than all single-task models, resistant to unbalanced classes and "illogical" responses from children.	High predictive performance (~87–90% accuracy/F1 across comfort votes); multi-task learning outperforms single-task and classical models; robust to class imbalance and noisy/illogical responses.	Increased architectural and training complexity; reduced interpretability compared with tree-based models; requires sizeable labeled datasets.
[38]	Public (various buildings from ASHRAE II: mainly office buildings and others—generator)	RF, Gradient Boosting, AdaBoost, SVM, kNN, ANN, Naive Bayes, DT, LR (regression and classification)	For TSV (classification) RF best: ~66.3% accuracy (3-class scale) and 61.1% (7-class), vs. PMV ~42–43% for 3 classes; key features: temperature, humidity, CLO, air velocity, age, MET.	RF achieved best performance (~66% vs. 42% PMV); effective for nonlinear relationships and multi-feature interactions; interpretable feature importance.	Moderate accuracy limits comfort prediction reliability; performance sensitive to data quality and scale definition; weaker generalization across buildings.
[43]	Residential (laboratory/field study—individual adults, quasi-home environment)	R LSTM (LSTM with L1 regularization and dropout), compared to RF based on aggregated features	R LSTM: average ~78% accuracy, weighted F1 ≈ 0.74, better AUC than RF; for most of the 14 individuals, accuracy improved from ~55–76% to ~66–86%, demonstrating the advantage of s equential DL for personal preference.	Sequential DL (R-LSTM) improves personal comfort prediction (~10 pp gain); captures temporal dynamics and individual variability; better AUC than RF.	Moderate overall accuracy (~78%); higher training cost and parameter sensitivity; limited interpretability for end-user feedback.
[48]	Public (various types of buildings with ASHRAE II—unit)	SVM, RF, DT, kNN, Gaussian Naive Bayes, LR, MLP; additionally MARM (associations) + various probabilistic calibration methods	After filtering out biased votes, weighted F1 for different labels (TSV, TPV, TAV, TCV) increases by ~1–5.5 p.p. (e.g., TPV: RF ~61.9% F1, +~4.3 p.p.); calibration reduces ECE by up to 80.66% (e.g., NB from ~0.20 to ~0.04), increasing the reliability of probabilistic predictions.	Calibration greatly improves reliability (ECE ≈ 80%); filtering biased votes enhances F1 by 1–5 pp; RF and NB yield stable probabilistic outputs.	Gains modest in absolute accuracy; preprocessing and calibration add complexity; model performance depends on label consistency.
[42]	Residential (home dwellings—251 AC in southern China)	LR, kNN, Naive Bayes, SVM, RF; class balancing methods (undersampling, SMOTE)	Three-class problem (warmer/no change/cooler): best RF with macro accuracy ~67.8% (range 55–68% between models); in the two-class version (raise vs. lower setting), accuracy ~72.1–79.9%, and for selected stable devices with additional features even ~87.3%; ~30–50 episodes are sufficient to stabilize the user preference model.	Adequate accuracy (up to 87%) with few training episodes (~30–50); RF robust across class setups; class balancing improves stability.	Performance varies widely by device; limited generalization; relies on sufficient interaction data for personalization.

Table 5. *Cont.*

Title	Object Type	Algorithms	Result	Advantages	Limitations
[41]	Public (university educational building—office and teaching building SDE4)	Build2Vec (BIM graph embedding) + RF; comparison with previous RF models based on "classical" temporal environmental features	Baseline models: accuracy ~0.58–0.72; the Build2Vec model (embedding + heart rate + body temperature) achieves accuracy ~0.86 (train ~0.92, val ~0.93), an improvement of ~14–28 p.p. compared to the baseline, which shows the high value of spatio-temporal features.	Significant accuracy gain (+14–28 pp) with Build2Vec embeddings (~0.86 accuracy); effectively captures spatio-temporal and physiological patterns.	Requires detailed BIM and sensor integration; higher computational cost; potential overfitting due to complex feature space.
[40]	Public (air-conditioned offices—university offices in Malaysia)	DT, RF, linear SVM, Neural Network (MLP), Naive Bayes, kNN; models with/without spatial parameters (WD, ACD)	With spatial parameters: for TSV RF ~71.1% accuracy (+38.6 p.p. compared to the model without SP); for TP RF ~70% accuracy (on average +6.8 p.p.); for TC SVM the largest relative increase (+50%), RF the highest accuracy ~54.1%; for TA, all algorithms 87.5–91% accuracy; for TSV RF with SP: MAE ≈ 0.39, RMSE ≈ 0.72 (MAE decrease ~41%, RMSE ~26% vs. PMV).	Spatial parameters greatly enhance performance (e.g., +38–50 pp); RF consistently highest accuracy (~70–91%); substantial error reduction vs. PMV.	Increased data collection complexity; model gains depend on spatial feature quality; limited portability to non-AC environments.
[47]	Public (office space—single office in EnergyPlus simulation)	SARSA off-policy (model-free RL) with ε-greedy and experience "refresh" (average Q table)	RL learns comfort limits in 4–16 working days, reducing the number of user interactions with the thermostat by up to ~70–80% when default settings are far from preferences; for tolerant/cool-loving users, heating energy reduction of ~35–40%, with a decrease in "exceedance" (time outside the comfort range), at the cost of higher costs for high-temperature preferences.	Rapid learning of user comfort limits (4–16 days); large reduction in interactions (~70–80%) and heating energy (~35–40%).	Performance sensitive to user preference profiles; higher energy use for warm-bias users; requires extended exploration phase.
[45]	Public (office buildings—13 LEED and conventional offices in Amman)	SVR, RF, ANN, XGBoost	Best: RF with $R^2 \approx 0.967$, MAE ≈ 0.021, RMSE ≈ 0.073 and XGBoost with $R^2 \approx 0.964$, MAE ≈ 0.010, RMSE ≈ 0.076; both methods clearly superior to ANN ($R^2 \approx 0.922$) and SVR ($R^2 \approx 0.801$); high accuracy of PMV prediction enables integration with MPC for HVAC comfort and savings.	Very high predictive accuracy ($R^2 \approx 0.96$–0.97); RF and XGBoost outperform others; suitable for MPC integration to optimize HVAC comfort and energy use.	Model performance dataset-specific; lower interpretability than physical models; requires comprehensive feature calibration.
[39]	Residential (apartments for elderly people living independently)	28 separate shallow neural networks (ANN, 1 hidden layer)—one per participant; PMVc reference (PMV model)	Personal models: average ~74% accuracy, Kap-pa ≈ 0.61, AUC ≈ 0.83 (accuracy range 33–100%); compared to PMVc, improvement in accuracy by ~49 p.p., Kappy by ~151% (from ~0.24), AUC by ~34% (from ~0.62); adding a subjective health assessment results in a small, statistically weak improvement in averages, but significantly increases accuracy for some individuals.	Strong individual adaptation (~74% accuracy, +49 p.p. vs. PMVc); substantial improvement in agreement (Kappa ≈ 0.61) and discrimination (AUC ≈ 0.83); flexible to include subjective factors.	High inter-individual variability (33–100%); requires personal calibration data; limited generalization across users.

Table 5. *Cont.*

Title	Object Type	Algorithms	Result	Advantages	Limitations
[44]	Public (elementary schools—5 naturally ventilated schools, 14 classrooms)	Multi-class SVM (one vs. rest) with linear, polynomial, and RBF kernels; class weighting, permutation feature importance	Strong spatial variability: F1 for TSV ~41–71%, TPV ~30–64%, TCV ~37–74% between schools; models trained separately for schools are usually better than the "All Schools" model (local increases in TCV accuracy by 10–71 p.p.); models on ASHRAE II (adults) have F1 ~0.78–0.86; generalizing models (trained on n − 1 schools) give stable but lower F1 (TSV ~0.53–0.59).	Captures spatial variability across classrooms; locally trained models show large accuracy gains (up to +71 p.p.); interpretable feature importance via SVM weights.	Limited generalization across schools; moderate F1 scores (~0.4–0.7); performance sensitive to data imbalance and local conditions.
[49]	Public (offices—office users in the survey; laboratory—recordings of behavior)	I3D (Two Stream Inflated 3D Con-vNet), SlowFast Network; adaptive vs. non-adaptive binary classification	Average accuracy of recognizing 14 adaptive behaviors: ~95% (I3D) and ~96% (SlowFast), for most gestures >90%; inference time <0.05 s/video; adaptive vs. non-adaptive classification: accuracy ~83.7% (I3D) and 85.4% (SlowFast). The results demonstrate the possibility of reliable, real-time acquisition of discomfort signals for further preference modeling and HVAC control.	Very high recognition accuracy (>90%); real-time performance (<0.05 s/video); enables automatic detection of discomfort behaviors for HVAC feedback.	Requires large labeled video datasets; high computational cost for training; limited to visible, camera-detectable actions.
[46]	Public (university building—air-conditioned rooms, laboratories, test room)	SVR (RBF), DT, bagged Ensemble Tree, ANN (1 hidden layer, Bayesian regularization); NCA feature selection, BO tuning	ANN best: for PMV MSE $\approx 2.5 \times 10^{-6}$, MAE ≈ 0.001, $R^2 \approx 100\%$ (100% of samples within ± 0.1 PMV); for SET MSE ≈ 0.0013, MAE ≈ 0.016 °C, $R^2 \approx 99.98\%$ (100% of samples within ± 0.5 °C); for TS accuracy $\approx 98.7\%$, F1 ≈ 0.96, AUC classes > 0.97. The models were used for automatic setpoint correction while maintaining PMV within the range of −0.5 to +0.5 and estimated energy savings of ~3–6% for each 1 °C change in setting.	Extremely high prediction accuracy ($R^2 \approx 1.0$); effective for PMV/SET control and automatic setpoint adjustment; measurable energy savings (~3–6% per °C).	Computationally intensive tuning (BO, NCA); limited interpretability of ANN decisions; performance reliant on high-quality sensor data.

3.3. Seasonal Activity Recognition

The field of research concerning the recognition of seasonal activity and energy load in smart building environments is currently encompassing a broad spectrum of issues. The scope of the model extends from the behavioral patterns of individual elderly people to load forecasting on the scale of commercial buildings and entire power systems. In the domain of habit monitoring, approaches based on non-invasive environmental sensors and sequential ML models dominate. The recognition of seasonal sleep patterns in elderly residents of smart homes is achieved through K-means of attributes describing sleep parameters (sleep time, falling asleep, waking up, restlessness) and subsequent classification of days using DTs, which achieve an accuracy of 87–95% and allow subtle differences between seasons to be identified [51]. The uniformity and stability of daily routines over a prolonged period of time are evaluated by segmenting daily location sequences and hierarchically clustering segments, where an increasing number of clusters and RMSE errors, MAE of

the start, and duration of the segments are an indicator of the "dysregulation" of behavior, which allows more stable periods from periods of disturbance in older people to be distinguished [52]. Moreover, a preliminary investigation conducted in single-person senior apartments has demonstrated that a simple statistical method based on thresholds can effectively distinguish between "normal" and atypical behavior in defined time windows (e.g., meals), generating relatively few false alarms and achieving substantial acceptance from caregivers [53].

At the level of more complex sequential patterns, ambient intelligence research utilizes deep models to explore routines and their violations. In multi-resident systems, sequential neural networks, including MLP, 1D/2D convolutional networks, various variants of RNN (Elman, GRU, LSTM), and CNN + LSTM hybrids, are used to recognize the activities of daily living (ADL) of two residents based on the states of ambient sensors. The results demonstrate that CNN + LSTM architectures achieve accuracies of 80–90% and clearly outperform classic HMM or CRF models, confirming the relevance of modeling long-term temporal dependencies [54]. Even further-reaching are approaches that combine deep NILM (seq2point), MLP, LSTM, and ILP models with multi-day activity logs. The induction of logical rules, for instance, delineating "morning routines," is followed by the detection of missing steps or atypical insertions. The studies achieve notably high F1 scores based on both real and synthetic data, demonstrating that combining representation learning with formal logical rules increases interpretability and usefulness in health monitoring [55]. On a city scale, instead of direct prediction, sequential and periodic pattern mining (e.g., CM-SPAM, PFPM) is used on check-in data to identify repetitive visit sequences and seasonal cycles, which provides insight into residents' mobility patterns and can provide context for energy planning and urban services [4].

The second, highly significant line of research focuses on the prediction of the electrical load—from individual homes to nationwide systems—in which a data-driven approach and sequential models predominate. A systematic review of STLF methods for the residential sector demonstrates that virtually all studies are based on historical load profiles (often enriched with calendar and weather features), with conventional statistical models (AR, ARIMA/SARIMA) gradually giving way to AI and ML methods (ANN, SVR, other supervised models) and deep LSTM and CNN architectures. For housing aggregates, typical MAPE errors are typically in the range of 5–10%, while for individual households, they are most often in the range of 5–20%, reflecting greater variability in behavior [1]. A case study of 709 households demonstrated that even a simple ANN network with a single hidden layer can significantly outperform the classic ARIMA, achieving an average MAPE of around 1.8% compared to 2.61%, especially in conditions of irregular, non-linear patterns and after disturbances, after which the neural network recovers its accuracy more quickly [56]. Against this background, HousEEC stands out. In this model, a DRNN with two input branches—processing classic time-weather features and a rich set of time-frequency features generated by TSFRESH—achieves very good results on Pecan Street data (RMSE of 0.44 kWh, MAE 0.23 kWh, $R^2 = 0.90$ for individual houses and errors of a few percent for an aggregate of 300 households). It clearly outperforms both classic ML algorithms and competing deep architectures [57]. Furthermore, studies on forecasting consumption in a single-family home using LSTM demonstrate that separating modeling into the device level (consumption and operating time of individual receivers) and the aggregate level (total power) allows for very low MSE values and more accurately captures the impact of user behavior on the load profile [58].

Another important trend pertains to the incorporation of behavioral pattern recognition into load forecasting within HEMS. In an approach based on the clustering apartment profiles using DBSCAN, k-means, and the PCC correlation algorithm (with delays of up

to 168 h), households are first grouped according to the similarity of their load patterns. Subsequently, separate sequence-to-sequence LSTM models are trained for each group, with the architecture being extended with fully connected layers, dropout, and stable SoftMax. This solution reduces the RMSE and gives most users MAPE errors below 10%, which shows how strongly the inclusion of behavior profiles improves the quality of forecasts [2]. In a commercial smart building, the weekly load profile (hourly) forecast compares classic DMD with a series of sequential models (RNN, LSTM, BiLSTM, GRU) and a hybrid RNN + BiLSTM. After tuning the hyperparameters, it is the hybrid model that achieves the lowest RMSE error, improving both DMD and single recurrent architectures [59]. On the scale of power systems, however, load profile propagation methods are used, in which weekly maximums predicted by classical ML models (e.g., MLP, SVR, RF, GBRT) are converted into an hourly profile based on historical patterns. The parallel utilization of 2D convolutional networks for direct STLF allows for very low MAPE and RMSE values with an acceptable compromise between learning complexity and accuracy [60].

Hybrid models combining signal decomposition with recurrent networks form a substantial subgroup. In a study based on the UCI dataset "Individual household electric power consumption", the EMD-BI-LSTM method decomposes the global active power series into IMF components and residuals and then forecasts each component separately using a bidirectional LSTM network. Following the consideration of meteorological conditions and appropriate delays, an MAPE of 0.28% and an RMSE of 0.31 are obtained. These values are significantly better than in the case of the classic BiLSTM without decomposition and similar hybrids with GRU [61].

At the same time, a rich set of methods for detecting energy and operational anomalies is being developed. The data mining framework for office buildings in Chongqing combines the construction of typical daily TELP profiles (using dimensionality reduction, DBSCAN, and k-means) with their classification using CART trees based on seasonal and weather characteristics and then interprets deviations from the predicted class as potential anomalies [62]. A large portfolio of telecommunications facilities used a system based on SAX transformation, in which each day is encoded as a symbolic word over a small alphabet of power levels, and "synthetic truth" is created by selecting the most frequent words covering the majority of observations. Semi-supervised SAX-CART and SAX-MLP models use context (calendar, weather, facility type) to classify symbols or estimate consumption, identifying anomalies as rare symbols or points with large errors [63]. In an academic building equipped with multiple HVAC units, an autoencoder with a TiFe improves the reconstruction of normal load profiles. This approach exhibits enhanced resilience to the presence of anomalies in the training set and facilitating the interpretation of unusual periods of operation and groups of devices with similar patterns. However, the emphasis was placed on an evaluation of the quality of reconstruction rather than classic detection metrics [64]. At the level of the office building portfolio, algorithms for the automatic extraction of installation operating hours (double gradient, cumulative histogram, start gradient) are also used, which, after prior clustering of daily profiles using the DBA k-means method, allow the definition of operational measures of operating hours. Their incorporation as features in EUI regression using RF and Gradient Boosting enhances the capability to explain consumption and indicates a potential for reduction of several to over ten percent [65].

Another, but related, topic concerns user interaction with building thermal conditions. In a study dedicated to window state detection during the heating season, the "open/closed" problem is formulated as a binary classification task based on 15 min aggregates of the temperature, humidity, and CO_2 concentration. A comparison of classical classifiers (LDA, SVM with RBF kernel, RF) with recurrent networks (LSTM, GRU)

demonstrates a clear advantage of sequential models, which, with appropriately selected feature transformations, achieve an F1 score of around 0.78 and can correctly detect about 84–88% of window opening times with a relatively small number of false alarms [66]. Combined with the results for routine recognition, load forecasting, and anomaly detection at various spatial scales, all of the discussed works show that modern smart building systems are increasingly integrating information about user behavior, environmental conditions, and energy profiles. The methods employed by these systems are based on sequential and hybrid models, enabling simultaneous improvements in comfort, safety, and energy efficiency. The visual classification of algorithms (types and methods) for seasonal activity recognition is presented on Figure 3. As shown in Table 6, a comprehensive collection of the most relevant publications on seasonal activity recognition has been compiled. This compendium includes key parameters, methodologies, object types, and performance indicators.

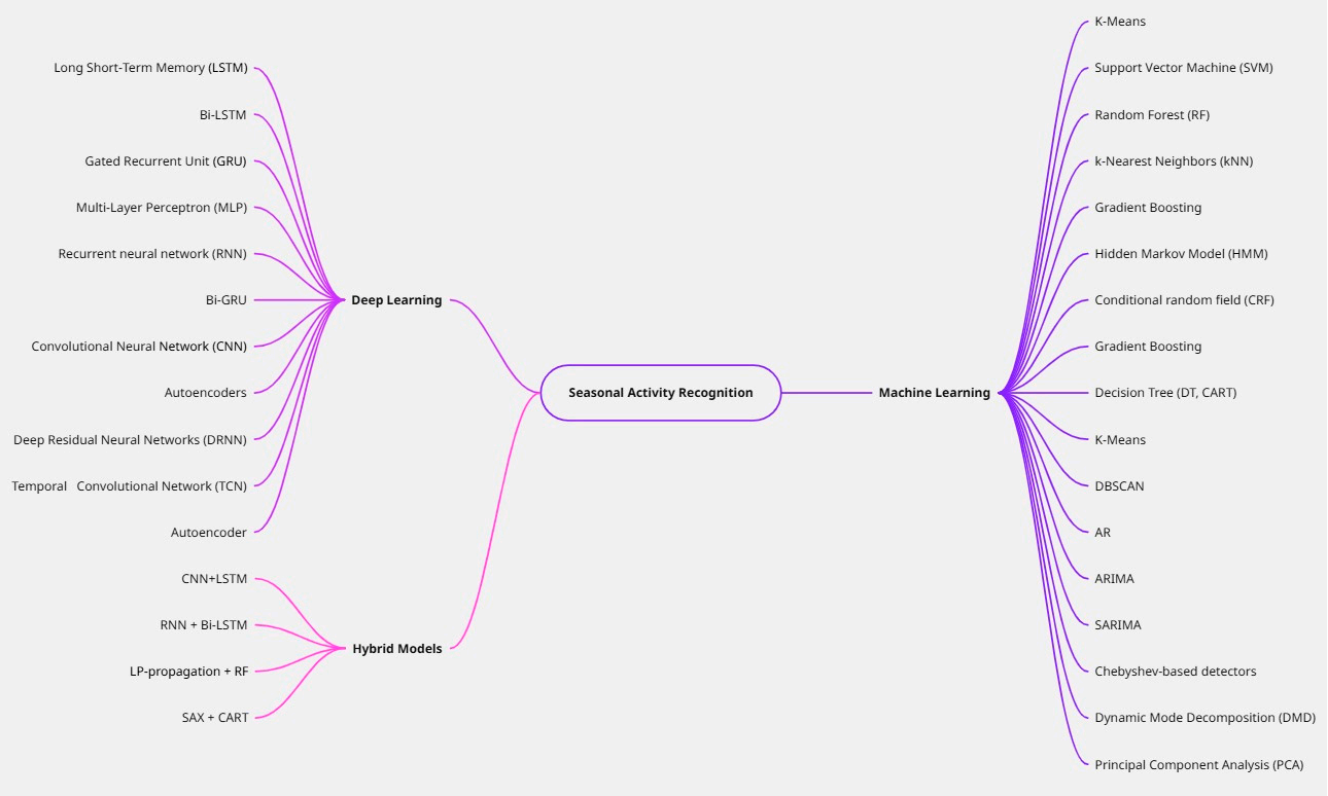


Figure 3. Algorithm classification for Seasonal Activity Recognition.

Table 6. The list of selected publications: Seasonal Activity Recognition.

Title	Object Type	Algorithms	Result	Advantages	Limitations
[51]	Residential (single-resident smart apartments for older adults)	K-means, DTs	Silhouette 0.4–0.5 (good clusters); DT sleep behavior classification accuracy 87–95% in test.	Good clustering quality (silhouette 0.4–0.5); high classification accuracy for sleep patterns (87–95%); simple and interpretable models.	Limited scalability to multi-resident settings; moderate cluster separability; relies on consistent sensor behavior.
[66]	Residential (apartment bedroom)	LDA, SVM (RBF), RF, LSTM, GRU	Best LSTM/GRU: F1 $\approx$ 0.78; classic methods F1 $\approx$ 0.70–0.72; detects 84–88% of window opening periods with few false positives.	High detection accuracy (F1 $\approx$ 0.78); effective temporal modeling with LSTM/GRU; low false-positive rate for window events.	Requires sequential sensor data; higher computational cost; classical models yield moderate performance (F1 $\approx$ 0.70).

Table 6. *Cont.*

Title	Object Type	Algorithms	Result	Advantages	Limitations
[54]	Residential (two smart apartments, two residents each)	MLP, 1D/2D CNN, Elman RNN, GRU, LSTM, hybrid CNN + LSTM, HMM, CRF, KNN	Best 1D CNN + LSTM: accuracy ≈ 86.3% (apt A, 30 days) and 88–92% (apt B), clearly above HMM/CRF.	High recognition accuracy (≈86–92%); effective temporal–spatial pattern extraction with hybrid CNN + LSTM; robust across apartments.	Larger computational demand; requires extensive labeled data; performance dependent on sensor placement consistency.
[59]	Public (commercial smart office building)	DMD (Hankel), RNN, LSTM, BiLSTM, GRU, hybrid RNN + BiLSTM	RMSE: DMD 3.99; best tuned hybrid RNN + BiLSTM RMSE 3.83; better than RNN 3.91, LSTM 4.39, GRU 5.01.	Improved prediction accuracy with hybrid RNN + BiLSTM (lowest RMSE ≈ 3.83); captures bidirectional temporal dynamics; outperforming traditional DMD and single RNN variants.	High model complexity and training cost; risk of overfitting on limited data; marginal gain over simpler architectures.
[60]	Public (national power system/ aggregated load)	MLP, SVR, RF, GBRT, 2D CNN, LP-propagation + RF	STLF: best 2D CNN MAPE 0.81%, RMSE 75.6 MW, WI 0.999; LP + RF for weekly peaks: MAPE ≈ 3.86%, RMSE 372.7 MW.	Very high short-term load forecasting accuracy (MAPE ≈ 0.8%); 2D CNN captures spatial–temporal dependencies; LP + RF effective for peak load estimation.	High computational requirements; reduced accuracy for long-term or irregular peaks; requires large, high-resolution datasets.
[4]	Public (city-scale residents, Foursquare users)	CM-SPAM sequential pattern mining, PFPM periodic pattern mining	Descriptive: frequent sequential rules and periodic visit patterns; reports support/period, not accuracy.	Identifies meaningful sequential and periodic behavioral patterns at city scale; scalable to large user datasets; useful for descriptive mobility analysis.	Lacks predictive evaluation (no accuracy metrics); sensitive to data sparsity and noise; provides limited temporal generalization.
[65]	Public (office buildings, Building Data Genome 2)	DBA k-means, Double Gradient, Cumulative Histogram, Gradient Beginning, RF, Gradient Boosting	Best operation hours: CH (R^2 = 0.58, K-L = 0.143); regression EUI RF + KPI: R^2 increases from ≈0.35 to 0.41; simulated savings 2.9–10.6%.	Improved EUI prediction with combined RF + KPI (R^2 = 0.41); effective extraction of operational patterns; potential energy savings (2.9–10.6%).	Moderate overall accuracy (R^2 < 0.6); performance varies by feature design; limited transferability across building types.
[62]	Public (office buildings)	DBSCAN, k-means, CART DT	TELP classification CART accuracy 82–92%; Dunn index used; two-step clustering extracts patterns/outliers better than single k-means/GMM.	High TELP classification accuracy (82–92%); two-step clustering improves pattern detection and outlier identification; simple and interpretable models.	Sensitive to parameter tuning; limited scalability to very large datasets; lower performance for overlapping activity clusters.
[64]	Public (academic building, multi HVAC)	Autoencoder with TiFe attention layer	TiFe model reconstructs profiles better, more anomaly-robust than plain autoencoder; no classical recall/precision, shown qualitative improvement.	Enhanced reconstruction quality and anomaly robustness with TiFe attention; captures complex temporal–feature dependencies; suitable for unsupervised monitoring.	No quantitative accuracy metrics; high architectural complexity; interpretability limited compared to simpler models.
[55]	Residential (single-family homes, real + synthetic)	Seq2point NILM, MLP, LSTM, ILP (ILASP, ASP)	Routine deviation detection: real F1 to 0.93 (recall 0.88–0.95, specificity 0.96–0.99); synthetic F1 to 0.99–1.00.	Very high detection accuracy (F1 ≈ 0.93–1.00); robust across real and synthetic datasets; effective for identifying routine deviations.	Dependent on high-quality appliance data; limited interpretability of deep models; potential overfitting to synthetic patterns.

Table 6. *Cont.*

Title	Object Type	Algorithms	Result	Advantages	Limitations
[52]	Residential (smart homes, older adults)	SEEDS-based segmentation, segment features, hierarchic region growth clustering	NC, RMSE, MAE grow with routine randomness (synthetic data); method separates disorder levels better than baselines; no classic accuracy/F1.	Effectively distinguishes routine disorder levels; robust to varying temporal granularity; interpretable segmentation-based approach.	No standard accuracy metrics; performance declines with high routine randomness; limited validation on real-world data.
[53]	Residential (single-resident smart flats for older adults)	Chebyshev inequality-based anomaly detector	668/716 events (≈93%) normal; 421 notifications (373 positive, 48 negative); user acceptance 4/5 (Likert).	High normal event detection rate (~93%); simple and computationally efficient; good user acceptance (4/5).	Limited anomaly type differentiation; reliant on predefined thresholds; lacks detailed performance metrics (e.g., precision/recall).
[63]	Public (telecom sites: central offices, data centers, offices)	SAX transform, SAX-CART, SAX-MLP	SAX-CART: symbol classification ≈ 81%; SAX-MLP: MAPE ≈ 14.5%; combined analysis improves sensitivity/interpretability of anomaly detection.	Good classification performance (≈81%) and interpretable symbolic representation; improved anomaly sensitivity through combined SAX analysis.	Moderate regression accuracy (MAPE ≈ 14.5%); depends on discretization quality; limited adaptability to highly dynamic signals.
[1]	Residential (single homes, clusters, aggregates—review)	AR, ARIMA/SARIMA, bayesian/Gaussian, FFT, ANN, SVR, LSTM, CNN, hybrids	Most aggregate models MAPE < 5–10% (high accuracy); single homes 5–20% MAPE; trend toward DL (LSTM, CNN, hybrids).	High forecasting accuracy at aggregate level (MAPE < 10%); broad applicability across model types; deep learning shows superior trend capture.	Lower accuracy for single-home data (MAPE 5–20%); complex models require large datasets; limited interpretability of hybrid DL approaches.
[57]	Residential (Pecan Street homes)	Deep Residual Neural Network (DRNN), SVR, RF, Gradient Boosting, SARIMA, FNN, LSTM, GRU, TCN	Individual: RMSE 0.44 kWh, MAE 0.23 kWh, R^2 = 0.90; aggregate 8% error (hourly), 2% (daily); much better than benchmarks/alternatives.	Very high predictive accuracy (R^2 = 0.90; error ≤ 8%); robust across temporal scales; DRNN outperforms classical and deep baselines.	High computational and training cost; requires extensive labeled data; limited interpretability of deep residual layers.
[56]	Residential (709 households, Ireland)	ARIMA, ANN	Mean MAPE: ANN ≈ 1.8% vs. ARIMA ≈ 2.61%; sample user ARIMA MAPE ≈ 689% vs. ANN ≈ 174%; ANN recovers accuracy faster after disturbances.	Very low average MAPE (≈1.8%); ANN robust to anomalies and faster recovery after disturbances; clearly outperforms ARIMA.	Occasional extreme individual errors; requires careful neural tuning; less transparent than statistical models.
[2]	Residential (2337 apartments, China)	S2S-LSTM, DBSCAN, k-means, PCC clustering, fully-connected DNN, dropout, softmax	S2S-LSTM: RMSE ≈ 0.95, 63% users with MAPE < 10%; method 2: RMSE ≈ 0.57, 84% < 10% MAPE, >94% <20% MAPE.	High accuracy for large residential dataset (up to 84% < 10% MAPE); effective temporal modeling with S2S-LSTM; robust clustering via hybrid methods.	Computationally intensive; sensitive to hyperparameter tuning; limited interpretability of deep components.
[58]	Residential (single detached house)	LSTM (dual model: devices and total)	Device model: test loss 0.0163, MSE 0.0163 (energy) and 1.0464×10^{-6} (usage time); total model: MSE ≈ 0.0237; lower MSE (0.0168) than other DL approaches.	Low prediction error (MSE ≈ 0.016–0.024); dual LSTM captures both device-level and total consumption; outperforms alternative DL models.	Requires detailed device data; moderate model complexity; limited generalization to multi-resident environments.
[61]	Residential (single household, UCI dataset)	EMD, BI-LSTM, BI-LSTM plain, EMD-BI-GRU, EEMD-LSTM	EMD-BI-LSTM: R^2 = 0.84, RMSE 0.31, MAE 0.20, MAPE 0.28%; BI-LSTM plain: R^2 = 0.70, RMSE 0.43, MAPE 0.42%; EMD-BI-GRU: R^2 = 0.79, RMSE 0.35, MAPE 0.30%.	Excellent forecasting performance (R^2 = 0.84, MAPE ≈ 0.28%); EMD decomposition enhances temporal feature extraction; hybrid models outperform plain architectures.	Complex multi-stage processing; higher computational cost; reduced transparency of hybrid mechanisms.

3.4. Sensor Based User Activity Recognition

Modern HAR systems in smart homes have evolved from simple probabilistic models to complex deep and hybrid architectures. In the context of supporting older adults, a modified Naive Bayes classifier has proven to be more effective than DTs and SVMs in recognizing basic ADL based on events from environmental sensors [67]. However, as data complexity has increased, deep sequential networks have become predominant. In the context of the CASAS datasets, it has been demonstrated that LSTM architectures (particularly the Casc-LSTM cascade model) outperform classical HMM and CRF methods in modeling spatiotemporal dependencies [68]. Analogous results were obtained on the Orange4Home dataset, where appropriately selected RNNs and autoencoders, combined with feature selection, achieved the highest accuracy in reconstructing residents' routines [69]. A recent new development in this area is the use of the self-attention mechanism. Transformer-based models tested on the CASAS Aruba dataset are able to integrate context from multiple sensors more effectively than previous CNN- or LSTM-based solutions [70].

Alongside the struggle for accuracy, the XAI of systems has become a key challenge. As part of LSTM activity recognition, an XAI framework based on LIME, Anchors, and SHAP methods was developed, with SHAP providing explanations that were most understandable to experts and preferred by users [71]. The alternative DeXAR approach treats sensor data as semantic images analyzed by CNN and introduces a new prototype method that generates explanations more consistent with common sense than popular saliency mapping techniques [72].

A significant implementation problem is the adaptation of systems to new environments in the absence of labels. Bootstrap methods "from scratch" use unsupervised learning of "action units" and active labeling, which allows useful accuracy to be achieved in a new home after just a few weeks [6]. The concept of the HAR system "life cycle" extends this approach with continual learning and long-term analysis of behavioral changes, which is crucial for long-term operation [7]. In extreme cases, fully unsupervised methods based on coupled HMM models can recognize more than a dozen types of activities with high accuracy (87–92%) without any manual supervision, using only non-invasive sensors and Bluetooth beacons [73].

Innovative architectures are utilized in complex multi-user scenarios. The LAHAR system employs an LLM to infer relationships between sensors and multiple users, effectively separating intertwined activity timelines [74]. In turn, the integration of activity recognition with behavioral analytics allows not only to classify current activities, but also to predict the user's future steps. A unified framework based on kernel fusion and frequent sequence rules improves the prediction accuracy by several percentage points [5]. Activity information can also control other systems, such as HEMS, where the HAR module provides context for the Deep Reinforcement Learning (Soft Actor-Critic) agent, leading to real savings [3].

In the field of wearables and IoT, the balance between accuracy and energy consumption is crucial. Hybrid CNN + RNN models (variants with LSTM, GRU, BiLSTM) achieve results described as "outstanding" on inertial datasets (e.g., PAMAP2), outperforming classic ML algorithms [75]. From a system perspective, the AE-HAR model introduces a dynamic division of computations between the device and the cloud, reducing energy consumption by over 90% without compromising recognition quality [76]. The visual classification of algorithms (types and methods) for human activity recognition is presented in Figure 4. As shown in Table 7, a comprehensive collection of the most relevant publications on sensor-based activity recognition has been compiled. This compendium includes key parameters, methodologies, object types, and performance indicators.

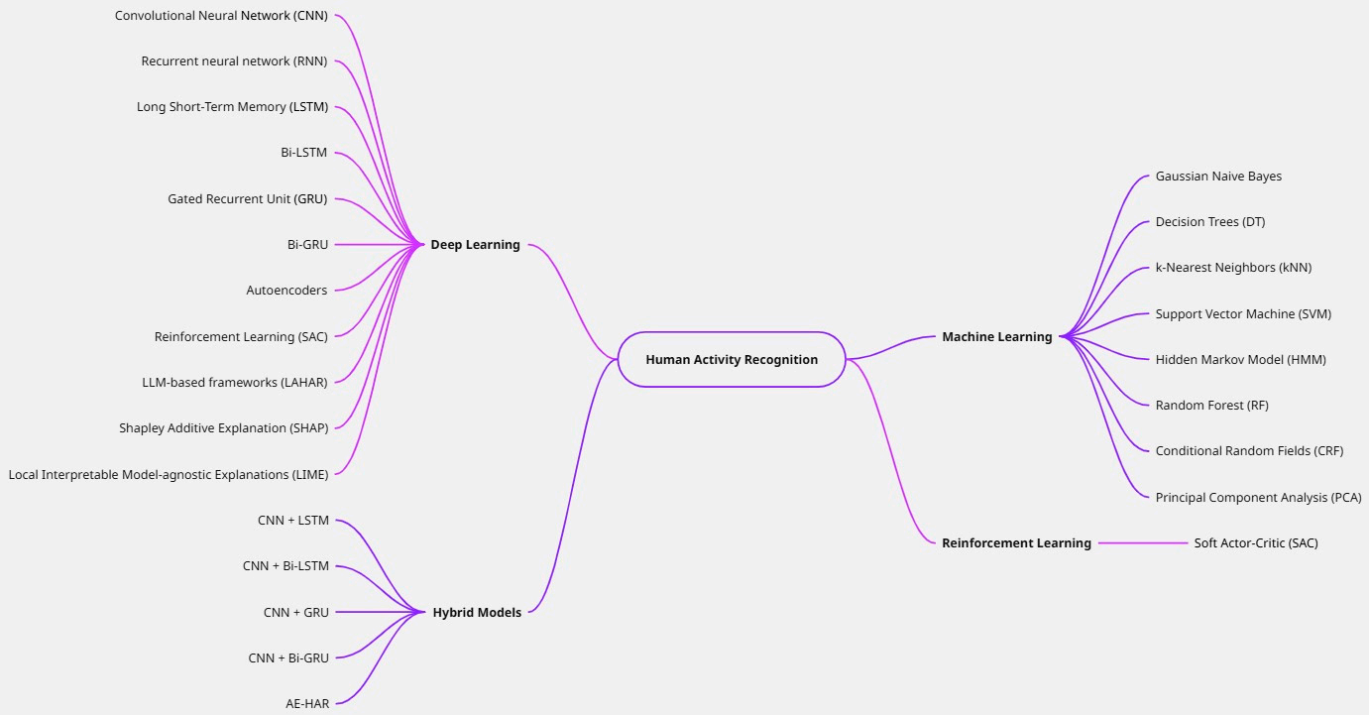


Figure 4. Algorithm classification for human activity recognition.

Table 7. The list of selected publications: Sensor-Based User Activity Recognition.

Title	Object Type	Algorithms	Result	Advantages	Limitations
[67]	Residential (smart home for the elderly)	Modified Naive Bayes, comparisons with DTs, SVM, HMM	The modified Naive Bayes achieved the highest accuracy and better quality measures than the compared classical ML algorithms.	High accuracy; simple, fast, and effective for sensor-based activity data; low computational cost.	Limited handling of complex temporal dependencies; assumes conditional independence; dependent on sensor quality and coverage.
[68]	Residential (CASAS flats with multiple tenants)	Uni-LSTM, Bi-LSTM, Cascaded LSTM (Casc-LSTM), comparisons with HMM, Naive Bayes, CRF and other ML	All LSTMs outperformed classical methods, and Casc-LSTM achieved the highest accuracy on the CASAS datasets.	Superior accuracy across CASAS datasets; Casc-LSTM effectively models long-term and bidirectional dependencies; outperforms classical ML baselines.	High computational and training cost; requires large labeled datasets; reduced interpretability compared to simpler models.
[71]	Residential (smart home with environmental sensors)	LSTM for HAR + methods XAI: LIME, Anchors, SHAP, LIME+	SHAP achieved approximately 92% "sensibility" of explanations in the experts' assessment, the explanations were preferred by ~83% of users and increased confidence in the model.	High expert-rated interpretability (~92% sensibility); SHAP improves user trust (~83% preference); combines accurate HAR with explainable insights.	Added computational overhead from XAI methods; explanation quality varies by context; requires expert evaluation for validation.
[72]	Residential (smart home with environmental and wearable sensors)	CNN on semantic sensor images + XAI: Grad-CAM, LIME, prototype method	CNN achieved high ADL recognition accuracy; the prototype method provided the best, most understandable explanations, surpassing Grad-CAM and LIME.	High ADL recognition accuracy; prototype method offers clear, human-interpretable explanations; effective integration of wearable and environmental data.	Requires image transformation of sensor data; computationally demanding; explanation methods sensitive to model complexity.

Table 7. *Cont.*

Title	Object Type	Algorithms	Result	Advantages	Limitations
[6]	Residential (CASAS smart homes with multi-week logs)	Unsupervised representation learning and clustering of action units + supervised classical models with active learning	After ~4 weeks of bootstrapping, the system achieves useful accuracy in recognizing the most common activities, clearly surpassing simple baselines; accuracy increases with observation time and number of labels.	Achieves competitive accuracy with minimal labeling; performance improves over time through active learning; suitable for long-term deployment.	Requires extended bootstrapping (~4 weeks); limited accuracy for rare activities; dependent on consistent sensor activity patterns.
[74]	Residential (ARAS multi-person flat)	LAHAR framework based on a LLM for assigning events to individuals and inferring activities	LAHAR achieves state-of-the-art accuracy with high temporal resolution and maintains high performance in multi-person scenarios.	State-of-the-art accuracy with fine temporal resolution; effective in multi-resident settings; leverages LLM contextual reasoning for activity inference.	High computational and data requirements; potential privacy concerns; limited transparency of LLM decision processes.
[5]	Residential (two real smart homes)	"Kernel fusion" in the classifier for HAR + frequent sequence rules for predicting subsequent actions	Kernel fusion improves the accuracy of activity recognition by an average of 13.82% and the F measure of predicting future actions by 6.76%.	Significant accuracy gain with kernel fusion (+13.8%); improved future action prediction (+6.8% F-measure); integrates recognition and prediction effectively.	Increased computational cost; requires careful kernel selection; performance may degrade with noisy or sparse sensor data.
[69]	Residential (smart home Orange4Home)	Autoencoders, RNNs and their variants; comparisons with RF and PCA	The best RNN (all-features) and autoencoder (feature reduction) significantly outperform the baselines, achieving the highest ADL accuracy on Or-ange4Home.	Highest ADL recognition accuracy; RNNs effectively capture temporal dynamics; autoencoders enable efficient feature reduction.	High training complexity; requires large labeled datasets; limited interpretability of deep latent features.
[73]	Residential (accommodation for two persons—SDHAR-HOME)	Two hidden Markov models (HMM): resident location + activity	Accuracy ranging from 86.78% to 91.68% was achieved for 16 activities, outperforming previous unsupervised approaches without the need for manual labeling.	High recognition accuracy (≈87–92%); effective joint modeling of location and activity; eliminates need for manual labeling.	Assumes clear sensor-activity mapping; limited scalability to large households; sensitive to sensor noise and missing data.
[3]	Residential (smart home with DER and ILM)	HAR module based on ILM + deep reinforcement learning Soft Actor-Critic (SAC) for energy control	SAC with activity information provides clear energy cost savings and additional revenue from network services while maintaining high user satisfaction.	Integrates HAR with energy control for cost savings and added revenue; maintains high user comfort; adaptive through deep reinforcement learning.	High model complexity; requires continuous activity recognition input; potential stability issues in long-term learning.
[76]	Public/personal IoT (users with mobile phones/ smartwatches in everyday environments)	AE HAR hybrid model: lightweight shallow on-device classifier + more accurate deep model in the cloud controlled by heuristics	HAR accuracy > 93% with >94% reduction in device energy consumption compared to continuous use of deep models in the cloud.	High HAR accuracy (>93%) with major energy savings (>94%); efficient hybrid edge–cloud design; adaptable to mobile IoT contexts.	Requires stable connectivity for cloud offloading; heuristic control may limit generalization; potential privacy concerns for transmitted data.
[70]	Residential (smart home CASAS Aruba)	Transformer with self-attention mechanism, compared with kNN, SVM, LogitBoost, CNN, LSTM	Transformer achieves higher HAR accuracy than classical ML and other deep models on CASAS Aruba, providing a noticeable improvement in accuracy.	Self-attention captures long-range temporal dependencies; effective on complex CASAS data.	Computationally demanding; requires large training datasets; limited interpretability of attention weights.
[75]	Public/personal (users with wearable sensors—PAMAP2)	Hybrid models CNN + LSTM, CNN + BiLSTM, CNN + GRU, CNN + BiGRU	All CNN + RNN hybrids achieve very high accuracy, F-score, sensitivity and specificity; the best configurations outperform classic ML and CNN alone (outstanding effectiveness).	Very high accuracy and robustness across metrics; hybrids effectively combine spatial and temporal features; superior to standalone CNN and ML models.	High computational and memory cost; sensitive to hyperparameter tuning; reduced interpretability compared to simpler architectures.

Table 7. *Cont.*

Title	Object Type	Algorithms	Result	Advantages	Limitations
[7]	Residential (general class of smart homes)	Framework model: bootstrapping with unsupervised/active HAR, continual learning, sequence analysis; no single specific algorithm	Conceptual work—no specific accuracy values; shows that a well-designed HAR life cycle allows for satisfactory effectiveness to be maintained with minimal supervision over the long term.	Enables long-term adaptability with minimal supervision; integrates unsupervised and continual learning; supports scalable HAR deployment.	Conceptual validation only (no accuracy metrics); lacks quantitative benchmarking; implementation complexity across system stages.

3.5. Energy Disaggregation (Home Appliances Activity Recognition)

Contemporary research in the domain of energy-consumption monitoring in residential buildings focuses on two main areas: NILM and ADL. In the domain of classifying current activities and identifying devices, both classic ML algorithms and advanced deep networks play an important role. Research conducted on the UK-DALE dataset has demonstrated that in tasks involving the prediction of the type of device in operation and the associated activity of household members, models based on neural networks, such as FFNN and LSTM, achieve higher accuracy (above 0.9) compared to SVM, although their ability to be generalized to new buildings remains a challenge requiring local retraining [8]. A comparative analysis of classical methods reveals analogous conclusions. With regard to the identification of the operating status (ON/OFF), DT and kNN algorithms achieved an F1-Score above 95% based on the RAE and REDD datasets. This confirms the usefulness of simpler models in specific applications of HEMS [10].

In the context of DL, particular attention is paid to sequential architectures. The employment of GRU and TCN models within the WaveNILM architecture has facilitated high precision in device identification and the estimation of their energy flexibility based on the AMPDs2 dataset, with WaveNILM demonstrating an advantage over GRU in terms of training speed and fitting accuracy [9]. Research on the S2P architecture represents a development of this trend. An innovative approach to S2Pe prediction, tested on the ENERTALK and UK-DALE datasets, demonstrated higher effectiveness and better generalization than standard S2P and S2S methods, especially when using GRU units in a multi-task model [77]. In turn, the employment of the WaveNet architecture, based on dilated convolutions and residual blocks, allowed for a significant reduction in MAE errors compared to classic CNN and LSTM networks. Moreover, it was demonstrated that direct binary classification is to be preferred to regression thresholding in ON/OFF-state-detection tasks [78].

Furthermore, recent advancements indicate a paradigm shift towards Transformer-based and GNN architectures. These architectures effectively address the limitations of RNNs in capturing long-range temporal dependencies and inter-appliance correlations. Studies using self-attention mechanisms with temporal pooling have demonstrated superior feature-extraction capabilities. These studies have achieved an F1-Score of 92.96% on the UK-DALE dataset, even in noisy environments with overlapping loads [79]. Concurrently, graph-based approaches such as GraphNILM allow for modeling complex interactions between devices, significantly reducing memory consumption while maintaining a disaggregation accuracy suitable for edge deployment [80].

Alongside the development of the model architecture, privacy preservation has emerged as a critical requirement for real-world deployment. To mitigate the risks of centralized data collection, FL frameworks have been introduced. Novel approaches, such as FUML, effectively address the challenge of data heterogeneity across different households. FUML employs a three-stage learning process and has been shown to outperform existing federated methods. It achieves performance levels comparable to centralized training while ensuring that sensitive raw data remains local [81].

Finally, enhanced granularity in detecting usage patterns facilitates the transition from passive monitoring to active, adaptive building control. Recent research highlights the potential of NIOM to drive demand-response strategies without the need for intrusive sensors. Energy-state-analysis algorithms can infer occupancy patterns directly from smart meter data and distinguish between active usage and baseloads with high reliability. This occupancy-driven approach enables scalable load-management systems that adapt to household-specific behaviors, effectively bridging the gap between non-intrusive monitoring and automated building control [82]. Furthermore, NILM-based solutions that automate energy services can predict heat pump consumption with an average daily error of less than 3.5%, offering a reliable foundation for automated demand-side management strategies [83].

At the same time, generative methods and systems dedicated to environments with renewable energy source integration are being developed. The employment of VAE within the VAE–NILM architecture, enriched with IBN-Net blocks and skip connections, facilitated the effective reconstruction of power-consumption profiles and state detection, reducing the MAE error by 18–33% compared to reference methods [84]. In more complex scenarios involving distributed energy resources such as PV or EV, the utilization of the ResNet-seq2seq architecture with a post-processing mechanism effectively overcomes the problem of model degradation in deep networks, achieving a significant reduction in errors for devices with highly variable characteristics [14]. These predictive methods are complemented by unsupervised approaches, such as a hierarchical decision system integrating fuzzy set theory and association analysis, which enables the automatic labeling of devices without the need for prior training, although its effectiveness may depend on the availability of detailed electrical parameters and the regional specificity of the equipment [85]. The visual classification of algorithms (types and methods) for energy disaggregation is presented on Figure 5. As summarized in Table 8, a comprehensive collection of the most relevant publications on energy disaggregation-based activity recognition has been compiled. This compendium includes key parameters, methodologies, object types, and performance indicators.

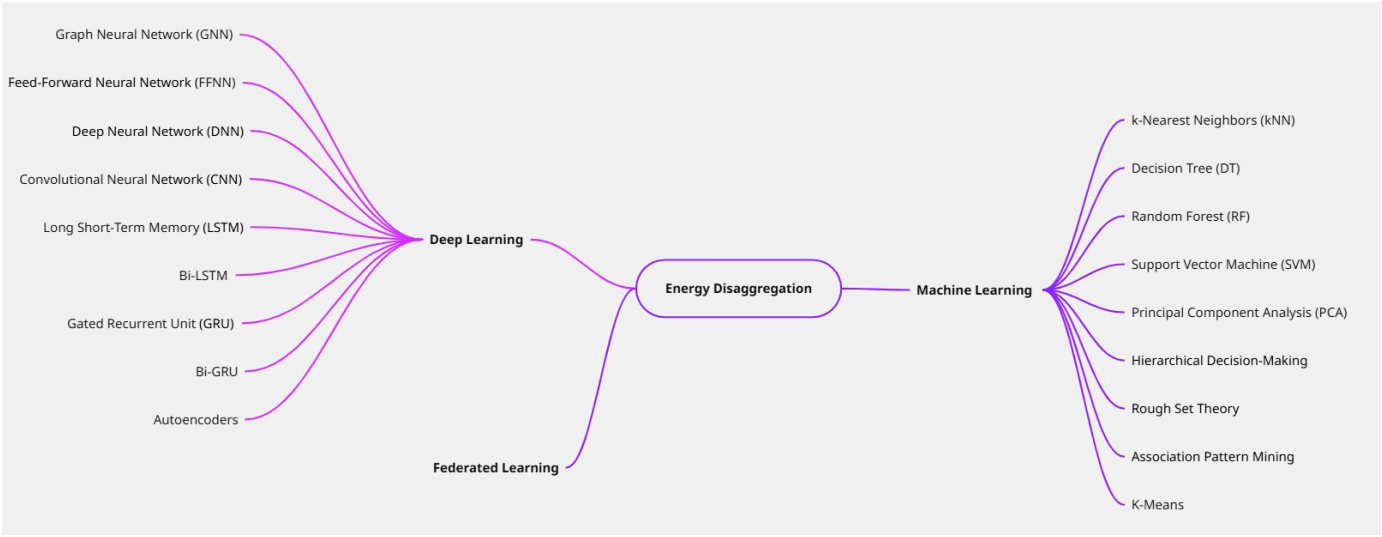


Figure 5. Algorithm classification for energy disaggregation.

Table 8. The list of selected publications: Energy Disaggregation.

Title	Object Type	Algorithms	Result	Advantages	Limitations
[8]	Residential (Smart Home; UK-DALE collection, houses no. 1 and 5)	FFNN, LSTM, SVM	The accuracy of neural network models (FFNN, LSTM) is above 0.9, and for SVM it is around 0.8. Tests on a new building showed a decrease in accuracy to 0.61.	High disaggregation accuracy ($\geq$0.9) for trained homes; neural models outperform SVM; effective for detailed appliance-level monitoring.	Significant accuracy drop on unseen buildings ($\approx$0.61); limited generalization; requires extensive labeled training data.
[9]	Residential (AMPDs2 collection)	WindowGRU (Gated Recurrent Unit), WaveNILM (Temporal Convolutional Network—TCN)	WaveNILM achieved an estimated accuracy of 91.19% (whole house), outperforming the GRU model (approx. 69.9%). Accuracy for devices ranged from 46% to 97%.	High overall disaggregation accuracy (~91%); WaveNILM effectively captures temporal dependencies; superior to GRU baseline.	Variable device-level performance (46–97%); high computational cost; requires precise temporal labeling.
[85]	Residential (6 private houses in China, REDD collection)	Hierarchical Decision-Making, Rough Set Theory, Association Pattern Mining	In the private collection, seven out of nine devices were correctly labeled. In the REDD collection, accuracy was lower due to a lack of data on the power factor and regional differences.	Effective device identification in detailed datasets (7/9 correctly labeled); interpretable hierarchical decision process; combines multiple knowledge-based methods.	Lower accuracy based on sparse or regional datasets; depends on power factor data; limited scalability to diverse appliance sets.
[10]	Residential (RAE and REDD collections)	PCA, kNN, DT, RF	F1-Score above 95% (DT on RAE: 95.21%, kNN on REDD: 95.70%).	Very high classification performance (F1 > 95%); simple, interpretable ML models; effective across multiple datasets (RAE, REDD).	Limited generalization to unseen appliances; feature selection critical; reduced robustness to noisy signals.
[77]	Residential (ENERTALK and UK-DALE collections)	CNN, GRU (architecture S2Pe)	SAE (aggregation error) often below 0.1; F1-Score above 93% for selected devices in the GRU model. Better results than standard S2P/S2S.	Very low aggregation error (SAE < 0.1); high device-level accuracy (F1 > 93%); GRU architecture outperforms standard S2P/S2S approaches.	Model performance varies by appliance type; computationally demanding training; requires synchronized high-frequency data.
[84]	Residential (UK-DALE collection—3 houses, REFIT collection)	VAE, VAE-NILM (IBN-Net and skip connections)	MAE error reduction by 18% (UK-DALE) and 33% (REFIT); F1-Score increase by over 11% compared to reference methods (CNN, S2P).	Significant error reduction (MAE 18–33%); improved F1 by >11% over CNN/S2P; VAE-NILM effectively captures complex appliance patterns.	High training complexity; requires extensive data preprocessing; limited interpretability of latent features.
[78]	Residential (REFIT collection—20 houses)	CNN, BLSTM (Bidirectional LSTM), BGRU, WaveNet	WaveNet achieved the lowest MAE errors (e.g., approx. 5.4 W). The direct binary classifier achieved a higher F1 score than regression result thresholding.	Very low MAE (~5.4 W) with WaveNet; direct binary classification improves F1 performance; effective modeling of temporal dependencies.	Computationally intensive; large data requirements; threshold-based regression less reliable for rare appliance states.
[79]	Residential (UK-DALE, REDD)	Transformer-NILM, Deep Neural Networks (Feature Extraction)	F1-Score of 92.96% and SAE of -0.036; superior feature extraction in noise.	High accuracy (F1 $\approx$ 93%); robust to noise with effective feature extraction; outperforms conventional deep NILM models.	High computational and memory demands; requires large training datasets; limited interpretability of attention layers.
[80]	Residential (UK-DALE, REDD)	GraphNILM, Graph Neural Network (Edge NILM deployment)	Comparable accuracy to SOTA (e.g., 86% for AC) with significantly reduced memory.	Competitive accuracy (~86%) with much lower memory usage; efficient for edge NILM deployment; scalable to resource-constrained devices.	Requires graph construction and tuning; potential performance drop for complex loads; limited interpretability of graph embeddings.
[81]	Residential (Heterogeneous Data)	FUML, Federated Learning, Privacy-Preserving NILM	MAE improvement up to 67.7% for washers; comparable to centralized training.	Large MAE improvement (up to 67.7%); maintains accuracy comparable to centralized models; ensures data privacy across households.	Communication overhead in federated setup; sensitive to data heterogeneity; increased training complexity.

Table 8. *Cont.*

Title	Object Type	Algorithms	Result	Advantages	Limitations
[82]	Residential (Smart Meter Data)	Occupancy-Driven DR, Energy State Analysis, K-Means clustering	Precise interfered occupancy patterns from smart meter data and scalable DR	Accurately infers occupancy from energy data; enhances demand response scalability; no need for additional sensors.	Dependent on high-resolution meter data; privacy concerns from inferred behaviors; limited activity granularity.
[83]	Residential (Heat Pump Data)	ML-based NILM, Decision Tree	Heat pump consumption prediction error < 3.5% (daily average).	Very low prediction error (<3.5%); high reliability for daily heat pump monitoring; suitable for real-world energy management.	Device-specific calibration required; limited transferability to other appliances; depends on consistent operational cycles.
[14]	Residential (REDD and Pecan Street collections—PV/EV)	ResNet-seq2seq (Residual Networks + Sequence-to-Sequence)	MAE reduction compared to CNN-seq2seq: 55% (PV), 69% (EV), 57% (fridge), 85% (microwave).	Substantial MAE reduction (55–85%) across devices; ResNet-seq2seq enhances temporal feature extraction; strong generalization to diverse loads.	High architectural complexity; long training times; requires large, device-specific datasets.

4. Gaps and Future Works

While the data-driven paradigm has fundamentally transformed building engineering and AI and ML methods have demonstrated significant potential in optimizing building-management systems, existing solutions continue to encounter inherent limitations that impede their scalable implementation in real-world scenarios. A critical challenge remains the lack of cross-domain robustness, resulting in limited generalizability across different facilities. The prevailing body of research validates algorithms on homogeneous datasets or within single buildings, leading to significant performance degradation when deployed in new operational environments. The "cold start" problem, along with the necessity of acquiring historical training data for every new installation, constitutes both an economic and technical barrier that even modern transfer learning techniques have not fully resolved [8,28,32,44]. Furthermore, the deficit of high-quality ground truth data remains a bottleneck for supervised learning methods. In uncontrolled residential environments, acquiring precise activity labels from occupants is an invasive and error-prone process. This necessitates the exploration of more autonomous semi-supervised methodologies [6,31].

A significant technological obstacle is the misalignment of spatiotemporal resolution between model requirements and actual metering infrastructure. In the domain of NILM, high predictive accuracy is often correlated with high-frequency sampling (e.g., 1 Hz), which contradicts the standard of AMI that typically provides data at 15- or 60-min intervals. This aggregation results in the loss of short-duration event signatures, thereby limiting the effectiveness of energy disaggregation [10,14,84]. Analogously, in the field of thermal-comfort modeling, conventional methodologies frequently neglect the spatial heterogeneity of a zone, treating a room as a homogeneous thermodynamic point and ignoring local microclimates determined by proximity to transparent surfaces or diffusers [40,41]. Moreover, the stochastic nature of human behavior often exceeds the capabilities of deterministic models, which frequently treat the household as a monolithic entity. The neglect of multi-resident scenarios results in input data ambiguity, where overlapping activities are difficult for conventional neural networks to separate [54,73,74]. Furthermore, static models rarely account for environmental non-stationarity, including concept drift resulting from seasonality or the evolution of occupant habits, which leads to the gradual degradation of prediction accuracy over time [7,51]. Finally, a disconnect is observable between statistical ML metrics and their operational utility within control loops. High-accuracy indicators (e.g., Accuracy) do not guarantee the optimization of objective functions such as energy

efficiency [29], and the lack of interpretability of complex "black box" models diminishes end-user trust and hinders the verification of control decisions [35,71].

Alongside these data and modeling challenges, there are distinct gaps in algorithmic robustness and market mechanism design. Advanced methodologies, such as HRL and HPRL, have been proposed to manage large state spaces and improve sample efficiency. However, traditional algorithms struggle to process hybrid discrete–continuous action spaces without simplifying the environment model. This compromises optimization accuracy in complex scenarios, such as island groups. In the context of market-based energy, management frameworks often fail to account for user defaults or allow for flexibility in bid adjustments within trading cycles, which limits their practical applicability. Additionally, integrating multi-energy carriers requires complex joint trading mechanisms that account for multi-time scale characteristics. Neglecting the differences in transmission speed among electricity, gas, and heat often results in trading failures [86,87].

Overcoming the identified limitations requires a fundamental paradigm shift in the design of HEMS—a transformation from reactive systems toward proactive, cognitive, and transparent solutions. Future research directions should focus on the implementation of dynamic home-arrival-time prediction models that, by integrating historical behavioral patterns with stochastic variables, enable the predictive conditioning of building zones, thereby eliminating energy losses associated with standby operation [88,89]. Simultaneously, the evolution of HBI interfaces is trending toward the elimination of burdensome explicit feedback in favor of implicit feedback inference. State-of-the-art algorithms should analyze natural user interactions with building actuators (e.g., windows, thermostats) as surrogates for comfort assessments, which constitutes a logical progression of research on adaptive behavior detection [90,91]. The availability of such granular behavioral data justifies the broader implementation of Deep Reinforcement Learning, allowing agents to autonomously learn control strategies via trial-and-error without the need for defining rigid expert rules [92–94]. A prerequisite for the long-term stability of such systems is the development of lifelong learning mechanisms capable of detecting and adapting to concept drift. These algorithms must precisely discriminate between one-time anomalies [95] and permanent shifts in usage profiles [96], ensuring continuous operation without maintenance intervention [7]. As system autonomy increases, transparency becomes a critical requirement, necessitating the advancement of XAI methods dedicated to building energy applications. Generating human-understandable justifications for control decisions is essential for increasing the acceptability of automation and building trust in the human–machine relationship [97,98]. Finally, in light of challenges related to data sovereignty and transmission latency, edge computing utilizing FL is becoming a technological imperative. This architecture enables the training of personalized models locally on edge devices, reconciling high predictive efficacy with rigorous privacy protection for residents [99,100], thereby addressing the growing security requirements in IoT ecosystems [76].

Future research should focus on several key directions. First, challenges related to real-world deployment, including hardware variability and dynamic operating conditions, should be addressed through the development of lightweight Gaussian Process approximations and integration with edge computing, as well as adaptive model-update mechanisms to preserve accuracy under evolving building dynamics [75]. In parallel, the further refinement of the dual LSTM architecture is needed by means of systematic hyperparameter optimization and architectural exploration, together with the incorporation of regularization techniques and external variables such as weather conditions and occupancy patterns to enhance robustness and predictive performance [47].

Additionally, future studies should incorporate the perspectives of healthcare staff to better distinguish seasonal sleep variations from health-related changes and extend the

analysis to multi-resident smart homes as well as diverse geographical regions to improve generalizability [40]. Predictive performance in scenarios involving thermal discomfort should also be improved through the application of advanced control strategies, including reinforcement learning and model predictive control, while extending the proposed methodologies to a broader range of measurement challenges beyond thermal comfort alone [34]. Finally, further work should investigate the integration of blockchain technologies with IoT systems for the secure handling of occupancy data, advance sensor fusion and optimal sensor placement strategies, and explore occupant localization together with activity forecasting to enhance both energy efficiency and privacy in smart building environments [23].

While the authors acknowledge the critical risks related to occupant privacy, unauthorized monitoring, and misuse of behavioral data, a detailed discussion of security and privacy-preserving mechanisms is beyond the scope of this review.

5. Conclusions

This review provides a comprehensive synthesis of recent advances in using AI and ML to detect user activities and identify energy habits within HEMS. A review of the extant literature reveals a rapidly growing volume of research focused on leveraging data-driven approaches that enable proactive energy control, occupant-cantered system adaptation, and demand-side flexibility. Utilizing a taxonomy based on detected phenomena (occupancy, thermal comfort, seasonal patterns, activity recognition, and energy disaggregation) and modeling architectures, this study illustrates how AI and ML techniques are transforming HEMS from passive monitoring to intelligent, predictive decision-support systems capable of learning and adapting to evolving user behaviors.

Key findings indicate that DL models, particularly LSTM, GRU, and hybrid CNN-LSTM architectures, consistently achieve the highest accuracy in activity detection and load forecasting, especially when applied to time-series energy consumption data enriched with contextual information, such as occupancy states, weather conditions, and thermal parameters. Sensor-based activity-recognition systems demonstrate occupancy detection accuracies ranging from 95 to 99%, while energy disaggregation—NILM methods successfully identify individual appliance operations without requiring intrusive monitoring. Hybrid systems that combine neural network predictions with rule-based logic, physics-informed models, or optimization algorithms demonstrate enhanced robustness and interpretability in real-world deployments. Furthermore, it has been demonstrated that models which have been trained on a range of building types are capable of exhibiting sufficient generalization capabilities to facilitate the deployment of behavioral pattern-detection methodologies in both residential and commercial contexts.

Nevertheless, a number of critical challenges limit the widespread deployment of these technologies. Most models remain trained on localized, single-building datasets, which severely restricts cross-building transferability and scalability to diverse household compositions and geographic contexts. The interpretability and explainability of DL predictions remain largely unaddressed, limiting user trust and regulatory compliance. Furthermore, integration with existing building automation systems and IoT platforms remains inconsistent, and few studies comprehensively address data privacy, security, and federated learning approaches necessary for deployment at neighborhood or district scales. The absence of standardized datasets, benchmarking protocols, and performance metrics impedes rigorous comparative assessment and reproducibility of findings across the field.

It is recommended that future research should prioritize developing lightweight, computationally efficient, and explainable models suitable for real-time deployment on edge devices with limited resources. Promising research directions include advancing federated

learning frameworks for privacy-preserving distributed training across multiple households, integrating activity- and habit-detection models with digital twin platforms and automated HVAC-lighting control systems, and incorporating uncertainty quantification to enhance robustness under varying conditions. Furthermore, research should extend beyond energy consumption optimization to include broader occupant-cantered metrics encompassing thermal comfort, indoor air quality, and behavioral adaptability to support holistic residential building operation.

The integration of detected activities and habits with DR programs, time-of-use tariff optimization, and renewable energy self-consumption management represents a significant opportunity for reducing residential sector energy consumption and facilitating grid-level demand-side flexibility. By combining real-time activity recognition with predictive analytics and optimization, AI- and ML-based HEMS can not only reduce operational costs and peak demand but also support the integration of distributed renewable generation and energy storage systems. Multimodal sensor fusion, performance metrics, open datasets, and industry–academia collaboration will be essential for building trustworthy, transparent, and scalable activity detection and habit identification systems that align with the transition toward low-carbon, resilient, and occupant-cantered smart buildings. The convergence of these technologies directly supports broader sustainability objectives and positions intelligent, data-driven energy management as a cornerstone of future smart building ecosystems.

Author Contributions: Conceptualization, J.G.; methodology, J.G.; validation, J.G., A.O. and L.W.; formal analysis, J.G. and L.W.; investigation, D.L. and J.G.; resources, D.L.; data curation, D.L. and J.G.; writing—original draft preparation, J.G., D.L., F.D. and A.O.; writing—review and editing, J.G., F.D., D.L. and A.O.; supervision, J.G. and L.W.; project administration, J.G.; funding acquisition, F.D., D.L. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering of the AGH University of Krakow as part of a research subsidy for young scientists (Dean's grants) in 2025.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: During the preparation of this manuscript, the authors used assistive tools the ChatGPT 5 (version 5.2) as well as the SCOPUS-AI for purpose of synthesizing literature. In addition, the Open Writefull tool (version 2025.50.0) was used to verify the grammatical and stylistic correctness of the text. The authors have reviewed and edited the output and takes full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

AC	Air Conditioning
ADL	Activities of Daily Living
AE-HAR	Accuracy and Energy Aware-HAR
AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
AO-ANN (BP)	Adam Optimizer–Artificial Neural Network–Back-Propagation Network
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
AUC	Area Under the Receiver Operating Characteristic (ROC) Curve

BBQ	Bayesian Binning into Quantiles
Bi-GRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long Short-Term Memory
BIM	Building Information Model
BLE	Bluetooth Low Energy
BO	Bayesian Optimization
Build2Vec	Build2Vec (Building to Vector)
CART	Classification And Regression Trees
CDLSTM	Convolutional Deep Long Short-Term Memory
CLO	Clothing Insulation
CNN	Convolutional Neural Network
CO2	Carbon Dioxygen Concentration
CRF	Conditional Random Fields
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DeXAR	Deep Explainable Sensor-Based Activity Recognition
DL	Deep Learning
DMD	Dynamic Mode Decomposition
DNN	Deep Neural Network
DR	Demand Response
DRNN	Deep Residual Neural Network
DSM	Demand Side Management
DT	Decision Tree
ECE	Expected Calibration Error
EMD	Empirical Mode Decomposition
ExKMC	Explainable K-Means Clustering
EV	Electric Vehicle
EUI	Energy Use Intensity
FFNN	Feed-Forward Neural Network
FL	Federated Learning
FUML	Federated Unilateral-Mutual Learning
GBDT	Gradient Boosting Decision Tree
GCN-LSTM	Graph Convolutional Network Long Short-Term Memory
GNN	Graph Neural Network
Gpcalib	Gaussian Process Calibration
Grad-CAM	Gradient-weighted Class Activation Mapping
GRU	Gated Recurrent Unit
HAR	Human Activity Recognition
HBI	Human-Building Interaction
HEMS	Home Energy Management System
HMM	Hidden Markov Model
HPRL	Hybrid Policy-Based Reinforcement Learning
HRL	Hierarchical Reinforcement Learning
HVAC	Heating, Ventilation, and Air Conditioning
IoT	Internet of Things
ILP	Inductive Logic Programming
K-means	K-means Clustering
kNN	k-Nearest Neighbors
LAHAR	Locality-Aware Human Activity Recognition
LDA	Linear Discriminant Analysis
LIME	Local Interpretable Model-agnostic Explanations
LLM	Large Language Model
LR	Linear Regression/Logistic Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error

MAPE	Mean Absolute Percentage Error
M-ARM	Multidimensional Association Rule Mining
MCC	Matthews Correlation Coefficient
MET	Metabolic Rate
ML	Machine Learning
MLP	Multilayer Perceptron
MPC	Model Predictive Control
MSE	Mean Square Error
MTL	Multi-Task Learning
NB	Naïve Bayes
NCA	Neighborhood Components Analysis
NILM	Non-Intrusive Load Monitoring
NIOM	Non-Intrusive Occupancy Monitoring
NN	Neural Network
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PIR	Passive Infrared
PMV	Predicted Mean Vote
PV	Photovoltaic
RBF	Radial Basis Function
RF	Random Forest
RFECV	Recursive Feature Elimination with Cross-Validation
RH	Relative Humidity
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
S2Pe	Sequence-to-Point End
S2S	Sequence-to-Sequence
S2S-LSTM	Sequence to Sequence Long Short-Term Memory
SAE	Signal Aggregate Error
SARMIA	Sesonal Auto Regressive Integrated Moving Average
SARSA	State-Action-Reward-State-Action
SAX	Symbolic Aggregate approXimation
SEEDS	Smart Energy Efficiency Data Set
SET	Standard Effective Temperature
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
STLF	Short-Term Forecasting
SVM	Support Vector Machine
SVR	Support Vector Regression
TAV	Termal Acceptability Vote
TCN	Temporal Convolutional Network
TCV	Thermal Comfort Vote
TiFe	Time Features
TL	Transfer Learning
TPV	Thermal Preference Vote
TSV	Thermal Sensation Vote
UWB	Ultra-Wide Band
VAE	Variational Autoencoder
XAI	Explainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting
XGB-RFECV	eXtreme Gradient Boosting Recursive Feature Elimination with Cross-Validation

References

1. Rodrigues, F.; Cardeira, C.; Calado, J.M.F.; Melicio, R. Short-Term Load Forecasting of Electricity Demand for the Residential Sector Based on Modelling Techniques: A Systematic Review. *Energies* **2023**, *16*, 4098. [CrossRef]
2. Fan, L.; Li, J.; Zhang, X.P. Load prediction methods using machine learning for home energy management systems based on human behavior patterns recognition. *CSEE J. Power Energy Syst.* **2020**, *6*, 563–571. [CrossRef]
3. Wu, Z.; Chen, X.; Lin, Y.; Wen, J.; Chen, Y. A smart home energy management system based on human activity recognition and deep reinforcement learning. *Energy Build.* **2024**, *325*, 114951. [CrossRef]
4. Ma, C.; Vu, H.Q.; Wang, J.; Trieu, V.H.; Li, G. Understanding Residents' Behavior for Smart City Management by Sequential and Periodic Pattern Mining. *IEEE Trans. Comput. Soc. Syst.* **2024**, *11*, 1260–1276. [CrossRef]
5. Fatima, I.; Fahim, M.; Lee, Y.K.; Lee, S. A Unified Framework for Activity Recognition-Based Behavior Analysis and Action Prediction in Smart Homes. *Sensors* **2013**, *13*, 2682–2699. [CrossRef] [PubMed]
6. Hiremath, S.K.; Nishimura, Y.; Chernova, S.; Plötz, T. Bootstrapping Human Activity Recognition Systems for Smart Homes from Scratch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 119. [CrossRef]
7. Crippa, P.; Biagetti, G.; Hiremath, S.K.; Plötz, T. The Lifespan of Human Activity Recognition Systems for Smart Homes. *Sensors* **2023**, *23*, 7729. [CrossRef]
8. Franco, P.; Martinez, J.M.; Kim, Y.C.; Ahmed, M.A. IoT Based Approach for Load Monitoring and Activity Recognition in Smart Homes. *IEEE Access* **2021**, *9*, 45325–45339. [CrossRef]
9. Mahmud, A.A. Artificial Intelligence (AI)-based identification of appliances in households through NILM. In Proceedings of the 2022 IEEE 4th Global Power, Energy and Communication Conference, GPECOM 2022, Cappadocia, Turkey, 14–17 June 2022; pp. 414–426. [CrossRef]
10. Lemes, D.A.M.; Cabral, T.W.; Fraidenraich, G.; Meloni, L.G.P.; De Lima, E.R.; Neto, F.B. Load Disaggregation Based on Time Window for HEMS Application. *IEEE Access* **2021**, *9*, 70746–70757. [CrossRef]
11. Sayed, A.N.; Himeur, Y.; Bensaali, F. From time-series to 2D images for building occupancy prediction using deep transfer learning. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105786. [CrossRef]
12. Abdel-Razek, S.A.; Marie, H.S.; Alshehri, A.; Elzeki, O.M. Energy Efficiency through the Implementation of an AI Model to Predict Room Occupancy Based on Thermal Comfort Parameters. *Sustainability* **2022**, *14*, 7734. [CrossRef]
13. Dutta, J.; Roy, S. OccupancySense: Context-based indoor occupancy detection & prediction using CatBoost model. *Appl. Soft Comput.* **2022**, *119*, 108536. [CrossRef]
14. Zhang, Y.; Qian, W.; Ye, Y.; Li, Y.; Tang, Y.; Long, Y.; Duan, M. A novel non-intrusive load monitoring method based on ResNet-seq2seq networks for energy disaggregation of distributed energy resources integrated with residential houses. *Appl. Energy* **2023**, *349*, 121703. [CrossRef]
15. Ohara, T.; Zheng, C.Y.; Murata, S.; Wada, C. Inducing unstable walking conditions through visual and auditory stimuli. *J. Phys. Ther. Sci.* **2024**, *36*, 330–336. [CrossRef]
16. (PDF) Support Vector Machine-A Survey. Available online: https://www.researchgate.net/publication/275331297_Support_vector_machine-A_survey (accessed on 12 January 2026).
17. Graves, A. *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer Nature: Durham, NC, USA, 2012; Volume 385. [CrossRef]
18. de Ville, B. Decision trees. *Wiley Interdiscip. Rev. Comput. Stat.* **2013**, *5*, 448–455. [CrossRef]
19. Kramer, O. K-Nearest Neighbors. In *Dimensionality Reduction with Unsupervised Nearest Neighbors*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 13–23. [CrossRef]
20. Dey, R.; Salemt, F.M. Gate-variants of Gated Recurrent Unit (GRU) neural networks. *Midwest. Symp. Circuits Syst.* **2017**, *2017*, 1597–1600. [CrossRef]
21. Singh, J.; Banerjee, R. A study on single and multi-layer perceptron neural network. In Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019, Erode, India, 27–29 March 2019; pp. 35–40. [CrossRef]
22. Logistic Regression: Concept and Application. Available online: https://www.researchgate.net/publication/289651384_Logistic_Regression_Concept_and_Application (accessed on 12 January 2026).
23. Salehinejad, H.; Sankar, S.; Barfett, J.; Colak, E.; Valaee, S. Recent Advances in Recurrent Neural Networks. *arXiv* **2018**, arXiv:1801.01078. [CrossRef]
24. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef]
25. Üstün, B.; Melssen, W.J.; Buydens, L.M.C. Visualisation and interpretation of Support Vector Regression models. *Anal. Chim. Acta* **2007**, *595*, 299–309. [CrossRef]
26. Yang, B.; Haghighat, F.; Fung, B.C.M.; Panchabikesan, K. Season-Based Occupancy Prediction in Residential Buildings Using Machine Learning Models. *e-Prime-Adv. Electr. Eng. Electron. Energy* **2021**, *1*, 100003. [CrossRef]

27. Kanthila, C.; Boodi, A.; Beddiar, K.; Amirat, Y.; Benbouzid, M. Building Occupancy Behavior and Prediction Methods: A Critical Review and Challenging Locks. *IEEE Access* **2021**, *9*, 79353–79372. [CrossRef]
28. Khalil, M.; McGough, S.; Pourmirza, Z.; Pazhoohesh, M.; Walker, S. Transfer Learning Approach for Occupancy Prediction in Smart Buildings. In Proceedings of the 12th International Renewable Engineering Conference (IREC), Amman, Jordan, 14–15 April 2021. [CrossRef]
29. Esrafilian-Najafabadi, M.; Haghighat, F. Impact of occupancy prediction models on building HVAC control system performance: Application of machine learning techniques. *Energy Build.* **2022**, *257*, 111808. [CrossRef]
30. Feng, C.; Mehmani, A.; Zhang, J. Deep Learning-Based Real-Time Building Occupancy Detection Using AMI Data. *IEEE Trans. Smart Grid* **2020**, *11*, 4490–4501. [CrossRef]
31. Mena-Martinez, A.; Delgado, M.D.; Alvarado-Uribe, J.; Ceballos, H.G. A Real-Life Evaluation of Supervised and Semi-Supervised Machine Learning Approaches for Indirect Estimation of Indoor Occupancy. *IEEE Access* **2024**, *12*, 118673–118693. [CrossRef]
32. Stjelja, D.; Jokisalo, J.; Kosonen, R. Scalable Room Occupancy Prediction with Deep Transfer Learning Using Indoor Climate Sensor. *Energies* **2022**, *15*, 2078. [CrossRef]
33. Fatehi, N.; Politis, A.; Nazari, M.H. AI-Enabled Anomaly-Aware Occupancy Prediction in Grid-Interactive Efficient Buildings. In Proceedings of the 2023 North American Power Symposium, NAPS 2023, Asheville, CA, USA, 15–17 October 2023. [CrossRef]
34. Khan, I.; Zedadra, O.; Guerrieri, A.; Spezzano, G. Occupancy Prediction in IoT-Enabled Smart Buildings: Technologies, Methods, and Future Directions. *Sensors* **2024**, *24*, 3276. [CrossRef] [PubMed]
35. Prabhakaran, K.; Dridi, J.; Amayri, M.; Bouguila, N. Explainable K-Means Clustering for Occupancy Estimation. *Procedia Comput. Sci.* **2022**, *203*, 326–333. [CrossRef]
36. Tekler, Z.D.; Chong, A. Occupancy prediction using deep learning approaches across multiple space types: A minimum sensing strategy. *Build. Environ.* **2022**, *226*, 109689. [CrossRef]
37. Han, X.; Hu, Z.; Li, C.; Wu, J.; Li, C.; Sun, B. Prediction of human thermal comfort preference based on supervised learning. *J. Therm. Biol.* **2023**, *112*, 103484. [CrossRef]
38. Luo, M.; Xie, J.; Yan, Y.; Ke, Z.; Yu, P.; Wang, Z.; Zhang, J. Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II. *Energy Build.* **2020**, *210*, 109776. [CrossRef]
39. Martins, L.A.; Soebarto, V.; Williamson, T.; Pisaniello, D. Personal thermal comfort models: A deep learning approach for predicting older people's thermal preference. *Smart Sustain. Built Environ.* **2022**, *11*, 245–270. [CrossRef]
40. Alam, N.; Zaki, S.A.; Ahmad, S.A.; Singh, M.K.; Azizan, A.; Othman, N. Machine learning approach for predicting personal thermal comfort in air conditioning offices in Malaysia. *Build. Environ.* **2024**, *266*, 112083. [CrossRef]
41. Abdelrahman, M.M.; Chong, A.; Miller, C. Personal thermal comfort models using digital twins: Preference prediction with BIM-extracted spatial–temporal proximity data from Build2Vec. *Build. Environ.* **2022**, *207*, 108532. [CrossRef]
42. Luo, M.; Jiang, K.; Wang, J.; Feng, W.; Ma, L.; Shi, X.; Zhou, X. Data-driven thermal preference prediction model with embodied air-conditioning sensors and historical usage behaviors. *Build. Environ.* **2022**, *220*, 109269. [CrossRef]
43. Chennapragada, A.; Periyakoil, D.; Das, H.P.; Spanos, C.J. Time series-based deep learning model for personal thermal comfort prediction. In *e-Energy 2022—Proceedings of the 2022 13th ACM International Conference on Future Energy Systems*; ACM: New York, NY, USA, 2022; pp. 552–555. [CrossRef]
44. Lala, B.; Kala, S.M.; Rastogi, A.; Dahiya, K.; Yamaguchi, H.; Hagishima, A. Building Matters: Spatial Variability in Machine Learning Based Thermal Comfort Prediction in Winters. In Proceedings of the 2022 IEEE International Conference on Smart Computing, SMARTCOMP 2022, Espoo, Finland, 20–24 June 2022; pp. 342–348. [CrossRef]
45. Boutahri, Y.; Tilioua, A. Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings. *Results Eng.* **2024**, *22*, 102148. [CrossRef]
46. Kumar, T.M.S.; Kurian, C.P. Real-time data based thermal comfort prediction leading to temperature setpoint control. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 12049–12060. [CrossRef]
47. Elehwany, H.; Ouf, M.; Gunay, B.; Cotrufo, N.; Venne, J.S. A reinforcement learning approach for thermostat setpoint preference learning. *Build. Simul.* **2024**, *17*, 131–146. [CrossRef]
48. Xiong, R.; Shi, Y.; Jing, H.; Liang, W.; Nakahira, Y.; Tang, P. Calibrating subjective data biases and model predictive uncertainties in machine learning-based thermal perception predictions. *Build. Environ.* **2024**, *247*, 111053. [CrossRef]
49. Liu, Y.; Li, H.; Zheng, P.; Wu, H.; Wang, Z.; Li, Z.; Lin, B. Non-invasive measurements of thermal discomfort for thermal preference prediction based on occupants' adaptive behavior recognition. *Build. Environ.* **2023**, *228*, 109889. [CrossRef]
50. Lala, B.; Rizk, H.; Kala, S.M.; Hagishima, A. Multi-Task Learning for Concurrent Prediction of Thermal Comfort, Sensation and Preference in Winters. *Buildings* **2022**, *12*, 750. [CrossRef]
51. Shahid, Z.K.; Saguna, S.; Ahlund, C. Recognizing Seasonal Sleep Patterns of Elderly in Smart Homes Using Clustering. In Proceedings of the IEEE Consumer Communications and Networking Conference, CCNC, Las Vegas, NV, USA, 6–9 January 2024; pp. 490–498. [CrossRef]

52. Mohan, P.; Lee, B.; Chaspari, T.; Ahn, C.R. Assessment of Daily Routine Uniformity in a Smart Home Environment Using Hierarchical Clustering. *IEEE J. Biomed. Health Inf.* **2021**, *25*, 3197–3208. [CrossRef] [PubMed]
53. ZShahid, K.; Saguna, S.; Ahlund, C. Detecting Anomalies in Daily Activity Routines of Older Persons in Single Resident Smart Homes: Proof-of-Concept Study. *JMIR Aging* **2022**, *5*, e28260. [CrossRef] [PubMed]
54. Natani, A.; Sharma, A.; Perumal, T. Sequential neural networks for multi-resident activity recognition in ambient sensing smart homes. *Appl. Intell.* **2021**, *51*, 6014–6028. [CrossRef]
55. Duppe, B.; Meiser, M.; Anisimov, A.; Antakli, A.; Muaz, M.; Zinnikus, I. Combining Machine Learning With Inductive Logic Learning To Detect Deviations From Daily Routines In Ambient Intelligent Environments. In *ACM International Conference Proceeding Series*; ACM: New York, NY, USA, 2021; pp. 310–317. [CrossRef]
56. Tarmanini, C.; Sarma, N.; Gezegin, C.; Ozgonenel, O. Short term load forecasting based on ARIMA and ANN approaches. *Energy Rep.* **2023**, *9*, 550–557. [CrossRef]
57. Kiprijanovska, I.; Stankoski, S.; Ilievski, I.; Jovanovski, S.; Gams, M.; Gjoreski, H. HousEEC: Day-Ahead Household Electrical Energy Consumption Forecasting Using Deep Learning. *Energies* **2020**, *13*, 2672. [CrossRef]
58. Severiche-Maury, Z.; Uc-Rios, C.E.; Arrubla-Hoyos, W.; Cama-Pinto, D.; Holgado-Terriza, J.A.; Damas-Hermoso, M.; Cama-Pinto, A. Forecasting Residential Energy Consumption with the Use of Long Short-Term Memory Recurrent Neural Networks. *Energies* **2025**, *18*, 1247. [CrossRef]
59. Revati, G.; Palak, M.; Suryawanshi, U.; Sheikh, A.; Bhil, S. Load Profile Prediction in Smart Building using Data Driven Approaches. In Proceedings of the 2021 31st Australasian Universities Power Engineering Conference, AUPEC 2021, Perth, Australia, 26–30 September 2021. [CrossRef]
60. Bendaoud, N.M.M.; Farah, N.; Ahmed, S.B. Applying load profiles propagation to machine learning based electrical energy forecasting. *Electr. Power Syst. Res.* **2022**, *203*, 107635. [CrossRef]
61. Mounir, N.; Ouadi, H.; Jrhilifa, I. Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system. *Energy Build.* **2023**, *288*, 113022. [CrossRef]
62. Liu, X.; Ding, Y.; Tang, H.; Xiao, F. A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. *Energy Build.* **2021**, *231*, 110601. [CrossRef]
63. Mascali, L.; Schiera, D.S.; Eiraudo, S.; Barbierato, L.; Giannantonio, R.; Patti, E.; Bottaccioli, L.; Lanzini, A. A machine learning-based Anomaly Detection Framework for building electricity consumption data. *Sustain. Energy Grids Netw.* **2023**, *36*, 101194. [CrossRef]
64. Pydi, D.P.; Advaith, S. Attention boosted autoencoder for building energy anomaly detection. *Energy AI* **2023**, *14*, 100292. [CrossRef]
65. Tian, Z.; Zhang, X.; Shi, X.; Han, Y. Mining operation hours on time-series energy data to identify unnecessary building energy consumption. *J. Build. Eng.* **2023**, *63*, 105509. [CrossRef]
66. de Rautlin de la Roy, E.; Recht, T.; Zemmari, A.; Bourreau, P.; Mora, L. Deep learning models for building window-openings detection in heating season. *Build. Environ.* **2023**, *231*, 110019. [CrossRef]
67. Onthoni, D.D.; Sahoo, P.K. Artificial-Intelligence-Assisted Activities of Daily Living Recognition for Elderly in Smart Home. *Electronics* **2022**, *11*, 4129. [CrossRef]
68. Liciotti, D.; Bernardini, M.; Romeo, L.; Frontoni, E. A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing* **2020**, *396*, 501–513. [CrossRef]
69. Mihoub, A. A Deep Learning-Based Framework for Human Activity Recognition in Smart Homes. *Mob. Inf. Syst.* **2021**, *2021*, 6961343. [CrossRef]
70. Huang, X.; Zhang, S. Human Activity Recognition based on Transformer in Smart Home. In *ACM International Conference Proceeding Series*; ACM: New York, NY, USA, 2023; pp. 520–525. [CrossRef]
71. Das, D.; Nishimura, Y.; Vivek, R.P.; Takeda, N.; Fish, S.T.; Plötz, T.; Chernova, S. Explainable Activity Recognition for Smart Home Systems. *ACM Trans. Interact. Intell. Syst.* **2023**, *13*, 1–39. [CrossRef]
72. Arrotta, L.; Civitarese, G.; Bettini, C. DeXAR: Deep Explainable Sensor-Based Activity Recognition in Smart-Home Environments. In *Proceedings ACM Interact Mob Wearable Ubiquitous Technol*; ACM: New York, NY, USA, 2022; Volume 6. [CrossRef]
73. Gómez-Ramos, R.; Duque-Domingo, J.; Zalama, E.; Gómez-García-Bermejo, J. An Unsupervised Method to Recognise Human Activity at Home Using Non-Intrusive Sensors. *Electronics* **2023**, *12*, 4772. [CrossRef]
74. Chen, X.; Cumin, J.; Ramparany, F.; Vaufreydaz, D. Towards LLM-Powered Ambient Sensor Based Multi-Person Human Activity Recognition. In Proceedings of the International Conference on Parallel and Distributed Systems—ICPADS, Belgrade, Serbia, 10–14 October 2024; pp. 609–616. [CrossRef]
75. Abbaspour, S.; Fotouhi, F.; Sedaghatbaf, A.; Fotouhi, H.; Vahabi, M.; Linden, M. A Comparative Analysis of Hybrid Deep Learning Models for Human Activity Recognition. *Sensors* **2020**, *20*, 5707. [CrossRef]
76. Jha, D.N.; Chen, Z.; Liu, S.; Wu, M.; Zhang, J.; Morgan, G.; Ranjan, R.; Li, X. A Hybrid Accuracy- and Energy-Aware Human Activity Recognition Model in IoT Environment. *IEEE Trans. Sustain. Comput.* **2023**, *8*, 1–14. [CrossRef]

77. Ayub, M.; El-Alfy, E.S.M. Contextual Sequence-to-Point Deep Learning for Household Energy Disaggregation. *IEEE Access* **2023**, *11*, 75599–75616. [CrossRef]
78. Jiang, J.; Kong, Q.; Plumbley, M.D.; Gilbert, N.; Hoogendoorn, M.; Roijers, D.M. Deep Learning-Based Energy Disaggregation and on/off Detection of Household Appliances. *ACM Trans. Knowl. Discov. Data* **2021**, *15*, 1–21. [CrossRef]
79. Azad, M.I.; Rajabi, R.; Estebsari, A.; Azad, M.I.; Rajabi, R.; Estebsari, A. Nonintrusive Load Monitoring (NILM) Using a Deep Learning Model with a Transformer-Based Attention Mechanism and Temporal Pooling. *Electronics* **2024**, *13*, 407. [CrossRef]
80. Shang, R.; Chen, S.; Chen, Z.; Lu, C.T. GraphNILM: A Graph Neural Network for Energy Disaggregation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Singapore, 2024; Volume 14646, pp. 431–443. [CrossRef]
81. Huang, G.; Zhou, Z.; Lei, S.; Li, Y.; Yang, Q.; Guo, C.; Lin, Z. Federated Unilateral-Mutual Learning for Privacy-Preserving Nonintrusive Load Monitoring. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 9001612. [CrossRef]
82. Ramani, V.; Zhang, Y.; Sonta, A. Non-intrusive energy state monitoring for occupancy driven load management of residential units. *J. Phys. Conf. Ser.* **2025**, *3140*, 062029. [CrossRef]
83. Ferrez, P.; Gabioud, D.; Roduit, P. NILM-Based Solutions for the Automation of Energy Services in Residential Buildings. In *International Conference on Smart Cities and Green ICT Systems, SMARTGREENS—Proceedings*; SciTePress: Setúbal, Portugal, 2024; pp. 85–95. [CrossRef]
84. Langevin, A.; Carbonneau, M.A.; Cheriet, M.; Gagnon, G. Energy disaggregation using variational autoencoders. *Energy Build.* **2022**, *254*, 111623. [CrossRef]
85. Liu, B.; Liu, W.; Luan, W.; Yu, Y.; Zhao, B.; Wang, Y. Automatic Appliance Labeling for Unsupervised NILM Based on Hierarchical Decision-Making. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2515813. [CrossRef]
86. Zhang, N.; Yan, J.; Hu, C.; Sun, Q.; Yang, L.; Gao, D.W.; Guerrero, J.M.; Li, Y. Price-Matching-Based Regional Energy Market With Hierarchical Reinforcement Learning Algorithm. *IEEE Trans. Ind. Inf.* **2024**, *20*, 11103–11114. [CrossRef]
87. Yang, L.; Li, X.; Sun, M.; Sun, C. Hybrid Policy-Based Reinforcement Learning of Adaptive Energy Management for the Energy Transmission-Constrained Island Group. *IEEE Trans. Ind. Inf.* **2023**, *19*, 10751–10762. [CrossRef]
88. Dobbs, J.R.; Hencey, B.M. Model Predictive HVAC Control with Online Occupancy Model. *Energy Build.* **2014**, *82*, 675–684. [CrossRef]
89. Kim, M.L.; Park, K.J.; Son, S.Y. Occupancy-Based Energy Consumption Estimation Improvement through Deep Learning. *Sensors* **2023**, *23*, 2127. [CrossRef]
90. Lv, Z.; Zhao, G.; Xu, Z.; Wu, J.; Zhou, Y.; Liu, K. Cross-Domain Transfer Learning Method for Thermal Adaptive Behavior Recognition with WiFi. *arXiv* **2024**, arXiv:2410.21827. [CrossRef]
91. Fang, Y.; Lim, Y.; Ooi, S.E.; Zhou, C.; Tan, Y. Study of Human Thermal Comfort for Cyber–Physical Human Centric System in Smart Homes. *Sensors* **2020**, *20*, 372. [CrossRef]
92. Gao, G.; Li, J.; Wen, Y. Energy-Efficient Thermal Comfort Control in Smart Buildings via Deep Reinforcement Learning. *arXiv* **2019**, arXiv:1901.04693. [CrossRef]
93. Ding, X.; An, Z.; Rathee, A.; Du, W. A Safe and Data-efficient Model-based Reinforcement Learning System for HVAC Control. *IEEE Internet Things J.* **2024**, *12*, 8014–8032. [CrossRef]
94. Ding, Z.-K.; Fu, Q.-M.; Chen, J.-P.; Wu, H.-J.; Lu, Y.; Hu, F.-Y. Connection Science Energy-efficient control of thermal comfort in multi-zone residential HVAC via reinforcement learning Energy-efficient control of thermal comfort in multi-zone residential HVAC via reinforcement learning. *Conn. Sci.* **2022**, *2022*, 2364–2394. [CrossRef]
95. Prabowo, A.; Chen, K. Continually learning out-of-distribution spatiotemporal data for robust energy forecasting. *arXiv* **2023**, arXiv:2306.06385.
96. Noorchenarboo, M.; Grolinger, K. Explaining deep learning-based anomaly detection in energy consumption data by focusing on contextually relevant data. *Energy Build.* **2025**, *328*, 115177. [CrossRef]
97. Wastensteiner, J.; Weiss, T.M.; Haag, F.; Hopf, K. Explainable AI for tailored electricity consumption feedback—An experimental evaluation of visualizations. *arXiv* **2022**, arXiv:2208.11408.
98. Clement, T.; Truong, H.; Nguyen, T.; Kemmerzell, N.; Abdelaal, M.; Stjelja, D. Beyond explaining: XAI-based Adaptive Learning with SHAP Clustering for Energy Consumption Prediction. *arXiv* **2024**, arXiv:2402.04982. [CrossRef]
99. Wang, H.; Si, C.; Zhao, J. A Federated Learning Framework for Non-Intrusive Load Monitoring. *arXiv* **2021**, arXiv:2104.01618. [CrossRef]
100. Zhang, Y.; Tang, G.; Huang, Q.; Wang, Y.; Wu, K.; Yu, K.; Shao, X. FedNILM: Applying Federated Learning to NILM Applications at the Edge. *IEEE Trans. Green. Commun. Netw.* **2021**, *7*, 857–868. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.