Energy & Buildings 202 (2019) 109383

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

A review of the applications of artificial intelligence and big data to buildings for energy-efficiency and a comfortable indoor living environment

Muhammad Uzair Mehmood[a], Daye Chun[b], Zeeshan[a], Hyunjoo Han[c], Gyuyeob Jeon[c], Kuan Chen[d,*]

[a] Department of Nuclear and Energy Engineering, Jeju National University, Jeju 63243, South Korea
[b] Waseda University, Tokyo 169-8050, Japan
[c] Department of Architectural Engineering, Jeju National University, Jeju 63243, South Korea
[d] Department of Mechanical Engineering, University of Utah, Salt Lake City, UT 84112, USA

ARTICLE INFO

Article history:
Received 29 May 2019
Revised 14 August 2019
Accepted 23 August 2019
Available online 24 August 2019

Keywords:
Artificial intelligence
Big data
Green buildings
Sustainable architecture

ABSTRACT

After decades of evolution and improvements, Artificial Intelligence (AI) is now taking root in our daily lives, and is starting to profoundly influence the fields of architecture and sustainability. The applications of AI to sustainable architecture include energy-efficient building design, forecasting and minimizing energy consumption, strategizing for mitigating impacts on environment and climate, and enhancements in the safety and comfort of the living environment. Due to the significant increases in internet speed and accessibility and the drops in computer prices and data storage costs in recent years, Big Data (BD) nowadays plays an important supplementary role to AI. Algorithms and computer codes have been developed for data mining and analysis. BD rejuvenates AI methods and applications in many areas, including sustainable architecture. The present paper starts with an introduction to AI history and techniques. This is followed by a discussion on how AI and BD can be used to design and operate energy–efficient commercial buildings and residential houses, followed by a review of recent applications of AI and BD to energy-efficient buildings with an emphasis on the use of machine learning (ML) and large databases. Future research topics are suggested at the end of this paper. It is reemphasized in the present paper that AI, when combined with BD, can tremendously increase the energy efficiency and cost effectiveness of buildings which are designed to provide occupants with a comfortable indoor living environment.

1. AI evolution, classification, and techniques

1.1. The AI evolution

Minsky and McCarthy described AI as "the ability of a machine or a program to perform a task, which would require some kind of intelligence if it was carried out by a human being" [1]. Wang thought that AI could be defined on the basis of structure, behavior, capabilities, function and principles [2]. Nilsson defined AI as the "activity devoted to making machines intelligent, and intelligence is the quality that enables an entity to function appropriately and with foresight in its environment" [3]. Lacking a precise and universally accepted definition might in fact help the advancement of the field of AI. Some of the capabilities of AI systems that can be associated with human intelligence are problem solving, knowledge representation, reasoning, learning, and to some lesser extent, social intelligence, and creativity.

The concept of AI is based on the assumption that the human thought process can be mechanized. Even before the industrial era, speculations of AI could be seen in different civilizations. However, its first practical application was seen during World War II. Alan Turing, a noted British mathematician and computer scientist, and his teammates created the Bombe machine to decipher the Enigma code, leading to the foundation of ML (Machine Learning).

In 1956 at Dartmouth College, the term "Artificial Intelligence" was coined for the first time by American computer scientist John McCarthy, and it was formally accepted as an academic discipline [4]. During the early years of AI evolution, the programs developed by AI were quite astonishing. Its applications at that time included solving algebra word problems, proving theorems in geometry, and learning to speak English. In the same time period, AI research had received much acclamation and funding from government agencies [5]. It was around that time when the world's first full-scale

* Corresponding author.
E-mail address: chen@mech.utah.edu (K. Chen).

https://doi.org/10.1016/j.enbuild.2019.109383

Nomenclature

AEL	Allowable Exposure Limit
AGI	Artificial General Intelligence
AHP	Analytic Hierarchy process
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning
ATES	Aquifier Thermal Energy Storage
BD	Big Data
BIM	Building Information Model
BMS	Building Management System
BREEAM	Building Research Establishment Environmental Assessment Method
BTDFs	Bidirectional Transmittance Distribution Functions
CBECS	Commercial Building Energy Consumption Survey
CFS	Complex Fenestration System
CHP	Combined Heat and Power
CRBM	Conditional Restricted Boltzmann Machine
CVA	Canonical Variate Analysis
DL	Deep Learning
DOE	Department of Energy
DPN	Deep Belief Network
EFB	Energy Efficient Building
EIS	Energy Information System
ELM	Extreme Learning Machine
ESB	Empire State Building
FCRBM	Factored Conditional Restricted Boltzmann Machine
FL	Fuzzy Logic
GA	Genetic Algorithm
GHG	Greenhouse Gas
GIS	Geographic Information System
GOFAI	Good Old Fashioned Artificial Intelligence
GP	Gaussian Process
IT	Information Technology
LCOE	Levelized Cost of Electricity
LEED	Leadership in Energy and Environmental Design
LiDAR	Light Detection and Ranging
MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Splines
ML	Machine Learning
MPC	Model Predictive Control
NN	Neural Network
NREL	National Renewable Energy Laboratory
NSRD	National Solar Radiation Database
PV	Photovoltaic
RBFN	Radial Basis Function Network
RNN	Recurrent Neural Network
SA	Simulated Annealing
SBRS	Sustainable Building Rating System
SVR	Support Vector Regression
WABOT	WAseda roBOT
XCON	eXpert CONfigurer

intelligent robot was made in the WABOT project in Japan [6]. However, AI is a difficult field and researchers in earlier times failed to appreciate its complexity. Their optimism raised the expectations about AI too high, and the expected advancements failed to materialize. In the 1970s, AI faced several critiques and financial setbacks, resulting in significant reduction in research interest and funding [5]. This was known as the first AI winter.

AI research was brought back to life by the adoption of the Expert Systems during the early 1980s. In the beginning, they were a huge success. In one case, an expert case called XCON saved 40 million dollars annually in 1986. However, this system eventually became too expensive to maintain. The main disadvantages of Expert Systems include the complications in updating the systems and their brittleness. This was the start of the second AI winter. Yet, after this period, the field of AI didn't completely disappear. Instead, it continued to grow under aliases such as computational intelligence, cognitive systems, etc. In the last decade, the modern age of AI has finally arrived in which the ever-improving ability of fast computation, especially the power of parallel computing, has given birth to new subfields such as Deep Learning (DL), Machine Learning (ML), Artificial General Intelligence (AGI), and Big Data (BD).

1.2. Classification of AI

AI can be categorized either on the basis of functionality, or on the basis of their evolution with time. Based on its functionality, AI can be split into narrow AI and general AI. Narrow AI can be seen in many daily practices because of its vast number of applications which continue to grow every day. Some of these tasks are organizing calendars-both personal and business, responding to customer queries, carrying out visual inspections of different infrastructures, flagging different kinds of inappropriate contents online, and so on. On the other hand, general AI (or general artificial intelligence) is very different and is very close to the type of intelligence found in humans. Such a system, after learning and accumulating experience, can perform tasks that require more precision. Up until now, general AI could only be seen in movies, and AI experts are trying to make this dream a reality soon.

Another way to categorize AI schemes is based on the revolution of AI. The earliest progression in AI since its origin is described as the old fashion of AI, or in other words, Good-Old-Fashioned Artificial Intelligence (GOFAI). GOFAI was a paradigm that was given birth by the Dartmouth conference in 1956. Mostly based on hand-coded symbolic manipulations, GOFAI was very similar to what modern programming looks like. It is also known as symbolism for its attempt to describe intelligence in symbolic terms [7]. Another of its common euphemisms is Symbolic AI. The purpose of Symbolic AI was to produce human-like intelligence in a machine, whereas most modern research in the field of AI focuses on specific sub-problems.

The most successful form of Symbolic AI is the Expert Systems. In artificial intelligence, an expert system is a computer system that matches or surpasses the decision making ability of a human expert [8]. Expert Systems were introduced in 1965 [9], and were among the first truly successful forms of AI software [10].The goal of these kinds of systems was to make the information required for the system to work explicitly rather than implicitly. In other words, the rules guiding the operation of the Expert Systems were to be defined in a format that was intuitive and easy to understand. Therefore, it can be edited even by the domain experts and not just IT (Information Technology) experts. The main advantages of this explicit representation were rapid development and ease of maintenance.

One of the major disadvantages of the Expert Systems is the knowledge acquisition problem. Other problems include the difficulty of integration, access to large databases, and performance issues [11]. Various researchers gave different reasons as to why GOFAI couldn't stand the test of time. The main argument of these critiques stated that the representations used in early GOFAI weren't meaningful representations of the real world [12]. Another argument stated that the emphasis on formal logics and deductive reasoning ignored other methods of reasoning.

The modern subcategories in the field of AI include ML (machine learning) and DL (deep learning). Although the terms of AI, ML and DL are used interchangeably in the literature, there is a clear distinction among these three. DL is a subset of ML, which is in turn a subset (or branch) of AI. One aspect that distinguishes ML from traditional knowledge graphs and expert systems is its ability to learn and improve from provided data on its own. ML can even create its own algorithms. Arthur Samuel, one of the pioneers of ML, defined it as "A field of study that gives computers that ability to learn without being explicitly programmed" [13]. DL is a further subset of ML, and generally it refers to deep artificial neural networks and sometimes deep reinforcement learning. In DL, computers learn those techniques that come naturally to humans. This is the key technology behind voice recognition, image recognition, driverless cars, and so on.

1.3. AI techniques

A variety of AI techniques have been developed and used in many different areas. These techniques are introduced and briefly discussed below in alphabetical order.

1.3.1. AHP

One popular decision-making scheme is the Analytic Hierarchy Process that was introduced by Saaty [14]. The merits of the scheme include its good mathematical properties, the easiness of obtaining the required input data, and the fact that it is a decision-making tool which can solve complex problems. The hierarchical structure of the AHP consists of criteria, sub criteria, objectives and alternatives.

In the AHP, the complex problems on hand are first broken down into component parts that are arranged into multiple hierarchical levels. In the next step, the decision makers compare each cluster in a pairwise fashion based on their experience and knowledge. Some degree of inconsistency may occur as the comparisons are performed through personal judgments.

The biggest advantage of the AHP scheme is its final operation called the consistency verification. Once the comparisons have been carried out at all the levels of the hierarchy and proved by the consistency verification, an overall priority ranking is developed which is based on each attribute's priority and its corresponding criterion priority. Fig. 1 shows the structure of a standard AHP scheme.

1.3.2. FL

Being one of the easiest to understand among the various AI techniques, the merits of fuzzy logic (FL) are numerous. It is based on natural language, and the merits include its capability to be built on top of the experience of experts, ability for modeling non-linear functions, tolerance of imprecise data, and finally, its capability to blend with conventional control techniques. The three main steps of a Fuzzy system are: Fuzzification, Interference and Defuzzificaion [15].

1.3.3. GA

Genetic algorithms are techniques based on the concept of survival of the fittest. The possible solutions to a problem are called "people", and their evolution in time is observed. The genetic algorithm utilizes three principal operators: selection, crossover and mutation [16].

A fitness function assesses the capability of each individual to solve the problem, and this is carried out after every generation in the reproduction process. The proportion to which any individual is reproduced is directly related to its fitness. In other words, the more suitable an individual or a solution is, the higher are its chances to be carried over to the next generation by its offspring. To find the potential answer to an issue, a GA performs specific reproduction of people. Following the principle of survival of the fittest, only the populace with higher chances of survival proceeds to the next generation. Genetic operators play the role of forming new and better offspring. The algorithm continues up to a specific number of generations unless an optimal solution to the given problem has been found. Fig. 2 shows the structure of a standard GA.

1.3.4. Neural network (NN)

An NN (or Artificial Neural Network) is a computer system or algorithm that is composed of artificial neurons that can be used to model the human brain and nervous system. The first trainable NN called the Perceptron was demonstrated by Frank Rosenblatt at Cornell in 1957. In an NN, weight factors are often used in the modeling of the relationship between an input and an output vector. During training, the error of the model is calculated and the weights are updated using stochastic gradient descent or other schemes to minimize the model error [17]. An RNN is an extension to the multi-layered perceptron NNs [18–21]. It models temporal dependencies in target data through feedback connections [17]. NN techniques are better at modeling multivariable problems compared to other models. In a multivariable problem, variables exhibit complex relationships among themselves. NN techniques can extract nonlinear relationships from the variables by learning from the training data. The most successful application of a NN has been in the forecasting of electric load in buildings.

1.3.5. Simulated annealing (SA)

Simulated Annealing (SA) is another AI technique for search or optimization. It is a heuristic search method originally developed by chemists and metallurgists to calculate the most stable state of a chemical system [22,23]. Unlike most iterative-improvement methods, SA allows less favorable solutions to be accepted in the search process, thus avoiding the search being trapped in local minima. Chen and Lin [23] applied both GA and SA for the optimization of multiple-module thermoelectric coolers, and found that SA required less computational time and effort.

All the above AI techniques have their strengths and weaknesses. GA is better at handling dynamic problems when compared with AHP or FL. However, fuzzy technique is useful for non-linear systems as it is better at dealing with uncertainty and subjectivity. The AHP, being simple and systematic, can be very useful for decision support systems. SA can avoid the search being trapped in local minima. NNs are commonly used in ML and DL recently.

2. Big data and artificial intelligence

Big Data (BD) refers to huge amounts of data, both structured and unstructured, that can be used to extract useful information. Owing to several reasons such as an exponential increase in global data creation, a massive increase in the computing power that is available for data processing, and advancements in the quantitative analysis and mathematical algorithms that are being used in AI and IT, BD has been used to create a synergy with AI in recent years that can help realize goals in both business and research [24].

BD was first characterized by three variables, namely volume, velocity, and variety, in 2001 [25]. Since then, many more variables have been added to the definition of BD [26]. Recently, the characterization involves a total of 8 Vs; volume, velocity, value, variety, variability, virality, viscosity, and veracity [27].

BD is usually too large or complex to be handled by traditional data processing software and algorithms. At present, the amount of data being created is huge; 2.5 quintillion bytes of data is created every day [28]. AI methods such as DL also utilize huge amounts of data. However, there are key differences between AI and BD. BD does not act on results to make decisions, it only tries to find accurate results. AI, on the other hand, is all about decision

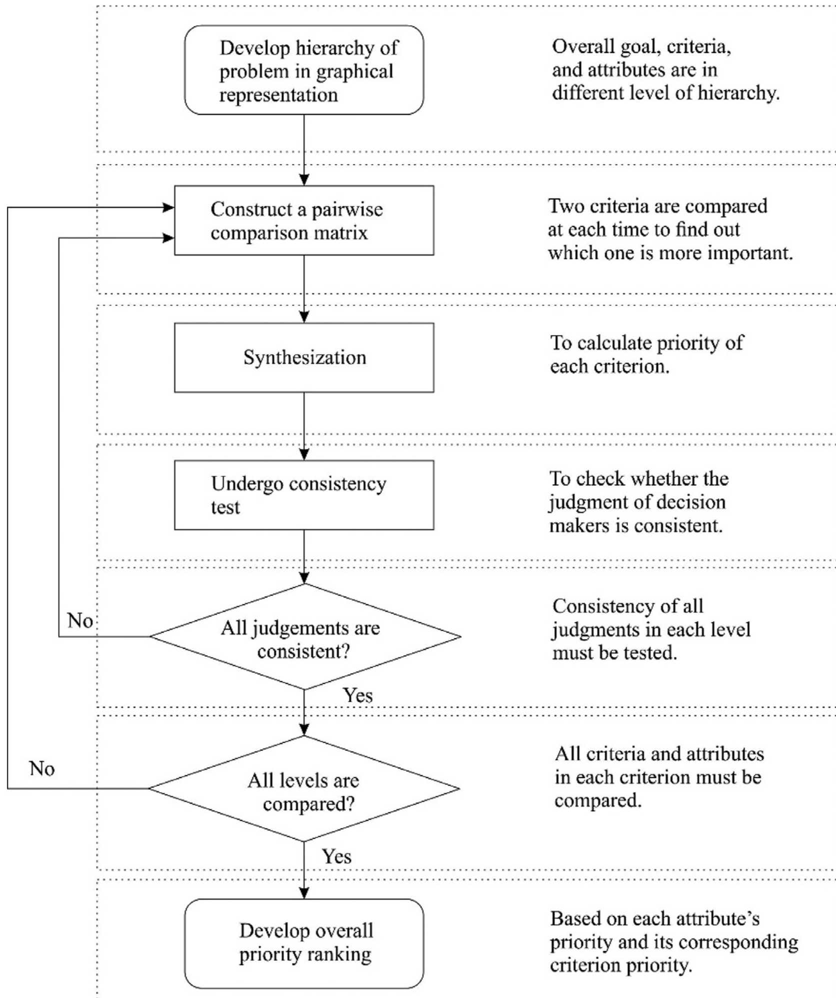

Fig. 1. AHP scheme flowchart [40].

making and learning to make better decisions. BD is primarily about gaining insight, while the aim of AI is trying to accomplish tasks previously done by humans but with increased efficiency and reduced errors. Even with the opportunities and vast amount of applications in different fields including biology, finance, education, and eCommerce, the algorithms initially designed on the models of computation are no longer valid for BD. Similarly, BD is sometimes confused with IoT. Both BD and IoT are concerned with the collection of data. However, there are key differences such as the sources of the data, the time constraints and what it revolves around. In the current context, the main focus is on the use of BD and AI to improve energy efficiency and comfort in buildings.

The BD markets are structured into three layers: the infrastructure layer which consists of hardware components, the data organization and management layer which consists of software components, and finally the services layer which deals with the applications of BD [29]. Fig. 3 shows an overview of the three layers of BD.

3. Applications of AI and BD to energy-efficient buildings

The schematic in Fig. 4 shows how AI and BD can be applied to the design and operation of energy–efficient buildings (EFBs). The AI platform may consist of many fast computer processors and various AI algorithms, computer codes, and simulation packages. Also included in the AI platform are monitors and software for data display and visualization, such as EIS.

After analyzing the relevant data mined from BD and/or received real-time from the sensors built in the EFB, the AI platform selects one or more AI techniques and algorithms to operate individual building components to achieve the optimal operating conditions. The BD for EFBs typically includes local and global weather data, building codes and historical data, databases developed by societies, research groups, and commercial companies (e.g., ASHRAE, LEED, DOE, Google) related to energy research and sustainable architecture.

To take full advantage of AI, the heating/cooling/illumination systems of a building should be very energy-efficient and versatile, and the functions and properties of the major components of an EFB should be highly adjustable. In recent years, 95%-efficient gas furnaces have gained popularity, and micro CHP (Combined Heat and Power) systems have been commercially available for quite some time [30]. Most of today's new buildings use built-in temperature, light, humidity, and motion-detection sensors to minimize energy consumption without compromising the quality of the indoor environment.

The EFB which works best with AI should include a building envelope of variable direction of heat flow as well as variable specific heat and surface absorptivity, windows and skylines whose opacity is adjustable, a solar energy collection and utilization unit for maximum solar energy harness and optimal usage of the entire solar spectrum, and generators and heat transfer devices (e.g., heat exchangers, thermosyphones and thermal diodes) to utilize local

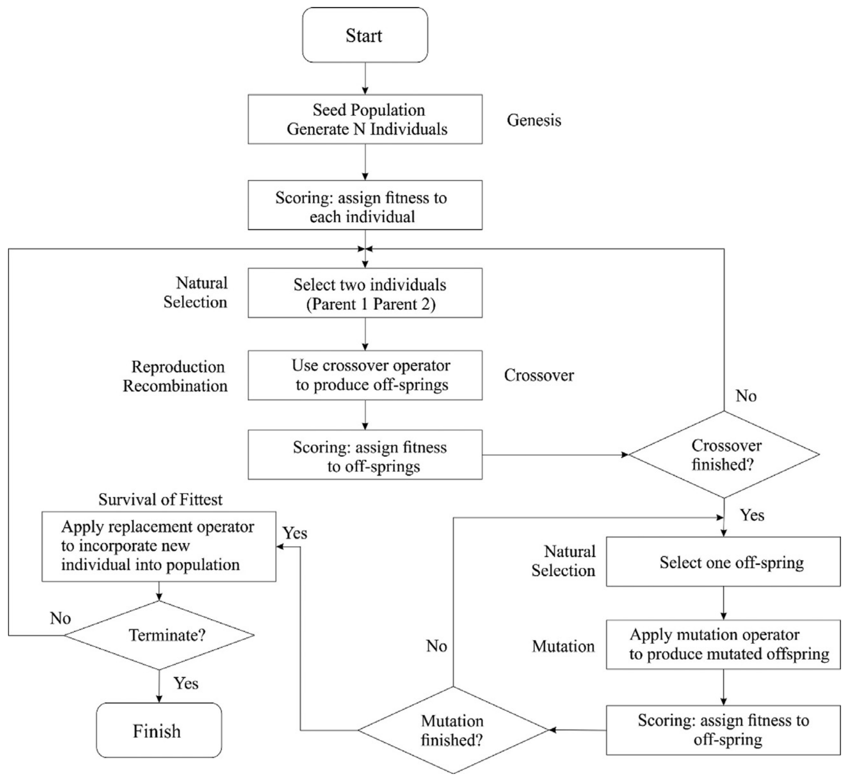

Fig. 2. Genetic algorithm flowchart [40].

Fig. 3. Three layers of big data [29].

wind, geothermal, and other energy sources to reduce building energy consumption.

The AI platform can work alone to operate and control the major components of EFBs using instantaneous and in-situ data measured by built-in sensors. However, it is more energy-efficient and effective to use BD to decide when and how to turn on and adjust the functions and features of individual building components. For example, the glass opacity or the shade and shutters of a smart window, or the movement of an awning can be controlled by a light sensor which continuously monitors the light incident on the building's surface. In this case, AI just uses a simple built-in algorithm to check the light intensity and decide if it is necessary to turn lights on and how to adjust the actuators of the smart window or the awning. The better way for the AI platform to operate

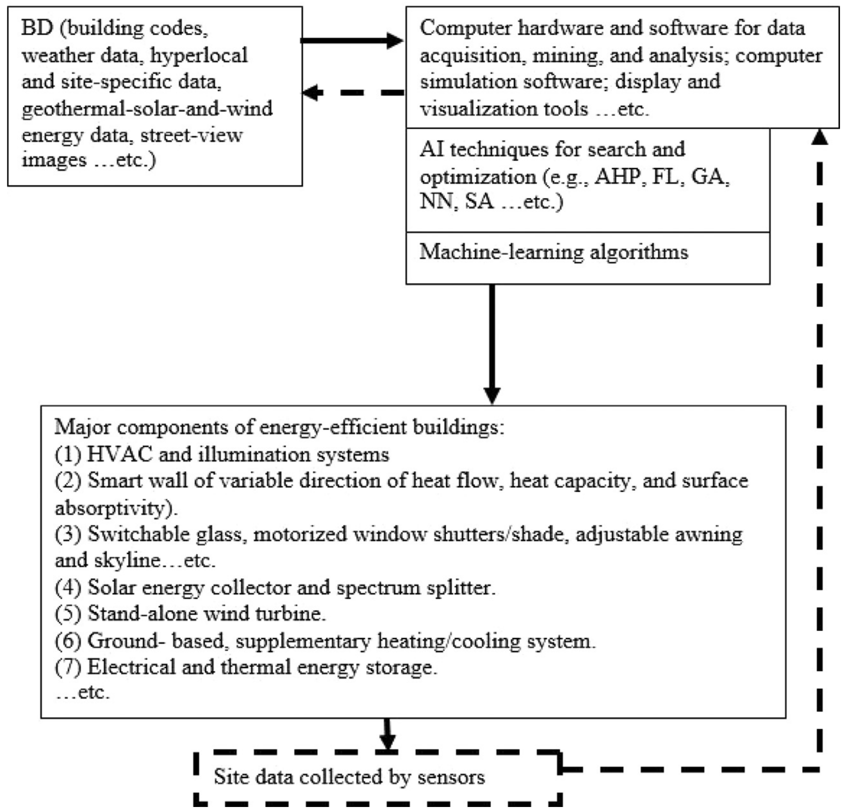

Fig. 4. Workflow schematic of AI and BD for energy efficient buildings.

smart windows and awnings is to use BD to turn on the sensors only occasionally, or don't use any sensors at all if the database has accumulated enough site-specific data to adjust the smart window and the awning autonomously.

Another example of the use of AI and BD for EFBs is the "smart wall". When Chen, Chun, and co-workers [30–32] first tested the energy-efficient construction modules they invented, they proposed to use crude weather data and/or sensors mounted on one module to control the directions of heat flow, specific heats, and surface absorptivities of all the modules that form the building envelope. The direction of heat flow only needs to be alternated a couple of times a year (when the bi-directional thermal diode in the smart construction module is switched from winter-heating to summer-cooling mode, or the reverse). The module's surface absorptivity must be changed about twice daily (from highly reflective during the daytime to highly absorptive at night for summer cooling, or from highly reflective at night to highly absorptive during daytime for winter heating). The adjustment of the specific heat is more complicated, as it depends on the heat input to the wall, the cooling/heating demand, and other factors. A robust AI technique is needed to determine the optimal adjustment of the specific heat of the module. However, when AI and BD work together and the BD has collected enough site-specific data, the AI platform can learn the patterns of daily and annual weather changes, heat transfer to and from building surfaces, building occupancy, and heating/cooling demands to adjust the specific heat without the need to frequently check the heating/cooling conditions and the demands and the occupancy of the building.

The problem of solar heating in the summer can be alleviated by using a thermodiode system. The bi-directional thermodiodes developed by Chun, Chen, and co-workers [31–33] can reduce solar heating in the summer and harness solar energy in the form of heat during the winter season.

Sustainable buildings can reduce the negative impact on the environment. For the design of such buildings, various software packages (e.g., EnergyPlus, DOE-2 and Green Building Studio) have been developed to accurately simulate the building's energy consumption. However, these simulation tools are only used to validate the final design of the building. Sustainable building design will be revolutionized if AI and BD are used for the exploration of multiple design possibilities rather than the validation of the final design [34].

Heating demand in buildings is often modeled through various energy simulation packages such as BLAST, DOE2.1, eQUEST, and EnergyPlus [35]. These computer codes are typically governed by partial differential equations that are derived from energy and mass balance considerations. These codes can help the user understand how different heat and mass transfer processes affect building heating and cooling loads. The major flaw of these codes is that they cannot account for the complex interactions between the energy systems in a building, and often the simplifications and assumptions used in these models result in a loss of accuracy [36]. Instead of using these simulation models, statistical models and ML algorithms can be used with much improved accuracy. The main advantage of using ML is that ML learns from the behavior of the system or the observed data. It emphasizes prediction accuracy rather than model accuracy [37]. DL uses multiple layers of abstraction [38] and thus can model more complex functions. This is why it has demonstrated improved performance over ML in many applications [39].

4. Recent applications of AI and BD to energy efficient buildings

In this section, a brief review of recent applications of AI and BD to energy-efficient buildings is presented, with emphasis on the use of ML, DL, and BD.

4.1. Design of EFBs

Different AI schemes have been studied with the aim of minimizing the detrimental effects of buildings on the environment, reducing overall costs and increasing energy efficiency. Ghada presented a critical review of current AI-based practices that have been employed to obtain optimal green solutions for different models used in green building architecture [40]. It was pointed out in said paper that the three most commonly used AI techniques are GA, FL, and AHP.

Kim et al. [41] utilized data mining technology to extract interrelationships and patterns of interest from a large database. The case study conducted in their paper revealed that data mining-based energy modeling could help project teams discover useful patterns to improve the energy efficiency of building design during the design phase.

Considerable work has been done in the field of Architectural Intelligence. As the awareness of the drastic impacts of high energy consumption of buildings increases, research interest on the construction of sustainable buildings also increases. Sustainable architecture seeks to mitigate the impacts of construction on the environment, and buildings constructed in this regard are interchangeably termed smart buildings or green buildings. However, there is a key difference between these two as pointed out by Zakari [42]. Green buildings are those that have been rated by a Sustainable Building Rating System (SBRS). Leadership in Energy and Environmental Design (LEED) and Building Research Establishment Environmental Assessment Method (BREEAM) are two of the world's most popular rating systems for assessing sustainable buildings. On the other hand, a smart building might reduce the impact on the environment, but it will not be considered a green building unless it has been rated.

According to the World Business Council for Sustainable Development, buildings consume a big portion (around 40 percent) of the total energy generated by most countries. Despite the introduction of alternative green energy resources, the International Energy Agency predicts that even by 2030, three-fourths of energy sources will still be carbon dioxide-related. Nowadays, a lot of efforts are being made to reduce the energy consumption in buildings through the use of sustainable building design.

The Empire State Building (ESB) in New York City will be 90 years old in 2021. However, not many people know about its achievement in reducing the energy consumption in recent renovations, and becoming a role model for intelligent buildings. Using a direct digital control system, numerous different parameters such as room temperature, indoor air quality, electricity, cooling/heating conditions, etc. are self-monitored, and minimum energy is wasted. At the same time, smart telemetry provides the ESB's tenants with real time information to control and optimize the power consumption [43].

The recent renovation of ESB had a capital expenditure of 550 million dollars, but the results proved it to be worth the investment. Initially, the renovation was expected to reduce the energy expenditure by 38 percent and reach a return of 4.4 million dollars as annual energy savings. Surprisingly, the project surpassed its expectations with every passing year. In its third consecutive year of operation, the savings from the innovation project were 16% above its original target, and saved a total of 7.5 million dollars [44]. The usage of AI-enabled technology in the ESB played a great role in all its achievements. ML algorithms drive the predictive analytics for forming more efficient power consumption strategies. This enables on demand maintenance of the building, and tremendously reduces the costs of human-heavy operations.

Another AI application to the ESB lies in the preventive measures provided by AI technology. The ESB can collect tremendous amount of operational data, and recognize any issues before the first signs start to appear, and subsequently take preventive measures. This results not only in financial benefits, but also ensures efficient problem detection, improves in-building safety, and can prevent causalities from incidents such as electrical fires. Finally, the ecological benefits from the AI-enabled building infrastructure have considerable importance. According to a new study on green certified buildings, for every dollar saved on energy expenditures by smart buildings, 77 cents are saved in health and climate benefits [45]. From this report, one can conclude that AI applications to the ESB produced a 3.4 million-dollar benefit in health and climate values along with the return in energy savings.

4.2. Prediction and reduction of building energy consumption

Tsanas and Xifara [46] developed a statistical ML framework to study the effect of variables, such as wall area and glazing area, on the heating and cooling loads of residential buildings. They concluded that the use of ML to estimate a building's heating and cooling loads gave accurate results as long as the requested query bore resemblance to the data used to train the model.

Various NN models [18,37,47–52] have been used to predict building energy consumption in recent years. These AI methods could achieve very accurate energy consumption forecasts over a one-hour to one-week time horizon [51,53,54]. Chou and Bui used various AI techniques to estimate building heating and cooling loads. The prediction models were constructed using 768 experimental datasets from the literature. Comparison results showed that the ensemble approach (SVR + ANN) and SVR were the best models for predicting the cooling load and the heating load, respectively, with mean absolute percentage errors below 4%. [55]

In their study of ML approaches for estimating commercial building energy consumption, Robinson et al. [21] found that gradient boosting regression models performed the best, and could make predictions that are on average within a factor of 2 from the true energy consumption values. Roy et al. explored many different ML techniques for predicting the heating and cooling loads of residential buildings. Advanced techniques such as Multivariate Adaptive Regression Splines (MARS) and Extreme Learning Machine (ELM) were studied and compared with conventional AI methods like linear regression, neural network, Gaussian processes, and Radial Basis Function Network [56].

Deng et al. [57] tested different predictive modeling approaches for US commercial building energy use based on the Building Energy Consumption Survey (CBECS) 2012 microdata. In this paper six regression or ML techniques were applied and compared for prediction performance. It was found that ML algorithms did not always outperform the linear regression. The mixed results suggested careful consideration in applying advanced predictive algorithms to the CBECS dataset.

Rahman and Smith employed a ML model to predict the fuel consumption in commercial buildings [51]. The goal of this study was to successfully predict fuel consumption in buildings one year ahead of time. It was found that Neural Networks and Gaussian process (GP) regression performed better than linear regression and ridge regression. The fuel consumption estimates for multiple climate zones over a one year period were made in this paper.

Fan et al. [58] used DL to predict short-term building cooling loads. Deep learning has great potential for developing prediction models. Additionally, when used in an unsupervised manner, it can help extract meaningful features from a large bulk of raw data. This study investigated both of these aspects. The DL model used in the study showed great performance for predicting cooling loads 24 h in advance. A review of the use of unsupervised ML techniques for non-residential building performance control and analysis can be found in Miller et al.'s paper [59].

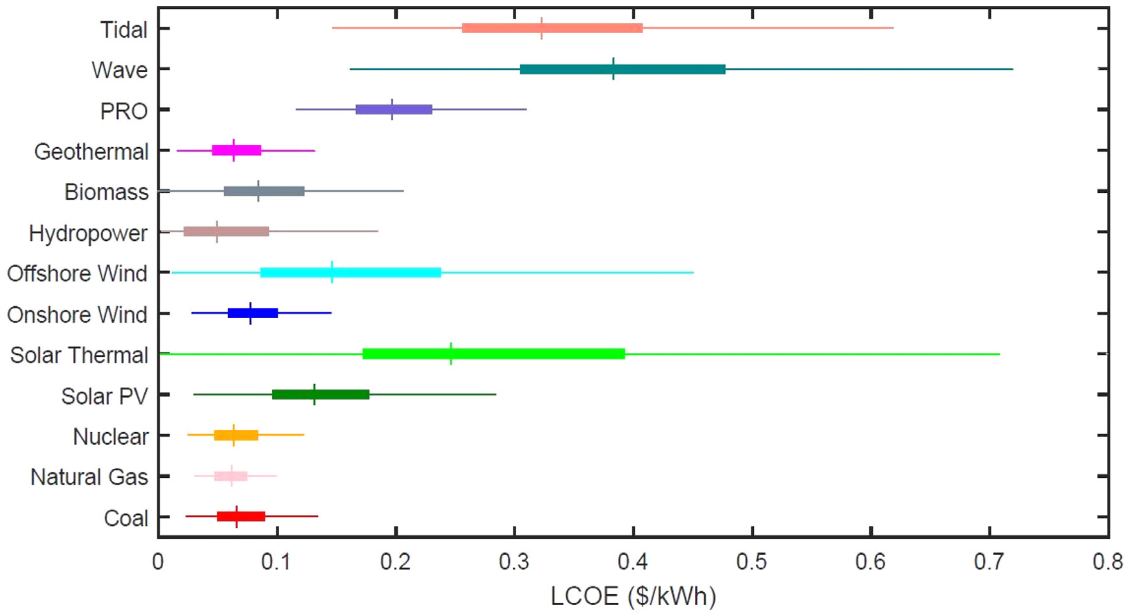

Fig. 5. LCOE values of different technologies without carbon pricing [63].

Yang et al. presented an adaptive ANN which can predict the unexpected behavior of incoming data and adapt to it accordingly. Two models, accumulative training and sliding window training, were tested against simulated and measured data. The sliding window technique had better performance in the case of real measurements. For simulated data, both techniques showed similar performances [53].

Gonzalez and Zamarreno used a feedback ANN to predict short term electric load consumption in buildings. The biggest advantage of this model lies in its simplicity. It used a minimal amount of resources and yet its precision was comparable to other methods used for forecasting [49]. Edwards et al. [76] tested seven different ML techniques on different data sets, and discussed the advantages, disadvantages, and technical benefits for each technique when applied to the prediction of future hourly residential electrical consumption.

Chae et al. proposed a short-term building energy usage forecasting model based on an Artificial Neural Network (ANN) model with Bayesian regularization algorithm to investigate the effects of network design parameters, such as time delay, number of hidden neurons, and training data, on the model capability and generality [60]. The model was used for day-ahead electricity usage of buildings in a 15-minute resolution.

Dedinec et al. [61] used a deep belief network to predict electricity power consumption in the Macedonia area over a 24-hour period. The model showed a significant improvement in MAPE when compared with the traditional NN model. Also, the model shows an improvement in the MAPE when compared with the forecasting data provided by the Macedonian system operator. Mocanu et al. [62] employed 2 DL models (CRBM and FCRBM) to estimate the electricity consumption in a residential building. They presented five different scenarios with varying resolutions, from that of one minute to a weekly resolution. The FCRBM model showed a significantly higher performance and cut the prediction error by half when compared to the traditional ANN.

Rahman, Srikumar and Smith proposed two deep RNN models to make medium- to long-term predictions about electricity consumption in commercial and residential buildings [39]. Several AI techniques were applied to the prediction of hourly electricity consumption as well as the aggregated hourly electricity consumption. Compared to the three-layered MLP model, the RNN models showed improved performance in predicting electric load profiles. However, they did not perform well in forecasting the aggregate load profiles over a one-year timespan. The RNN was also used by Rahman and Smith [17] to predict the heating demand in buildings. The deep RNN model again showed improved performance in predicting the heating loads when compared to the three-layered MLP model. These AI predictions can help the design of a stratified thermal storage tank by the successful estimation of the performance characteristics of a CHP unit.

Rahman, Srikumar and Smith also applied RNN to predicting commercial electricity loads at a high resolution of 10-minute intervals. This RNN model successfully incorporated the sharp discontinuities observed in the electric profile and the long-term temporal dependencies related to electricity consumption, which are two of the biggest challenges for high-resolution predictions. As a result, the model showed better performance when compared to the MLP model and the one-hour model in predicting electricity consumption [63].

Yildiz et al. conducted a review and analysis of regression and machine learning models on the electricity load forecasting of commercial buildings. Their study indicated that Artificial Neural Networks with Bayesian Regulation Backpropagation had the best overall root mean squared and mean absolute percentage error performance [64].

The evaluation of economic aspects of energy resources holds significant importance as the overall cost of any energy system can determine the feasibility of its operation under particular circumstances. To compare the economic competitiveness of newly emerging renewable energy technologies with conventional fossil fuel-based energy systems, Tran performed various case studies using the Monte Carlo approach and incorporated global sensitivity and uncertainty analysis into LCOE calculations [65]. The results proved that the uncertainties in the input data can have a considerable effect on the LCOE values. Due to high uncertainties in the case of emerging renewable energy technologies, the LCOE of fossil fuel systems is lower as compared to the former. However, after including the carbon pricing for different energy systems, the renewable energy systems were found to have better economic viability. Shown in Figs. 5 and 6 are the LCOEs of different electricity generating technologies including and without including the

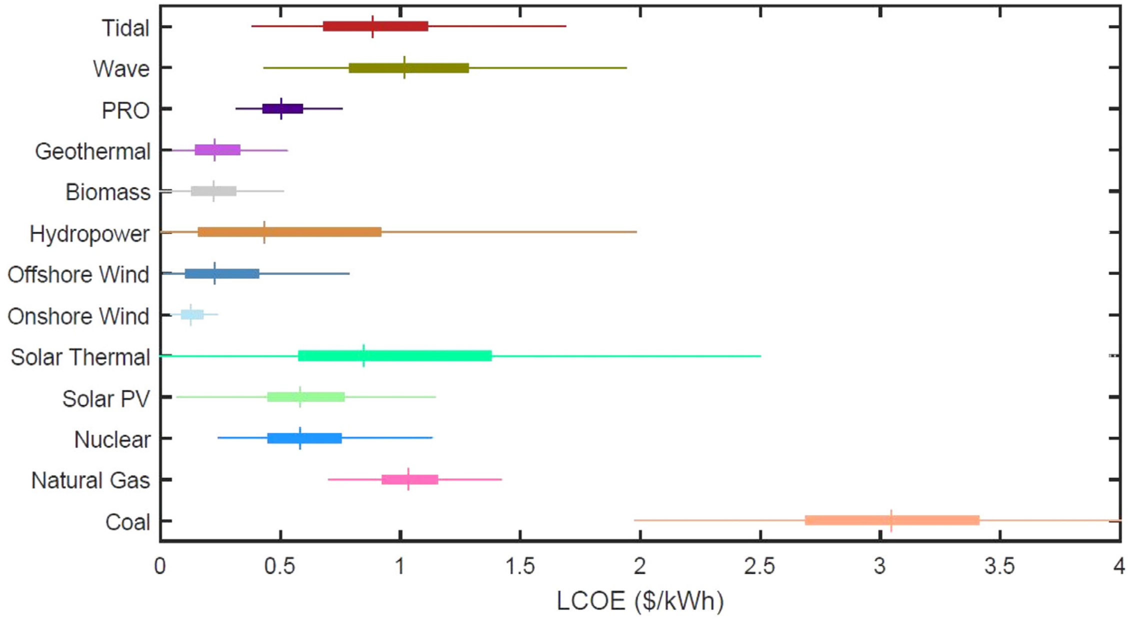

Fig. 6. LCOE values of different technologies with carbon pricing [63].

carbon pricing. This comparison is presented to show the increases in electricity costs of different technologies when carbon pricing is included in energy cost calculations.

One method of reducing energy consumption in food processing facilities is by means of heat recovery. Legorburu presented a bottom-up modular framework to model the energy consumption of a cannery [66]. The results of this study showed that by adding a globally optimized heat recovery system, the total gas consumption could be reduced by 6% annually. With the optimal assumptions of electricity and gas costs, the system can be applied for a payback period of 15 years. The modeling could become more complete by adding the use of EnergyPlus. EnergyPlus is a well-known simulation software for modeling the energy consumption in buildings. With its inclusion, the model by Legorburu can become a more realistic food processing facility model.

According to the U.S Department of Energy, commercial office buildings contribute a major portion of total energy consumption. Even with the use of energy efficient materials and advanced Building Management Systems (BMS), reliable building operation is not always guaranteed. Most new buildings consume energy at levels higher than the specifications, resulting in system failure. In this regard, Wu [67] presented a new ML approach to produce accurate forecasts for building energy consumption. This model applies Support Vector Regression (SVR) on the historical energy use of the building. The data used by this system also includes building energy data, weather data and power grid data. Fig. 7 shows the flowchart of the model. To ensure the reliability of ML, an automated online evaluator continuously monitors different external and internal conditions of the building, such as electricity load, peak load, temperature, humidity, and building work and maintenance schedules. The biggest achievement of this work was the improvement in the efficiency and reliability of the building without investing huge amounts of additional capital. The SVR model for predicting the energy demand was accurate. In addition, as the automated online evaluation works in parallel with ML, it aids in tuning the building's energy systems and operation schedules.

One of the main causes of the production of GHG from human activities can be attributed to over-cooling, over-heating and over-lighting in buildings [68]. Project Dasher, a prototype building site, was considered for addressing these issues [69]. The sensor networks in the building were integrated with the IP networks to monitor the real time information of power consumption. Point cloud data was generated to place the sensor data streams into their physical context. Such a scan revealed about 1.3 billion points. For a more intelligent analysis and visualization of this data, a Building Information Model (BIM) was created using Autodesk Revit. Unfortunately, software for the automated visualization of the point cloud data is still not available. All of the data gathered by this means, if accessible to the building occupants, can be used for informed decision making, thus enabling them to participate in improving the overall building efficiency. Moreover, implementing these kinds of visualization strategies on a city level could contribute to improving the accountability of governing agencies.

Li presented an intelligent data analysis method for the modeling and prediction of daily electricity consumption [70]. Statistical methods and an auto-regression model were employed to extract the features of daily energy consumption. This method can detect abnormalities in the system by detecting the outlier values based on Canonical Variate Analysis (CVA). Fig. 8 shows an outline of this method. To verify the prediction of the model, data were captured from a company based in Birmingham, U.K for a complete year. It was concluded that the model was efficient enough in providing the building operators with information regarding any kind of abnormal energy consumption and correcting the problems.

Energy dashboards, alternatively termed as an Energy Information System (EIS), are a display and visualization tool that uses the EIS data and technology to provide critical information to users [71]. Some of the main features of this system include benchmarking, anomaly detection, base-lining, load shape optimization, and energy rate analysis, as well as retrofit and retro commissioning savings [72]. The application of an EIS is intended towards BD for high performance buildings. Real time evaluation of heat pumps [73], AEL forced air, chillers, and energy recovery ventilation systems can be accomplished by using energy dashboards.

EnergyPlus is a flagship simulation software which has been used for the assessment of energy consumption in both residential and commercial buildings. Up to 3000 input parameters that are representative of various building properties can be handled by this software. Reference building models are also available. However, accurate data is not available for most of these parameters.

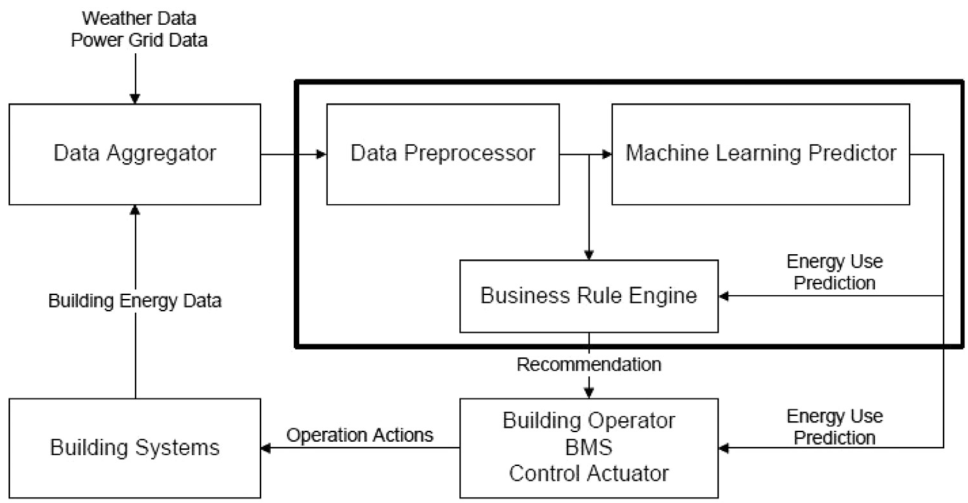

Fig. 7. Predictive building energy optimization workflow [65].

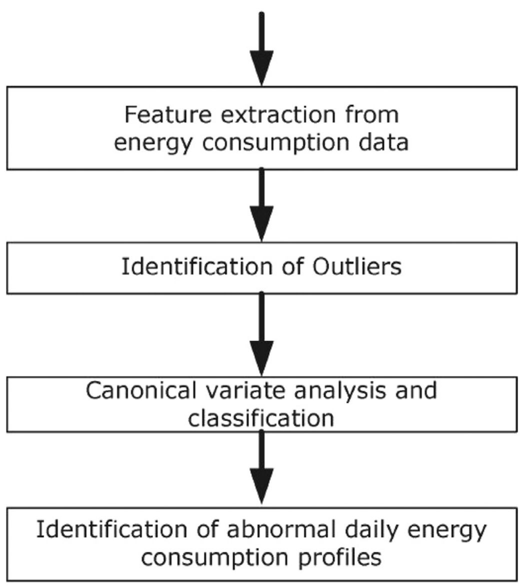

Fig. 8. Outline of the identification of outliers and modeling of the daily energy consumption profiles [67].

There is often a mismatch between the product specifications and the manufacturer's supplied data. For this reason, a building energy model must be tweaked by a building modeling expert which often results in the consumption of much time and resources. One solution is a software called Autotune [74,75], which was developed to save time by automating the process of tuning building energy models. To achieve the goal of automation, sensor data along with the building model are provided to the Autotune software. The machine learning agents handle these data and give a tuned model of the building as output [76]. The system in its late development stage is capable of tuning the models in its various forms as provided by DOE [77] and can work with the different climate zones and subzones specified by ASHRAE. This approach proves to be a great challenge for data storage, transfer and analysis as the data is generated for a whole year at 15-minute intervals and is constituted of up to 180 output variables. The bulk of raw data amounts to 270 Terabytes. This issue is tackled by evolving the system architecture and summarizing the statistical data. Moreover, the final product is envisioned to be a service rather than a software in the sense that the template of the building will be provided by the user as a prototype, and the Autotune system will return the model tweaked for the optimized solution to the design problem.

A further improvement to the Autotune project was recently made by New et al. [78]. With the help of some of the fastest computing resources in the world, 130,000 AI algorithms were tested in parallel to calibrate building energy models to the measured data. The algorithm surpasses the 30% error requirement stated in the ASHRAE guidelines for hourly utility data with an average error rate of 3.6% for 20,000 buildings. Robust calibration algorithms, ensemble approaches, or post-processing techniques can help in further reducing the input side error, which, in turn, can provide one of the most reliable software models for buildings.

The reference model built with the help of BD by the Building Energy Management System can be used as the MPC for heat pumps. This can be applied to maintaining of an ATES system, which is the reason for the rapid growth of sustainable office buildings in the Netherlands. An ATES system uses a well pump to extract cold groundwater for cooling, which can be used along with a heat pump if required. Thermal energy balance of the system is necessary for the optimal operation of an ATES. A case study for a building similar to a Dutch office building enabled the model to be universally applied to most Dutch office buildings [79]. The performance of this model is affected by sensor accuracy, the number of assumptions made, and the coherence of the dataset used.

4.3. Solar energy utilization

Illumination and lighting in buildings can consume a considerable portion of the electricity used (typically up to 10% of the total electric consumption). Therefore, improvement in the methods of illumination can lead to improved energy efficiency in buildings. Solar energy is the biggest free resource of energy. Two of its direct applications come in the forms of daylighting and solar heating. Solar energy is also the sole energy source for electricity generation using PV technique or solar-powered heat engines. Chun, Chen and co-workers have proposed the use of different wavelengths of the solar spectrum for sanitization, illumination, electricity generation, and heating applications, respectively.

Before using solar energy for any practical application, its potential in the area of application needs to be assessed. Different studies that focus on the forecasting of the Global Solar Potential

Table 1
Different studies performed for the forecasting of Global Solar Potential.

Study	Year	Network type	Location of station (s)	$MAPE_{max}$ (%)
Gopinathan et al. [80]	1995	Empirical models	Spain	7.97
Mohandes et al. [81]	1998	MLP	Saudi Arabia	12.61
Togrul et al. [82]	1999	Regression analysis	Turkey	9.80
Mohandes et al. [83]	2000	RBF	Saudi Arabia	10.09
Reddy et al. [84]	2002	MLFF	India	10.20
Reddy et al. [84]	2002	MLFF	India	12.50
Sozen et al. [85]	2004	MLP	Turkey	6.73
Sozen et al. [86]	2005	MLP	Turkey	6.78
Rehman and Mohandes [87]	2008	MLP	Saudi Arabia	4.49
Azadeh et al. [88]	2009	MLP	Iran	6.70
Behrang et al. [89]	2010	MLP	Iran	5.21
Behrang et al. [89]	2010	RBF	Iran	5.56

have been performed by researchers [80–89]. The significance of these models lie in their convenience, as the parameters used in these models such as humidity, air temperature, sunshine hours, and wind speed, are generally accessible. Moreover, in the absence of monitoring stations, such as in the Middle East and Turkey, the results from these models can be used for designing solar energy systems. Table 1 shows the comparison between various techniques used in these studies.

A direct application of solar energy is daylighting. Various well-designed daylighting illumination systems were cited in the literature and are commercially available. For active daylighting, Chun presented a solar tracking system that is capable of aggressively harvesting sunlight and can be used for illumination in multiple rooms of a building [90]. The novelty of the system is the inclusion of auxiliary illumination by artificial means for the continuous operation of the system in case the sunlight is insufficient. Parans offered a similar lighting system whose efficiency and energy-saving capabilities have been studied [91]. For passive daylighting, façade systems play an important role in providing indoor illumination. The use of recorded photometric data is still rare in practice because of its complexity. Jan de Boer studied the modeling of bidirectional transmittance distribution functions (BTDFs) of complex fenestration systems (CFS) [92]. This model was independent of any specific lighting simulation software and could be used as a standalone tool. Alongside the improvement of façade systems, this model also allowed for the detailed evaluation of complex façade systems.

Application of wavelet-based data compression techniques significantly reduces the data volume. This method was validated against different numerical and analytical test cases. Additionally, it can be included in simulation programs in the format of an Application Programming Interface (API). Upon validation, this method was integrated into RADIANCE and SUPERLITE. The viability of these systems in any region, from both economic and performance perspectives, can be assessed with the Global Solar Potential forecasting techniques.

PV systems are currently one of the leading renewable energy systems. To ensure the viability of PV systems, irradiance data are required. Generally, such data are available at macro scale levels. For example, photovoltaic companies in the United States rely on the data provided by the National Renewable Energy Laboratory's (NREL's) National Solar Radiation Database (NSRD) [93]. The resolution of this database is about four kilometers. However, in any urban environment, the shadowing effects of buildings and trees cannot be ignored. Bowles studied an irradiance model to simulate the solar availability at a localized level [94]. The model used the data from a Geographic Information System (GIS) coupled with aerial Light Detection and Ranging (LiDAR) topography data. The space resolution of this modeling work is three meters and can be further improved with more accurate LiDAR data. Bowles's study led to a more accurate estimation of the LCOE value of a PV system.

4.4. New BD for EFB design and operation

Other large databases which can be used to design and operate EFBs in the near future may include 3D topographic data with very high space resolutions, as well as hyperlocal weather data and detailed landscape images collected by companies such as Google. Google Street View allows designers to examine the surroundings of a potential construction site with 3D virtual reality. This application of Street View is very useful for assessing the shadowing of surroundings, air flow, noise level, and other site-specific information important to EFBs.

Weather Intelligence Plus combines data from national weather stations, more than 270,000 personal weather stations, satellite data, radar data, and more to provide hyperlocal accuracy into the current and projected weather in a very small geographical area [95]. This database is currently used to schedule smart sprinkler systems. Its usage can be easily extended to the scheduling and control of individual buildings' temperature, humidity, and illumination conditions.

The Global Forest Watch was launched by the World Resources Institute. It has been integrated with the Google Earth engine, which can provide information about deforestation around the world. The collaboration of these two tools can reveal deforestation hotspots in the world. Google has used network cloud-based servers to create the global forest map that processed 650,000 images by using 10,000 computers working together [96].

5. Summary

This paper starts with an introduction to AI and BD, followed by an illustration of how AI and BD can be and have been applied to energy–efficient commercial and residential buildings while maintaining a high-quality indoor environment. The second half of the paper is a review of recent publications on the application of AI and BD to energy-efficient buildings, with emphasis on the use of ML and large databases for improved search as well as optimization speed and accuracy. The results presented in this paper clearly indicate the increasingly important role AI and BD will play in the future for the design and operation of energy-efficient buildings with a comfortable indoor living environment. Suggested research directions in this area may include fast data mining and optimal weighting of weather data from different sources such as satellite data and personal weather stations, more user-friendly interfaces to suit different users' preferences and operating conditions and patterns, and applications of AI and BD to smart homes and buildings with highly adjustable building components.

Declaration of Competing Interest

None.

Acknowledgements

This research was supported by National Research Foundation of Korea (NRF-2016R1D1A1B04934265, NRF- 2018R1A2B2008542, and NRF-2017R1A2A1A05001461).

References

[1] What is AI? everything you need to know about artificial intelligence | ZDNet, (n.d.). https://www.zdnet.com/article/what-is-ai-everything-you-need-to-know-about-artificial-intelligence/ (Accessed 18 March 2019).
[2] P. Wang, What do you mean by "AI"? in: Proc. 2008 Conf. Artif. Gen. Intell. 2008 Proc. First AGI Conf., Amsterdam, The Netherlands, The Netherlands, IOS Press, 2008, pp. 362–373. http://dl.acm.org/citation.cfm?id=1566174.1566207.
[3] P. Stone, R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager, J. Hirschberg, S. Kalyanakrishnan, E. Kamar, S. Kraus, K. Leyton-Brown, D. Parkes, W. Press, A. Saxenian, J. Shah, M. Tambe, A. Teller, Artificial intelligence and life in 2030: one hundred year study on artificial intelligence, in: Rep. 2015-2016 Study Panel, Stanford Univ, Stanford, CA, 2016, p. 52. https://ai100.stanford.edu.
[4] J. McCarthy, M. Minsky, N. Rochester, C. Shannon, A proposal for the Dartmouth summer research project on artificial intelligence, Uma Ética Para Quantos? XXXIII (2012) 81–87, doi:10.1007/s13398-014-0173-7.2.
[5] D. Crevier, AI: The Tumultuous History of the Search for Artificial Intelligence, Basic Books, Inc., New York, NY, USA, 1993.
[6] S. Zeghloul, M.A. Laribi, J.-P. Gazeau (Eds.), Robotics and Mechatronics Proceedings of the 4th IFToMM International Symposium on Robotics and Mechatronics, 20, 1st ed., 2016.
[7] Good Ol' Fashioned AI, (n.d.). https://www.cs.swarthmore.edu/~eroberts/cs91/projects/ethics-of-ai/sec3_1.html (Accessed 18 March 2019).
[8] P. Jackson, Introduction to Expert Systems, 3rd ed., Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1998.
[9] Mark Kantrowitz, A timeline of artificial intelligence., (1994). http://biology.kenyon.edu/slonc/bio3/AI/TIMELINE/timeline.html (Accessed 18 March 2019).
[10] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed., Prentice Hall Press, Upper Saddle River, NJ, USA, 2009.
[11] K. Simon, C. Malcolm, An Introduction to Knowledge Engineering, Springer-Verlag, Berlin, Heidelberg, 2006.
[12] D. McDermott, A critique of pure reason, Comput. Intell. 3 (1987) 151–160, doi:10.1111/j.1467-8640.1987.tb00183.x.
[13] A.L. Samuel, Some studies in machine learning using the game of checkers, IBM J. Res. Dev. 3 (1959) 210–229.
[14] R.W. Saaty, The analytical hierarchy Process- What and why it is used, Math. Model 9 (1987) 161–176 https://doi.org/10.1016/0270-0255(87)90473-8.
[15] M. Paulescu, P. Gravila, E. Tulcan-Paulescu, Fuzzy logic algorithms for atmospheric transmittances of use in solar energy estimation, Energy Convers. Manag. 49 (2008) 3691–3697, doi:10.1016/j.enconman.2008.06.025.
[16] S.A. Kalogirou, Optimization of solar systems using artificial neural-networks and genetic algorithms, Appl. Energy 77 (2004) 383–405, doi:10.1016/S0306-2619(03)00153-3.
[17] A. Rahman, A.D. Smith, Predicting heating demand and sizing a stratified thermal storage tank using deep learning algorithms, Appl. Energy. 228 (2018) 108–121, doi:10.1016/j.apenergy.2018.06.064.
[18] L.E. Atlas, M.J. Damborg, R.J. Marks, D.C. Park, M.A. El-Sharkawi, Electric load forecasting using an artificial neural network, IEEE Trans. Power Syst. 6 (2002) 442–449, doi:10.1109/59.76685.
[19] H.X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (2012) 3586–3592, doi:10.1016/j.rser.2012.02.049.
[20] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: a review, Renew. Sustain. Energy Rev. 23 (2013) 272–288, doi:10.1016/j.rser.2013.03.004.
[21] C. Robinson, B. Dilkina, J. Hubbs, W. Zhang, S. Guhathakurta, M.A. Brown, R.M. Pendyala, Machine learning approaches for estimating commercial building energy consumption, Appl. Energy 208 (2017) 889–904, doi:10.1016/j.apenergy.2017.09.060.
[22] A.E.W. Jones, G.W. Forbes, An adaptive simulated annealing algorithm for global optimization over continuous variables, J. Glob. Optim. 6 (1995) 1–37, doi:10.1007/BF01106604.
[23] K. Chen, G.T. Lin, Optimization of multiple-module thermoelectric coolers using artificial-intelligence techniques, Int. J. Energy Res. 26 (2002) 1269–1283, doi:10.1002/er.849.
[24] Datamation, Big data vs. artificial intelligence, (n.d.). https://www.datamation.com/big-data/big-data-vs.-artificial-intelligence.html (Accessed 22 March 2019).
[25] D. Laney, 3D Data management: controlling data volume, Velocity, and Variety (2009), doi:10.1016/j.infsof.2008.09.005.
[26] Enterprise Strategy Group, The 6 vs: The BI/Analytics game changes so microsoft changes excel, (n.d.). https://www.esg-global.com/blog/the-6-vs-the-bianalytics-game-changes-so-microsoft-changes-excel (Accessed 26 March 2019).
[27] W. Vorhies, How many "V's" in big data ? The characteristics that define big data, Data Sci. Cent. (2015) 1–8 https://www.datasciencecentral.com/profiles/blogs/how-many-v-s-in-big-data-the-characteristics-that-define-big-data. (Accessed March 26, 2019).
[28] IBM, 10 Key marketing trends for 2017 and ideas for exceeding customer expectations, (n.d.). https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN (Accessed 22 March 2019).
[29] J. Wu, S. Guo, J. Li, D. Zeng, Big data meet green challenges: greening big data, IEEE Syst. J. 10 (2016) 873–887, doi:10.1109/JSYST.2016.2550538.
[30] H.J. Han, Y.I. Jeon, S.H. Lim, W.W. Kim, K. Chen, New developments in illumination, heating and cooling technologies for energy-efficient buildings, Energy 35 (2010) 2647–2653, doi:10.1016/j.energy.2009.05.020.
[31] W. Chun, K. Chen, H.T. Kim, Performance study of a bi-directional thermodiode designed for energy-efficient buildings, J. Sol. Energy Eng. 124 (2002) 291, doi:10.1115/1.1498849.
[32] W. Chun, K. Chen, Test results of a bi-directional thermodiode system for solar energy utilization, Sol. Energy 73 (2002) 269–280, doi:10.1016/S0038-092X(02)00064-6.
[33] K. Chen, R.W. Shorthill, S.S. Chu, P. Chailapo, and S. Narasimhan, An Energy - Efficient Construction Module of Variable Direction of Heat Flow, Heat Capacity, and Surface Absorptivity, US Air Force Report WL-TR-95-3045, 1995.
[34] AI.Business, Machine learning and energy efficient building design, (n.d.). http://ai.business/2017/03/23/machine-learning-and-energy-efficient-building-design/ (Accessed 18 March 2019).
[35] D.B. Crawley, J.W. Hand, M. Kummert, B.T. Griffith, Contrasting the capabilities of building energy performance simulation programs, Build. Environ. 43 (2008) 661–673, doi:10.1016/j.buildenv.2006.10.027.
[36] H.S. Rallapalli, A Comparison of EnergyPlus and eQUEST Whole Building Energy Simulation Results for a Medium Sized Office Building, Arizona State University, 2010, doi:10.1017/CBO9781107415324.004.
[37] L. Breiman, Statistical modeling: the two cultures, Stat. Sci. 16 (2001) 199–231.
[38] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press, 2016.
[39] A. Rahman, V. Srikumar, A.D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Appl. Energy 212 (2018) 372–385, doi:10.1016/j.apenergy.2017.12.051.
[40] G. Elshafei, A. Negm, AI technologies in green architecture field: statistical comparative analysis, Procedia Eng. 181 (2017) 480–488, doi:10.1016/j.proeng.2017.02.419.
[41] H. Kim, A. Stumpf, W. Kim, Analysis of an energy efficient building design through data mining approach, Autom. Constr. 20 (2011) 37–43, doi:10.1016/j.autcon.2010.07.006.
[42] B. Faizah Mohammed, A.B. Yunus Yusuf, Are smart buildings same as green certified buildings? a comparative analysis, Int. J. Sci. Res. Publ. 4 (2014) 2250–3153 www.ijsrp.org.
[43] IOT Evolution, How ai AI makes the empire state building smart, (n.d.). https://www.iotevolutionworld.com/iot/articles/437918-how-ai-makes-empire-state-building-smart.htm (Accessed 18 March 2019).
[44] Clinton Foundation, Innovative empire state building program cuts $7.5 M in energy costs over past three years, (n.d.). https://www.clintonfoundation.org/press-releases/innovative-empire-state-building-program-cuts-75m-energy-costs-over-past-three-years (Accessed 21 March 2019).
[45] P. MacNaughton, X. Cao, J. Buonocore, J. Cedeno-Laurent, J. Spengler, A. Bernstein, J. Allen, Energy savings, emission reductions, and health co-benefits of the green building movement review-article, J. Expo. Sci. Environ. Epidemiol. 28 (2018) 307–318, doi:10.1038/s41370-017-0014-9.
[46] A. Tsanas, A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy Build 49 (2012) 560–567 https://doi.org/10.1016/j.enbuild.2012.03.003.
[47] D. Kriesel, A brief introduction to Neural networks, Univ. Bonn Semin. Proc. (2005) http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf.
[48] K. Yun, R. Luck, P.J. Mago, H. Cho, Building hourly thermal load prediction using an indexed ARX model, Energy Build 54 (2012) 225–233, doi:10.1016/j.enbuild.2012.08.007.
[49] P.A. González, J.M. Zamarreño, Prediction of hourly energy consumption in buildings based on a feedback artificial neural network, Energy Build 37 (2005) 595–601, doi:10.1016/j.enbuild.2004.09.006.
[50] W. Charytoniuk, M.S. Chen, Nonparametric regression based short-term load forecasting, IEEE Trans. Power Syst. 13 (1998) 725–730, doi:10.1109/59.708572.
[51] A. Rahman, A.D. Smith, Predicting fuel consumption for commercial buildings with machine learning algorithms, Energy Build 152 (2017) 341–358, doi:10.1016/j.enbuild.2017.07.017.
[52] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy Build 37 (2005) 545–553, doi:10.1016/j.enbuild.2004.09.009.
[53] J. Yang, H. Rivard, R. Zmeureanu, On-line building energy prediction using adaptive artificial neural networks, Energy Build 37 (2005) 1250–1259, doi:10.1016/j.enbuild.2005.02.005.
[54] Y. Heo, V.M. Zavala, Gaussian process modeling for measurement and verification of building energy savings, Energy Build 53 (2012) 7–18, doi:10.1016/j.enbuild.2012.06.024.
[55] J.S. Chou, D.K. Bui, Modeling heating and cooling loads by artificial intelligence for energy-efficient building design, Energy Build 82 (2014) 437–446, doi:10.1016/j.enbuild.2014.07.036.
[56] S. Sekhar Roy, R. Roy, V.E. Balas, Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid

model of MARS and ELM, Renew. Sustain. Energy Rev. 82 (2018) 4256–4268, doi:10.1016/j.rser.2017.05.249.
[57] H. Deng, D. Fannon, M.J. Eckelman, Predictive modeling for US commercial building energy use: a comparison of existing statistical and machine learning algorithms using CBECS microdata, Energy Build 163 (2018) 34–43, doi:10.1016/j.enbuild.2017.12.031.
[58] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, Appl. Energy. 195 (2017) 222–233, doi:10.1016/j.apenergy.2017.03.064.
[59] C. Miller, Z. Nagy, A. Schlueter, A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings, Renew. Sustain. Energy Rev. 81 (2018) 1365–1377, doi:10.1016/j.rser.2017.05.124.
[60] Y.T. Chae, R. Horesh, Y. Hwang, Y.M. Lee, Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, Energy Build 111 (2016) 184–194, doi:10.1016/j.enbuild.2015.11.045.
[61] A. Dedinec, S. Filiposka, A. Dedinec, L. Kocarev, Deep belief network based electricity load forecasting: an analysis of Macedonian case, Energy 115 (2016) 1688–1700, doi:10.1016/j.energy.2016.07.090.
[62] E. Mocanu, P.H. Nguyen, M. Gibescu, W.L. Kling, Deep learning for estimating building energy consumption, Sustain. Energy, Grids Networks 6 (2016) 91–99, doi:10.1016/j.segan.2016.02.005.
[63] A. Rahman, Applications of Machine Learning, Deep Learning and Reduced-Order Models, Mechanincal Engineering department, The University of Utah, 2018.
[64] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renew. Sustain. Energy Rev. 73 (2017) 1104–1122, doi:10.1016/j.rser.2017.02.023.
[65] T.T.D. Tran, A.D. Smith, Incorporating performance-based global sensitivity and uncertainty analysis into LCOE calculations for emerging renewable energy technologies, Appl. Energy. 216 (2018) 157–171, doi:10.1016/j.apenergy.2018.02.024.
[66] G. Legorburu, A.D. Smith, Energy modeling framework for optimizing heat recovery in a seasonal food processing facility, Appl. Energy. 229 (2018) 151–162, doi:10.1016/j.apenergy.2018.07.097.
[67] L. Wu, G. Kaiser, D. Solomon, R. Winter, A. Boulanger, R. Anderson, Improving efficiency and reliability of building systems using machine learning and automated online evaluation, in: 2012 IEEE Long Isl. Syst. Appl. Technol. Conf. LISAT 2012, 2012, pp. 1–6, doi:10.1109/LISAT.2012.6223192.
[68] U.S.E.I. Administration., Assumptions to the annual energy outlook, 0554 (2009).
[69] A. Khan, K. Hornbæk, Big data from the built environment, (2011) 29. doi:10.1145/2025528.2025537.
[70] X. Li, C.P. Bowers, T. Schnier, Classification of energy consumption in buildings with outlier detection, IEEE Trans. Ind. Electron. 57 (2010) 3639–3644, doi:10.1109/TIE.2009.2027926.
[71] K. Marini, R. Ghatikar, R. Diamond, Using dashboards to improve energy and comfort in federal buildings, LBNL-4283E (2011). doi:10.2172/1007230.
[72] J. Granderson, M.A. Piette, G. Ghatikar, P. Price, Preliminary findings from an analysis of building energy information system technologies, 2009 Natl. Conf. Build. Comm., 2009.
[73] David Allen Lotz, Energy dashboard for real-time evaluation of a heat pump assisted solar thermal system, Open Access Theses 790 (2016) https://docs.lib.purdue.edu/open_access_theses/.
[74] J. Sanyal, J. New, Simulation and big data challenges in tuning building energy models, in: 2013 Work. Model. Simul. Cyber-Physical Energy Syst. MSCPES 2013, 2013, pp. 1–6, doi:10.1109/MSCPES.2013.6623320.
[75] J.R. New, J. Sanyal, M. Bhandari, S. Shrestha, Autotune e+ building energy models, IBPSA-USA J. 5 (1) (2012) 270–278.
[76] R.E. Edwards, J. New, L.E. Parker, Predicting future hourly residential electrical consumption: a machine learning case study, Energy Build 49 (2012) 591–603, doi:10.1016/j.enbuild.2012.03.010.
[77] K. Field, M. Deru, D. Studer, Using DOE commercial reference buildings for simulation studies, Proc. SimBuild 4 (1) (2010) 85–93.
[78] J. New, J. Sanyal, B. Slattery, A. Gehl, W. Miller, A. Garrett, Big data mining for assessing calibration of building energy models, Int. J. Comput. Softw. Eng 3 (2018), doi:10.15344/2456-4451/2018/136.
[79] J. Hoving, G. Boxem, W. Zeiler, Model predictive control to maintain ATES balance using heat pump, in: 12th IEA Heat Pump conference (HPC2017), 2017, pp. 1–10.
[80] K.K. Gopinathan, A. Soler, Effect of sunshine and solar declination on the computation of monthly mean daily diffuse solar radiation, Renew. Energy. 7 (1996) 89–93, doi:10.1016/0960-1481(95)00108-5.
[81] M. Mohandes, S. Rehman, T.O. Halawani, Estimation of global solar radiation using artificial neural networks, Renew. Energy 14 (1998) 179–184, doi:10.1016/S0960-1481(98)00065-2.
[82] I.T. Togrul, H. Togrul, D. Evin, Estimation of monthly global solar radiation from sunshine duration measurement in Elazig, Renew. Energy 19 (2000) 587–595, doi:10.1016/S0960-1481(99)00084-1.
[83] M. Mohandes, A. Balghonaim, M. Kassas, S. Rehman, T.O. Halawani, Use of radial basis functions for estimating monthly mean daily solar radiation, Sol. Energy 68 (2000) 161–168, doi:10.1016/S0038-092X(99)00071-7.
[84] K.S. Reddy, M. Ranjan, Solar resource estimation using artificial neural networks and comparison with other correlation models, Energy Convers. Manag 44 (2003) 2519–2530, doi:10.1016/S0196-8904(03)00009-8.
[85] A. Sözen, E. Arcaklioglu, M. Özalp, Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data, Energy Convers. Manag 45 (2004) 3033–3052, doi:10.1016/j.enconman.2003.12.020.
[86] A. Sözen, E. Arcaklioğlu, M. Özalp, N. Çağlar, Forecasting based on neural network approach of solar potential in Turkey, Renew. Energy 30 (2005) 1075–1090, doi:10.1016/j.renene.2004.09.020.
[87] S. Rehman, M. Mohandes, Artificial neural network estimation of global solar radiation using air temperature and relative humidity, Energy Policy 36 (2008) 571–576, doi:10.1016/j.enpol.2007.09.033.
[88] A. Azadeh, A. Maghsoudi, S. Sohrabkhani, An integrated artificial neural networks approach for predicting global radiation, Energy Convers. Manag 50 (2009) 1497–1505, doi:10.1016/j.enconman.2009.02.019.
[89] M.A. Behrang, E. Assareh, A. Ghanbarzadeh, A.R. Noghrehabadi, The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data, Sol. Energy 84 (2010) 1468–1480, doi:10.1016/j.solener.2010.05.009.
[90] W. Chun, A hybrid solar lighting device using solar tracking dish concentrators and auxiliary lighitng sources, Korean Patent No. 1013049590000 (2013) http://kpat.kipris.or.kr/kpat/biblioa.do?method=biblioFrame.
[91] D. Lingfors, T. Volotinen, Illumination performance and energy saving of a solar fiber optic lighting system, Opt. Express. 21 (2013) A642 –55, doi:10.1364/OE.21.00A642.
[92] J. de Boer, Modelling indoor illumination by complex fenestration systems based on bidirectional photometric data, Energy Build 38 (2006) 849–868 https://doi.org/10.1016/j.enbuild.2006.03.018.
[93] National Solar Radiation Database, National Renewable Eneryg Laboratory., (n.d.). https://nsrdb.nrel.gov/ (Accessed 25 April 2019).
[94] T.G. Bowles, Simulating irradiance variability in an Urban Environment using geographic information systems and quantifying impacts on the energy and economic performance of rooftop photovoltaics MS thesis, Mechanical Engineering department, The University of Utah, 2017.
[95] Sarah M., What is weather intelligence^{TM} plus?, (n.d.). https://support.rachio.com/hc/en-us/articles/360002590533 (Accessed 14 May 2019).
[96] U. Wang, How Google is using big data to protect the environment, Guard (2016) https://www.theguardian.com/sustainable-business/2016/oct/12/google-environmental-sustainability-data-kate-brandt.