sustainability

MDPI

Review

Knowledge Acquisition and Representation for High-Performance Building Design: A Review for Defining Requirements for Developing a Design Expert System

Seung Yeoun Choi [1] and Sean Hay Kim [2,*]

[1] Hanil Mechanical & Electrical Consultant, Seoul 07271, Korea; seungyeoun.choi@himec.co.kr
[2] School of Architecture, Seoul National University of Science and Technology, Seoul 01811, Korea
* Correspondence: seanhay.kim@seoultech.ac.kr

Abstract: New functions and requirements of high performance building (HPB) being added and several regulations and certification conditions being reinforced steadily make it harder for designers to decide HPB designs alone. Although many designers wish to rely on HPB consultants for advice, not all projects can afford consultants. We expect that, in the near future, computer aids such as design expert systems can help designers by providing the role of HPB consultants. The effectiveness and success or failure of the solution offered by the expert system must be affected by the quality, systemic structure, resilience, and applicability of expert knowledge. This study aims to set the problem definition and category required for existing HPB designs, and to find the knowledge acquisition and representation methods that are the most suitable to the design expert system based on the literature review. The HPB design literature from the past 10 years revealed that the greatest features of knowledge acquisition and representation are the increasing proportion of computer-based data analytics using machine learning algorithms, whereas rules, frames, and cognitive maps that are derived from heuristics are conventional representation formalisms of traditional expert systems. Moreover, data analytics are applied to not only literally raw data from observations and measurement, but also discrete processed data as the results of simulations or composite rules in order to derive latent rule, hidden pattern, and trends. Furthermore, there is a clear trend that designers prefer the method that decision support tools propose a solution directly as optimizer does. This is due to the lack of resources and time for designers to execute performance evaluation and analysis of alternatives by themselves, even if they have sufficient experience on the HPB. However, because the risk and responsibility for the final design should be taken by designers solely, they are afraid of convenient black box decision making provided by machines. If the process of using the primary knowledge in which frame to reach the solution and how the solution is derived are transparently open to the designers, the solution made by the design expert system will be able to obtain more trust from designers. This transparent decision support process would comply with the requirement specified in a recent design study that designers prefer flexible design environments that give more creative control and freedom over design options, when compared to an automated optimization approach.

Keywords: design expert system; knowledge base; high performance building; knowledge acquisition; knowledge representation

check for updates

Citation: Choi, S.Y.; Kim, S.H. Knowledge Acquisition and Representation for High-Performance Building Design: A Review for Defining Requirements for Developing a Design Expert System. *Sustainability* **2021**, *13*, 4640. https://doi.org/10.3390/su13094640

Academic Editor: Ali Bahadori-Jahromi

Received: 17 March 2021
Accepted: 16 April 2021
Published: 21 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Recently, requirements and constraints that should be satisfied during the planning and designing of high-performance buildings (HPBs) have been increasing. For instance, the location of ground heat exchangers for ground heat pumps (e.g., whether they should be positioned around or under the building) should be decided as early as the planning phase since their total length may not be reasonable if they clash over the building foundation; it should be analyzed whether a cool roof is advantageous for reducing the cooling

load or whether it would increase the heating load depending on roof footprint and neighbor's shade.

Furthermore, certification requirements and regulations are being enforced more strictly due to the increasing performance threshold of HPBs (e.g., the increasing renewable energy ratio). Consequently, often, the initial design approved by the client in the earlier phases may not be valid in later phases, especially for long-term construction projects.

With the increase in client expectations for HPBs and public preference for sustainability, various cutting-edge systems and integrated measures have been proposed. Accordingly, even experienced HPB designers may have to research new systems, learn their uses, and analyze their suitability for the project at hand.

While leading design firms could provide economically reasonable specifications of HPBs on the basis of the facility's size, purpose, and budget, architects who do not belong to such firms or who have been managing small projects could lag behind such firms in terms of HPB design experience, unless they are trained or advised by experienced professionals. Furthermore, since recently introduced measures for HPBs, including those pertaining to renewable and hybrid systems, require their deep integration with legacy systems, they should be considered from a very early phase of the project to hedge risks and increase the project's feasibility.

Often, consultants are hired for advising designers and clients with up-to-date HPB experience and knowledge. However, on the basis of their experience with similar projects, the consultants may also have to investigate performance opportunities, research appropriate legal and engineering measures, run performance simulations, and consult domain experts. Designers of small projects whose budget does not permit the hiring of HPB consultants should will have to proceed through trial and error by themselves.

For users who require design or engineering advice but lack the budget to directly hire consultants and resources to research effective measures, computer aids and information services can be affordable to offer a reasonable solution. For instance, a Korean consulting firm provides a web service offering real estate feasibility assessment for clients desiring to know the possible total floor area and number of floors for a given site and estimates of the construction cost and expected rent [1]. Although the initial mass and space program proposed by this web service may not provide information as accurate and detailed as that presented by an architect, it can help the client make a go/no-go decision.

An expert system is a computer system that emulates a consultant or advisor. It provides tangible and affordable solutions in an appropriate context at reasonable cost, and ultimately to offer an alternative to or aid a human consultant in solving problems whose category and scope have already been defined. Thus, routine practices and trials-and-errors often work well for the expert system.

Specifically, a design expert system provides design advice to designers and clients in every design phase of HPBs [2,3], like a building energy consultant. Similar to a generic expert system, the design expert system has two distinguishable cores: a knowledge base and an inference engine. While the knowledge base represents facts and information about the world in digital form, the inference engine, whose inference process is typically based on IF-THEN rules, applies logic to the knowledge base and deduces new knowledge. In other words, the interference engine interprets the design problem and determines the most appropriate solution for simple answers (e.g., what is the allowable U-value of exterior walls of apartments in Seoul?) by exploring the knowledge base; for broader advice requests (e.g., what is the most economical air-conditioning system for a five-story commercial building in a shopping district where district heating is provided?), it synthesizes ontologies from the knowledge base by reasoning about logic and rules.

2. Study Objectives and Research Flow

As the expert system attempts to solve a complex task like a human expert, the successful development of both knowledge base and inference engine depends on the acquisition of relevant knowledge within a reasonable problem boundary, expressive representation of

the knowledge, and lastly the development of an appropriate digital formalism that computers can easily understand and implement. Furthermore, since explicit representation of the acquired knowledge is essential for reasoning to draw inferences and extract new knowledge, the knowledge representation formalism should allow the reasoning process itself to be influenced by new inferences during meta-level reasoning.

In the design of a knowledge representation formalism, there is a trade-off between expressivity and practicality [4]. Apart from IF-THEN rules, recently, more formal and domain specific knowledge representation formalisms that pursue enhancing both expressivity and practicality, such as gbXML, have been developed for use in various domains where knowledge base systems are employed. While some formalisms tend to be proofs of concept academically, some have been integrated with applications deployed in corporate environments and actually used for a decision support tool for actual engineering problems [5].

In the long run this research aims to develop a HPB design expert system that can provide design advice easily at low cost to designers and clients. To this end, this study encompasses base research because the development of the design expert system can be initiated by selecting or developing appropriate knowledge representations (KR) once the specifications and requirements about HPB design KRs are first established—these being required to develop the expert system as the first step. Therefore, the final purpose of this study is to establish the knowledge definition and category required for existing HPB design and find the method to acquire the most appropriate knowledge and the representation criteria for the HPB design expert system.

In summary, this study seeks to achieve the objectives presented above through the following procedures.

1. Step 1: The roles and functions of the knowledge base (KB) and inference engine (IE) in general expert systems were reviewed and the functions and processes of the general knowledge formulation were described (Section 3.1).
2. Step 2: An overview of design expert systems and decision support systems in the AEC (Architecture, Engineering, and Construction) industry was established (Section 3.2).
3. Step 3: The HPB design phase was reviewed to identify design problems that could be solved by the HPB design expert system, and the required design knowledge was summarized for each design phase (Section 4).
4. Step 4: Based on the classification by [5], the knowledge acquisition (KA) and knowledge representation (KR) classification criteria were redefined according to the HBP design processes and required knowledge specifications (Sections 5 and 6).
5. Step 5: The literature on HPB design was reviewed and classified according to the KA and KR reset criteria. In particular, to ensure the research trend and currency of the latest HPB design-related studies, the literature after 2010 was set as the main analysis target (Sections 5 and 6).
6. Step 6: Finally, the specifications and requirements of the HPB design KR were discussed for the development of the future HPB design expert system, and the technical requirements of the system were presented to complete this study (Sections 7 and 8).

3. Design Expert System and Decision Support System

3.1. Expert System and Knowledge Formulation

Figure 1 depicts a typical component configuration of the expert system. When a user provides an engineering problem, the inference engine analyzes the problem and synthesizes new answers or results by applying rules and logic to given facts and information contained in the knowledge base. Since a knowledge base should cover the scope of engineering problems and the range of relevant information within which the expert system can provide solutions, the corresponding raw data and information should be formulated to be affordable and in a tangible format so that the inference engine can easily access the portable knowledge.

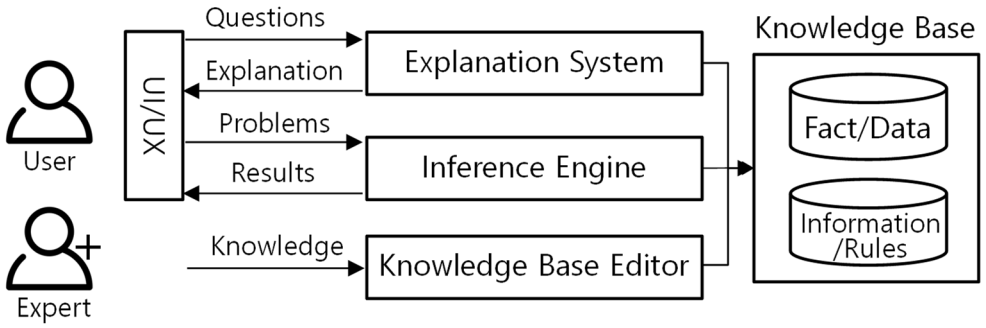


Figure 1. Typical system architecture of expert system.

Knowledge formulation can be addressed as a three-phase process: knowledge acquisition, the mechanics associated with structuring knowledge, and knowledge porting [6].

Traditionally knowledge acquisition is known as the process of extracting knowledge from domain experts. Knowledge can be accumulated through interviews and surveys of experts by a knowledge engineer, or through lab sessions conducted by an expert with a computer-based knowledge acquisition tool. Additionally, as computer modeling and simulations, which used to only be research tools, have become tangible technology to general users, computer-based acquisition has become a major knowledge accumulation method. Above all, as big data observations and measurements are quality sources of knowledge, data analytics and machine learning are emerging methods of knowledge acquisition.

Structuring knowledge involves arranging acquired raw data and information to form a semantic structure. It also includes the refinement of the object's definition, semantics, specifications, and relationships contained in the raw knowledge into a tangible and consistent format. Generally, in the structuring of knowledge, there is considerable focus on the mechanism used for organizing the raw data and the extraction of refined knowledge.

Knowledge formation is improved by porting knowledge that has been obtained from a search for refined knowledge relevant to the engineering problem under consideration. Eventually, porting knowledge is the process that delivers off-the-shelf solutions or synthesizes solutions. Popular representation formats of portable knowledge include rules, frames, cognitive maps, and case-based reasoning [5]. However, with the maturing of human–computer interactions and cognitive science, advances in computer-based analysis and the resulting big data analytics have extended the types of knowledge representation formats into, for example, Semantic Webs and (dynamic) data structures.

3.2. Expert System and Decision Support System for Building Design

Since the 1980s, many expert systems for architectural design have been developed. The studies published mainly deliberated on how computation tools could support architects in building layout [7] with automated design process [8]. Some studies provided a systemic process for interpreting design codes as design knowledge [9] and code compliance check [10]. These studies tended to follow regular specifications of expert systems, including knowledge acquisition, knowledge base construction, and use cases of the design expert system. However, most of the first-generation architectural design expert systems could not evolve into a deployable software at the industry environment owing to technical difficulties in implementing initial concepts, and in expanding the scope of the knowledge toward a level that practitioners prefer.

A comparatively fewer number of expert systems have been used in infrastructure industries, including the construction, transportation, and real estate industries [5]; although, some of early expert systems were applied to structural design [11]. There are several intrinsic reasons for automated expert systems not proliferating in the AEC industry; most AEC businesses depend on human experience and expertise. This is attributed to the AEC industry's history, namely, a majority of construction projects have been rooted in a custom context, resulting in proprietarization and localization often being the only working solu-

tions. Above all, most decision-makers are more familiar with asking for experience and expertise from consultants or predecessors. Consequently, in the AEC industry, although raw and field knowledge should be collected, structured, refined, and polished into a portable form, much of the knowledge still remains in verbal records and rule of thumbs.

In addition to decisions made in the AEC industry mainly depend on localization and heuristics, the decisions made in the early phases of AEC project (i.e., conceptualization and planning) seriously influence the project's total cost, schedule, and facility's performance later on. Thereby a number of studies pertaining to make scientific and objective decisions at the early phase have started to be published since 2000; their spectrum covers aspects as diverse as performance-based designs, integrated design methods, the primary decision-making process, optimization algorithms, modeling environments, simulation tools, the system framework, and the design platform. In one design decision support system with an objective similar to that of a design expert system, decision-making processes and algorithms whose operations are analogous to that of an inference engine of expert system can be loaded on it. However, details of the knowledge acquisition process and the standardization, normalization, and formatting of the acquired knowledge tend to be omitted or poorly described in the aforementioned studies.

The scope of literature review presented here extends up to the development of models, simulations, algorithms, processes, frameworks, and platforms for HPB design decision-making. First, design phases and the design problems of each phase on which the studies focused were categorized. Subsequently, the manner of knowledge acquisition and ordering of raw information and coarse knowledge was investigated, and it was examined how they were structured and represented as tangible portable knowledge for use in HPB design decision-making.

4. Classification of Design Problems for High Performance Building

4.1. Building Design Phases

An expert system would function well and perform as expected when the problem domain is properly scoped and well defined [12]. In this section, major HPB design problems and decisions where an expert's advice is required are summarized for each design phase. Additional details about energy design tasks and issues in each design phase can be found in [13,14].

4.1.1. Conceptual Planning Phase

In accordance with the owner's project requirement (OPR) and client preferences, building volumetry, form, footprint, number of floors, and facility programming are determined in the conceptual planning phase. Typically, in this phase, the focus is on the economy of the construction project rather than the building performance, and only statutory requirements relating to the building performance for the urban plan and facility type (such as municipal energy code) are investigated. This is because the project may be shelved if its economy does not meet the owner's expectations. Thus, the conceptual design is set by considering the maximum economy at the given conditions and constraints.

4.1.2. Schematic Design Phase

Further details of the building's baseline, including geometry, structure and construction, and facility programming, are developed. Discussions on the project's energy goal commence in this phase. Furthermore, the use of natural and renewable energy and measures to decrease the building's energy demand should be determined in this phase. Since most of the required performance agendas are already prescribed at the conceptual planning phase, performance opportunities and risks within the boundary of the preset budget and timeline should be quantitatively discussed in this phase.

4.1.3. Design Development Phase

The building's plans/sections/elevations, space program, fabric and facades, and construction material and finishes are determined. All primary engineering configurations and specifications are also determined. Engineering includes structure, civil, landscaping, mechanical, electrical, plumbing, telecommunication, and fire protection engineering. As well as measures to reduce building's energy demand, measures to reduce energy use should be developed from this phase.

Because all the energy-sensitive design variables are selected and their values are determined and recorded in the binding legal documents at this phase, modifying the project's preset energy goals is very difficult in later design phases. Therefore, primary design decisions and design changes must be made in this phase to avoid escalation of the project cost.

4.1.4. Construction Document Phase

All the engineering and construction details are described for the baseline at the design development phase. From the perspective of energy use, these details include the specifications, capacity, efficiency, and temporal performance of systems and devices. In particular, their operation and control strategies for the purpose of the space are designed in this phase, which can have an unexpectedly large influence on the long-term energy performance of the building. This phase is the very last design stage where only minor fixes and change events are marginally allowed at relatively low cost.

Design decisions to increase the energy performance of the building pertain to two aspects: (i) lower energy demand and (ii) lower energy use. The former objective can be achieved by so-called passive measures that reduce heat gain and loss and by using natural energy. The latter objective can be achieved through active measures that increase the system efficiency, produce energy, and improve the energy transport process. First, the energy demand should be minimized, and the capacity and complexity of the energy system should then be minimized. With this minimized baseline, temporal energy use can be optimized eventually.

In later design phases, design changes to decrease both energy demand and energy use become more difficult to introduce and their opportunity costs increase. In particular, since the values of most design variables, which are sensitive to the energy performance of the building, are determined prior to the construction document phase, experience and expertise in HPBs are very important in the early design phases. Thereby, most HPB design support tools primarily target phases as early as the schematic design phase.

4.2. Types of HPB Design Problems

HPB design problems in the surveyed literature can be largely classified into analysis problems and synthesis problems, and further detailed classifications are shown in Table 1. The original structure of Table 1 has been borrowed and modified from [15].

Table 1. Types of design problems for high performance building.

PD-I. Analysis problems
PD-I-1. Understanding and comparison: identifying difference or similarity of expected outcomes
PD-I-2. Classification: categorizing based on observables
PD-I-3. Interpretation: inferring situation description from numeric and text data
PD-I-4. Prediction: Inferring likely consequences of given situations
PD-II. Synthesis problems
PD-II-1. Configuration: configuring collections of objects under considerations in framework or platform
PD-II-2. Planning and design: configuring collections of objects under considerations in relatively large search spaces
PD-II-3. Process design: planning with strong temporal and/or spatial constraints

4.2.1. HPB Analysis Problems

Analysis problems should first be understood and then classified and decomposed into subproblems to understand their mechanism and behavior. Eventually analysis problems aim to infer useful rules and/or consequences and to predict the expected benefits and/or risks.

Although an analysis problem does not generate or suggest design alternatives, it provides designers or decision-makers knowledge of where to start, what to do, and how to do. Classification is the most fundamental analysis for design problems. Classification not only enables the identification of the prime features of a group, but also facilitates feature comparisons between groups, thereby it contrasts the difference and indicates what to select first. Eventually this classified design information can serve as a design guideline to provide normalized and static design knowledge. Exemplary studies concerning classification include case classification based feature similarity [16], classification of energy sensitive building characteristics [17–22], style classification, and prediction of residential buildings [23].

The two representative interpretation problems include selection of important variables and design pattern analysis. Technically the former is called sensitivity analysis, and the latter is called rule extraction or rule mining. While sensitivity analysis indicates which design variables decision-makers should first concentrate on, the rules indicate the values of design variables and the conditions required in a specific situation and in a generic context. In particular, the rules can be both qualitative and quantitative. For example, statutory rules can specify that the living room shall be sunlit and ventilated (qualitative), while also stipulating that the illuminance and ventilation rate of the living room shall be at least 300 lux and 21.6 CMH (m^3/h) per capita, respectively (quantitative). Exemplary studies concerning interpretation include sensitivity analysis [24–26], ontology based information extraction [27], main effect [28], design of experiments [29], text mining [30,31], and pattern matching extraction [27].

Prediction problems are more case-specific and solution-oriented than interpretation problems; the system provides a specific value to a decision-maker's specific question. Most prediction problems reported in the literature have been solved using mathematical and statistical calculations, while some other prediction problems have been solved by employing logical inference. Exemplary studies concerning prediction include energy demand prediction at early stage [32], metamodel of energy consumption at early stage [25,33], statistical prediction [34,35], and machine learning based predictive model [36].

4.2.2. HPB Synthesis Problems

In general engineering, synthesis problems disassemble and assemble resources in combinations under different conditions and constraints and render the resource as narrow as that from in a single functional aspect, or as broad as that in a semantic framework.

The HPB design configuration problem aims to structure collections of objects under consideration in a framework or platform. This type of problem often aims to develop a new modeling platform if the existing modeling environment cannot meet a new performance aspect and/or further integration needs, or if the existing modeling environment should be revised or reconstructed. Exemplary studies concerning design configuration include exterior lighting design tool [37], design expert system [38–41], open knowledge base for HPB [42], optimization framework for HPB [43], BIM interoperability specification [44], BIM-DOE design framework [29], BIM based optimization framework for HPB [45], sustainable BIM [41,46,47], semantic web framework [48], ontology framework for sustainable structure [49], performance assessment ontology [50], simulation based framework [24,51], façade design framework [52–54], single house design framework [55], BMS framework for design and operation [56], qualitative assessment tool for building envelope [57], multicriteria decision making framework [16], case based reasoning framework for HPB [31], and cross-domain building data share platform [58].

The HPB planning and design problem aims to arrange collections of objects under consideration in search spaces to meet the project objective. This is a typical engineering

problem that searches for the optimal formation of building element layout or geometry, system configuration, construction and material properties, and other architectural and system features in the given context, conditions, and constraints. The optimal formation is searched for either in an off-the-shelf platform or in a customized or new environment. Exemplary studies concerning optimal formation include exterior lighting design [37], façade design [48,58], daylighting design [28,43,45], visual comfort design [52], energy performance design [24,29,43,45,46,51,52,55], life cycle cost optimization [45], refurbishment cost optimization [39], servicescape design [40], and construction cost optimization [55].

The HPB process design problem aims to set up collections of objects and functions under consideration in search spaces with temporal order. This is also a typical engineering problem in building operation problems, but it is somewhat rare in building design problems. Because this type of problem searches for an optimal sequence and schedule of the objects that are already optimally located in terms of physical space, it is frequently observed that the optimally located objects can be modified to make their process model optimal for better lifecycle performance. Process design problems reported in the literature are not many, and exemplary studies concerning process design include interoperability specification development [44], data acquisition process model [56], robustness-based decision making approach [59], and HPB design process modeling [60].

5. Knowledge Acquisition

Knowledge engineering refers to an activity to construct an ontological structure of knowledge by means of heuristic rules, formula, semantic models, and other formalisms on the basis of data acquired via observations, interviews, experiments, literature surveys, analytics, and researches. Knowledge acquisition is therefore the first step of the knowledge engineering. In the HPB design literature, they can be classified into heuristic and computer-aided acquisition methods, as shown in Table 2.

Table 2. Knowledge acquisition sources and methods.

KA-I. Heuristic acquisition
KA-I-1. Unstructured interviews with experts and users
KA-I-2. Structured surveys with experts and users
KA-I-3. Protocol, code, guideline analysis
KA-I-4. Customary practice, business convention, de-facto standards
KA-II. Computer-aided acquisition
KA-II-1. Web search
KA-II-2. Databases, digital libraries, ISPs
KA-II-3. Transactional and operational data (i.e., measured data)
KA-II-4. Simulated data

5.1. Heuristic Acquisition

Classical ways of collecting knowledge have been face-to-face interviews and surveys with domain experts. Business analysts and knowledge engineers collect clients' requirements and documents, conceptualize a domain expert's understanding and knowledge, and finally implement plans to satisfy the client requirements using knowledge engineering tools such as taxonomy, schematics, and maps. Thus, the most basic heuristic knowledge acquisition method is conducting a structured/unstructured interview (Table 3). A structured interview involves a wide range of systematic questionnaire tools that the knowledge collector can use to interview domain experts, such as surveys, lists, maps, and an early prototype system for feedback.

Protocol analysis involves the examination of verbal and signage rules for handling domain problems that are customarily or privately used by domain experts, while code analysis involves the examination of visual, schematic, and written rules, tutorials, manuals, and guidelines. Similarly, expertise also can be acquired by using business conventions, customary practices, and de facto standards. A major difference between these knowledge

sources and protocols and codes is that the knowledge sources are verbal and behavioral experiences that have been accumulated by practitioners, the so-called rule of thumb. Although these types of knowledge may not be stated in a written document, they present a live body of knowledge at a currently running business, which is often very useful to construct a semantic model. In many cases in the literature, however, the data originated from customary practices and conventions tended to be stated without particular remarks concerning its sources.

Table 3. Literature by heuristic knowledge acquisition.

KA-I-1.	Unstructured interviews	Energy saving features [61–63], Solar design [64]
KA-I-2.	Structured surveys	AHP [41,65], systemic group survey [49,57]; CBA [16]; CBR [30,31]; Energy saving features [61,62]; Usability testing [24] Building energy performance gap [66]
KA-I-3.	Protocol, code, guideline analysis	Energy code [27,52]; BREEAM [47]
KA-I-4.	Customary practice, convention, de-facto standards	Daylight knowledge base [28]; Servicescape rule extraction [40]; Façade design rule [52]; Clash database schema and feature definitions [67]

5.2. Computer-Aided Acquisition

Since 1990, the explosive use of the internet and database services has expedited the transformation of knowledge acquisition from off-line and monodirectional sources to online and interactive sources (Table 4). Apart from professional opinions and expertise and conventional practices, to which considerably effort and expense should be devoted for knowledge acquisition, most data and information that used to be available in books have been published online. Moreover, more online sources have become shared at no or low cost, resulting in reduced expenditure for data collectors.

Table 4. Literature by computer-aided knowledge acquisition.

KA-II-1.	Web search	Semantic web [42,48,49]
KA-II-2.	Databases and ISPs	Manufacturers' DB [55], Green material DB [47], GIS and climate service [48], Energy benchmark [19,20,36]
KA-II-3.	Transactional and operational data	Energy audit [17,18], Operation data [68–70]
KA-II-4.	Simulated data	Design of experiment [54], Simulated energy data [25,29,32–34,71], Optimization [37,39,43,45]

5.2.1. Web Search

The scope and class of data and information available on the web is more than those of data and information in classic media. The main advantage of online information is that they are "alive" and "self-censored"; because web sites are managed by special author groups (e.g., the EnergyPlus user group) or by anonymous public (e.g., Wikipedia), their information is continuously updated if any glitches are observed. Some online open communities (e.g., eng-tips.com) offer professional knowledge and troubleshooting, which could previously have been obtained only through highly systematic interviews with field experts. Furthermore, personal social network sites can offer meaningful background information. However, basically, all the web sources should be cross-validated.

5.2.2. Public Databases and Information Service Providers

Public design data such as climate and population data were previously obtainable only by making an application to the public data warehouse. By contrast, currently, long-

term public data, including microclimate, floating population, and energy use per industry sector, are released periodically for access by the general public. Furthermore, national energy statistics derived from authority-driven research projects (e.g., CBECS [72]) have also been released to the public.

Commercial information service providers (ISPs) provide customized and streamlined design information. For instance, some online weather ISPs provide application programming interfaces (APIs) for both real-time and historic weather, and even simulation weather files for specific locations. Some geographic information systems (GIS) provide periodic floating population and traffic data for the last decade, which can be useful for assessing project feasibility in the planning phase.

5.2.3. Transactional and Operational Data

As an increasing number of activities and interactions migrate online, the accumulated transactional data become a good source of design information. For instance, an increase in the sale of meal kits implies that kitchen and food preparation spaces can be reduced in small residential buildings.

Sensor and monitors also provide operational big data. These operational data are very useful not only to optimize building operation, but also to evaluate whether a building is used for the designed purpose during post-occupancy evaluations. In particular, actual operational data often suggest values for new engineering design standards such as the required ventilation rate per capita.

5.2.4. Simulated Data

Although observed data can be the best source of live design knowledge, they are hardly under control. Because the state of an actual facility is transient only in a certain range, demand peaks that are typically used for the system design may not appear in an observed data repository. Furthermore, there is no way to calculate the design tolerance on the basis of the likelihood of specific events (e.g., TAC 2.5%) with observed data. In this case, samplings from a formally or well-established model (i.e., simulation mode) are very useful to obtain nominal data.

6. Knowledge Representation

Knowledge representation is the explicit formalization of premises and phenomena of the problem universe in a "machine-readable" language. It describes what entity determines (or causes) what consequences by "reasoning" about the problem space, embodying a formality that can be used to easily design and build a complex system.

As intelligent reasoning is the coherent and logical flow of inductive (or deductive) thinking, it is inextricably intertwined with the related knowledge structure. Thereby computer-aided measures must provide information on how acquired knowledge is framed, how the user's question can be interpreted, decomposed, and resynthesized within the framed knowledge, and finally how analytically the answer can be inferred from the framed knowledge.

However, knowledge representation is not a data structure as Randall Davis indicated [73]. While every knowledge representation requires a data structure, the representational property is in the correspondence to actual features in the world and in the constraint that correspondence imposes [73]. In other words, while a fragmentary knowledge can be represented by a data structure, an entire body of knowledge must be understood within the problem semantics and context, even if the representation of the entire body does not appear different from a simple data structure.

Therefore, this study includes, but not by way of limitation, data structures and equations in a design knowledge representation scheme. This is because even such a simple and plain representation can provide great insights over the problem space. Table 5 lists the knowledge representation schemes employed in the reviewed HPB design studies, and they range from schemes based on a simple logic to complex models.

Table 5. Knowledge representation schemes.

KR-I. Logic, equations and rules
KR-I-1. First order logic and formula KR-I-2. If-then rules and case statements KR-I-3. Fuzzy logic
KR-II. Semantics and ontology
KR-II-1. RDF, OWL and SWRL KR-II-2. Proprietary sematic model and ontology KR-II-3. Building information model
KR-III. Analysis model and simulation
KR-III -1. Analysis model KR-III -2. Building performance simulation
KR-IV. Machine learning
KR-IV-1. Regression KR-IV-2. Classifiers KR-IV-3. Neural networks KR-IV-4. Trees KR-IV-5. Primary factors KR-IV-6. Clusters KR-IV-7. Pattern and association rule extraction KR-IV-8. Multi agent systems KR-IV-9. Optimization
KR-V. Case-based reasoning
KR-VI. Group decision making

6.1. Knowledge Representation by Logic, Formula and Rules

Logic, formulas, and rules (Table 6) can be used to represent the HPB design knowledge in a highly refined and succinct format; they provide users with directions or guidance about what to do in a certain context and certain conditions, or directly provide answers to users' questions.

Table 6. Knowledge representation by logic, formula and rules (KR-I).

KR-I-1.	First order logic and formula	Design indicator of energy performance [71], Robustness and performance indicator [59] Building geometry indicator [74]
KR-I-2.	If-then rules	Design guideline [54], Servicescape design rule [40], Building design alternative rule [52]
KR-I-3.	Fuzzy logic	Design logic [28,38], Performance level assessment [41], Consensus scheme [57]

6.1.1. First Order Logic (FOL) and Formulas

First order logic describes theorems with predicates and quantification. A predicate of the FOL takes an entity of the problem universe as input and describes what it is. For instance, in the first-order logic "X is Y," the variable X can be instantiated (e.g., dog is an animal), the variable Y (a function of X) can be conditioned (e.g., species of the dog) or quantified (e.g., for every dog). Furthermore, relationships between predicates can be stated with a hypothesis and a conclusion (e.g., if Bingo is a dog, then Bingo is an animal).

While first-order logic describes mathematical axioms, a formula is a more practical formalism to identify what entity can be equated to another by a mathematical expression. They are typically written as a logic, arithmetic, algebraic, or closed-form expression. They are advantageous for expressing the "primary" behavior of the problem space with single letters instead of words or phrases. Thus, a formula can capture the principle of a large and complex system more easily.

For instance, the *shape factor* (Equation (1)) is the ratio of a building's total external surface area to its internal volume. According to the shape factor formula, a building design with a higher shape factor results in higher heat loss and gain through the envelope. Using the formula, a designer can determine, with a quantitative basis, whether or not building mass should be made more compact in order to reduce the heat transfer (Equation (1)—*Shape factor* [75]).

$$Shape\ Factor\ (SF) = \sum \frac{S_i}{V} \left[\mathrm{m}^2/\mathrm{m}^3\right] \quad (1)$$

where S denotes surface area [m^2] and V denotes internal volume [m^3].

6.1.2. IF-THEN Rule

Rules are useful to engineer and store information in a generalized form because they are declarative and procedural. Owing to their simple assertion and their easy implementation for reasoning about assertions, the rule-based expert system is the earliest and most widely known knowledge-based system.

According to [76] representing knowledge by using rules has three advantages: (i) easier acquisition and maintenance (domain experts can define and manage the rules themselves without programming), (ii) transparency (it is explainable how the inference was made and how it leads to the conclusion), and (iii) new knowledge discovery (reasoning about and synthesizing rules often reveals new knowledge).

For instance, the following rule [54] indicates the desired change of illuminance per unit distance between an object and a device. Referring this rule (Rule 1), a designer can simply control the expected illuminance level of his/her design by adjusting the distance between the opening and the sensor.

Rule 1. A design rule to change illuminance level [54].

IF SensorPriority is High AND SensorType is Illuminance AND IlluminancePerformance is TooLow:
(a) IF distanceFromGoal is Far, THEN DesiredChange is "Increase Illuminance by a Large Amount";
(b) IF distanceFromGoal is Close, THEN DesiredChange is "Increase Illuminance by a Small Amount".

6.1.3. Fuzzy Logic

The above mentioned first-order logic, formula, and rules define a Boolean operation for each. However, most knowledge is vague and imprecise in nature. For instance, for a rule "IF temperature is hot, THEN turn the fan on," the perception of "hot" differs among occupants. Fuzzy logic has the capability to recognize, represent, manipulate, interpret, and use data and information that are vague and lack certainty [77].

For instance, Figure 2 depicts a fuzzy model for assessing green building performance by evaluating environmental, economic, and social level inputs, which are described in linguistic terms. As linguistic terms should have a subjective meaning for every respondent, fuzzy membership functions are formulated to "quantify" both inputs and outputs and to "normalize" the performance level.

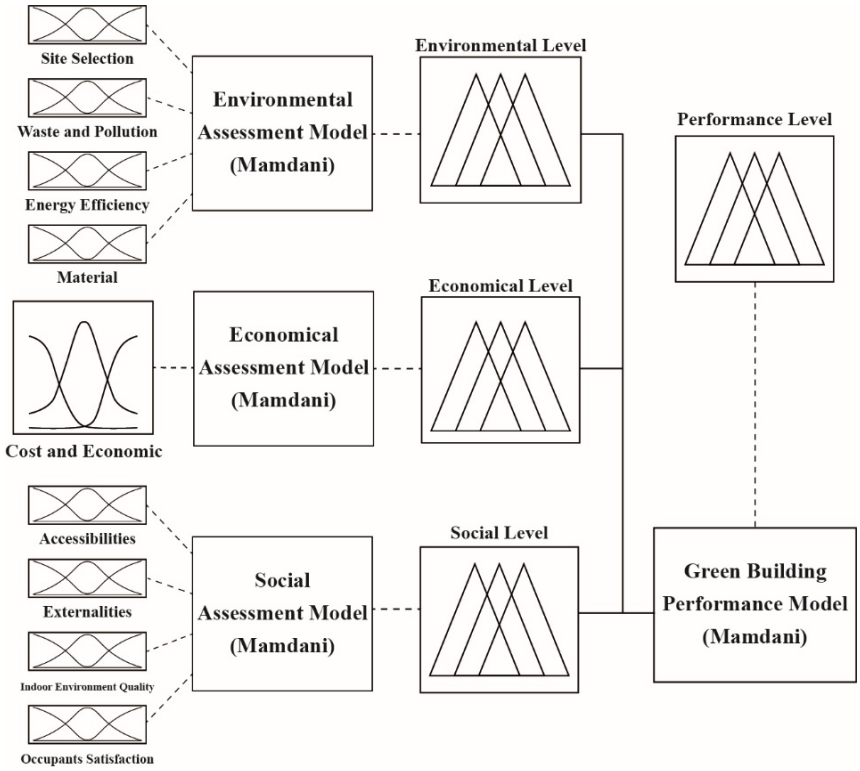


Figure 2. Fuzzy model for green building performance assessment (modified from [41]).

6.2. Knowledge Representation by Semantics and Ontology

In general, the design of a building advances as the alternatives are compared and the drawbacks are overcome while focusing on the advantages. Thus, multiple alternatives are prepared, analyzed, and integrated to be then iterated in the initial design phase. However, only the selected alternatives have their advantages and disadvantages analyzed and their features singled out in the later phase. As users should understand the characteristics of design problems and make complex predictions involving factors such as synergy depending on the decision made rather than relying on refined knowledge such as KR-I (Section 6.1), it has become increasingly necessary to represent knowledge with "schema" and "ontology" as listed in Table 7.

Table 7. Knowledge representation by semantics and ontology (KR-II).

KR-II-1.	Sematic model and ontology	Performance assessment ontology, simulation Domain Model ontology, semantic Sensor Network ontology [50], SEMERGY building model [48], Information extraction ontology [27], gbXML for as-built building scanning [78], SysML for sustainable building design [79,80], Hierarchical graph-based building model [81], Relational database schema for transformation of BIM [82], Product-oriented knowledge base and optimization process model [83]
KR-II-2.	RDFS, OWL and SWRL	Structural design rule [49], ifcOWL [50,84], Semantic BIM [55], SEMERGY building model [48], Linked data ontology [42], Energy efficiency knowledge base [85]
KR-II-3.	Building information model	BIM integrated with performance based design [29,44,45,47]

6.2.1. Semantic Model and Ontology

A semantic model is a set of descriptions and specifications that represent information about the subjects of the problem universe, structure/hierarchy/relations of the subjects,

and their properties and functions. Ontology is a schema or method to construct a semantic model by specifying how categories and concepts are defined, what properties and relations are drawn between the concepts, how data and entities substantiate the concept, and any other significance to explain the problem universe.

As indicated, users can find many semantic modeling languages in the field; some of them are general-purpose standard languages such as Integrated DEFinition Methods (IDEF) and Unified Modeling Language (UML), while some others are proprietary languages for building HPB domain ontologies such as gbXML for green buildings, CIS/2 for structural steel projects, and Systems Modeling Language (SysML) for systems engineering.

For instance, Figure 3 depicts a building semantic model [42] that is made by proprietary ontology. It contains building-related information from various sources, including the web, and describes properties, behavior, and relations determined from performance evaluations; performance-related information is acquired, formatted, and then systemically stored in this model. By synthesizing known knowledge and queries over the semantic model, a designer can acquire new design knowledge—for instance, the area-averaged U-value of all glazings can be calculated by querying the aperture areas and the windows installed on them.

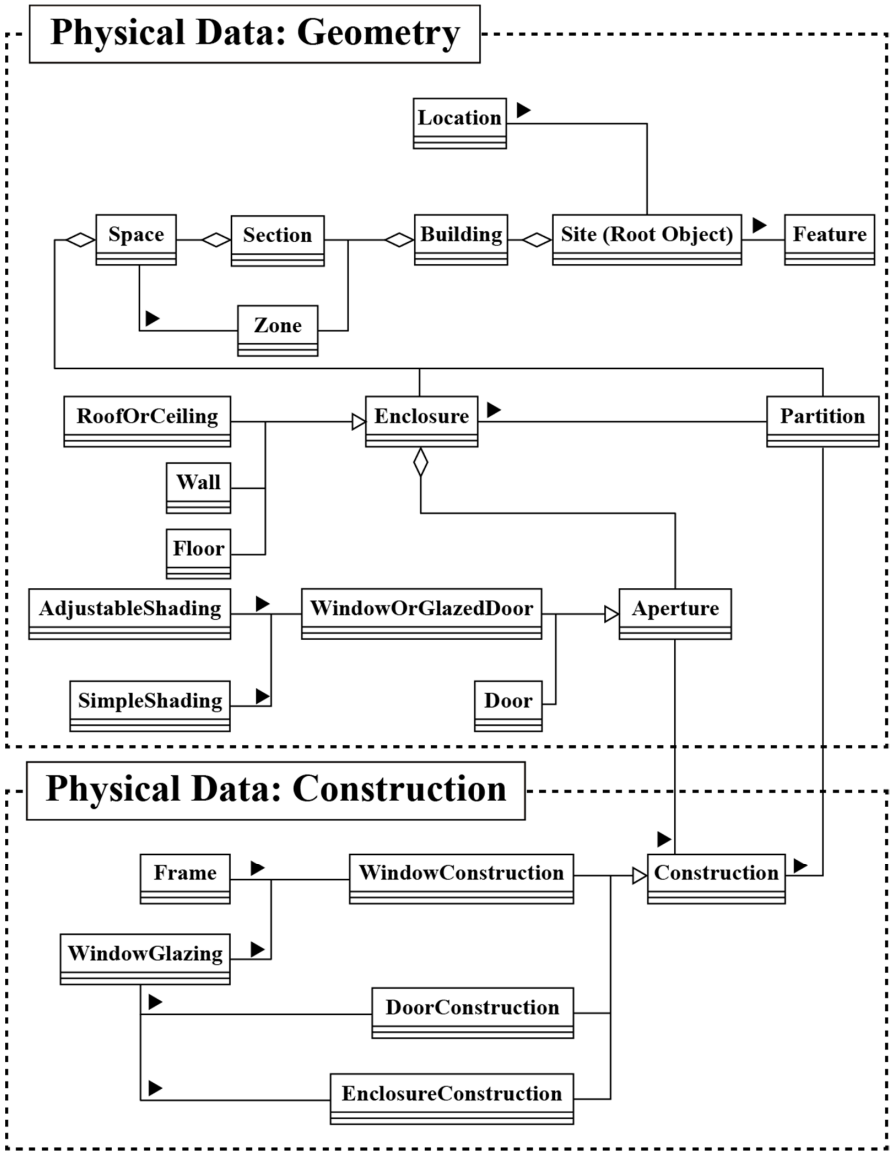


Figure 3. Part of the SEMERGY building model (modified from [48]).

6.2.2. Resource Description Framework (RDF), Web Ontology Language (OWL) and Semantic Web Rule Language (SWRL)

The RDF is a metadata model that defines a data model to store web resources. It uses the form subject–predicate–object, known as triple. As RDF stores these triples and then connects them, RDF is a graph data model. Thus, RDF data are directed and labeled multigraphs, and they are known to be better suited for representing real-world situations than other data models.

However, with only the RDF as the data model, it is difficult to represent real-world knowledge because the data model is merely a container. The structure and properties of data entities and relations between those entities should be declared additionally to capture real-world activities and insights. An RDF schema (RDFS) was developed for a schema language to connect and describe the behavior and relation of web resources stored in the RDF data model. When RDF data models are created by different teams for different uses at different times, RDFS works as a universal translator between data models (so-called common vocabularies). Accordingly, RDFS is in object oriented that describes classes of objects in nature.

OWL is also a schema language that enables not only data modeling but also automated reasoning. Compared with RDFS, OWL offers a larger vocabulary (to define classes and relations), allows users to tailor the modeling option for achieving a faster query search and conceptual reasoning (e.g., by imposing constraints), and eventually optimizes for particular applications based on computational realities.

SWRL is a markup language for publishing and sharing rule bases on the web. It combines OWL (-DL) with a subset of the rule markup language in order to standardize inference rules (i.e., to increase implementability). Its rules are expressed in terms of the OWL scheme (classes, properties, and individuals) and Horn-like rules at the cost of decidability and practical implementation.

For instance, the following OWL and SWRL in Rule 2 describe a design rule for a sustainable structure with low embodied energy [49]; the rule signifies that the total embodied CO2e of a concrete column equals the volume of the column multiplied by the embodied CO2e per unit volume. Using this rule, a designer can assess the total embodied CO2e of a concrete column, and then optimize its dimension and composition in order to have less embodied CO2e.

Rule 2. Design rule to calculate total embodied CO2e of a concrete column [49].

$$\text{Column(?C)} \wedge \text{Volume(?C,?CV)} \wedge \text{hasConcrete(?C,?Con)} \wedge \text{Concrete(?Con)} \wedge \text{EmbodiedCO2e(?Con,?ECO2)} \wedge \text{swrlb:multiply(?TECO2,?CV,?ECO2)} \rightarrow \text{TotalECO2e(?C,?TECO2)}$$

6.2.3. Building Information Modeling (BIM)

As semantic model is not a data model that simply stores information but represents the domain knowledge and demonstrates use cases. BIM is a special and standard version of a semantic model used in the AEC industry, and it contains spatial relationships, quantities, properties, and time and cost information of building components as well as 3D geometry. In particular, BIM is a shared knowledge representation about a facility for forming a reliable decision basis during the life cycle of the facility.

Since HPB design should consider the total cost of ownership from very early design phases, the number of HPB design studies mentioning BIM is increasing. For instance, [44] proposed a BIM-based process model for design check and energy performance evaluation (Figure 4). By following this interoperability process, all stakeholders related to HPB conceptual design can understand their roles and exchange artifacts and information to identify all possible design alternatives.

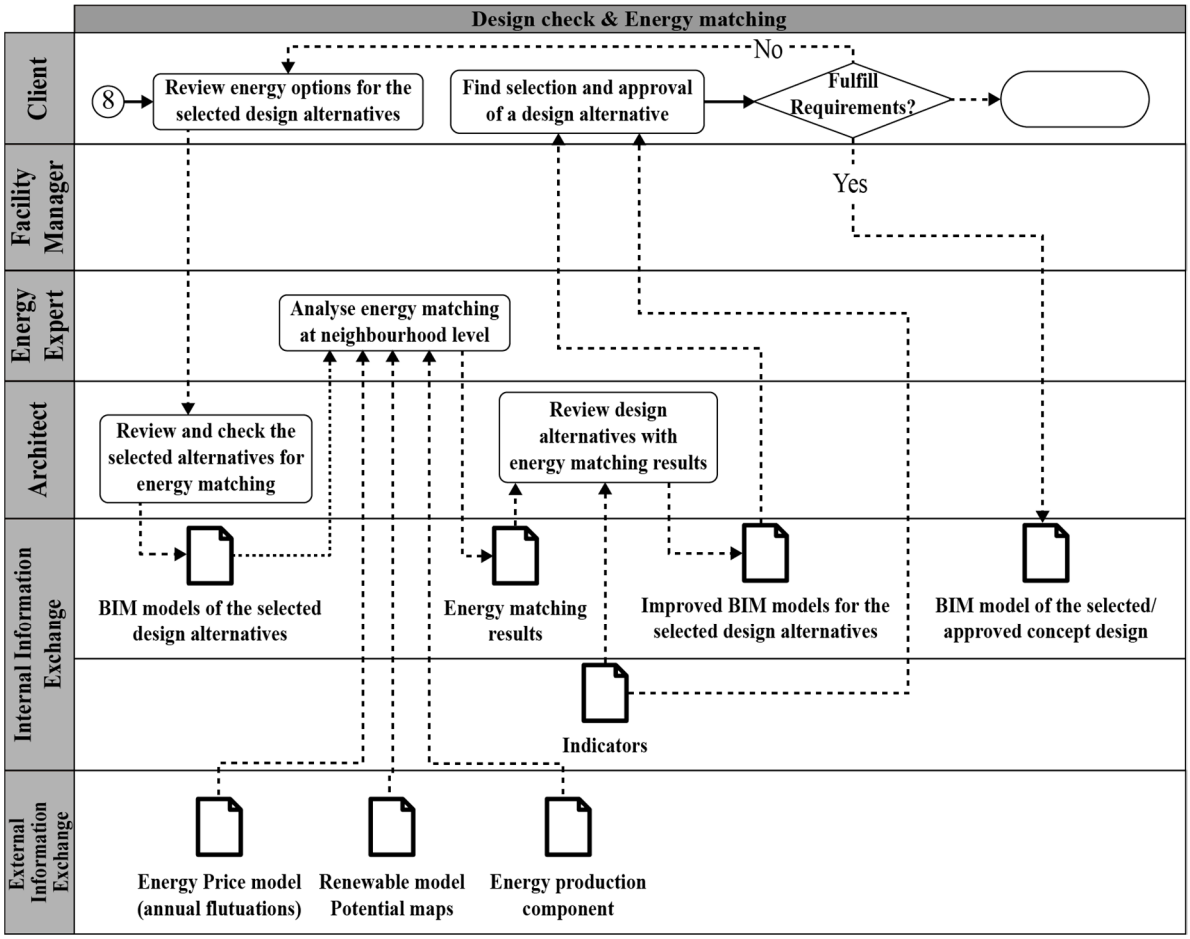


Figure 4. Process map of design check and energy benchmark at concept design stage (modified from [44]).

6.3. Knowledge Representation by Analysis Model and Simulation

HPB decision-makers employ performance analysis to quantitatively evaluate an environmental impact of their alternatives and predict the performance of their design. This is because a building analysis and performance model (Table 8) represents the building physics in the built environment, systems, surrounding environment, and occupant reactions and responses. Thus, the principal mechanism and dynamics of a building performance model should be modeled at an adequate level with the appropriate type of mathematical equations and expressions under multiple dimensions, but assuming nominal semantics between model entities.

Table 8. Knowledge representation by analysis model and simulation (KR-III).

KR-III-1.	Analysis model	Inverse gray box reduced order model [86], Reduced order model [87], VRF and air handler [88]
KR-III-2.	Building performance simulation	Daylight simulation [28,38], Energy simulation [24,51,52,55], IAQ [29,51]

6.3.1. Analysis Model

The first principles are employed to understand the physical behavior of the built environment and systems, interactions between a building and its surrounding, and the resulting dynamic and steady states of the building, such as the energy balance equation (Equation (2)). As more aspects such as visual comfort and hygrothermal performance are involved in the HPB design, more terms should be included in the first principles, which renders abstract building physics more complex.

While analyses of several performance aspects pertain to practical and popular use cases, or to special and professional use cases that have been packaged and deployed in off-the-shelf simulation tools, numerical analyses reported in the literature have been focused more on new (combined and comprehensive) performance aspects that expand the parameters consideration for making HPB design decisions. See Equation (2)—Energy balance equation for a zone:

$$\begin{aligned} A{\cdot}L{\cdot}\rho{\cdot}C_p{\cdot}\tfrac{dT}{dt} &= \sum Q_{cond} + \sum Q_{conv} + \sum Q_{LWR} + \sum Q_{SWR} + \sum Q_{air\ exchange} \\ &\quad + \sum Q_{internal\ gains} \\ A{\cdot}L{\cdot}\rho{\cdot}C_p{\cdot}\tfrac{dT}{dt} &= \tfrac{k}{L}{\cdot}A{\cdot}\Delta T + h_{rad}{\cdot}A{\cdot}F{\cdot}\Delta T + \sum Q_{SWR} + \dot{V}{\cdot}\rho{\cdot}C_p{\cdot}\Delta T \\ &\quad + \sum Q_{internal\ gains} \end{aligned} \quad (2)$$

6.3.2. Building Performance Simulation

Analysis models of building and systems have been effective in proving the concept of new building technologies in the research field. As HPB design is more focused on the comprehensive performance of the entire building, designers have to analyze multiple performance aspects of the building concurrently. Accordingly, analysis tools that were primarily used for research have been packaged and released as standalone performance simulations for enabling general users to make design decisions. The users of the simulation tools are architects and practitioner engineers rather than researchers. The popularity of off-the-shelf simulation tools and platforms have relieved architects and practitioner engineers of the burden of manual acquisition of appropriate and reasonable model parameters by using built-in defaults and libraries.

The main contribution of off-the-shelf simulation tools to knowledge formulation is that they directly offer condensed knowledge about libraries (material, construction, profiles, etc.), design defaults, the standard simulation process and its analysis, and boundary conditions. Consequently, designers, as general users, can concentrate on evaluating the performance of their alternatives instead of modeling, validating, and troubleshooting.

6.4. Knowledge Representation by Machine Learning Algorithms

Often, the underlying trends of and effects of one parameter on another parameter are empirically obtained through parametric simulations. These analysis insights can serve as design information, and they are eventually accumulated and refined as HPB design knowledge.

Even without an analysis and simulation model based on first principles, data-driven models based on raw data can be formulated using machine learning algorithms. Machine learning algorithms are largely divided into supervisory learning, unsupervisory learning, and reinforced learning. These algorithms "learn" sensible or latent relations between features or between inputs and outputs, and they then construct a derived model that contains knowledge about the problem space. The derived model can serve as a surrogate for an analytical model, and it can offer trends and estimations, classify causality relations, and even directly suggest design rules and solutions. The categories of machine learning algorithms listed in Table 9. were major driving forces for the development of data-driven models in previous studies.

Table 9. Knowledge representation by machine learning (KR-IV).

KR-IV-1.	Regression	Regression [25,33,34,36], Multiple regression [89], Support vector regression [90,91]
KR-IV-2.	Classification	SVM [36], Random Forests, AdaBoost Decision Tree, SVM [92]
KR-IV-3.	Neural networks	Neuro fuzzy [32], ANN [36,90,91]
KR-IV-4.	Trees	Random forest [36], Classification and regression tree [20,91,92], Random Forests, AdaBoost Decision Tree [92], Conditional inference tree [14]
KR-IV-5.	Primary factors	Main effects [54], Factor effects [29], Sensitivity analysis [18,25,33,35,93]
KR-IV-6.	Clustering	k-means [16–19,94], KNN [95], k-shape clustering [96], DBSCAN [45]
KR-IV-7.	Pattern and association rule extraction	Text mining [30,31], Pattern matching extraction [27], Bayesian inference [86], Automated regulatory compliance check [97], Sequential pattern mining [98], Association rule learning [45]
KR-IV-8.	Optimization	Multi-criteria optimization [16,37,55], Passive performance optimization [43,99], BIM-based design optimization [45], Expert system for refurbishment [39]
KR-IV-8.	Multi agent systems	Design exploration [53]

6.4.1. Regressions

Regression analysis estimates the relationship between the outcome (i.e., dependent variable) and the input variables (i.e., independent variables, explanatory variables, and features). As supervisory learning algorithms, regressions are used for predicting and forecasting the outcome. This is done by identifying causal relationships between the outcome and a set of input variables by using quantitative terms. As the causal relationships are presented in an intuitive manner, unlike other black box algorithms, regressions are often very useful for designers to choose what design features they should initially focus on. For instance, the energy consumption of a building in an early design stage can be approximated using the number of stories (*NS*), floor area (*FA*), form ratio (*FR*), window-to-wall ratio (*WWR*), coefficient of performance (*COP*), and energy efficiency ratio (*EER*), as shown in Equation (3) (multivariable regression of energy consumption by early phase building variables [34]).

$$\begin{aligned} EC = A + \beta_1 * 1/FR + \beta_2 * 1/FR + \beta_3 * e^{(NS)} + \beta_4 * 1/WWR + \beta_5 \\ * \log(COP) + \beta_6 * \log(EER) \end{aligned} \quad (3)$$

6.4.2. Classification

Classification involves selecting the class or category to which an observation (of new data) belongs after the categories are identified by training instances in the raw data. It is a supervisory learning algorithm in that the category for a training instance is adjusted by evaluating its value against the expected outcome. By contrast, clustering is an unsupervisory learning algorithm because categories are determined according to inherent similarity or distance between data.

Classifier refers to a machine learning algorithm or a mathematical function that evaluates the unlabeled instance and then determines its category. Classifiers have a specific set of dynamic rules, which includes an interpretation procedure to handle vague or unknown values, all tailored to the type of inputs being examined [100]. Different types of classifiers correspond to different dynamic rules, and examples are linear classifiers, support vector machines, kernel estimation, neural networks, and decision trees. A classifier analyzes a set of quantifiable properties of the data set, which is called feature or explanatory variable.

Features can be categorical, ordinal, integer, or float, and they eventually decide the best performing classifier type or combination of classifiers. For instance, Figure 5 and Equation (4) illustrate that a Naive Bayesian classifier developed from an offline rule-based repository identifies the system type based on a new event from a building.

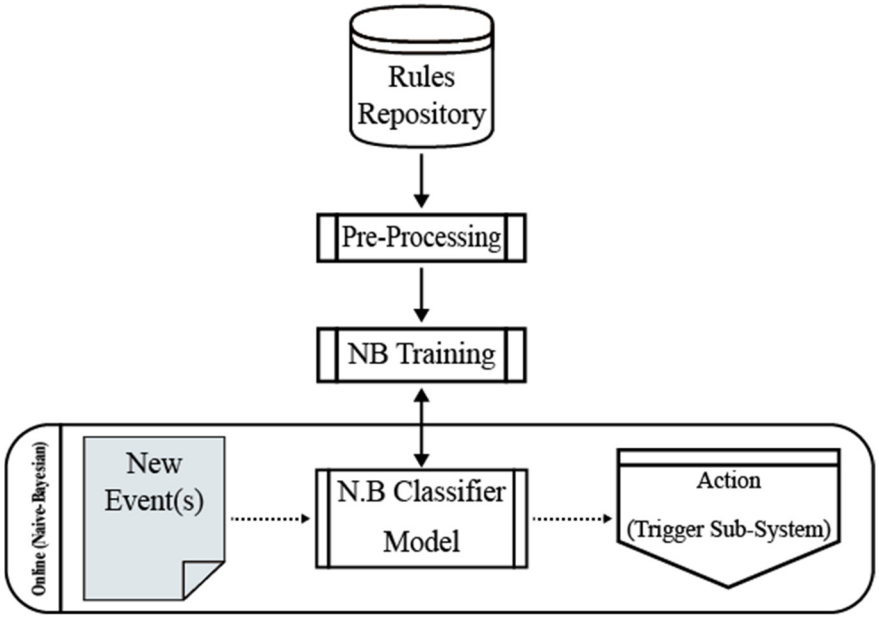


Figure 5. The Naive Bayesian classifier for selecting system type [101]

$$\begin{aligned} \mathrm{P}(\mathrm{Y} = y_k | x_{1\ldots} x_n) &= \frac{P(Y=y_k)\prod_i P(x_i|Y=y_k)}{\sum_j P(Y=y_j)\prod_i P(x_i|Y=y_j)}, \\ \underset{.MAP}{Y} &\leftarrow \operatorname{argmax} \mathrm{P}(\mathrm{Y} = y_k) \prod_i P(x_i|Y = y_k) \end{aligned} \tag{4}$$

6.4.3. Neural Networks

Basically, a neural network is a predictor, but can function a classifier depending on the type of target output. Thus, a neural network involves supervisory learning, and it attempts to minimize the averaged error between the network's output and the desired output. Neural networks are known to show high capability in capturing the nonlinear behavior of a system since hidden layers between the input and output layers, which are composed of multiple perceptrons, dissect the system into multiple dimensions, and their weights then connect those dimensions. Although overfitting can be a drawback of neural networks, theories and measures to handle overfitting have been intensively applied.

For instance, energy demands of buildings with different form factors, azimuth angles, transparency ratios, and insulation thicknesses have been trained and their meta-relations were then formulated into an ANN model (Figure 6). A designer can estimate the energy demand of alternatives without running load analysis or a simulation. However, unlike a regression model, the designer has no means to understand what design variables among the trained variables have the largest impact on the demand, without trials-and-errors.

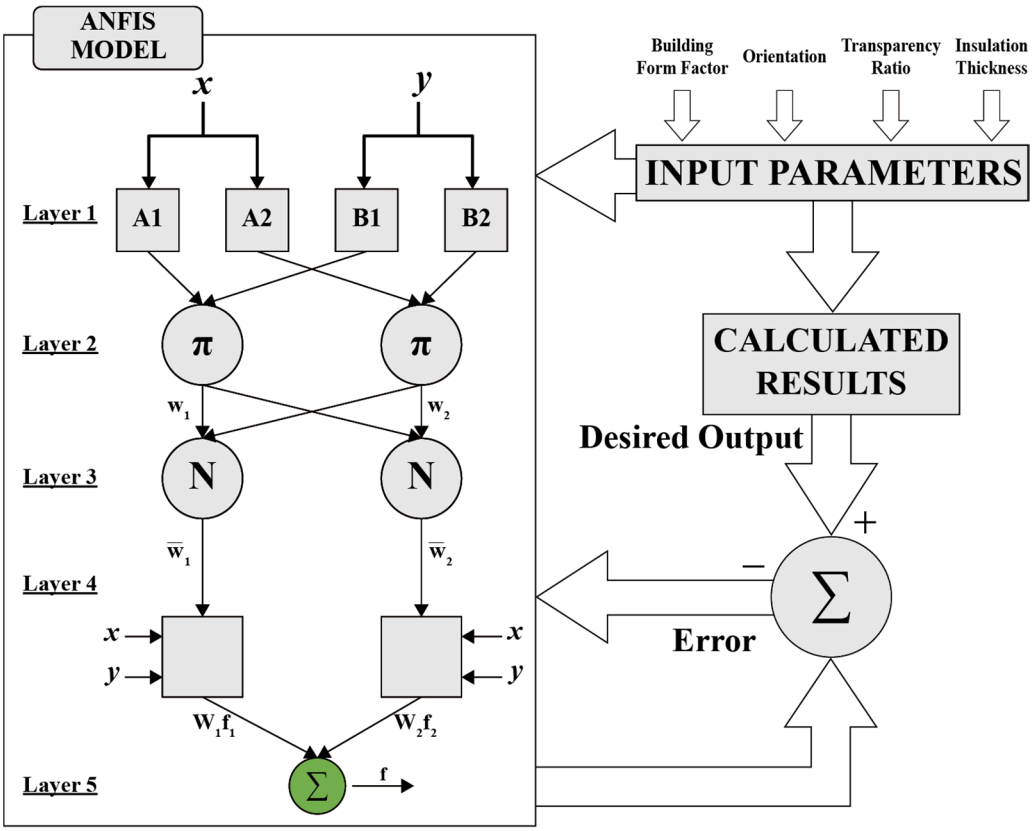


Figure 6. Energy demand model by ANN (modified from [32]).

6.4.4. Trees

A decision tree is a tree-shaped model in which classification rules and possible consequences are encoded. Thus, it represents the major and directional conditional causality and cascading inferences of the problem space.

As decision trees are transparent, it is simple to understand and interpret the knowledge that they represent. Furthermore, because they display all the possible decision scenarios, users can compare possible decisions from the worst to the best cases, which gives them confidence in their decisions as well as warns of the responsibility for a wrong decision. However, the instability and low accuracy of decision tree have made users resort to other classifier and predictor algorithms. Ensemble trees such as random forest have enhanced the classification and prediction performance of single trees, but with little transparency.

For instance, Figure 7 depicts a range of energy use per combination of building features using a classification and regression tree. The superiority of decision trees compared with other machine learning algorithms for making predictions results from their transparency, and a designer can make decisions in serial upon one single choice while still watching the big picture overview.

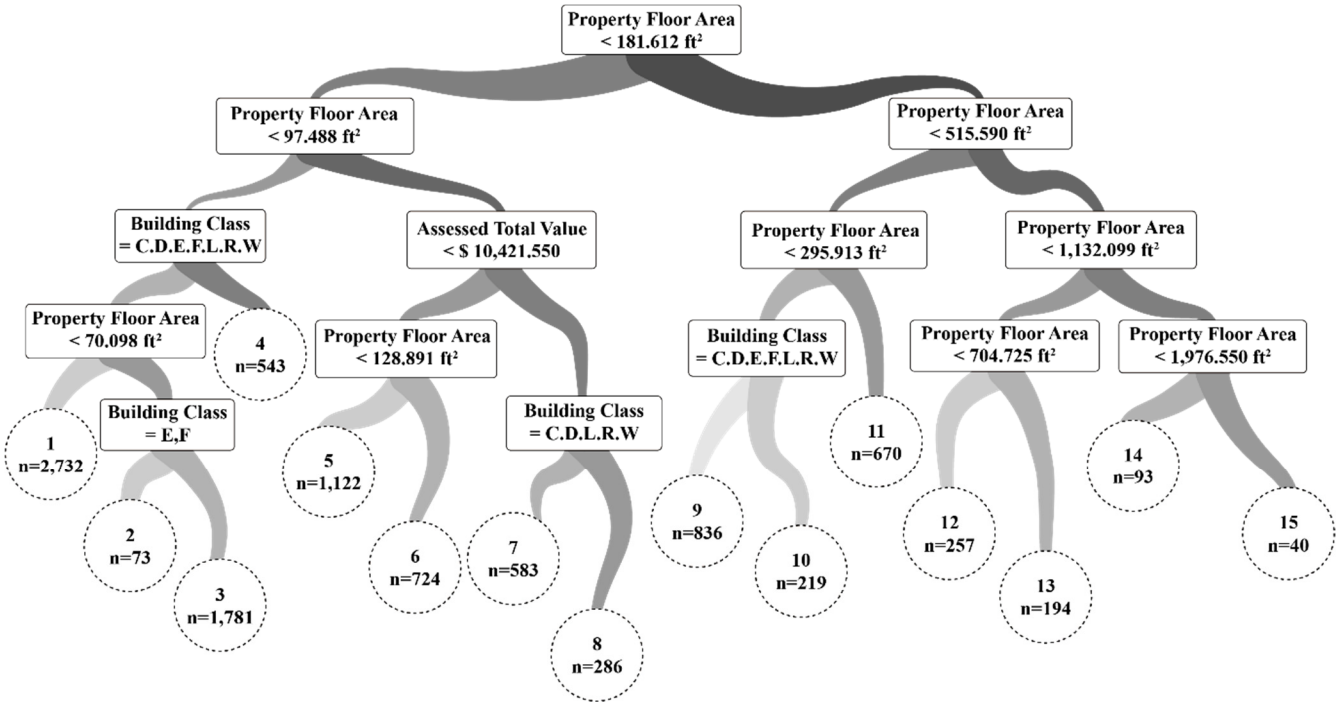


Figure 7. Classification and regression tree for total energy use of building feature groups (modified from [20]).

6.4.5. Primary Factors

Although any property or attribute of a problem can be a feature, actually, feature refers to a determinant characteristic of the problem as long as it is useful to understand the problem space. Feature engineering therefore aims to extract explanatory variables and their significant range according to the analysis purpose from raw data by using data mining, which is referred to as feature learning.

Often it is not mathematically and computationally convenient to extract useful features from convoluted real-world data. In this case, feature leaning can be used for explorative examinations, not by explicit algorithms. The features extracted from the real-world data are then used for dimension reduction of the problem space. Otherwise, they can be used to modify the heuristics to discern the major behavior of the problem space.

Types of feature learnings observed in the literature can be largely divided into sensitivity analysis and feature projection. Feature projection captures some structures underlying the high-dimensional input data and then transforms (or inflates) them into synthesized or reconstructed data of fewer dimensions. Sensitivity analysis is basically uncertainty analysis to identify how uncertainty in the output of a mathematical "model" can be divided and allocated to different sources of the uncertainty in its inputs [102]. It is different from feature projection in that sensitivity analysis tests the robustness of the established model to understand significant uncertainty, while feature projection mainly performs dimension reduction, which is in fact for constructing a more efficient model. Sensitivity analysis provides more domain-oriented knowledge concerning the problem space, such as a lineup of the most sensitive variables in a given uncertainty range. For instance, design parameters that influence the energy use, over-temperature penalty, and daylight performance can be chosen by sensitivity analysis (Figure 8). The designer can then adjust the value of the design parameters to fit the desired performance range.

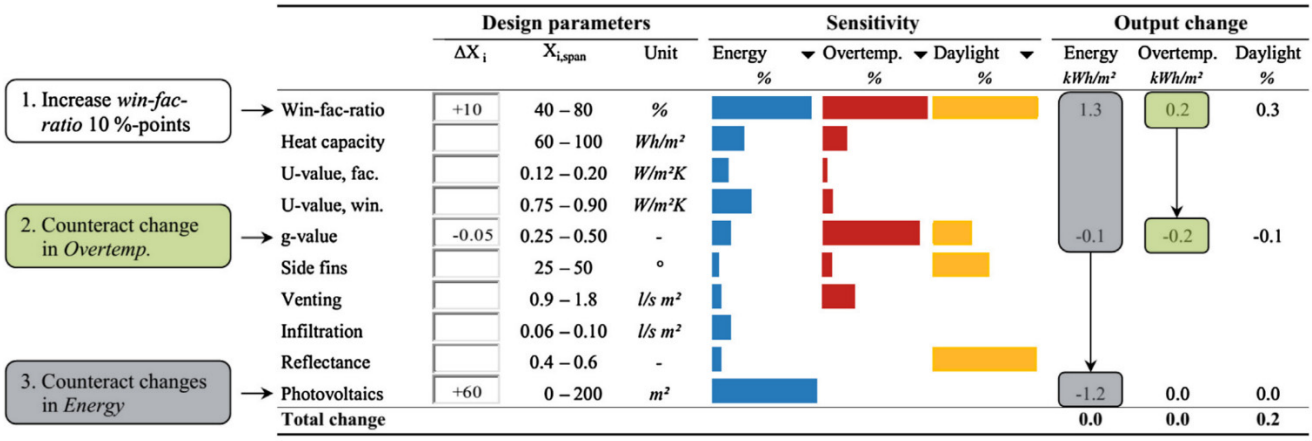


Figure 8. Sensitivity of design parameters over energy, thermal, and daylight performances [25].

6.4.6. Clustering

Clustering is the process of grouping a set of data with similar properties, such that one cluster should be differentiated from another cluster in terms of characteristics. It is a commonly employed statistical analysis technique that is called exploratory data mining. In other words, clustering is a front-end analysis to extract knowledge when not much information is available about the raw data of interest.

While classification employs predefined classes annotated with the target variable, clustering only identifies similarities between data and differences between clusters. Accordingly, a reasonable set of clusters for the raw data in the problem space provides a good base to take a divide-and-conquer approach, in which classifications can be performed for each cluster. Properties to determine clusters include distance between a cluster's member data, density of the data space, intervals between data or particular statistical distributions.

Typically, clustering is performed iteratively, and it involves trials and failures. Therefore, it is often required to remix, rearrange, and relabel the raw data, or remove outliers (i.e., preprocessing) until a satisfactory degree of clusters is achieved. Owing to this exploratory nature of clustering, different clustering algorithms should be repeatedly tested for a particular set of raw data, and clustering performed by tested algorithms should be evaluated and compared.

For instance, energy signatures of 56 office buildings prior- and post-retrofits were collected and divided into three clusters using k-means (Figure 9). Deb and Lee [18] identified that the gross floor area, non-air-conditioning energy consumption, average chiller plant efficiency, and installed capacity of chillers could be the best classification criteria (using the clustering algorithm) to allow the retrofit strategy to be varied per cluster.

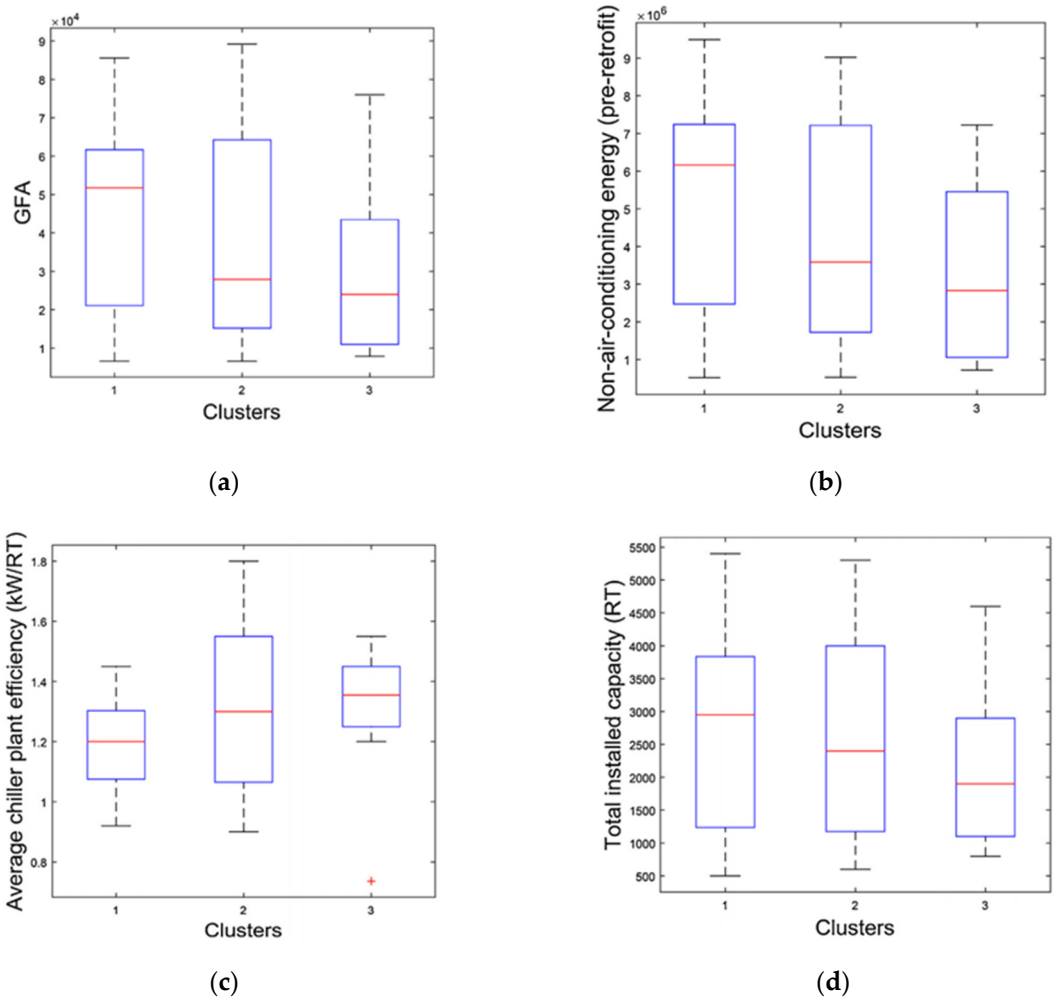


Figure 9. Building features that provide the best result for retrofit measure clustering (modified from [18]). (**a**) GFA; (**b**) Non-air-conditioning energy (pre-retrofit); (**c**) Average chiller plant efficiency (kW/RT); (**d**) Total installed capacity (RT).

6.4.7. Pattern and Association Rule Extraction

As indicated above, tangible rules that are able to describe mechanics of the problem space and system behavior are declarative and procedural knowledge. As long as governing rules can be heuristically extracted by subject matter experts and domain experts, quality and effectiveness of the heuristic rules should be dependent on individual expert's subjective experience and perspective. As acquisition of useful knowledge from the right experts is one of the hardest and the most tough processes for constructing a knowledge base, it also takes extraordinary long time for experts to build practical and refined domain knowledge.

Use of machine "learning" algorithms to mine useful information and purified knowledge out of big data can expedite the knowledge construction process, and also effectively filter out useful knowledge for solving a specific problem. Most of all, association rule mining is a machine learning algorithm for identifying strong rules that implicitly govern the problem space. It is an unsupervisory learning because it does not assume any structure, and also raw data are not annotated or labeled. In many cases, therefore, association rule mining algorithms limit minimum support and confidence thresholds to avoid the case that strong but statistically less meaningful rules are captured.

Figure 10 illustrates the sequence of extracting compliance check rule from energy conservation codes using text mining algorithm [27]. The automated rule extraction would

ease designer's manual effort in filtering out the requirements from regulatory documents and in annotating regulatory document.

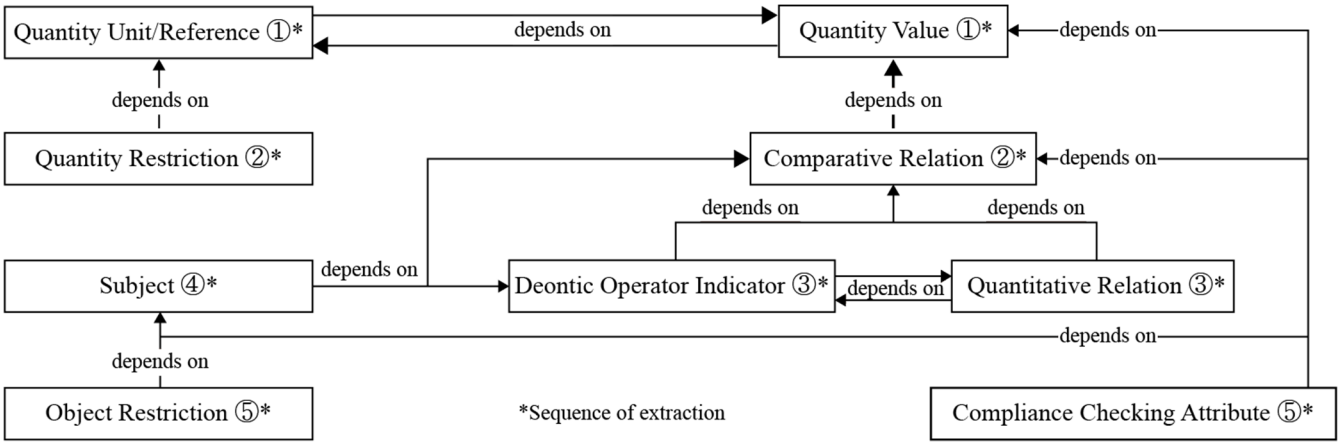


Figure 10. Sequence of compliance check rule extraction from energy conservation codes (modified from [27]).

6.4.8. Optimization

Design optimization involves selecting the optimal design by employing a mathematical description of the design problem along with mathematical optimization algorithms. The mathematical formulation of a design optimization problem is to search for a set of optimal model inputs that minimizes (or maximizes) the objective function for given constraints and ranges of variables; the objective function is the outcome performance of a parametric model such as energy use intensity (EUI).

In fact, the conventional manual process of searching for the optimal solution is not different from this automated mathematical programming; a designer evaluates multiple design alternatives, compares their expected performance and selection risks, and finally orders the most optimal design values. However, the designer cannot obtain a sufficient number of alternatives in a limited time and with limited resources. Consequently, manually chosen optimal designs tend to be another design that has been altered from a previous design under similar circumstances and context. By contrast, design optimization can evaluate more alternatives much faster.

Based on a building performance simulation that represents knowledge about physical behavior and interactions between the built environment and systems, the model formulation and evaluation framework for mathematical programming represent the HPB design knowledge about the desired and feasible level of optimality for designer. For instance, while energy cost as part of operation cost can be calculated from energy use resulted from simulations that are already defined by the legacy design knowledge, designer should define his/her own refurbishment cost calculation based on assumptions and approximations of the construction cost, financial cost add-ons, risk cost, and other costs. Then the optimization framework seeks for the global optimal cost comparing operation cost and refurbishment cost (Figure 11).

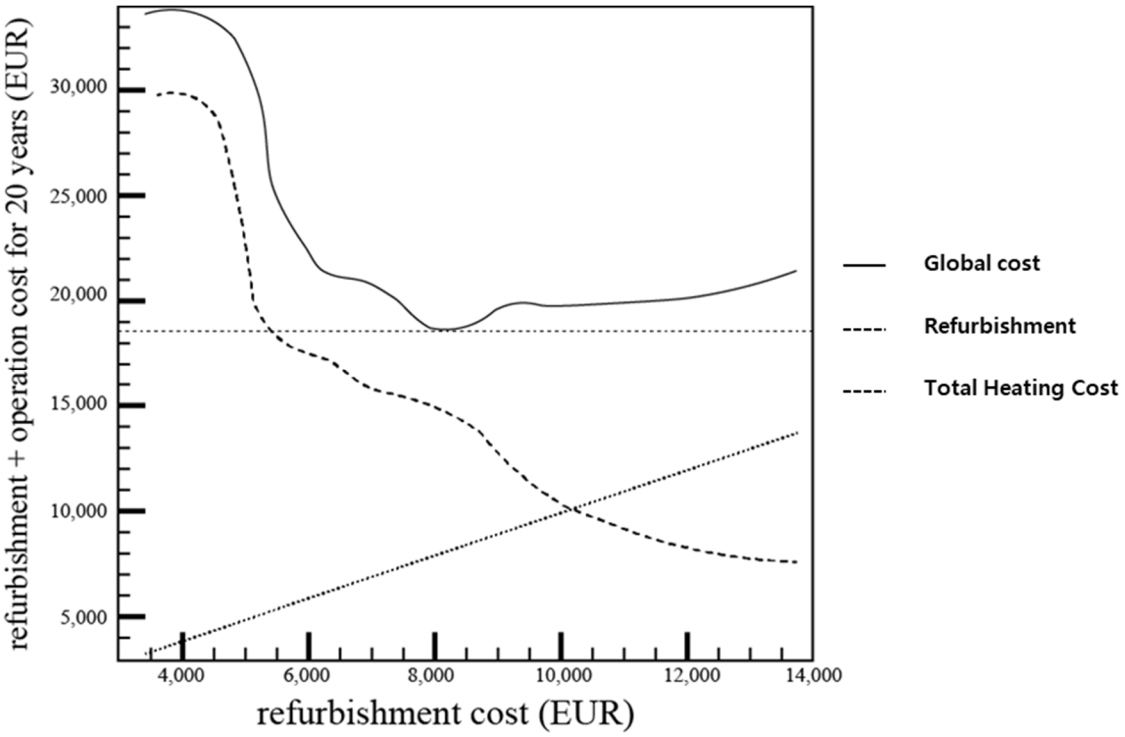


Figure 11. Cost optimization of refurbishment alternatives (modified from [39]).

6.4.9. Multi Agent Systems

When a mathematical model is not available although the problem space is known, multiple agents are able to find out explanatory insights about the problem space, and then interpret them into their collective action. Thus they behave closely to physical phenomena of the physically constrained world.

Moreover, when only simulated model of the problem space is given, but both analytical and numerical explorations are not feasible due to complexity and scale of the space in a limited timeline, multiple agents are able to collect state and dynamics within each one's boundary, collection of which becomes a surrogate of the simulation model.

Multiple agents try to solve given problem at their best although some may be successful and others are not. Compared to optimization, multi agent systems can prevent propagation of faults, self-recover and be fault tolerant in searching a solution, mainly due to the redundancy of components [103], thereby more agents can be dispatched to more vulnerable spots. Consequently, multiple agent system can represent the real-world scenario with a cheaper cost.

For instance, [53] introduced the MAS to explore generative façade designs that meet multiple performance requirement as well as to implement non-deterministic design process where the precise definition of local rules can be combined with analytical tools (Figure 12). Eventually this automated design process has assisted architects and their collaborators in generating higher performing design solutions with unique outlooks.

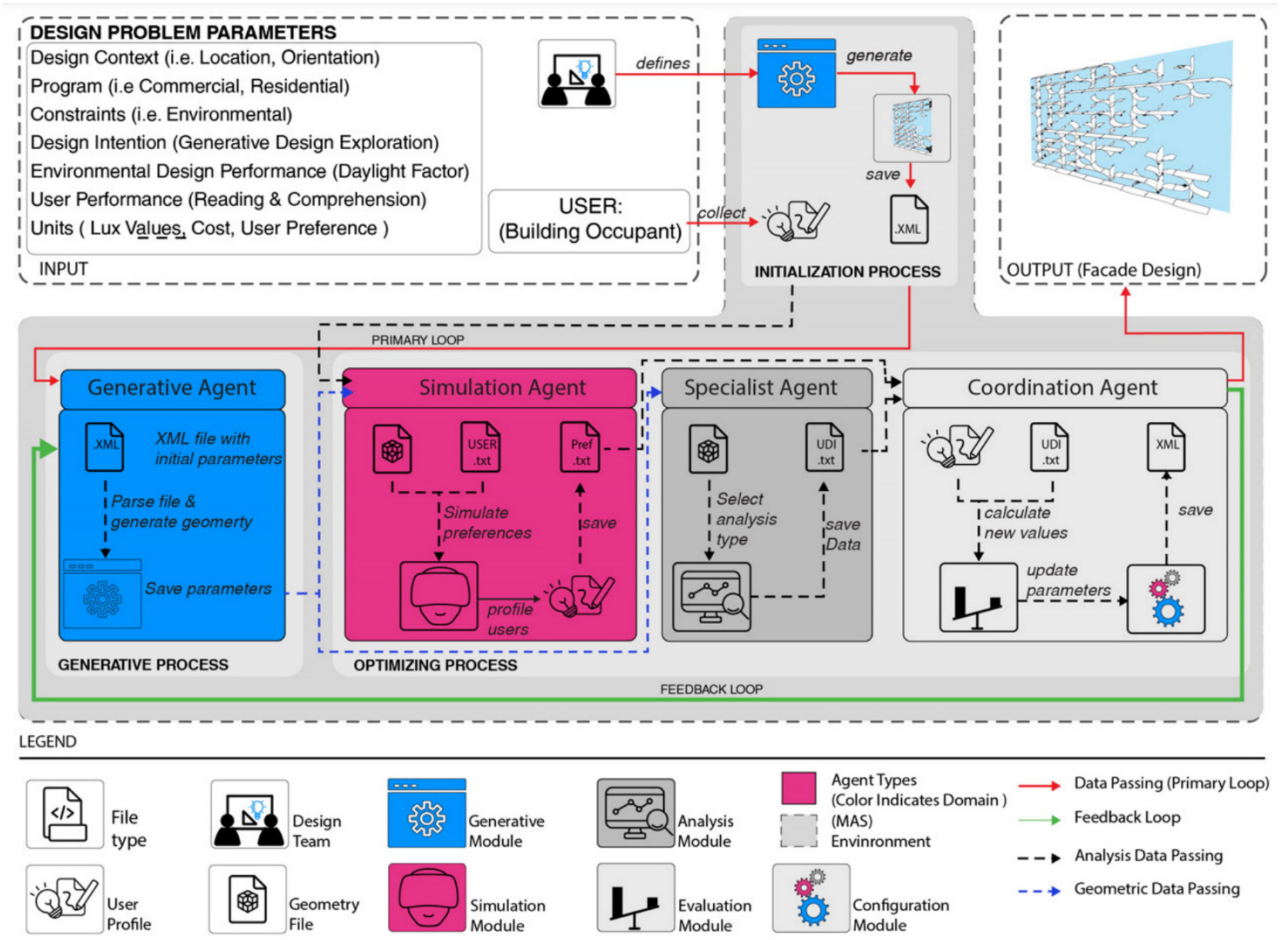


Figure 12. Façade geometry generation by MAS [53].

6.5. Knowledge Representation by Case-Based Reasoning

Case-based reasoning is the process of solving a new problem by referring to solutions of similar past problems, the same way that human reasoning is based on past cases and training. Compared with rule induction algorithms based on accumulated comprehensive cases, case-based reasoning makes local generalization with several episodes, instead of global generalization. However, it is not a rare situation in practice, as shown in Table 10, that data are too scarce for making a statistical inference; for such situations, case-based reasoning can provide a practical knowledge base built in a short term with a limited data set.

For instance, the green building experience-mining model [31] classified previous green building solutions according to criteria such as type, construction, and applied technologies, as depicted in Figure 13. This model enables the designer to easily investigate previous solutions and select the most appropriate benchmark.

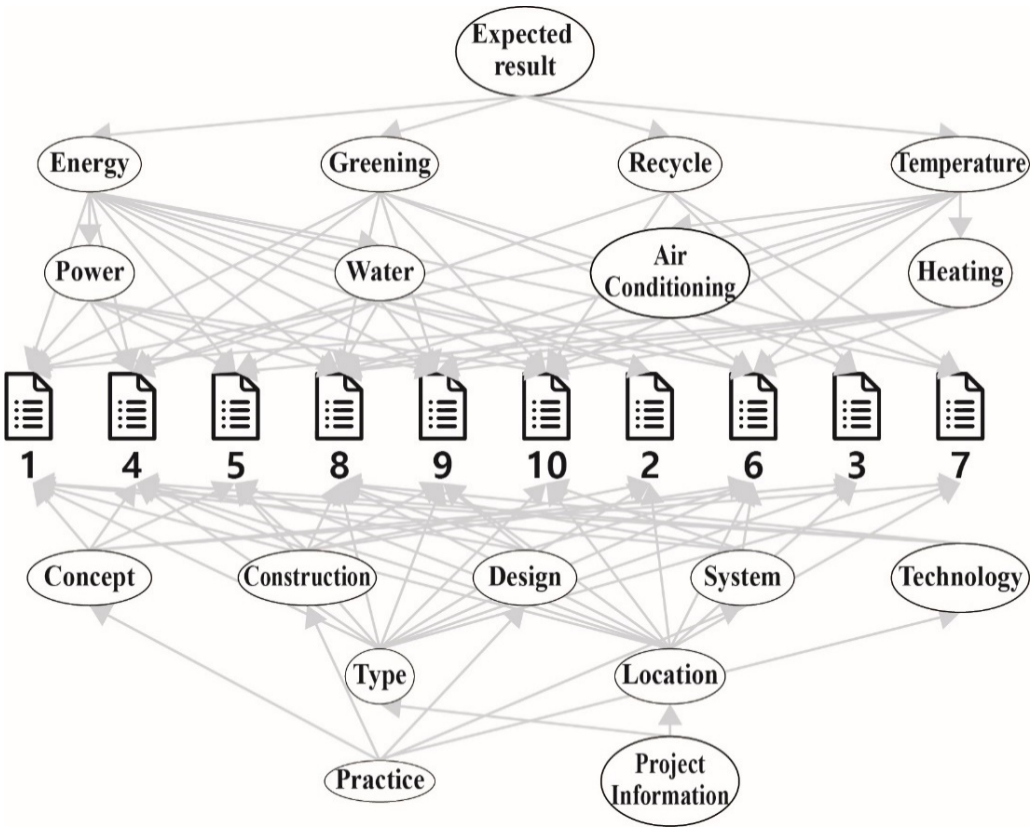


Figure 13. Case storage structure of the green building experience-mining model (modified from [31]).

Table 10. Knowledge representation by case-based reasoning (KR-V) and group decision making (KR-VI).

KR-V.	Case-based reasoning	Case representation and retrieval [16,30,31,104], Case based reasoning with expert system [97], Knowledge-based BIM [105]
KR-VI.	Group decision support system	AHP [57,104], Choosing by advantage [16], Score-driven decision support [65], Multi-criteria decision making [106–108]

6.6. Knowledge Representation by Group Decision Support

Among stake holders, making a decision on the basis of the collective intelligence of experts is perceived to be more comprehensive and less risky than a decision made by a single individual. Although it is true in most cases, when there is not enough time for deliberation, discussion, and dialogue, when there is not enough information to share and communicate, when each member does not have clear criteria or skills to vote for the preference, and/or when influencers of a group may dominate over others who can make a meaningful contribution, group decision can be less efficient and effective, and even biased.

A group decision support system is a structured and computer-based framework to assist a group of stakeholders in considering various courses of thinking and reasoning in a short timeframe, to reduce human-oriented errors and to formalize a priori experiences into quantified decision criteria.

For instance, data of green building features can be collected from experts in the field via pair-wise questionnaires. Then, the most important parameters of rating systems chosen on the basis of their subjective weights can be selected using the analytical hierarchy process (AHP), (Figure 14). Subsequently, fuzzy rules (i.e., performance evaluation knowledge) are used to quantify the performance level for each design case.

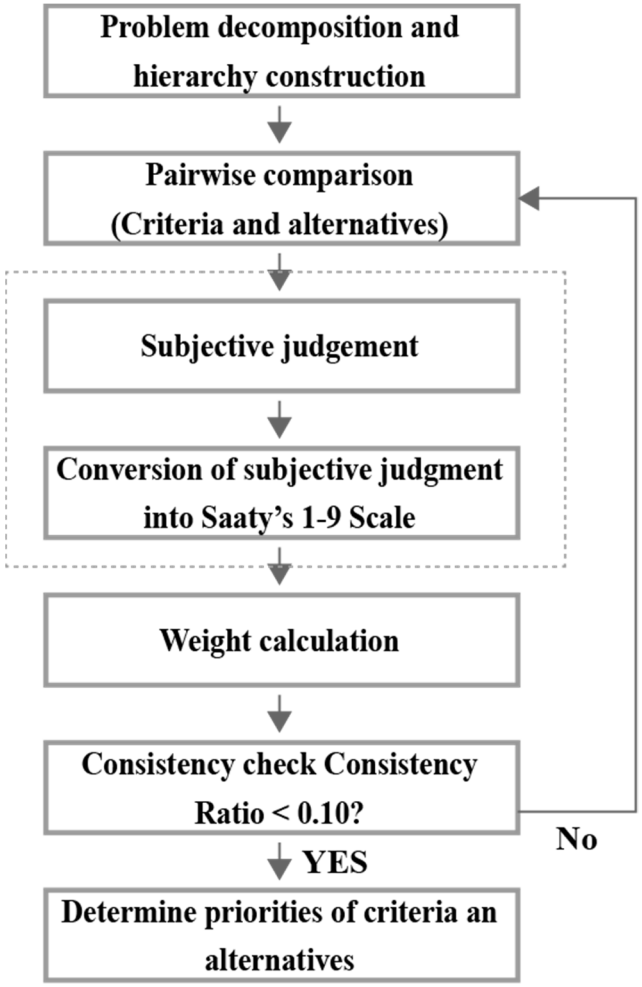


Figure 14. AHP process for assessing the performance level of green buildings [41].

7. Summary of HPB Design Knowledge Acquisition and Representation

The findings from literature review regarding HPB design knowledge acquisition (KA) are summarized below.

1. Traditional methods that enable knowledge acquisition through strategies such as interviews and surveys with experts are still widely used. Passive knowledge acquisition that extracts portable knowledge by discovering and analyzing knowledge from documented facts is also considerably used. However, studies that provide detailed explanations of the methods of extracting knowledge from dialogue with experts or their opinions or from design knowledge from document-formatted facts are still scarce.
2. There is a growing number of studies that extract knowledge from digital media, analyze big data, or employ computer-analyzed data as a knowledge acquisition source. However, the main actors who collect raw data, select data, and analysis methods, and conduct analysis are still domain experts or agents who should be verified by domain experts. Thus, the subjective opinions and heuristics of analyzers must be involved even in computer-aided acquisition.
3. Computer-aided acquisition has the advantage of enabling the extraction of portable knowledge from a relatively objective knowledge source that has wider viewpoints. In most cases, the perspective of domain experts can be further expanded by computer-aided acquisition and, in some cases, even the bias initially possessed by domain experts can be corrected. Thus, it is more reasonable to perform computer-aided

knowledge engineering directly by domain experts or to verify the (machine) extracted knowledge by domain experts.

The findings from literature review regarding HPB design knowledge representation (KR) are summarized below.

1. There were rules, frames, and a semantic network as in the conventional knowledge representation formalism of the expert system, and the knowledge representation types shown in the HPB design references were also originated from them.
2. Several studies employ KR-I, because the rules and logic are the methods to represent the primary factors of knowledge the most succinctly and clearly. Although all related knowledge cannot be represented in detail, the primary knowledge such as the governing equation is relatively easy to be acquired and the representation by the rules and logic has no room for misunderstanding or misinterpretation. In addition, the rules and logic are the knowledge representation that can be easily accepted and understood by designers, as the regulations that most affect the building design are the rule formats.
3. However, KR-I cannot fully represent the characteristics and state of the knowledge agent. In addition, it is difficult to systematically represent the knowledge about the activity between the knowledge agents, as well as their temporal and spatial interactions. In particular, there is an evident limitation for representing direct and indirect HPB design knowledge using KR-I for buildings, because the types of entities that are involved with the behavior of a large system and their interaction should be transient and diverse given the significant size of its system—which is the case with buildings. Thus, in recent years, studies have increasingly employed the KR-II semantic model and ontology as the framework of building design knowledge to represent detailed knowledge, such as the interaction between entities or change of state in building entities. Notably, building ontology can employ proprietary ontology specialized for a given problem or standard BIM. The former can only represent matters about a given problem efficiently, with the disadvantage of decreasing compatibility if the problem scope is expanded or changed. The latter has a drawback that the scope of the BIM may have a larger scope and more inappropriate detail than the scope and detail required by the design problem. Thus, it would be more appropriate to select the specific KR-III building performance model specifically for the corresponding performance problem.
4. KR-II and KR-III are rich models that can not only represent fine details, but also provide significant knowledge that can be used in design decision-making if the model is filled with significant input values to some extent. Ultimately, the reliability of the incomplete model's simulation results must be considerably low if the value that is directly related to a given problem is difficult to be acquired and/or if the input value sensitivity is coincidently large when the confidence on the model input value is low. In line with this, an increasing number of studies have reported on the analysis and interpretation methods that are useful to design decision-making, which are required when using KR-II and KR-III.
5. KR-I, II, and III are relatively refined knowledge representation formalisms. However, as the proportion of knowledge acquisition from the database, digital media, and monitoring data increases daily, the ratio of research that extracts design knowledge required from raw data using machine learning algorithms has increased steadily. In particular, HPB design knowledge has been increasingly acquired from data analytics for design insights that extract latent rules, hidden patterns, trends, and metamodels from the processed data, which are prepared by collecting discrete simulations or rule results. While the traditional analysis method comprises the extraction of underlying knowledge, such as governing equation or primary factor using regressions or sensitivity analysis, attempts to implement a data model or data structure of raw

data dynamics using supervised learning algorithms or to identify raw data structures and patterns using unsupervised learning algorithm have recently increased.

6. Although it is important for design problems to represent the problem space and conditions comprehensively and succinctly thereby helping designers to develop the solution efficiently, reaching the optimal solution is ultimately the designer's responsibility, as he/she is the one who employs all the time and effort to select the solution. Thus, studies on automated optimization framework and multi-agent systems which find design solutions under the condition and constraints that designers define, and then directly suggest the solutions have increased rapidly in the past 10 years. In this case, designers select the scope of the design knowledge which is already defined by the certified knowledge provider (e.g., off-the-shelf simulation) and thus focus more on how to implement the outcome out of the selected design knowledge.
7. To acquire significant interpretation via machine learning, a considerable amount of base data must be secured. Otherwise, a realistic alternative can be selecting the final solution through case-based reasoning or group decision-making using a limited number of cases. Thus, platforms where various group decision-making processes are implemented keep being reported in the recent literature.

8. Requirements of KA, KR, and Decision Support for Developing Design Expert System

The expert system should provide a solution to the design problems questioned by users. That is, it provides a simple answer for a simple question and context-driven advice that analyzes multiple scenarios based on conditions and assumptions for a complex problem.

However, "knowledgeable" user would first evaluate the solution provided by the system by referring to the analysis provided by the system along with the solution rather than only accepting the solution provided by the system without questioning. In other words, the expert system should be used to confirm the users' ideas or to induce and persuade users to a different direction by providing a basis for users to judge.

By opening the logic to the user from the analysis criteria and process, as well as the analysis results that lead to the solution, it is important for users to make a final decision on the reasoning and the use of inference procedure not by the system. That is, the process of finding the solution is more important—if decision makers experience finding the solution by themselves or at least witness the decision-making process, they are more confident with the conclusion because they understand how the system arrived at the conclusion. This claim is supported by a recent design study by Brown [109] that designers often prefer flexible design environments that allowed for more creative control and freedom over design options, when compared to an automated optimization approach. Thereby he suggested the further development that seeks to integrate human intuition along with computational feedback and guidance.

The results of the analysis on the trends in the literature regarding HPB designs in the past 10 years have revealed the requirements of KA, KR, and decision support that should be met by the HPB design expert system as followings.

1. Heuristics is still a good source of design knowledge. However, both direct and indirect knowledge regarding design decision-making should be acquired directly from domain experts or tools recommended by the experts. If experts cannot acquire knowledge by themselves, they should at least play the role of quality assurance (QA) for the acquired knowledge.
2. It is difficult to completely acquire all the valid knowledge that solves the design problem solely through the experts' heuristics. Thus, knowledge extracted through data analytics should work as a complementary tool, and the expert system should provide a platform to manage extraction, formation, synthesis, new insertion, and modification of knowledge for the both types of knowledge sources so as to achieve synergy between them. The expert knowledge particularly acquired by data analytics should be managed by open systems to be utilized in similar problems in the future after the knowledge is complete in both representation and formalization. It should

also be transparent and tangible for users to easily understand, select, and apply the knowledge acquired.

3. Primary knowledge extracted from raw data should first be represented in a format as portable and compact as possible so as to be applicable to a wide range of design contexts and situations such as formula and if-then rules, although it may not be sufficiently accurate and precise to a specific configuration of the design problem. It is noteworthy that, because the knowledge base consisting of only the primary knowledge cannot reflect the unique characteristics of a specific problem context, the secondary knowledge needs to be derived by applying the primary knowledge to the context frame after configuring separately the frame that reflects the design context. Thus, the primary knowledge should be a format whose knowledge structure is simple and that can describe knowledge one-dimensionally to be easily incorporated into the context frame such as model.
4. Instead of using a proprietary context frame, a framework such as a publicly validated BIM or certified simulation model should be used even if the scope is larger than a given scope. If a framework is newly built by users, subjective opinions of the user must be involved. Thus, the knowledge implemented by the framework is likely to include potential and constant biases. Thus, it may cause a distortion to the secondary knowledge derived from the framework.
5. Recently many studies proposed the system that enables directly obtaining the design solution such as optimization and the agent-based system, because the resource and time to execute the performance evaluation and analysis of the alternatives are not sufficient even if designers are well experienced in HPB. That is, if designers do not have sufficient expertise on HPB, an expert system should be able to provide HPB design knowledge suitable to the design context as well as directly offer the design alternative like a HPB consultant.
6. However, the purpose of employing HPB consultants by designers even with the cost is to take professional expertise and receive their advice on the latest HPB trend as well as advance the design alternatives in consultation with them. That is, the final decision is made with designers and clients by comparing the advantages, disadvantages, and economic feasibility of each alternative provided by the consultants, rather than blindly accepting the proposed design. Thus, in order for the solution provided by the expert system to be reliable, the expert system should not only provide a design solution directly but also transparently disclose which frame the primary knowledge is used in to reach the solution and how the solution is derived.

9. Conclusions

New functions and requirements of HPB being added daily and several regulations and certification conditions being reinforced steadily make it harder for designers to decide HPB designs alone. Thus, many designers wish to rely on HPB consultants for advice on HPB specifications, latest design trends, and design review conditions, and then they select the design alternatives and decide the final design in consultation with clients. However, as not all projects can afford consultants, designers must invest efforts and financial resources to acquire HPB expertise knowledge by themselves while bearing the risk of decision making alone.

We expect that, in the near future, computer aids or information services such as design expert systems can help designers by providing the role of HPB consultants even if it can be partially. The effectiveness and success or failure of the expert system depend on the credibility and efficacy of the solution provided by the expert system, which must be ultimately affected by the quality, systemic structure, resilience, and applicability of expert knowledge that is handled by the expert system.

Specifications and requirements of design knowledge acquisition and representation should be established before a suitable knowledge representation formalism is selected or developed for the knowledge base and inference engine of the design expert system.

Thereby this study aims to set the problem definition and category required for existing HPB designs, and to find the knowledge acquisition and representation methods that are the most suitable to the design expert system based on the literature review.

An analysis of the HPB design literature on building design expert systems and design decision makings from the past 10 years revealed that the greatest features of knowledge acquisition are the increasing proportion of computer-analyzed data from digital media and big data as the knowledge acquisition source, while the source of expertise in traditional expert systems was limited to expert heuristics. As for knowledge representation, its greatest feature is also the increasing proportion of studies on data structures and models using machine learning algorithms whereas rules, frames and cognitive maps are conventional representation formalisms of traditional expert systems. In particular, studies on extraction of not only literally raw data from observations and measurement but also latent rule, hidden pattern, and trends of discrete processed data, which are the results of simulations or rules, have also increased. The typical representation formalism of discovered design knowledge using data analytics comprises: (i) signatures or features such as primary factor or governing equation; (ii) data structure and ontology that quantifies system dynamics, such as a state machine, network, or graph; and (iii) distributions and profiles, such as pattern and cluster.

There is a clear trend in the literature that designers prefer the method that decision support tools find and propose a solution directly as optimizer or agent systems does. This is due to the lack of resources and time for designers to execute performance evaluation and analysis of alternatives by themselves, even if they have sufficient experience on the HPB. However, because the risk and responsibility for the final design should be taken by designers solely, they are afraid of convenient black box decision making provided by machines.

Indeed, if the cost allows, many designers would like to employ HPB consultants because they can advance the design alternatives based on the advices on HPB design that they receive. That is, the final decision is made with the designer and client by comparing the advantages, disadvantages, and economic feasibility of each alternative provided by the consultants, rather than blindly accepting the proposed design.

We expect that the design expert system should be able to provide HPB design knowledge suitable to the design context as well as provide design alternatives like a HPB consultant, but at a reasonable cost. If the process of using the primary knowledge in which frame to reach the solution and how the solution is derived are "transparently" open to the designers, the solution made by the design expert system will be able to obtain more trust from designers. Moreover, this transparent decision support process would comply with the requirement of a recent design study [109] that designers prefer flexible design environments that give more creative control and freedom over design options, when compared to an automated optimization approach.

Author Contributions: Conceptualization, S.H.K.; investigation, S.Y.C. and S.H.K.; data curation, S.Y.C. and S.H.K.; Writing—Original draft preparation, S.Y.C. and S.H.K.; Writing—Review and editing, S.Y.C. and S.H.K.; supervision, S.H.K.; project administration, S.H.K.; funding acquisition, S.H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Research Program funded by the Seoul National University of Science and Technology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Howbuild. 2021. Available online: https://sketch.howbuild.com (accessed on 3 March 2021).
2. Lutton, L. HYPEREX—A generic expert system to assist architects in the design of routine building types. *Build. Environ.* **1995**, *30*, 165–180. [CrossRef]
3. Flemming, U.; Coyne, R.; Glavin, T.; Rychener, M. *A Generative Expert System for the Design of Building Layouts*; Springer: Berlin/Heidelberg, Germany, 1986; pp. 811–821.
4. Knowledge Representation and Reasoning. Available online: https://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning (accessed on 3 March 2021).
5. Wagner, W.P. Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies. *Expert Syst. Appl.* **2017**, *76*, 85–96. [CrossRef]
6. Mitta, D.A. Formulation of expert system knowledge. *Proc. Hum. Factors Soc. Annu. Meet.* **1989**, *33*, 350. [CrossRef]
7. Adeli, H.; Hawkins, D.W. A hierarchical expert system for design of floors in highrise buildings. *Comput. Struct.* **1991**, *41*, 773–788. [CrossRef]
8. Bédard, C.; Gowri, K. Automating building design process with KBES. *J. Comput. Civ. Eng.* **1990**, *4*, 69–83. [CrossRef]
9. Rosenman, M.A.; Gero, J.S. Design codes as expert systems. *Comput. Aided Des.* **1985**, *17*, 399–409. [CrossRef]
10. Saouma, V.E.; Doshi, S.M.; Pace, M. Architecture of an expert-system-based code-compliance checker. *Eng. Appl. Artif. Intell.* **1989**, *2*, 49–56. [CrossRef]
11. Adeli, H. *Expert Systems in Construction and Structural Engineering*; Chapman & Hall Ltd.: London, UK; New York, NY, USA, 1988.
12. Adelman, L. Measurement issues in knowledge engineering. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 483–488. [CrossRef]
13. Hayter, S.J.; Torcellini, P.A.; Hayter, R.B. The energy design process for designing and constructing high-performance buildings. In Proceedings of the 7th REHVA World Congress and Clima, Naples, Italy, 15–18 September 2001; pp. 15–18.
14. Kim, S.H.; Nam, J. Can both the economic value and energy performance of small- and mid-sized buildings be satisfied? development of a design expert system in the context of Korea. *Sustainability* **2020**, *12*, 4946. [CrossRef]
15. Clancey, W.J. Heuristic classification. *Artif. Intell.* **1985**, *27*, 289–350. [CrossRef]
16. Arroyo, P.; Mourgues, C.; Flager, F.; Correa, M.G. A new method for applying choosing by advantages (CBA) multicriteria decision to a large number of design alternatives. *Energy Build.* **2018**, *167*, 30–37. [CrossRef]
17. Arambula, L.R.; Pernigotto, G.; Cappelletti, F.; Romagnoni, P.; Gasparella, A. Energy audit of schools by means of cluster analysis. *Energy Build.* **2015**, *95*, 160–171. [CrossRef]
18. Deb, C.; Lee, S.E. Determining key variables influencing energy consumption in office buildings through cluster analysis of pre- and post-retrofit building data. *Energy Build.* **2018**, *159*, 228–245. [CrossRef]
19. Papadopoulos, S.; Bonczak, B.; Kontokosta, C.E. Pattern recognition in building energy performance over time using energy benchmarking data. *Appl. Energy* **2018**, *221*, 576–586. [CrossRef]
20. Yang, Z.; Roth, J.; Jain, R.K. Due-B: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis. *Energy Build.* **2018**, *163*, 58–69. [CrossRef]
21. Wei, Y.; Zhang, X.; Shi, Y.; Xia, L.; Pan, S.; Wu, J.; Han, M.; Zhao, X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1027–1047. [CrossRef]
22. Benavente-Peces, C.; Ibadah, N. Buildings energy efficiency analysis and classification using various machine learning technique classifiers. *Energies* **2020**, *13*, 3497. [CrossRef]
23. Xia, B.; Li, X.; Shi, H.; Chen, S.; Chen, J. Style classification and prediction of residential buildings based on machine learning. *J. Asian Archit. Build. Eng.* **2020**, *19*, 714–730. [CrossRef]
24. Attia, S.; Gratia, E.; De Herde, A.; Hensen, J.L.M. Simulation-based decision support tool for early stages of zero-energy building design. *Energy Build.* **2012**, *49*, 2–15. [CrossRef]
25. Østergård, T.; Jensen, R.L.; Maagaard, S.E. Early building design: Informed decision-making by exploring multidimensional design space using sensitivity analysis. *Energy Build.* **2017**, *142*, 8–22. [CrossRef]
26. Pang, Z.; O'neill, Z.; Li, Y.; Niu, F. The role of sensitivity analysis in the building performance analysis: A critical review. *Energy Build.* **2020**, *209*, 109659. [CrossRef]
27. Zhou, P.; El-Gohary, N. Ontology-based automated information extraction from building energy conservation codes. *Autom. Constr.* **2017**, *74*, 103–117. [CrossRef]
28. Gagne, J.M.L.; Andersen, M.; Norford, L.K. An interactive expert system for daylighting design exploration. *Build. Environ.* **2011**, *46*, 2351–2364. [CrossRef]
29. Schlueter, A.; Geyer, P. Linking BIM and design of experiments to balance architectural and technical design factors for energy performance. *Autom. Constr.* **2018**, *86*, 33–43. [CrossRef]
30. Shen, L.; Yan, H.; Fan, H.; Wu, Y.; Zhang, Y. An integrated system of text mining technique and case-based reasoning (TM-CBR) for supporting green building design. *Build. Environ.* **2017**, *124*, 388–401. [CrossRef]
31. Xiao, X.; Skitmore, M.; Hu, X. Case-based reasoning and text mining for green building decision making. *Energy Procedia* **2017**, *111*, 417–425. [CrossRef]
32. Bektas Ekici, B.; Aksoy, U.T. Prediction of building energy needs in early stage of design by using ANFIS. *Expert Syst. Appl.* **2011**, *38*, 5352–5358. [CrossRef]

33. Hester, J.; Gregory, J.; Kirchain, R. Sequential early-design guidance for residential single-family buildings using a probabilistic metamodel of energy consumption. *Energy Build.* **2017**, *134*, 202–211. [CrossRef]
34. Pulido-Arcas, J.A.; Pérez-Fargallo, A.; Rubio-Bellido, C. Multivariable regression analysis to assess energy consumption and CO_2 emissions in the early stages of offices design in Chile. *Energy Build.* **2016**, *133*, 738–753. [CrossRef]
35. Amiri, S.S.; Mottahedi, M.; Asadi, S. Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S. *Energy Build.* **2015**, *109*, 209–216. [CrossRef]
36. Deng, H.; Fannon, D.; Eckelman, M.J. Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy Build.* **2018**, *163*, 34–43. [CrossRef]
37. Rocha, H.; Peretta, I.S.; Lima, G.F.M.; Marques, L.G.; Yamanaka, K. Exterior lighting computer-automated design based on multi-criteria parallel evolutionary algorithm: Optimized designs for illumination quality and energy efficiency. *Expert Syst. Appl.* **2016**, *45*, 208–222. [CrossRef]
38. Andersen, M.; Gagne, J.M.; Kleindienst, S. Interactive expert support for early stage full-year daylighting design: A user's perspective on Lightsolve. *Autom. Constr.* **2013**, *35*, 338–352. [CrossRef]
39. Csík, A. An Expert System for the Cost-Optimal Refurbishment of Buildings. *Adv. Mater. Res.* **2014**, *899*, 599–604. [CrossRef]
40. Min, D.A.; Hyun, K.H.; Kim, S.-J.; Lee, J.-H. A rule-based servicescape design support system from the design patterns of theme parks. Adv. *Eng. Informatics* **2017**, *32*, 77–91. [CrossRef]
41. Nilashi, M.; Zakaria, R.; Ibrahim, O.; Majid, M.Z.A.; Zin, R.M.; Chugtai, M.W.; Abidin, N.I.Z.; Sahamir, S.R.; Yakubu, D.A. A knowledge-based expert system for assessing the performance level of green buildings. *Knowl. Based Syst.* **2015**, *86*, 194–209. [CrossRef]
42. König, M.; Dirnbek, J.; Stankovski, V. Architecture of an open knowledge base for sustainable buildings based on Linked Data technologies. *Autom. Constr.* **2013**, *35*, 542–550. [CrossRef]
43. Konis, K.; Gamas, A.; Kensek, K. Passive performance and building form: An optimization framework for early-stage design support. *Sol. Energy* **2016**, *125*, 161–179. [CrossRef]
44. Arayici, Y.; Fernando, T.; Munoz, V.; Bassanino, M. Interoperability specification development for integrated BIM use in performance based design. *Autom. Constr.* **2018**, *85*, 167–181. [CrossRef]
45. Liu, S.; Meng, X.; Tam, C. Building information modeling based building design optimization for sustainability. *Energy Build.* **2015**, *105*, 139–153. [CrossRef]
46. Hiyama, K.; Kato, S.; Kubota, M.; Zhang, J. A new method for reusing building information models of past projects to optimize the default configuration for performance simulations. *Energy Build.* **2014**, *73*, 83–91. [CrossRef]
47. Ilhan, B.; Yaman, H. Green building assessment tool (GBAT) for integrated BIM-based design decisions. *Autom. Constr.* **2016**, *70*, 26–37. [CrossRef]
48. Pont, U.; Ghiassi, N.; Fenz, S.; Heurix, J.; Mahdavi, A. SEMERGY: Application of semantic web technologies in performance-guided building design optimization. *J. Inf. Technol. Constr.* **2015**, *20*, 107–120.
49. Hou, S.; Li, H.; Rezgui, Y. Ontology-based approach for structural design considering low embodied energy and carbon. *Energy Build.* **2015**, *102*, 75–90. [CrossRef]
50. Corry, E.; Pauwels, P.; Hu, S.; Keane, M.; O'Donnell, J. A performance assessment ontology for the environmental and energy man-agement of buildings. *Autom. Constr.* **2015**, *57*, 249–259. [CrossRef]
51. Petersen, S.; Svendsen, S. Method and simulation program informed decisions in the early stages of building design. *Energy Build.* **2010**, *42*, 1113–1119. [CrossRef]
52. Ochoa, C.E.; Capeluto, I.G. Advice tool for early design stages of intelligent facades based on energy and visual comfort approach. *Energy Build.* **2009**, *41*, 480–488. [CrossRef]
53. Gerber, D.J.; Pantazis, E.; Wang, A. A multi-agent approach for performance based architecture: Design exploring geometry, user, and environmental agencies in façades. *Autom. Constr.* **2017**, *76*, 45–58. [CrossRef]
54. Jaime, M.L.; Marilyne, A. A daylighting knowledge base for performance-driven facade design exploration. *J. Illum. Eng. Soc.* **2011**, *8*, 93–110.
55. Chardon, S.; Brangeon, B.; Bozonnet, E.; Inard, C. Construction cost and energy performance of single family houses: From integrated design to automated optimization. *Autom. Constr.* **2016**, *70*, 1–13. [CrossRef]
56. Oti, A.; Kurul, E.; Cheung, F.; Tah, J. A framework for the utilization of Building Management System data in building information models for building design and operation. *Autom. Constr.* **2016**, *72*, 195–210. [CrossRef]
57. Singhaputtangkul, N.; Low, S.P.; Teo, A.L.; Hwang, B.-G. Knowledge-based Decision Support System Quality Function Deployment (KBDSS-QFD) tool for assessment of building envelopes. *Autom. Constr.* **2013**, *35*, 314–328. [CrossRef]
58. O'Donnell, J.; Corry, E.; Hasan, S.; Keane, M. Building performance optimization using cross-domain scenario modeling, linked data, and complex event processing. *Build. Environ.* **2013**, *62*, 102–111. [CrossRef]
59. Homaei, S.; Hamdy, M. A robustness-based decision making approach for multi-target high performance buildings under uncertain scenarios. *Appl. Energy* **2020**, *267*, 114868. [CrossRef]
60. Korkmaz, S.; Messner, J.I.; Riley, D.R.; Magent, C. High-Performance Green Building Design Process Modeling and Integrated Use of Visualization Tools. *J. Arch. Eng.* **2010**, *16*, 37–45. [CrossRef]
61. Pieter De, W.; Godfried, A.; Marinus Van Der, V. Managing the selection of energy saving features in building design. Eng. *Constr. Archit. Manag.* **2002**, *9*, 192–208.

62. Wilde, P.D.; Voorden, M.V.D. Providing computational support for the selection of energy saving building components. *Energy Build.* **2004**, *36*, 749–758. [CrossRef]
63. Struck, C.; De Wilde, P.J.; Hopfe, C.J.; Hensen, J.L. An investigation of the option space in conceptual building design for advanced building simulation. *Adv. Eng. Inform.* **2009**, *23*, 386–395. [CrossRef]
64. Kanters, J.; Horvat, M.; Dubois, M.-C. Tools and methods used by architects for solar design. *Energy Build.* **2014**, *68*, 721–731. [CrossRef]
65. Ruggeri, A.G.; Calzolari, M.; Scarpa, M.; Gabrielli, L.; Davoli, P. Planning energy retrofit on historic building stocks: A score-driven decision support system. *Energy Build.* **2020**, *224*, 110066. [CrossRef]
66. Zou, P.X.; Wagle, D.; Alam, M. Strategies for minimizing building energy performance gaps between the design intend and the reality. *Energy Build.* **2019**, *191*, 31–41. [CrossRef]
67. Hsu, H.-C.; Chang, S.; Chen, C.-C.; Wu, I.-C. Knowledge-based system for resolving design clashes in building information models. *Autom. Constr.* **2020**, *110*, 103001. [CrossRef]
68. Fan, C.; Xiao, F. Mining Gradual Patterns in Big Building Operational Data for Building Energy Efficiency Enhancement. *Energy Procedia* **2017**, *143*, 119–124. [CrossRef]
69. Pickering, E.M.; Hossain, M.A.; French, R.H.; Abramson, A.R. Building electricity consumption: Data analytics of building operations with classical time series decomposition and case based subsetting. *Energy Build.* **2018**, *177*, 184–196. [CrossRef]
70. Fan, C.; Xiao, F.; Li, Z.; Wang, J. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. *Energy Build.* **2018**, *159*, 296–308. [CrossRef]
71. Granadeiro, V.; Correia, J.R.; Leal, V.M.S.; Duarte, J.P. Envelope-related energy demand: A design indicator of energy per-formance for residential buildings in early design stages. *Energy Build.* **2013**, *61*, 215–223. [CrossRef]
72. Michaels, T.; Leckey, J. Commercial Buildings Energy Consumption Survey. 2012. Available online: https://www.eia.gov/consumption/commercial/ (accessed on 3 March 2021).
73. Davis, R.; Shrobe, H.; Szolovits, P. What is a knowledge representation? *AI Mag.* **1993**, *14*, 17–33.
74. Chen, S.; Zhang, G.; Xia, X.; Setunge, S.; Shi, L. A review of internal and external influencing factors on energy efficiency design of buildings. *Energy Build.* **2020**, *216*, 109944. [CrossRef]
75. Depecker, P.; Menezo, C.; Virgone, J.; Lepers, S. Design of buildings shape and energetic consumption. *Build. Environ.* **2001**, *36*, 627–635. [CrossRef]
76. Grosan, C.; Abraham, A. Rule-Based Expert Systems. In *Contactless Human Activity Analysis*; Ahad, M.A.R., Mahbub, U., Rahman, T., Eds.; Springer Science and Business Media LLC: Berlin, Germany, 2011; Volume 17, pp. 149–185.
77. Fuzzy Logic. Available online: https://en.wikipedia.org/wiki/Fuzzy_logic (accessed on 3 March 2021).
78. Wang, C.; Cho, Y.K. Application of As-built Data in Building Retrofit Decision Making Process. *Procedia Eng.* **2015**, *118*, 902–908. [CrossRef]
79. Geyer, P. Systems modelling for sustainable building design. *Adv. Eng. Inform.* **2012**, *26*, 656–668. [CrossRef]
80. Kim, S.H. Automating building energy system modeling and analysis: An approach based on SysML and model transfor-mations. *Autom. Constr.* **2014**, *41*, 119–138. [CrossRef]
81. Ślusarczyk, G.; Łachwa, A.; Palacz, W.; Strug, B.; Paszyńska, A.; Grabska, E. An extended hierarchical graph-based building model for design and engineering problems. *Autom. Constr.* **2017**, *74*, 95–102. [CrossRef]
82. Solihin, W.; Eastman, C.; Lee, Y.-C.; Yang, D.-H. A simplified relational database schema for transformation of BIM data into a query-efficient and spatially enabled database. *Autom. Constr.* **2017**, *84*, 367–383. [CrossRef]
83. Montali, J.; Sauchelli, M.; Jin, Q.; Overend, M. Knowledge-rich optimisation of prefabricated façades to support conceptual design. *Autom. Constr.* **2019**, *97*, 192–204. [CrossRef]
84. Pauwels, P.; Krijnen, T.; Terkaj, W.; Beetz, J. Enhancing the ifcOWL ontology with an alternative representation for geometric data. *Autom. Constr.* **2017**, *80*, 77–94. [CrossRef]
85. Mousavi, A.; Vyatkin, V. Energy Efficient Agent Function Block: A semantic agent approach to IEC 61499 function blocks in energy efficient building automation systems. *Autom. Constr.* **2015**, *54*, 127–142. [CrossRef]
86. Pavlak, G.S.; Henze, G.P.; Hirsch, A.I.; Florita, A.R.; Dodier, R.H. Experimental verification of an energy consumption signal tool for operational decision support in an office building. *Autom. Constr.* **2016**, *72*, 75–92. [CrossRef]
87. Kim, D.; Bae, Y.; Yun, S.; Braun, J.E. A methodology for generating reduced-order models for large-scale buildings using the Krylov subspace method. *J. Build. Perform. Simul.* **2020**, *13*, 419–429. [CrossRef]
88. Legorburu, G.; Smith, A.D. Incorporating observed data into early design energy models for life cycle cost and carbon emissions analysis of campus buildings. *Energy Build.* **2020**, *224*, 110279. [CrossRef]
89. Catalina, T.; Iordache, V.; Caracaleanu, B. Multiple regression model for fast prediction of the heating energy demand. *Energy Build.* **2013**, *57*, 302–312. [CrossRef]
90. Chou, J.-S.; Bui, D.-K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy Build.* **2014**, *82*, 437–446. [CrossRef]
91. Ngo, N.-T. Early predicting cooling loads for energy-efficient design in office buildings by machine learning. *Energy Build.* **2019**, *182*, 264–273. [CrossRef]
92. Jun, M.; Cheng, J.C. Selection of target LEED credits based on project information and climatic factors using data mining techniques. *Adv. Eng. Inform.* **2017**, *32*, 224–236. [CrossRef]

93. Talami, R.; Jakubiec, J.A. Early-design sensitivity of radiant cooled office buildings in the tropics for building performance. *Energy Build.* **2020**, *223*, 110177. [CrossRef]
94. Miller, C.; Nagy, Z.; Schlueter, A. Automated daily pattern filtering of measured building performance data. *Autom. Constr.* **2015**, *49*, 1–17. [CrossRef]
95. Faia, R.; Pinto, T.; Abrishambaf, O.; Fernandes, F.; Vale, Z.; Corchado, J.M. Case based reasoning with expert system and swarm intelligence to determine energy reduction in buildings energy management. *Energy Build.* **2017**, *155*, 269–281. [CrossRef]
96. Yang, J.; Ning, C.; Deb, C.; Zhang, F.; Cheong, D.; Lee, S.E.; Sekhar, C.; Tham, K.W. k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy Build.* **2017**, *146*, 27–37. [CrossRef]
97. Beach, T.H.; Rezgui, Y.; Li, H.; Kasim, T. A rule-based semantic approach for automated regulatory compliance in the con-struction sector. *Expert Syst. Appl.* **2015**, *42*, 5219–5231. [CrossRef]
98. Yarmohammadi, S.; Pourabolghasem, R.; Shirazi, A.; Ashuri, B. A Sequential Pattern Mining Approach to Extract Information from BIM Design Log Files. In Proceedings of the 33rd International Symposium on Automation and Robotics in Construction (ISARC), International Association for Automation and Robotics in Construction (IAARC), Auburn, AU, USA, 18–21 July 2016; pp. 174–181.
99. Lapisa, R.; Bozonnet, E.; Salagnac, P.; Abadie, M. Optimized design of low-rise commercial buildings under various climates – Energy performance and passive cooling strategies. *Build. Environ.* **2018**, *132*, 83–95. [CrossRef]
100. Statistical Classification. Available online: https://en.wikipedia.org/wiki/Statistical_classification (accessed on 3 March 2021).
101. Shahi, A.; Sulaiman, N.; Mustapha, N.; Perumal, T. Naive Bayesian decision model for the interoperability of heterogeneous systems in an intelligent building environment. *Autom. Constr.* **2015**, *54*, 83–92. [CrossRef]
102. Saltelli, A.; Tarantola, S.; Campolongo, F.; Ratto, M. Global Sensitivity Analysis for Importance Assessment. *Sensit. Anal. Pract.* **2004**, *22*, 31–61. [CrossRef]
103. Multi-Agent System. Available online: https://en.wikipedia.org/wiki/Multi-agent_system (accessed on 3 March 2021).
104. Arroyo, P.; Tommelein, I.D.; Ballard, G. Comparing AHP and CBA as Decision Methods to Resolve the Choosing Problem in Detailed Design. *J. Constr. Eng. Manag.* **2015**, *141*, 04014063. [CrossRef]
105. Motawa, I.; Almarshad, A. A knowledge-based BIM system for building maintenance. *Autom. Constr.* **2013**, *29*, 173–182. [CrossRef]
106. Invidiata, A.; Lavagna, M.; Ghisi, E. Selecting design strategies using multi-criteria decision making to improve the sus-tainability of buildings. *Build. Environ.* **2018**, *139*, 58–68. [CrossRef]
107. Farsäter, K.; Strandberg, P.; Wahlström, Å. Building status obtained before renovating multifamily buildings in Sweden. *J. Build. Eng.* **2019**, *24*, 100723. [CrossRef]
108. Serrano-Jiménez, A.; Femenías, P.; Thuvander, L.; Barrios-Padura, Á. A multi-criteria decision support method towards selecting feasible and sustainable housing renovation strategies. *J. Clean. Prod.* **2021**, *278*, 123588. [CrossRef]
109. Brown, N.C. Design performance and designer preference in an interactive, data-driven conceptual building design scenario. *Des. Stud.* **2020**, *68*, 1–33. [CrossRef]