2018 Building Performance Modeling Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA
Chicago, IL
September 26-28, 2018

An Intelligent Knowledge-based Energy Retrofit Recommendation System for Residential Buildings at an Urban Scale

Usman Ali[1], Mohammad Haris Shamsi[1], Cathal Hoare[1], Eleni Mangina[2] and James O'Donnell[1]
[1]School of Mechanical & Materials Engineering, Energy Institute
University College Dublin (UCD), Ireland
[2]School of Computer Science and Informatics, University College Dublin (UCD), Ireland

ABSTRACT

Buildings play a significant role in driving the urban demand and supply of energy. Research conducted in the urban buildings sector indicates that there is a considerable potential to achieve significant reductions in energy consumption and greenhouse gas emissions. These reductions are possible through retrofitting existing buildings into more efficient and sustainable buildings. Building retrofitting poses a huge challenge for owners and city planners because they usually lack expertise and resources to identify and evaluate cost-effective energy retrofit strategies. This paper proposes a new methodology based on machine learning algorithms to develop an intelligent knowledge-based recommendation system which has the ability to recommend energy retrofit measures. The proposed methodology is based on the following four steps: archetypes development, knowledge-base development, recommendation system development and building retrofitting or performance analysis. A case study of Irish buildings dataset shows that the proposed system can provide effective energy retrofits recommendation and improve building energy performance.

INTRODUCTION

The built environment accounts for a significant portion of the overall energy demand and greenhouse gas emissions of any country. For instance, the energy consumption by the buildings sector in the US accounts for more than 70%of total consumption (EESI 2018). Similarly, in Europe, buildings are responsible for 40% of overall energy con-sumption (EU-Energy 2018). Furthermore, around 39%and 36% of the CO_2 emissions are associated with build-ings in the US and Europe respectively (EU-Energy 2018; EESI 2018). Also, statistics show that around 35% of buildings in the EU are over 50 years old (EU-Energy 2018). Hence, there exists a considerable potential to achieve significant reductions in energy consumption and emissions for the building sector; these reductions can be achieved through the transformation of existing energy systems into more efficient and sustainable systems. It has been suggested in the literature that, through the im-plementation of energy efficient measures, it is possible to reduce the overall energy consumption and CO_2 emis-sions in the EU by approximately 5% (EU-Energy 2018).

As buildings play a significant role in urban demand and supply of energy, improving the efficiency of existing building stock is vital to reducing the consumption and emissions. Existing buildings were built without sustainability or energy efficiency considerations in mind. The bigger challenge is to identify, implement, and apply the cost effective retrofit option for achieving better energy performance. Research conducted so far in the energy retrofit domain focuses on retrofit tools for commercial buildings (Lee et al. 2015), while limited work has been done in the area of retrofit tools for the urban residential sector.

In past few years, many national and international organizations have put great effort towards energy efficiency improvements in the existing buildings (Ma et al. 2012). However, the implementation of these retrofit strategies faces a lot of issues such as huge financial investments, governmental policies and lack of appropriate knowledge etc. A significant amount of literature has demonstrated the process to investigate and apply various energy efficiency techniques to improve existing buildings' energy performance (Asadi et al. 2012; Flourentzou and Roulet 2002). The results identified that proper implementation of retrofitting strategies could be helpful in reducing en-ergy consumption and greenhouse gas emissions if and only if appropriate information was available about the building stock (Mahlia, Razak, and Nursahida 2011; Hes-tnes and Kofoed 2002; Chidiac et al. 2011). The major-ity of the knowledge on retrofitting was acquired through conducting online surveys or provided by the construction companies. However, the information gathered is often case specific and is limited by it's applicability to different environments. As such, it becomes difficult for building owners to make informed and intelligent decisions as to which retrofit measure would result in a cost optimal and minimum energy solution.

Before applying the energy retrofits, knowledge of ex-isting building energy performance is required. There-fore, Energy Performance Certification (EPC) provides a means of energy rating individual buildings including res-idential, commercial or public sector. However, the EPC calculation process requires extensive information and de-pends on many factors including climate, location, build-ing envelope, construction materials and heating, cooling and ventilation systems (IEA 2010). The EPC also provides recommendation that help the building owners and occupiers in improving the energy efficiency of the building (UK and Government 2017). Generally, the EPC recommendation reports cover non-domestic buildings (UK and Government 2017). Recently, the Sustainable Energy Authority of Ireland (SEAI) has started a deep retrofit pi-lot programme to improve the energy efficiency in the housing sector. However, the absence of an efficient knowledge-base poses a lot of challenges and opportunities in the residential sector (SEAI 2017).

This paper proposes a generalized methodology for designing an intelligent knowledge-based energy retrofit recommendation system for the urban residential buildings. The main goal of this research is to develop a recommendation system that helps stakeholders to decide the best retrofit measures for improving the building energy performance. Recommendation systems are now being extensively used in a variety of areas including e-commerce websites, books, movies, music, research articles and other digital products (Park et al. 2012; Portugal, Alencar, and Cowan 2017). The following sections describe the key steps identified in the development of the retrofit recommendation system followed by the results and discussion and conclusions.

METHODOLOGY

The methodology reflects the energy use of a building through building energy and retrofit recommendation at an urban scale shown in Figure 1. Furthermore, the methodology identifies and recommends retrofit solutions in terms of energy and cost savings. This methodology can be categorized into the following four phases.

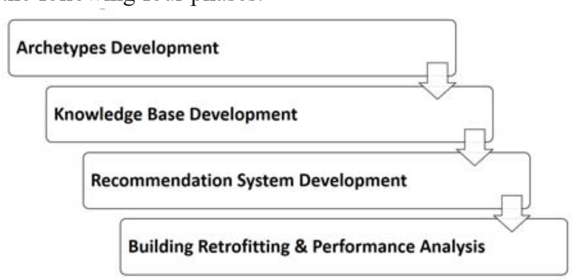

Figure 1: Methodology for an intelligent knowledge-based energy retrofit recommendation system.

Archetypes development

Buildings possessing similar characteristics are usually grouped together to represent a large building stock and are termed as archetypes or reference buildings. (Galante et al. 2012; Famuyibo, Duffy, and Strachan 2012). Building archetypes concept is most commonly used in energy modeling at the urban scale. The reference buildings concept has been used for building regulation (Recast 2010).

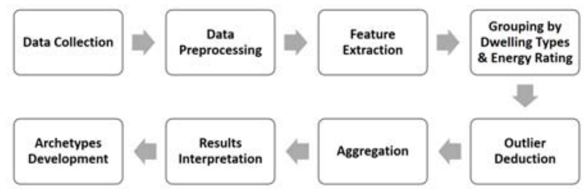

Figure 2: Methodology for archetypes development.

The typification of the building stock, as buildings typolo-gies or reference buildings, can be categorized through three different methodological approaches, namely, real example, real average and synthetical average building (de Vasconcelos et al. 2015; Corgnati et al. 2013; Ballar-ini, Corgnati, and Corrado 2014; Sousa Monteiro et al. 2015). In the real example building approach, the selec-tion of the building type is done by means of experience; the building type is selected on the basis of the experience of experts panel and other sources of information within an actual climatic context. The other information sources normally include the most commonly used materials, spe-cific size and construction period classes. In the real aver-age building approach, the selection of the building type is performed through the statistical analysis of a large building sample data. In synthetical average building approach, the archetype selection is performed by using the information on the most commonly used materials and systems. Several projects (de Vasconcelos et al. 2015), at the Euro-pean and international levels, are being developed in or-der to define the building stocks, for instance, TABULA (Loga, Diefenbach, and Stein 2012), ASIEPI (Intelligent Energy Europe 2018), BPIE (European Union 2018) and DOE (US Department of Energy 2018). In this paper, real average building approach is used to build the list of recommendations for any new building. The archetypes are developed using machine learning al-gorithms rather than traditional qualitative techniques to generate an accurate representation of the overall build-ing stock. The methodology of archetypes development includes the following sequential steps: data prepossess-ing, feature selection, outlier detection and aggregation as shown in Figure 2. Data preprocessing is a machine learn-ing technique that involves transforming raw or real-world data into an understandable format. In preprocessing, the data goes through a series of steps such as data cleaning, data integration, data transformation, data reduction and data discretization. Feature selection is the process of selecting a subset of most relevant variable or attributes for the model. The feature selection method aims to remove irrelevant and redundant attributes to get accurate results. Outlier detection is the process of identification of observations in the data that deviates by a significant amount from a given set of data. The most common outlier detection techniques are distance-based, density-based and lo-cal outlier factor (LOF). In this paper, the LOF algorithm is used for detecting the outliers (Breunig et al. 2000). The aggregation process involves categorizing the data into groups and then applying arithmetic or geometric mathematical operations. The obtained aggregated value represents the characteristics of one building archetype.

Knowledge-base development

A knowledge-based system (KBS) is a technique that reasons and uses a knowledge base to solve complex problems. In KBS, the data is stored mostly in the hierarchical or relational database. To enhance the practicality and obtain more realistic results, the expert opinion is also stored in the database such as retrofits measures obtained through existing published surveys. There are two types of knowledge-based systems based on the database type, namely, static and dynamic. In static KBS, all the knowledge is collected from the existing literature, survey, and reports. In dynamic KBS, all the knowledge is collected from real-time data such as smart meters. Static KBS suffices for the development of an overall building retrofit recommendation system. However, to design a deep retrofit recommendation system, a dynamic KBS is required as it would use enhanced details to provide recommendations. Dynamic KBS is often difficult to develop due to the non-availability of data. To counter this lack of data, urban building energy modeling (UBEM) is the best solution for their development.

Recommendation system development

There are three types of Recommendation Systems (RS), namely, content-based, collaborative and intelligent knowledge-based RS (Park et al. 2012; Lafta et al. 2016). The content-based RS is based on the description of an item and a profile of the users preferences. The collab-orative RS uses collaborative filtering methods that are based on collecting and analyzing a significant amount of information from users' behaviors, activities and prefer-ences. The intelligent knowledge-based RS is based on machine learning algorithms such as decision tree (DT), Nave Bayes (NB) and k-nearest neighbor (kNN) and deep learning algorithms for recommendation.
In this paper, a intelligent knowledge-based RS develop-ment approach is used that uses a predictive model. The methodology diagram is shown in Figure 3. The RS takes different inputs including a predictive model, archetypes recommendation, knowledge base and new building infor-mation. The RS' primary goal is to predict the energy rat-ing and recommend a list of retrofit measures for the new

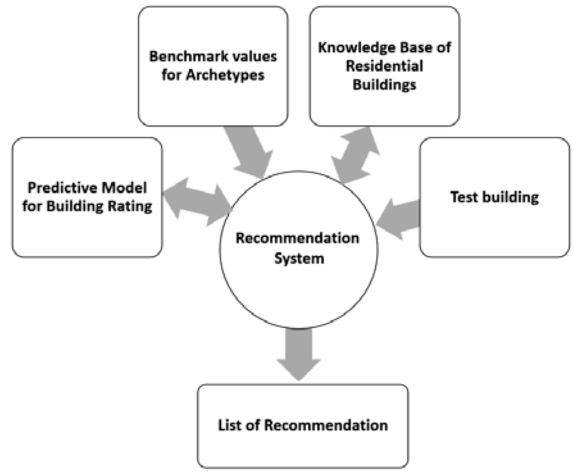

Figure 3: Methodology for recommendation system devel-opment.

buildings. For the predictive model, supervised machine learning deep learning algorithm is used.

A deep learning algorithm allows for higher levels of ab-straction with higher prediction accuracy. A deep learning algorithm is based on a hidden multi-layer of an artifi-cial neural network that is trained with a set of propaga-tion formulas. Generally, the deep learning architecture is composed of layers of parallel neurons with each layer's neuron connected with other layer's neurons forming a mesh of input, hidden, and output layers as depicted in Figure 4. The inputs for the algorithms include the train-ing dataset, hidden layer sizes, output labels and epochs for the dataset.

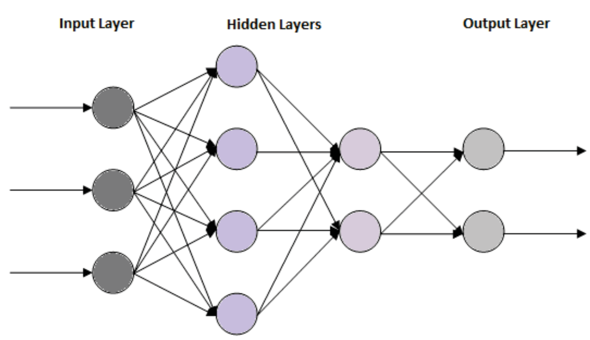

Figure 4: Deep learning architecture with input, hidden and output layers.

The methodology for developing the predictive model is shown in Figure 5 and includes steps such as data collec-tion for training and testing, data preprocessing, feature selection, outlier detection and deep learning modeling and testing. The validation of deep learning algorithms is done through cross-validation methods that divide the test and train data randomly into k equal size sub-samples. To examine the effectiveness of the deep learning pre-diction models, adopted performance indices such as the Root Mean Squared Error (RMSE), precision, recall, ac-curacy, and Classification Error (CE) are used; these are computed follows as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(c_i - \bar{c}_i)^2} \quad (1)$$

$$precision = \frac{TP}{TP+FP} \quad (2)$$

$$recall = \frac{TP}{TP+FN} \quad (3)$$

$$accuracy = \frac{TP+TN}{P+N} \quad (4)$$

$$CE = 1 - \frac{TP+TN}{P+N} \quad (5)$$

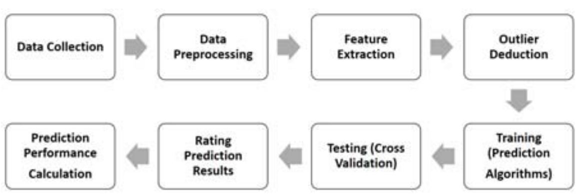

Figure 5: Methodology for deep learning predictive model for the recommendation system.

Building retrofitting and performance analysis

The final step in the process involves the retrofitting and performance analysis of the building on the basis of rec-ommendation results. To evaluate the potential for retrofit energy savings, it is necessary to analyze the current building energy performance. Therefore, the RS will eval-uate the existing building energy rating and recommend retrofit measures that will improve the building energy ef-ficiency and rating. Some specific examples of retrofit measures include replacing wall, ceiling, or adding roof insulation, upgrading the fuel type or the HVAC system, or upgrading the windows and walls with low U-values (Hong et al. 2015). However, the improvement in the energy rating is dependant upon the deployment of these retrofit measures for the specific building characteristics, such as building envelope, weather condition, heating, cooling and ventilation system properties.

RESULTS AND DISCUSSION

The main objective of this paper was to develop a build-ing energy retrofit recommendation system for an entire urban area. The devised methodology followed a two step process; the first step involved the calculation of the existing building performance in terms of energy rating, which was then used in the second step to suggest energy retrofits measures for improving the identified rating. The methodology presented above was applied to the publicly available Irish Building Energy Performance Certificate (EPC) data published by Sustainable Energy Authority of Ireland (SEAI). The EPC data contains the building's en-ergy performance or certificate rates in terms of the pri-mary energy consumption (kWh/m^2/year) and varies on a scale of A-G. An A-rated building has the highest en-ergy efficiency and will tend to have the lowest energy consumption and CO_2 emissions. On the other hand, a G-rated building is the least energy efficient. The EPC is calculated using Dwelling Energy Assessment Proce-dure (DEAP) software which is the official method used in Ireland for calculating the building energy rating of new and existing buildings. The EPC database contains more than 600,000 Irish buildings' data with 203 variables in-cluding building physics, energy, and CO_2 information. Dublin city was chosen as the use case for the evalua-tion of the recommendation system as the EPC building dataset contains 30% of Dublin's building stock (highest amongst all the cities in Ireland). Besides, the energy per-formance ratings exist for only 42% of Dublin's building stock, which calls for the need of a proper energy rating prediction framework.

In first step, archetypes development for the building retrofit recommendation, a total of 76 archetypes were de-veloped based on the dwelling types and building energy rating; these are tabulated in Table 1. After data collec-tion, the next step was to pre-process and organize the data such as replacing missing values by sample average, filtering out the irrelevant variables and less frequent val-ues using the standard deviation threshold. A total 73 rel-evant variables, out of a list of 203, are selected in the pre-processing and feature selection process. An outlier detection process was also run to enhance the quality of the filtered data. Outlier detection is an important step in machine learning that filters out the noisy or dis-similar information and eliminates all those inconsistencies from the data. The LOF algorithm was used for detecting the outliers by using euclidean distance with a lower bound (*MinPts*) of 10 and an upper bound (*MinPts*) of 20. This followed the aggregation process where the average val-ues of the relevant variables are considered with respect to building energy rating. The set of average values for

Table 1: *Number of buildings in each archetype based on dwelling types and building energy rating for the Dublin city*

Dwelling Types \ EPC Rating	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**Total**
Apartment	26	1023	666	288	155	76	68	2302
Basement Dwelling	0	5	14	11	22	15	40	107
Detached house	1042	1672	4966	3749	1931	925	1001	15286
End of terrace house	1373	1159	6719	4773	2527	1435	1625	19611
Ground-floor apartment	357	2626	5906	4405	3054	1228	1438	19014
House	39	488	1007	851	454	217	239	3295
Maisonette	66	758	1137	610	266	101	144	3082
Mid-floor apartment	1055	10012	10237	6110	2029	729	810	30982
Mid-terrace house	1831	3942	14402	9213	5454	2543	2510	39895
Semi-detached house	3177	2814	14425	15064	7508	3315	2187	48490
Top-floor apartment	359	2994	5472	4329	2655	945	1807	18561
Total	9325	27493	64951	49403	26055	11529	11869	200625

Table 2: *Confusion matrix of the deep learning predictive model for building energy rating prediction*

	true D	**true C**	**true E**	**true G**	**true F**	**true B**	**true A**	**class precision**
pred. D	4281	320	207	0	0	1	0	89.02%
pred. C	288	5878	1	0	0	227	0	91.93%
pred. E	287	0	2152	1	163	0	0	82.67%
pred. G	0	0	5	1020	70	0	0	93.15%
pred. F	2	0	185	123	891	0	0	74.19%
pred. B	0	218	0	0	0	2426	19	91.10%
pred. A	0	0	0	0	0	24	819	97.15%
class recall	88.12%	91.61%	84.39%	89.16%	79.27%	90.59%	97.73%	Accuracy= 89%

any archetype was used to develop the database for the recommendation system. Only static knowledge is used to develop the database.

In the RS development step, the deep learning algorithm was applied to the Dublin city EPC data. The modelling steps to create a learning model retain the order as used to create the archetypes. The learning model was then trained and tested using 10 cross-validations with ran-dom sampling. To improve the performance of the learn-ing algorithm, the adaBoost algorithm was used to per-form several simulations. To evaluate the performance of the model, learning algorithms were tested through sev-eral validity metrics such as the root mean squared error (RMSE), precision, recall, accuracy, and classification er-ror. The model was examined with a number of hidden layer neurons in the range between 50 and 140 to give a robust input to the model and evaluate the sensitivity of the deep learning algorithms. There doesn't exist gen-eral rule for selecting the number of hidden layers in deep learning. However, researchers have suggested that the hidden layers should be more than the number of inputs i.e. $n + 1$, or more than twice the number of inputs i.e. $2n + 1$, where n is the number of input variables (Hecht-Nielsen et al. 1988). The best results correspond to 74 hidden layers with 200,000 sample building data. The achieved model accuracy and RMSE were 89% and 0.343 calculated using the confusion matrix as shown in Table 2, which describes the performance of the formulated learn-ing model on the set of chosen test data.

Improving any building's energy performance requires the knowledge of existing energy rating of the building. The devised methodology lays out an informed process to identify the building's energy rating. Furthermore, the methodology helps in the development of benchmark buildings for comparing the energy performance of dif-ferent buildings. The developed archetypes are used as benchmarks and each archetype has a list of possible retrofit recommendations associated to improve the exist-ing building rating. For example, through improvements in the building material, for instances by upgrading the U-values of building elements such as windows, walls, roof and floor, a stakeholder can target any desired en-ergy rating. A detailed description of the recommended U-values with respect to the building dwelling type and energy rating is shown in Table 3. In this case, if the apart-ment building's current rating is "C", then building win-dows,floor, walls and roof U-values are 2.21, 0.13,0.43 and 0.09 (W/m^2K) respectively. To improve the build-

Table 3: List of U-value (W/m^2K) recommendations obtained using the Dublin city archetypes

Dwelling Types	Type	A	B	C	D	E	F	G
Apartment	Window	1.62	1.85	2.21	2.75	3.03	3.17	3.54
	Floor	0.04	0.07	0.13	0.18	0.30	0.42	0.57
	Walls	0.25	0.34	0.43	0.59	0.70	0.94	1.35
	Roof	0.06	0.08	0.09	0.12	0.24	0.35	0.87
Basement Dwelling	Window		2.01	2.87	3.23	3.58	3.35	3.76
	Floor		0.35	0.39	0.50	0.56	0.57	0.69
	Walls		0.27	0.72	0.96	1.38	1.37	1.66
	Roof		0.10	0.08	0.57	0.21	0.35	0.76
Detached house	Window	1.23	2.01	2.62	2.93	3.24	3.52	3.90
	Floor	0.14	0.35	0.50	0.58	0.66	0.71	0.76
	Walls	0.17	0.32	0.56	0.90	1.45	1.74	1.90
	Roof	0.14	0.22	0.32	0.43	0.71	0.97	1.43
End of terrace house	Window	1.29	1.94	2.67	3.02	3.25	3.59	4.08
	Floor	0.14	0.34	0.53	0.60	0.67	0.72	0.77
	Walls	0.17	0.32	0.44	0.73	1.34	1.75	1.93
	Roof	0.13	0.18	0.23	0.35	0.60	0.85	1.49
Ground-floor apartment	Window	1.32	1.94	2.50	2.79	2.93	3.17	3.52
	Floor	0.15	0.26	0.37	0.44	0.49	0.57	0.68
	Walls	0.20	0.37	0.51	0.67	0.83	1.03	1.40
	Roof	0.04	0.06	0.09	0.11	0.14	0.23	0.50
House	Window	1.54	1.87	2.54	3.01	3.43	3.54	4.09
	Floor	0.17	0.29	0.45	0.55	0.60	0.66	0.71
	Walls	0.19	0.30	0.54	0.92	1.47	1.77	1.82
	Roof	0.15	0.19	0.29	0.46	0.75	1.17	1.46
Maisonette	Window	1.28	2.16	2.66	2.83	2.95	3.13	3.50
	Floor	0.12	0.20	0.28	0.36	0.47	0.64	0.79
	Walls	0.19	0.39	0.52	0.65	0.91	1.11	1.62
	Roof	0.13	0.20	0.28	0.32	0.49	0.67	1.42
Mid-floor apartment	Window	1.34	2.07	2.60	2.81	3.06	3.30	3.61
	Floor	0.02	0.02	0.05	0.07	0.23	0.39	0.32
	Walls	0.19	0.44	0.58	0.72	0.92	1.13	1.45
	Roof	0.03	0.03	0.06	0.05	0.06	0.12	0.29
Mid-terrace house	Window	1.29	2.09	2.80	3.16	3.54	3.86	4.19
	Floor	0.14	0.33	0.44	0.51	0.56	0.60	0.63
	Walls	0.17	0.37	0.56	1.05	1.61	1.83	1.93
	Roof	0.13	0.20	0.27	0.47	0.85	1.34	1.79
Semi-detached house	Window	1.25	1.97	2.65	2.96	3.28	3.63	4.00
	Floor	0.14	0.35	0.50	0.55	0.63	0.68	0.72
	Walls	0.17	0.33	0.63	0.97	1.57	1.89	1.95
	Roof	0.14	0.21	0.32	0.44	0.67	0.94	1.47
Top-floor apartment	Window	1.26	1.99	2.52	2.73	2.92	3.20	3.63
	Floor	0.01	0.06	0.07	0.07	0.11	0.16	0.25
	Walls	0.18	0.38	0.51	0.61	0.81	1.08	1.57
	Roof	0.14	0.22	0.28	0.35	0.48	0.78	1.63

ing energy performance, the user has two options to upgrade to ”A” or ”B” Rating. To achieve sustainable building goals, the window U-values should change from 2.21 to 1.62 (W/m^2K) and wall U-values from 0.43 to 0.25 (W/m^2K) etc. Similarly, the user can implement any other retrofit measure such as using more efficient fuel or upgrading to an energy efficient heating, cooling, and ventilation system to improve the overall rating. The model accuracy of 89% signifies that with the implementation of any of the above measures, the energy use intensity would significantly reduce leading to an increase in the overall energy rating. For instance, the apartment building with ”C” energy rating has a primary energy consumption between 150 and 200 ($kWh/m^2/year$). The upgrade to an ”A” rated building would reduce the consumption to less than 50 ($kWh/m^2/year$) which corresponds to a reduction between 66% and 75% in the primary energy.

CONCLUSIONS

Determining the energy performance of any building is always a complex task as a number of interlinked fac-tors play a role in the performance evaluation. At a dis-trict scale, the complexity of the system is double folded and the performance evaluation process becomes resource intensive as a number of resources are required to per-form the analytical, extensive, and time-consuming simu-lations. The research conducted in this paper identifies a generalized methodology to identify the performance through machine learning approaches. The methodology identifies the existing building energy performance us-ing a predictive model and then recommends several en-ergy retrofit measures to increase that performance. As the results indicate, the proposed methodology is able to calculate the building’s current energy performance even with a limited knowledge of the building dynamics. For instance, the Irish EPC data used 200 variables to clas-sify buildings based on their energy rating. However, the proposed methodology required only 73 variables to pre-dict any building’s energy rating. The identified approach will allow stakeholder such as the building owners and city planners to make informed decisions when planning retrofit measures. As the obtained accuracy is quite high 89%, the measures suggested by the recommendation sys-tem will be realistic.

One of the major limitation of this methodology is the quality of data required for the knowledge base devel-opment. All recommendations should be present in the knowledge base before the implementation of the pro-posed methodology. Also, the system would never recom-mend any measures outside the available knowledge. Cur-rently, the methodology relies on static data for predicting the rating; the future work will consider the dynamic data obtained through urban energy simulation modeling ap-proaches. This would also allow for performing a parametric analysis to achieve more robust results. Furthermore, a cost analysis of implementing the retrofit measures would also be included in future work.

ACKNOWLEDGMENT

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under the SFI Strategic Partnership Programme Grant number SFI/15/SPP/E3125. The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Science Foundation Ireland.

REFERENCES

Asadi, Ehsan, Manuel Gameiro Da Silva, Carlos Henggeler Antunes, and Luís Dias. 2012. “Multi-objective optimization for building retrofit strategies: A model and an application.” *Energy and Buildings* 44:81–87.

Ballarini, Ilaria, Stefano Paolo Corgnati, and Vincenzo Corrado. 2014. “Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project.” *En-ergy Policy* 68:273–284.

Breunig, Markus M, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. “LOF: identifying density-based local outliers.” *ACM sigmod record*, Volume 29. ACM, 93–104.

Chidiac, SE, EJC Catania, E Morofsky, and S Foo. 2011. “Effectiveness of single and multiple energy retrofit measures on the energy consumption of office build-ings.” *Energy* 36 (8): 5037–5052.

Corgnati, Stefano Paolo, Enrico Fabrizio, Marco Filippi, and Valentina Monetti. 2013. “Reference buildings for cost optimal analysis: Method of definition and application.” *Applied energy* 102:983–993.

de Vasconcelos, Ana Brandão, Manuel Duarte Pinheiro, Armando Manso, and António Cabaço. 2015. “A Portuguese approach to define reference buildings for cost-optimal methodologies.” *Applied Energy* 140:316–328.

EESI. 2018. Building and Climate Change by Environmental and Energy Study Institute.

EU-Energy. 2018. Energy for Europe by European Commission.

European Union, EU. 2018. The Buildings Performance Institute Europe (BPIE).

Famuyibo, Adesoji Albert, Aidan Duffy, and Paul Strachan. 2012. “Developing archetypes for domestic dwellingsan Irish case study.” *Energy and Buildings* 50:150–157.

Flourentzou, Flourentzos, and C-A Roulet. 2002. “Elaboration of retrofit scenarios.” *Energy and Buildings* 34 (2): 185–192.

Galante, Annalisa, Marco Torri, et al. 2012. “A methodology for the energy performance classification of residential building stock on an urban scale.” *Energy and Buildings* 48:211–219.

Hecht-Nielsen, Robert, et al. 1988. “Theory of the backpropagation neural network.” *Neural Networks* 1 (Supplement-1): 445–448.

Hestnes, Anne Grete, and Niels Ulrik Kofoed. 2002. “Effective retrofitting scenarios for energy efficiency and comfort: results of the design and evaluation ac-tivities within the OFFICE project.” *Building and Environment* 37 (6): 569–574.

Hong, Tianzhen, Mary Ann Piette, Yixing Chen, Sang Hoon Lee, Sarah C Taylor-Lange, Rongpeng Zhang, Kaiyu Sun, and Phillip Price. 2015. “Commercial building energy saver: an energy retrofit analysis toolkit.” *Applied Energy* 159:298–309.

IEA. 2010. Energy Performance Certification of Buildings.

Intelligent Energy Europe, European Commission. 2018. Assessment and improvement of the EPBD Impact (for new buildings and building renovation)(ASIEPI).

Lafta, Raid, Ji Zhang, Xiaohui Tao, Yan Li, Vincent S Tseng, Yonglong Luo, and Fulong Chen. 2016. “An intelligent recommender system based on predictive analysis in telehealthcare environment.” *Web Intelli-gence*, Volume 14. IOS Press, 325–336.

Lee, Sang Hoon, Tianzhen Hong, Mary Ann Piette, and Sarah C Taylor-Lange. 2015. “Energy retrofit analy-sis toolkits for commercial buildings: A review.” *En-ergy* 89:1087–1100.

Loga, Tobias, N Diefenbach, and B Stein. 2012. Typology approach for building stock energy assessment. Main results of the TABULA project.

Ma, Zhenjun, Paul Cooper, Daniel Daly, and Laia Ledo. 2012. “Existing building retrofits: Methodology and state-of-the-art.” *Energy and buildings* 55:889–902.

Mahlia, TMI, H Abdul Razak, and MA Nursahida. 2011. “Life cycle cost analysis and payback period of light-ing retrofit at the University of Malaya.” *Renewable and Sustainable Energy Reviews* 15 (2): 1125–1132.

Park, Deuk Hee, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. 2012. “A literature review and clas-sification of recommender systems research.” *Expert Systems with Applications* 39 (11): 10059–10072.

Portugal, Ivens, Paulo Alencar, and Donald Cowan. 2017. The use of machine learning algorithms in rec-ommender systems: a systematic review.

Recast, EPBD. 2010. “Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast).” *Official Journal of the European Union* 18 (06): 2010.

SEAI. 2017. Dublin City Spatial Energy Demand Analysis.

Sousa Monteiro, Claudia, Carlos Cerezo, Andre´ Pina, and Paulo Ferr˜ao. 2015. “A method for the generation of multi-detail building archetype definitions: Application to the city of Lisbon.” *Proceedings of In-ternational Conference CISBAT 2015 Future Build-ings and Districts Sustainability from Nano to Urban Scale*. LESO-PB, EPFL, 901–906.

UK, Department for Communities, and Local Government. 2017. A guide to energy performance certificates for the construction, sale and let of non dwellings.

US Department of Energy, (DOE). 2018. Commercial Reference Buildings.

NOMENCLATURE

TP	number of true positives
FP	number of false positives
FN	number of false negatives
P	number of positives in ground truth
N	number of negatives in ground truth
K	training or testing samples
c_i	predicted value
c	target value
$MinPts$	minimum points