BIM-based machine learning engine for prediction of building energy consumption

Sisi Bie[1*] and Haijiang LI[1]
[1]BIM for Smart Engineering Centre, Cardiff University, Cardiff, Wales, UK
* email: Bies@cardiff.ac.uk

Abstract

Buildings consume a large amount of energy; it is therefore important to find a better way to manage building energy performance more effectively. The recent developments in data science and machine learning show a good potential to leverage the different sources of data to predict energy consumption. This paper reviews the current research in these areas and proposes a BIM (Building Information Modelling) based approach to provide data and information to train the relevant machine learning (ML) engine. The method has been demonstrated with a daylight illuminance prediction analysis. The presented preliminary result can potentially lead to a more generic tool, where rich data and information embedded in BIM models can be utilized for better decision making.

Keywords: BIM (Building Information Modelling), ML (Machine Learning), Building energy consumption.

1. Introduction

Building energy consumption takes a large part of the energy, for example it reported in Europe, the building energy use represents 40% of total energy consumption (Bull, Chang, & Fleming, 2012). The prediction of building energy consumption is therefore important to achieve energy conservation and reduce the environmental impact. However, there is lacing a better way to manage building energy performance more effectively.

BIM has been more widely used in the construction industry than ever before. More and more BIM models are accumulated for various industry organisations. The data embedded in BIM models include geometry, materials, energy, safety. Most of them are simply archived with little effort to find use out of it. The data, information and knowledge embedded in those BIM models can be leveraged to conclude knowledge to benefit the organisations, but how to leverage the data to get benefits remains unclear. The recent advances in machine learning could potentially help resolve this issue, e.g. machine learning methods have been used to clash detection and so on (Hu & Castro-Lacouture, 2018). artificial intelligence method could be very quick as it is based on 'searching' for pre-trained solutions. Also, comparing to the traditional way of building energy analysis, the values from the ML engine are more representative of actual building performance than simulation values (Chen, 2019). The training data sources used for ML can come from real historical data and simulation-based mocked data. Traditional building energy simulation during the design stage is time-consuming and often comes with biased errors. The BIM way of working is to integrate energy simulation data into BIM models; hence it can provide comprehensive life cycle data sets, which can create an automated energy controller in the lifecycle of a building. Therefore, the BIM can provide data to ML for solving the specific problems.

Usage of artificial light takes a large part of energy consumption in a building. Artificial lighting is controlled by the prediction of daylight illumination to reduce electric consumption. The traditional method is to place sensors to the reference point to collect data all the time. However, the sensor is an expenditure and can be easily removed from the reference point when collecting the daylight illuminance data. Also, the simulation model by the software can simulate the operation status of this building, but it needs to rebuild the model every time. Hence it cannot realise the real-time illuminance prediction engine.

This paper therefore proposes a BIM (Building Information Modelling) based approach to provide data and information to train the relevant machine learning (ML) engine, which is further used for real-time illuminance prediction. The overall contents include 4 sections, (1) Introduction; (2) Literature Review; (3) Development and Implementation; (4) Conclusion. The Literature Review covers BIM supporting building energy analysis, ML for building energy consumption and using BIM and ML supporting construction industry applications. the authors propose and develop a prototype of BIM-based ML engine and tested via a simple case study. Future work has been discussed at the end of the paper.

2. Literature Review

In this section, published articles are reviewed from major databases, including ScienceDirect, Scopus, Google Scholar. The searching dates are from 2010 to July 2019. Keywords used for searching were: 1) "Building information modelling" and "Building Energy" and "data"; 2) "Machine learning" and "Building Energy"; 3) "Building information modelling" and "Machine learning". The inclusion criteria are: 1) The studies published in English or in other languages with English abstracts; 2) Only search research articles; 3) All the keywords should be included. The exclusion criteria were: 1) Uncorrelated 2) Outdated before 2010. The result has been illustrated in three fields: 1) BIM applied to energy simulation 2) ML for building energy consumption;3) ML applied with BIM. It can be seen form figure 1 that the number of studies on ML for building energy consumption area are significantly increased over years, which can reflect the truth that energy is a big concern in the word. BIM-based building energy analysis has a moderate growth, comparing the ML for building energy analysis, the quantity of studies on this area is less. The findings are limited in BIM and ML combined area.

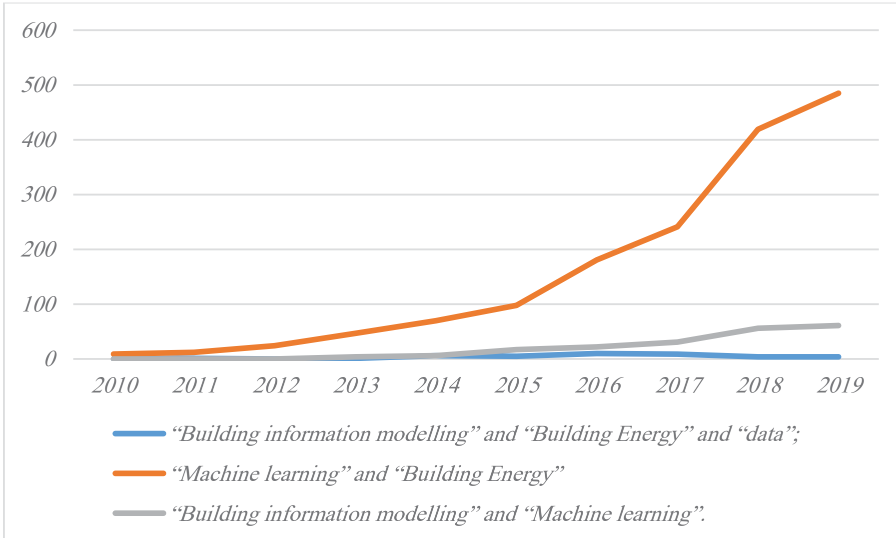

Figure 1: The Quantity of Publications in Different Domain

2.1 BIM supporting building energy analysis

BIM is a method monitoring the whole life cycle of a construction (Eleftheriadis, Mumovic, & Greening, 2017). BIM captures multi-dimensional CAD information (Eadie, Browne, Odeyinka, McKeown, & McNiff, 2013). This digital revolution boosts the development of the AEC industry. The data can be easily accumulated and collected by the integration of BIM models. The concept of BIM is referred by the Krygiel et al. (2008) to an integrated database which stored all parametric and interconnected information of the entire building, and design documents. It will be reflected instantly throughout the rest of the project in all views if there are any changes to an object in the model. Additionally, as more systems have been added to buildings, the more energy is demanded to operate them.

There are many contributions of BIM in building energy domain, such as automation of energy modelling, enhancing the existing libraries, and storing and organizing the building data (Kamel & Memari, 2019). The main format of the BIM file is gbXML and IFC, which are generated from the BIM tool such as Revit. The research in 2015 developed a ModelicaBIM library for BIM-based building energy simulation (Kim, Jeong, Clayton, Haberl, & Yan, 2015). The ModelicaBIM library was used to investigate the system interface between BIM and energy simulation. This system interface can semi-automatically translate the building models in BIM to building energy models.

2.2 ML supporting energy consumption prediction

Recently, the ML method is used in many areas, such as the prediction of building energy consumption. There are hundreds of studies on this domain. Aimed at reaching the goal of energy-saving and reduction of environmental influence, it is vital to predict the building energy consumption to enhance the performance of the building.

For example, in 2012, a review study summarized that there are four main categories of methods in building energy prediction: The engineering methods, Statistical methods regression, AI(Artificial intelligence) methods, Grey model (Zhao & Magoulès, 2012). The AI method mainly includes artificial neural networks, support vector machines, decision tree etc. From the review study in 2018, a summary of energy consumption prediction models was proposed from four aspects: scoping data properties, algorithms, and performance (Amasyali & El-Gohary, 2018).

The artificial neural network (ANN) is the popular AI method which was used in building energy management (Bilal et al., 2016). This method is performed well and effective in solving non-linear and complex problems. The input data is the most important factor in ANNs method. Because of the noise of raw data, the data pre-processing technologies are developed by many researchers. In 2018, research conducted the machine learning method to forecast the usage of energy in a building (Dan & Phuc, 2018). Their dataset comes from historical data and the building design parameters.

2.3 The BIM-based ML development

BIM's method is top-down modelling of information while machine learning is a data-driven bottom-up approach that can help to identify structure and semantics in data. Hence, BIM developed the 'knowledge discovery' by providing information right out of the box. There are many studies on the application of machine learning in building design. However, there are only a few pieces of research on knowledge discovery by using the advantage of BIM model.

Many ML methods were fine-tuned used in different fields. Looking back over the last ten years, the recognition of project progress in construction sites was conducted with ML methods based on the data from photos in sites and BIM models (Golparvar-Fard, Peña-Mora and Savarese, 2012). Then an integrated framework related on coordinating sensor and BIM element was proposed by Bogen etc. (Bogen, Rashid, East, & Ross, 2013), which was aimed to comparing the running state with the scheduled state of facilities in buildings. K-means clustering and hierarchical clustering were used for classifying the resource usage according to a resolution typical of human-specified schedules. Then the experiment in 2014 used deep learning method to classify 3D models under the environment of BIM, which had a good result (Qin, Li, Gao, Yang, & Chen, 2014). In 2015, the research was focused on semi-automation and automation of collecting and dealing with various photos from infrastructure construction sites (Teizer, 2015). It used the image recognition techniques to capture, analyse and record the process of construction. Based on the previous research, a study in 2016 got a good result of 3D façade modelling and materials recognition through photo recognition for as-built building (Yang, Shi, & Wu, 2016). The next year, a basic clash detection in the construction safety domain was proposed using ML (Tixier, Hallowell, Rajagopalan, & Bowman, 2017). At the same time, based on data in the construction material library (CML) and development of BIM, a web-based platform to simplify the process of data collection was used to annotate material patches according to BIM overlays (Han & Golparvar-Fard, 2017). In 2018, an application of machine learning used for semantic enrichment BIM models has been proposed for extensive data pre-processing (Bloch & Sacks, 2018a). The paper in the same year showed that the ML method is directly applicable to space classification problems(Bloch & Sacks, 2018b). Similarity, ML technologies could distinguish the relevant and irrelevant clashes which used to enhance the quality of clash detection (Hu & Castro-Lacouture, 2018). In 2019, the researcher proved that the predicted values from ML engine are more representative of actual building performance than simulation values (Chen, 2019). At the same time, a deep convolutional neural network was used in indoor localization, which could recognise the synthetic images from 3D indoor model recognise (Acharya, Khoshelham, & Winter, 2019). A research proposed an AI method to generate proper building design according to the requirement of client, but the researcher only realised automation of a window design(Karan & Asadi, 2019). Also, NLP and unsupervised learning were used to automatically distinguish whether it is the BIM case study(Jung & Lee, 2019). According to the previous works, the main application and implication of BIM-based ML are mentioned. This cross-domain requires further research into the potential of BIM-based ML using.

3. Development and Implementation

3.1 System design

Based on the reviewed published research and research gaps, a prototype of a combination of BIM-based ML engine for building energy consumption is proposed and showed in figure 2. BIM files produced by BIM tools, such as Revit, can be input in Graphical User Interface (GUI), according to the different file format. Then, the modified file can be mapped in the energy simulation engine, such as Energyplus. After that, the simulated energy data can be fed into the ML engine, which is the most key step in this model. So many ML algorithms can be used to make building energy prediction. In addition, the output of previous steps can be stored back in the smart BIM framework for controlling the whole life cycle of buildings. A smart BIM model is then developed from original BIM models.

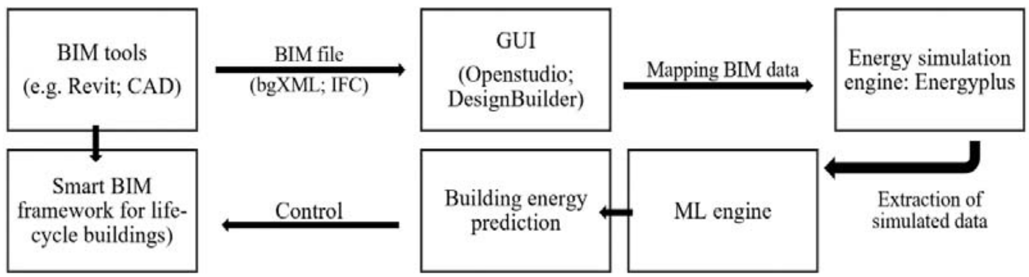

Figure 2: A prototype of BIM-based ML engine for energy consumption

3.2 Case study

The target of energy consumption contains many aspects, this case study starts from a simple target: daylight illuminance prediction. Designbuilder can easily export the idf file, which can be imported to Energyplus software directly. Energyplus here is being used for dataset simulation. Then, machine learning algorithms will be used to train the model. Once the machine learning model is trained, a real-time daylight luminance prediction can be made based on the model, given the future weather data for an existing building, which can be used to control the artificial light automatically in a building. Here are the steps in this case study.

3.2.1 Create BIM model using BIM software

A BIM model for a normal terraced house located in Cardiff was built in Designbuilder, which is shown in Figure 3. below. It is a two-layers classic single-family house in the UK. Here the living room on the ground floor has been used for daylight analysis.

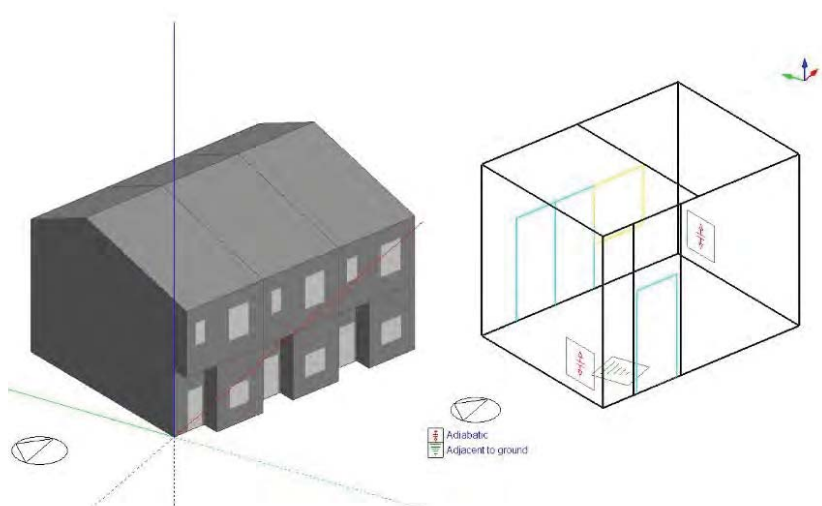

Figure 3: 1) the terraced single-family house; 2) the layout of the living room in this house.

A simulated sensor was set in the middle of the room, where the height is 0.8 meters. Then a one-year period from 1 January to 31 December was chosen to simulate. This house is located in the Cardiff area. The simulated result can be seen in figure 4. From April to October, the general lighting indicated by the blue line is reduced because of the longer daytime (yellow line) than other time, because of the typical UK weather format.

Figure 4: The visible one-year simulated result in Designbuilder.

3.2.2 Extract energy data from a BIM model

According to Energyplus input and output reference document, there are 10 output variables, that can be modified: hour of the day, outdoor Air Drybulb Temperature [C], Site Outdoor Air Humidity Ratio [kgWater/kgDryAir], Site Wind Speed [m/s], Site Diffuse Solar Radiation Rate per Area [W/m2], Site Direct Solar Radiation Rate per Area [W/m2], Site Solar Azimuth Angle [deg], Site Solar Altitude Angle [deg], Zone Windows Total Transmitted Solar Radiation Rate [W], Daylighting Reference Point 1 Illuminance [lux]. The output dataset of Energyplus can be shown in an xlsx or csv format file in excel, some of which are shown in figure 5 below.

	A	B	C	D	E	F	G	H	I	J
1	Date/Time	Environment:Site Outdoor Air Drybulb Temperature [C](Hourly)	Environment:Site Outdoor Air Humidity Ratio [kgWater/kgDryAir](Hourly)	Environment:Site Wind Speed [m/s](Hourly)	Environment:Site Diffuse Solar Radiation Rate per Area [W/m2](Hourly)	Environment:Site Direct Solar Radiation Rate per Area [W/m2](Hourly)	Environment:Site Solar Azimuth Angle [deg](Hourly)	Environment:Site Solar Altitude Angle [deg](Hourly)	3145:Zone Windows Total Transmitted Solar Radiation Rate [W](Hourly)	3145:Daylighting Reference Point 1 Illuminance [lux](Hourly)
2	01/01 01:00:00	8.525	0.006649	5.675	0	0	13.73314	-60.926	0	0
3	01/01 02:00:00	8.325	0.006496	5	0	0	39.31949	-56.7506	0	0
4	01/01 03:00:00	8.55	0.006536	5.8	0	0	59.28305	-49.6835	0	0
5	01/01 04:00:00	8.6	0.006679	6.575	0	0	74.78922	-41.1099	0	0
6	01/01 05:00:00	8.825	0.00705	8.2	0	0	87.64132	-31.9081	0	0
7	01/01 06:00:00	9.5	0.007469	6.45	0	0	99.14265	-22.6113	0	0
8	01/01 07:00:00	10.075	0.007538	5.25	0	0	110.1509	-13.5984	0	0
9	01/01 08:00:00	10.125	0.007731	4.35	0	0	121.2574	-5.20262	0	0
10	01/01 09:00:00	10.625	0.007902	4.475	14.25	11	132.8861	2.23384	2.509658	48.87171
11	01/01 10:00:00	10.725	0.007899	4.225	52.75	90.75	145.3104	8.338182	10.02342	185.5677
12	01/01 11:00:00	10.7	0.00814	3.35	81.75	224.75	158.6091	12.72221	16.34238	242.1758
13	01/01 12:00:00	10.625	0.007939	3.1	104.75	190.75	172.6063	15.04087	19.74009	278.5615
14	01/01 13:00:00	10.675	0.007755	3.475	116	128	186.8789	15.08278	20.48618	280.4431
15	01/01 14:00:00	11.15	0.007976	5.55	87.5	57.75	200.8946	12.84392	15.04453	201.7524
16	01/01 15:00:00	11.225	0.007829	5	46	0	214.2233	8.52915	7.800996	121.1626

Figure 5: Partly show the extracted dataset in Excel form

3.2.3 ML algorithms development

According to the hourly simulated dataset from Energyplus, the 9 input variables are chosen for the ML engine, while the only target variable is daylight illuminance. The figure 6 gives a clear view of 9 input variables and one output in ML engine. From the previous review, the main machine learning algorithms such as ANN and Random Forest (RF) used here only accept the consistent data type. Hence, the time step of record instance is hourly, the time can be assigned by 1 to 8760. There are a total of 8760 instances in this dataset. Weka is a commercial ML software based on C plus language, which is friendly with engineers. This dataset in excel can be fed into Weka software for normalization and training. 66% of the dataset is for training and the remainder are for testing.

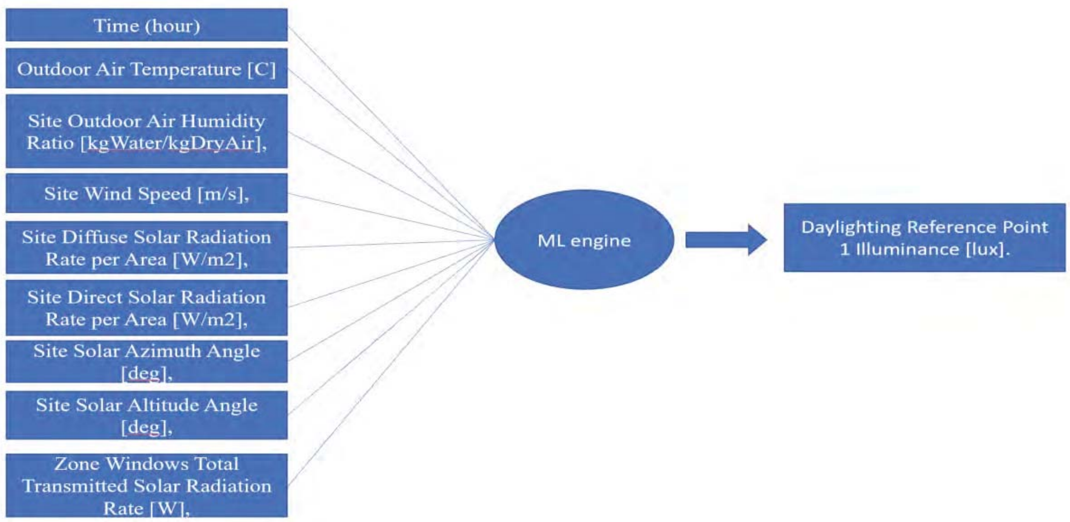

Figure 6: The input and output of machine learning engine

Multilayer Perceptron, Random Forest (RF), and Linear Regression are mainly algorithms used in this study. Multilayer Perceptron is a classic feedforward neuron network. Here the author uses one hidden layer and chooses 3,5 and 7 neurons per layer respectively. After comparing the results from different neurons per layer, the author finds that the neuron network with 5 neurons per layer performed better than other options here. Hence, the ANN built in this experiment has only one hidden layer and 5 neurons per layer. Random forest has a high degree of accuracy and the high speed of learning process. Here the bag size percent and batch size are both 100. The tree depth is unlimited here in each tree of the random forest. Linear regression is used for comparison with other two methods here. Cross-validation rank in 10 for detecting the errors. The experiment compares these three methods from five aspects: correlation coefficient, mean absolute error, root mean squared error, relative absolute error, root relative squared error. The result is illustrated in table 1. The linear regression has higher errors than others. The reason for this situation is that this dataset is highly scattered.

Table 1: The Result of Each Algorithms

	Correlation coefficient	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error	Total Number of Instances
Multilayer Perceptron (1 hidden layer, 5 neurons per layer)	0.9948	21.6228	30.5761	9.15%	10.51%	8760
Random Forest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1	0.9991	6.3013	12.5186	2.67%	4.30%	8760
Linear Regression	0.9531	52.6389	88.016	22.28%	30.26%	8760

3.2.4 Discussion

Form the result in table 1, firstly, the Correlation coefficient of these algorithms presented here nearly equals to 1. It indicates that the input variables have high Correlation with the output variables. The daylight illuminance is mainly affected by the time of day and other weather conditions, also there are other effects like surrounding buildings and trees. Secondly, there are other results in the table. 2, which include Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error. All of these errors are similar to measuring the error between real value and prediction value. The RF has obvious smaller errors than the other two algorithms, which shows a significant advantage than the other two in the regression problem for daylight illuminance prediction. Additionally, form the structure of the dataset, the number of 0 value of the daylight illuminance in the wintertime is more than in the summertime. It can reflect the weather condition in the UK, where the daytime in winter is shorter than summer. Thus, the prediction of future daylight illuminance through the weather forecast is feasible for an existing real building. It can also be used in other aspects of the energy area.

4. Conclusion

The authors presented a simple case with limited data in this study to demonstrate how to extract energy data from the BIM model to predict the real-time daylight illuminance using the ML engine in the life cycle of the building. It shows the feasibility of using the weather forecast to predict the future daylight illuminance for an existing real building. Additionally, it also presents the feasibility of deducing the implicit knowledge from BIM model. There are other effects of input variables that were not considered comprehensively, such as other shelters near the building etc. Based on the existing problem in the energy consumption filed, this engine can be extended to other aspects in the energy area. It is therefore can realize smart energy management in buildings.

References

Acharya, D., Khoshelham, K., & Winter, S. (2019). BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *150*(March), 245–258. https://doi.org/10.1016/j.isprsjprs.2019.02.020

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*(September 2017), 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., … Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

Bloch, T., & Sacks, R. (2018a). Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models. *Automation in Construction*, *91*(March), 256–272. https://doi.org/10.1016/j.autcon.2018.03.018

Bloch, T., & Sacks, R. (2018b). Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models. *Automation in Construction*, *91*(July 2017), 256–272. https://doi.org/10.1016/j.autcon.2018.03.018

Bogen, A. C., Rashid, M., East, E. W., & Ross, J. (2013). Evaluating a data clustering approach for life-cycle facility control. *Journal of Information Technology in Construction*, *18*, 99–118.

Bull, R., Chang, N., & Fleming, P. (2012). The use of building energy certificates to reduce energy consumption in European public buildings. *Energy and Buildings*, *50*, 103–110. https://doi.org/10.1016/j.enbuild.2012.03.032

Carnot, N., Koen, V., Tissot, B., Carnot, N., Koen, V., & Tissot, B. (2016). Modelling Behaviour. *Economic Forecasting*, (buildingSMART 2014), 103–132. https://doi.org/10.1057/9780230005815_5

Chen, S. (2019). *Enhancing Validity of Green Building Information Modeling with Artificial-neural-network-supervised Learning — Taking Construction of Adaptive Building Envelope Based on Daylight Simulation as an Example*. *31*(6), 1831–1845.

Dan, T. X., & Phuc, P. N. K. (2018). Application of Machine Learning in Forecasting Energy Usage of Building Design. *Proceedings 2018 4th International Conference on Green Technology and Sustainable Development, GTSD 2018*, 53–59. https://doi.org/10.1109/GTSD.2018.8595595

Eadie, R., Browne, M., Odeyinka, H., McKeown, C., & McNiff, S. (2013). BIM implementation throughout the UK construction project lifecycle: An analysis. *Automation in Construction*, *36*, 145–151. https://doi.org/10.1016/j.autcon.2013.09.001

Eleftheriadis, S., Mumovic, D., & Greening, P. (2017). Life cycle energy efficiency in building structures: A review of current developments and future outlooks based on BIM capabilities. *Renewable and Sustainable Energy Reviews*, *67*, 811–825. https://doi.org/10.1016/j.rser.2016.09.028

Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2012). Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. *Journal of Computing in Civil Engineering*, *29*(1), 04014025. https://doi.org/10.1061/(asce)cp.1943-5487.0000205

Han, K., & Golparvar-Fard, M. (2017). Crowdsourcing BIM-guided collection of construction material library from site photologs. *Visualization in Engineering*, *5*(1). https://doi.org/10.1186/s40327-017-0052-3

Hu, Y., & Castro-Lacouture, D. (2018). Clash Relevance Prediction Based on Machine Learning. *Journal of Computing in Civil Engineering*, *33*(2), 04018060. https://doi.org/10.1061/(asce)cp.1943-5487.0000810

Jung, N., & Lee, G. (2019). Automated classification of building information modeling (BIM) case studies by BIM use based on natural language processing (NLP) and unsupervised learning. *Advanced Engineering Informatics*, *41*(April), 100917. https://doi.org/10.1016/j.aei.2019.04.007

Kamel, E., & Memari, A. M. (2019). Review of BIM's application in energy simulation: Tools, issues, and solutions. *Automation in Construction*, *97*(October 2018), 164–180.

https://doi.org/10.1016/j.autcon.2018.11.008

Karan, E., & Asadi, S. (2019). Intelligent designer: A computational approach to automating design of windows in buildings. *Automation in Construction*, *102*(February), 160–169. https://doi.org/10.1016/j.autcon.2019.02.019

Kim, J. B., Jeong, W., Clayton, M. J., Haberl, J. S., & Yan, W. (2015). Developing a physical BIM library for building thermal energy simulation. *Automation in Construction*, *50*(C), 16–28. https://doi.org/10.1016/j.autcon.2014.10.011

Qin, F., Li, L., Gao, S., Yang, X., & Chen, X. (2014). A deep learning approach to the classification of 3D CAD models. *Journal of Zhejiang University SCIENCE C*, *15*(2), 91–106. https://doi.org/10.1631/jzus.c1300185

Teizer, J. (2015). Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites. *Advanced Engineering Informatics*, *29*(2), 225–238. https://doi.org/10.1016/j.aei.2015.03.006

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2017). Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining. *Automation in Construction*, *74*, 39–54. https://doi.org/10.1016/j.autcon.2016.11.001

Yang, J., Shi, Z. K., & Wu, Z. Y. (2016). Towards automatic generation of as-built BIM: 3D building facade modeling and material recognition from images. *International Journal of Automation and Computing*, *13*(4), 338–349. https://doi.org/10.1007/s11633-016-0965-7

Zhao, H., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592. https://doi.org/10.1016/j.rser.2012.02.049