Examining and Improving Accuracy in a Deep Learning-based Pipeline for the Prediction of Building Energy Demand

Manav Mahan Singh[1, *], Philipp Geyer[1, 2]
[1] KU Leuven, Belgium
[2] Technische Universität Berlin, Germany
[*] manavmahan.singh@kuleuven.be

Abstract. It requires energy prediction of several hundreds of building design combinations to explore the design space and evaluate the effect of design parameters which lead to the development of quick energy prediction tool. In the past, researchers have suggested the use of data-driven approaches such as machine learning (ML) to make quick energy prediction. However, the challenge is to develop an ML model which is generalisable enough to accommodate the variations of the early stage, such as different shapes. Component-based ML (CBML) is a solution which we are developing further in this research. This ML model uses several ML components composed in a system entering way to predict building energy demand. Hence, accuracy relies on the performance of each component. In this paper, we are using two ML technique to understand the accuracy of CBML model. First, ceiling analysis serves to examine the improvement in accuracy by examining and exchanging internal model structure. Second, the feature importance exercise delivers information on the importance of a feature in prediction accuracy. By these methods, we identify ML components which have a higher potential to improve the accuracy of overall predictions. The inclusion of new features is examined as one way of prediction improvement. The test case applies CBML on 300 random shapes. In the test case, infiltration, heating energy, and energy demand components have been identified to improve the accuracy more than other ML components. We introduced additional features to improve the accuracy of these models. The accuracy has increased from 0.971 to 0.981 after including new features.

1. Introduction

Machine learning (ML) models are suitable to make thousands of energy predictions in few seconds which is required to explore a multi-dimensional design space for an energy-efficient design (Tian *et al.*, 2018; Singaravel, Suykens and Geyer, 2019). However, the development of a generalisable ML model which predicts the energy demand of different shapes is challenging. We are using component-based machine learning (CBML) which is based on the concept of using several machine learning components arranged in a particular order to address a problem. CBML is based on the modular paradigm of systems engineering. In some ML applications such as text recognition, ML models are arranged in a linear order, commonly referred to ML pipelines (Brad Neuberg, 2017). In case of building energy predictions, we refer to CBML, in which the components (ML models) are arranged into levels following decomposition of a building in an engineering way. A building is composed of zones, and a zone is composed of building elements (Geyer and Singaravel, 2018). The components are reusable ML models, organised into three levels and used for different building shapes. In the CBML approach, it makes intermediate predictions of heat flows for each building element. Zone components aggregate heating and cooling energy demand. Finally, a building level component determines the energy demand for the whole building, as shown in Figure 1. This CBML forms a machine learning pipeline in which input from the lower-level model is utilised to make higher-level predictions. Hence, the inaccuracies in energy demand prediction are the result of errors in each component.

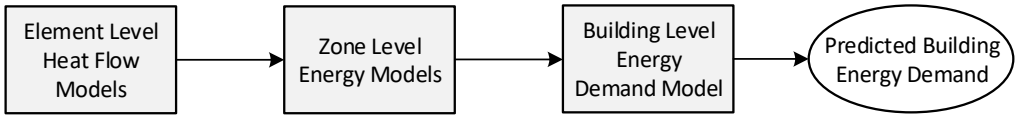


Figure 1: CBML based energy demand prediction approach

The objective of this paper is to explain the source of errors in overall predictions in the pipeline at the component level and improve its accuracy. Besides this component level error source, we are studying individual ML components to improve their performance. It is required to analyse which component has the biggest influence on the error and thus the best potential to improve the accuracy of energy demand predictions before focussing on component improvement. We are using a ceiling analysis tool for this purpose which allows identifying a component in ML pipeline with the highest potential to improve performance (Roncancio, Hernandes and Becker, 2013). It observes the improvement in accuracy of predictions dependent on replacement of model predictions by true values. Ceiling analysis measures the accuracy within the ML pipeline that can be achieved if the ML component makes exact predictions as true values. This provides evidence about the prediction improvement by component improvement.

Following the ceiling analysis, we have identified a few ML components in the pipeline, whose improvement increases the accuracy of predictions more than others. We worked on improving the performance of these ML components by introducing additional manually engineered features. The change in accuracy of ML components and overall accuracy of CBML is an important indicator to prove the relevance of feature engineering. The measurement of the importance of additional features allows for concluding that additional features are relevant. We are calculating the feature importance using an external model-independent technique known as permutation importance. In this technique, feature importance is measured by calculating a change in the model accuracy depending on the permutation of a feature supplied as noise.

In summary, the first step identifies important ML components and focusses on their individual need for improvement to increase accuracy. The second step introduces additional features to improve the performance and determines their impact using feature importance. Thus, it helps in identifying relevant features to improve the performance of individual ML components.

2. Background

It is important to focus on energy-efficient building design from an early stage as it has the potential to improve the performance with a lesser cost of change (Geyer and Schlüter, 2014). A generalisable ML model for different building shapes is necessary for its use at an early stage of design. Previous works related to this development do not assess their applicability of ML model on the building shapes which are not used for training. Such as Catalina et al., 2008 introduced features like time constant and shape factors to improve the generalisability of regression models for different building shapes, but the test data has very limited variability in terms of shape (Catalina, Virgone and Blanco, 2008). Asadi et al. developed regression-based models for variable building shapes (Asadi, Amiri and Mottahedi, 2014). Geyer and Singaravel introduced component-based ML model, but a decrease in accuracy is observed when tested on shapes which are independent of training data (Geyer and Singaravel, 2018). Thus, there is a small but growing set of ML methods to predict energy demand in design especially in early stages, in which important decisions are made.

In machine learning pipeline, such as the case of CBML, it is very important to understand the importance of individual components in the overall accuracy of predictions. Roncancio *et al.* have used it to determine the importance of individual components to detect human figures in an image (Roncancio, Hernandes and Becker, 2013). Other applications, such as character recognition in an image utilises three components arranged sequentially to identify characters. It has been found the first component which detects part of the image containing text will improve the accuracy of predictions much more than other components (Ng, 2019). Thus, it is useful to identify and work on the component, which improves the accuracy of predictions. In the case of CBML, there are many models arranged on three levels and working on all of them may not be useful and time-consuming. Since the concept of CBML or ML pipeline is rarely used in building energy prediction, ceiling analysis has not been used for detecting relevant components to improve accuracy.

We are using a technique of permutation importance to calculate the importance of features in individual ML components (Breiman, 2001). This technique is quite valuable in assessing the feature importance in model accuracy without retraining them. However, this technique has many limitations, such as it does not make a correct estimate of feature importance if features are correlated or the model features are changed (Hooker and Mentch, 2019).

3. Research Methodology

We are using component-based ML model for predicting building annual energy demand. CBML utilises the concept of decomposing a building into subordinate components in a system engineering way. The ML models for components are created and composed for prediction. A building is composed of zones which are composed of construction elements, such as walls, windows, roofs, floors, etc., as shown in Figure 3. The system flows in the composed structure serve for making energy predictions, as shown in Figure 2. The details of the initial features for each ML components are shown in Table 1. This composition allows the reusability of the ML model trained on one shape to the other shapes and thus the reuse of ML models in new design cases.

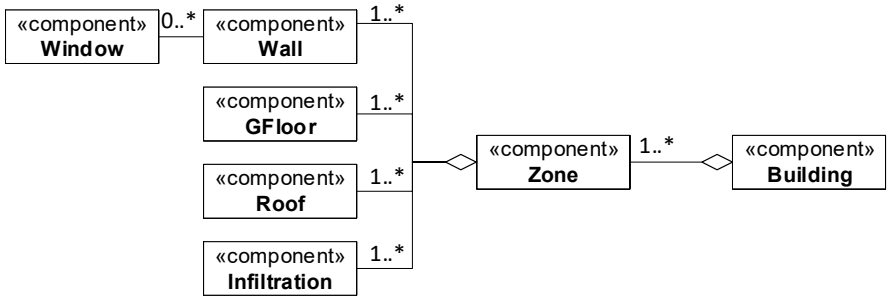


Figure 2: System based representation of BIM

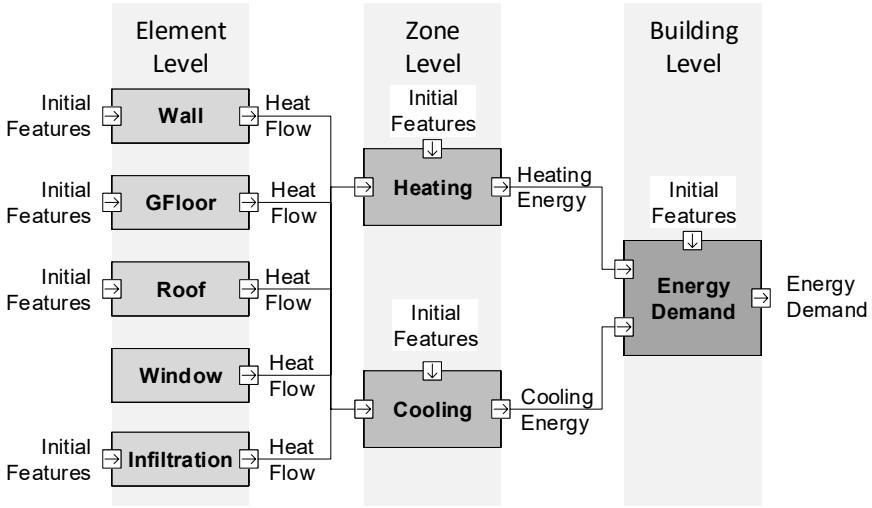


Figure 3: Structure of CBML

Table 1: Details of initial features for ML components

ML Component	Initial features
Wall	Area, Orientation, U Value
Ground Floor (GFloor) or Roof	Area, U Value
Window	Area, Orientation, U Value, g- Value
Infiltration	Zone Volume, Infiltration
Zone Heating and Cooling	Zone Area, Internal Heat Gain[1], Operating Hours, (All Heat Flows)
Building Energy Demand	Boiler Efficiency, Chiller COP, (Zone Heating and Cooling Energy)

[1] Internal heat gain refers to the sum of light and equipment heat gain.

3.1 Ceiling analysis

Ceiling analysis assesses the contribution of an individual model in an overall accuracy of machine learning pipeline by observing the change in accuracy when the model predictions are replaced by the true values (Ng, 2020). First, we calculate the baseline accuracy using the predictions from each component model, as shown in Figure 4. In the proposed CBML, there are three levels of models and a higher-level model utilises predictions from lower-level models. We carried out this exercise to identify the components which require focus to improve the performance of overall energy predictions. The baseline accuracy is calculated using the final predictions P. Then, we replace the predictions from a component model 1 by true values and make final predictions again P1. The accuracy which is possible to achieve by making component model 1 accurate is calculated using prediction P1. Similarly, for component model 2 is calculated using final predictions P2. If we replace prediction values by true values for all components, we will get true values T and accuracy will be 1. Thus, the change in accuracy when predicted values are replaced by true value shows a possible improvement in overall accuracy if the component makes perfect predictions and thus allows to determine the component's impact on accuracy.

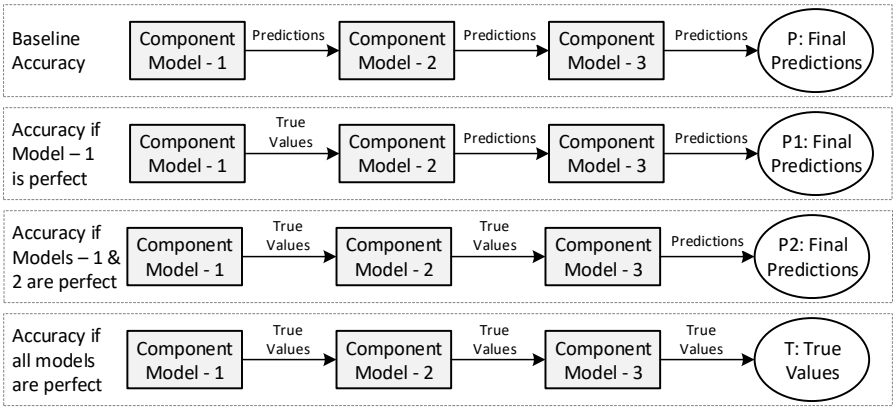


Figure 4: Explanation of ceiling analysis exercise

3.2 Feature importance

Feature importance for an ML model is estimated using a technique called permutation importance (Breiman, 2001; Altmann *et al.*, 2010). Information for one feature of the model is replaced by random noise, and a mean decrease in the accuracy of the model is recorded to calculate feature importance. The simplest way to introduce noise is to shuffle values for a feature, i.e. a feature is permuted. Since this technique requires randomisation, the task is repeated for 25 times to get reliable results. Feature importance (FI) is equal to a mean decrease in model accuracy when a feature is permuted. In this paper, we calculated this error on the test dataset. The algorithm to calculate feature importance based on the works of Fisher, Rudin and Dominici, 2018 is as follows:

1. Assume feature matrix is X, target prediction is y and trained ML component is f(X)
2. The model error $e = R^2\ (y, f(X))$
3. For each feature i = 1, . . ., n
    a. Generate feature matrix X^i by permuting i^{th} feature in the matrix X.
    b. Estimate error $e^i = R^2\ (y, f(X^i))$
    c. feature importance $FI^i = e^i - e$

3.3 Testcase and data generation

The implementation of an ML model pipeline for a medium-sized office building located in Munich demonstrates our approach. Building design parameters, mentioned in **Error! Not a valid bookmark self-reference.**, serve to generate the random combinations of building designs. Parametric energy simulations have been performed using the dynamic energy simulation tool EnergyPlus to get data for developing and testing the ML model. It is at this moment practically impossible to get energy consumption data for the building with varying shapes and required value of other design parameters; hence, we are relying on energy predictions from a well-established dynamic energy simulation tool (US Department of Energy, 2018). The tool is validated regularly against the performance of realized buildings, systems and components. The building shapes are generated randomly. Few of these shapes are shown in Figure 5. There are no internal zone divisions following one-zone-per-floor rule to create BEM (Sefaira, 2019). A total of 1000 samples are generated for training ML components, and 300 samples are generated for testing and reporting accuracy impact.

Accuracy - The accuracy of ML components and overall predictions is measured by the coefficient of determination R^2.

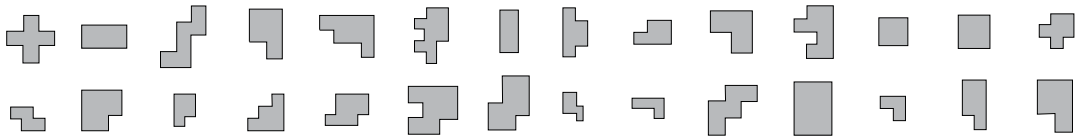


Figure 5: Randomly generated building shapes for testing energy prediction models

Table 2: Design parameters used in this study

Design Parameter	Min	Max	Design Parameter	Min	Max
Area	100	2000	u-value of Intermediate Floor	0.1	0.75
No. of Floors	1	5	Heat Capacity of Floor Slabs	800	1600
Floor Height	2.5	4.5	WWR_N/E/W/S	0.1	0.9
Orientation	0	180	Infiltration	0.2	0.8
u-value of Wall	0.1	0.6	Operating Hours	8	12
u-value of Ground Floor	0.1	0.6	Light Heat Gain	5	15
u-value of Roof	0.1	0.6	Equipment Heat Gain	5	15
u-value of Window	0.3	1.3	Chiller COP	3	5
g-value of Window	0.3	0.9	Boiler Efficiency	0.7	0.95

3.4 Training of ML components

Each ML component has a six-layer deep neural network architecture. It also utilises L2-regularisation and early stopping to restrict over-fitting. L2-regularisation prevent overfitting by adding a squared-term of model parameters to the cost function. Early stopping terminates the training process once there is no decrease in the cost. A total of five hyper-parameters are tuned for each ML component, i.e. the number of neurons in four hidden layers and the coefficient for L2 regularisation. We tuned these hyper-parameters for least validation-loss calculated on the validation data set (which is 20% of training data set). The values mentioned in Table 3 are used for different hyper-parameters.

Table 3: Values of different hyper-parameters

Parameter	Values
No. of neuron in hidden layer 1	500, 1000
No. of neuron in hidden layer 2	300, 500
No. of neuron in hidden layer 3	150, 300
No. of neuron in hidden layer 4	100, 150
Coefficient of L2-regularisation	0.0001, 0.00001

A total of 32 combinations of hyper-parameters has been tested to identify the least validation-loss has been selected for further use. ML models are developed using Keras library and TensorFlow backend (Chollet, 2015).

4. Results

The results section consists of two subsections. The first subsection describes the results of ceiling analysis exercise and identifies two components which improve the accuracy of predictions much more than other components. The section presents the results of feature importance analysis for the two models and introduces new features to improve the overall accuracy of models.

4.1 Ceiling analysis for CBML

The ceiling analysis measures a possible improvement in the accuracy of overall prediction if a component in the ML pipeline makes perfect predictions. Since, in the presented structure of CBML, there are models at three levels, the possible improvements are assessed at these three levels along with individual components. The accuracy of the baseline model is a coefficient of R^2 of 0.971. In brackets, it shows the difference from the best R^2 of 1, i.e. perfect prediction. There is a total improvement of 0.029, which is possible from the baseline model if all ML components make perfect predictions. If all the element level components make perfect predictions, the accuracy will improve to 0.984. If only the infiltration component perfectly predicts, the accuracy is 0.980. Thus, at the element level, infiltration is an important component and requires a focus for accuracy improvement.

Similarly, zone level components will improve the accuracy to 0.992. However, care must be taken to interpret this improvement. This value assumes that all the element level models are also making perfect predictions. Thus, the improvement which will be achieved after making zone level models perfect is only (0.992-0.984 = 0.008). The heating model contributes more than cooling model in this improvement. The remaining error of 0.008 is caused by building energy demand model.

As a result of this analysis, we will focus on three models to improve the performance, which are infiltration, zone heating and building energy demand model. These three models can improve the accuracy of overall prediction by 0.009, 0.006 and 0.008, respectively.

Table 4: Ceiling analysis for CBML components

	Model	**Accuracy (R2)**	
	Baseline	0.971 (0.029)	
Element Level Model	Wall Heat Flow	0.973	0.984 (0.016)
	Window Heat Flow	0.975	
	GFloor Heat Flow	0.974	
	Roof Heat Flow	0.974	
	Infiltration Heat Flow	0.980	
Zone Level Model	Zone Heating Energy	0.990	0.992 (0.008)
	Zone Cooling Energy	0.984	
Building Level Model	Building Energy Demand	1.000	

4.2 Feature Importance for ML components

The three identified relevant components identified in the previous section form the basis of the feature importance analysis. Improving these three models has the potential to increase the

accuracy of the overall prediction more than others. We calculated the feature importance in these two models and then introduced additional features to improve the model performance.

Table 5: List of features in infiltration heat flow, zone heating energy and energy demand model

Model	Initial Features	Additional Features
Infiltration Heat Flow	Zone Volume, Infiltration	Zone Area, Internal Heat Gain, Operating Hours, Total Heat Capacity[1]
Zone Heating Energy	Zone Area, Internal Heat Gain, Operating Hours, Wall Heat Flow, Window Heat Flow, GFloor Heat Flow, Roof Heat Flow, Infiltration Heat Flow	Zone Volume, Total Heat Capacity[1]
Building Energy Demand	Boiler Efficiency, Chiller COP, Heating Energy, Cooling Energy	Total Floor Area

[1] Total heat capacity is the sum of heat capacities of all the elements of the zone including wall, roof, floors and internal mass.

It should be noted that in this section, we are reporting the accuracy of individual ML components in this section as opposed to the previous section, in which, the accuracy of overall prediction is reported. After introducing additional features, the performance of the infiltration model improves from 0.976 to 0.981. Similarly, zone heating energy model shows an improvement from 0.986 to 0.989 and building energy demand model improves from 0.992 to 0.996. Improvement in these models results in the improvement in building energy demand prediction from 0.971 to 0.981. We analysed the importance of additional features along with the initial feature, and the results are presented in Figure 6. We found that among additional features, zone area has the highest importance in infiltration heat flow model, zone volume is important feature in zone heating energy model, and total floor area improves the performance of building energy demand model.

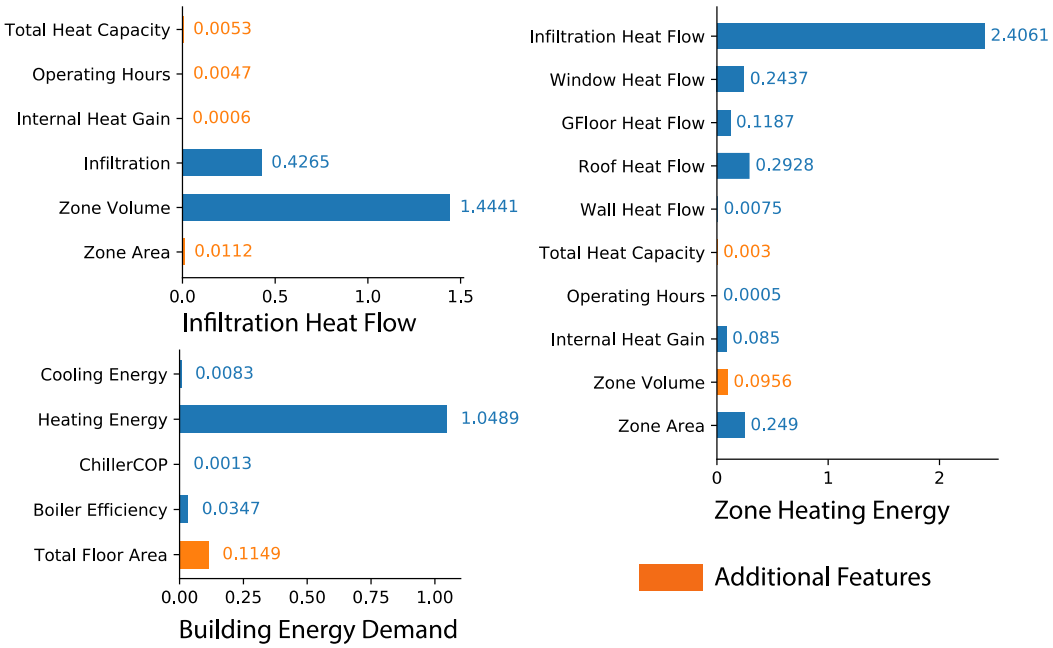


Figure 6: Feature importance for infiltration, zone heating, and building energy demand model

5. Discussion

The approach of ceiling analysis allows identifying the source of errors in the ML-based energy prediction pipeline. ML components are organised into three levels in CBML based energy prediction model. An improvement of 0.029 is possible in energy prediction if all ML components make perfect predictions (refer Figure 7). Improving all ML components to make perfect predictions will require significant computing resources, and it may not be useful to achieve such a small improvement at the cost of computing resources. However, it is useful to analyse the ML components which area causing a maximum part of this error and focus on a few components only and where to invest resources to improve accuracy. Out of this 0.029, 0.984-0.971 = 0.013 (45%) is caused by errors in element level models, 0.992-0.984 = 0.008 (28%) is caused by errors in zone level models, and the remaining, 1-0.992 = 0.008 (28%) is caused by errors building energy demand model. These values are a mere indication of possible improvement in CBML based predictions, but not be possible to achieve as it will require to develop ML components making perfect predictions. We improved a few of ML components, namely, infiltration heat flow, zone heating energy, building energy demand model. A few additional features are included to train these ML components again. These additional features result in an overall improvement of 0.981-0.971 = 0.01 (34%) against the possible improvement of 0.029. This example shows how to determine the components with the highest potential to improve the prediction accuracy by the implementation of additional features.

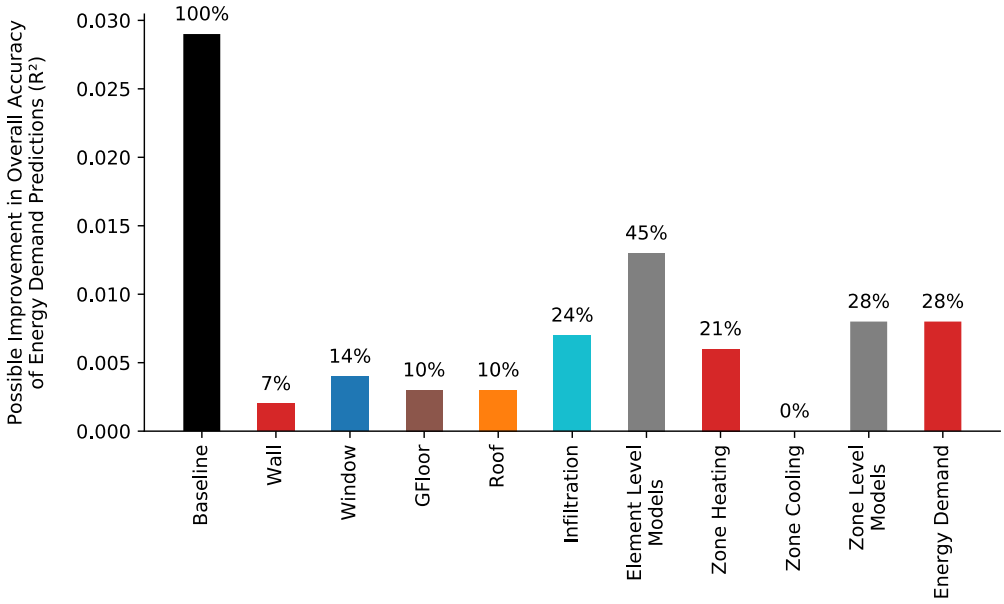


Figure 7: Improvement possible in overall accuracy corresponding to each ML component

Feature importance analysis for these components serves to assess the contribution of additional features in the accuracy of three ML components. The feature importance analysis suggests that additional features improve the prediction accuracy slightly. However, the results of feature importance must be interpreted with extreme care because of the following limitations of the technique. First, the feature importance is not equal to a decrease in the accuracy if the feature is removed. It is just equal to a decrease in accuracy when a feature is passed as random noise. Second, correlated features reduce the importance of each other. For example, in the infiltration heat flow model, zone volume and zone area are correlated. This model shows the accuracy of

0.981. After removing zone volume from the feature list, the accuracy reduces to 0.9136, but feature importance of zone area becomes 0.8146 from 0.0112. Thus, the value of feature importance is only indicative; it is a relative measure if there is no change in the model.

6. Conclusions

The approach of CBML allows the development of energy prediction models for a case where the building cannot be described by a set of parameters such as different building geometries or different zoning conditions etc. However, this approach presents a challenge of training several ML components and structuring them to make energy demand predictions. Since the accuracy of energy demand predictions in CBML based model depends on the prediction accuracies of individual ML components, this study has presented an approach to examining the source of error and reducing it. This gives valuable information for future development of future CBML models by providing information which are the crucial components.

The approach of ceiling analysis is useful to identify components of CBML that cause the highest share of the error. In the presented case, there is a total of 9 ML components, structured in three levels to make energy demand predictions. However, just improving three models improves the accuracy from 0.971 to 0.981. This increase is 34% of the possible improvement of 0.029. It should be noted that the full improvement, i.e. achieving the accuracy of $R^2 = 1$, is not possible. Hence, 34% of improvement possible, just by improving 3 out of 9 ML components, is significant. This method significantly reduces the efforts required for improving all ML components by feature engineering. An approach of feature engineering and importance analysis serves to improve individual ML components. In the test case, only three of the nine ML components required this process. Additional features have been introduced in these ML components, which results in the mentioned improvement of these components. Feature importance suggests that few additional features have quite significant importance, but few of them have relatively low importance. However, a technique of feature importance has limitations for related features which need to be considered. In sum, the methodology provides an efficient approach to improve the performance of CBML based energy prediction models using ceiling analysis and feature importance.

Acknowledgements

The authors want to acknowledge the support of Deutsche Forschungsgemeinschaft (DFG) for funding the research through research unit FOR 2363. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI.

References

Altmann, A. *et al.,* (2010). 'Permutation importance: a corrected feature importance measure', *Bioinformatics*, 26(10), pp. 1340–1347. doi: 10.1093/bioinformatics/btq134.

Asadi, S., Amiri, S. S. and Mottahedi, M., (2014). 'On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design', *Energy and Buildings*, 85, pp. 246–255. doi: 10.1016/j.enbuild.2014.07.096.

Brad Neuberg, (2017). *Creating a Modern OCR Pipeline Using Computer Vision and Deep Learning*, *dropbox.tech*. Available at: https://dropbox.tech/machine-learning/creating-a-modern-ocr-pipeline-using-computer-vision-and-deep-learning (Accessed: 8 May 2020).

Breiman, L., (2001). 'Random Forests', *Machine Learning*, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324.

Catalina, T., Virgone, J. and Blanco, E., (2008). 'Development and validation of regression models to predict monthly heating demand for residential buildings', *Energy and Buildings*, 40(10), pp. 1825–1832. doi: 10.1016/j.enbuild.2008.04.001.

Chollet, F., (2015). *Keras*. Available at: https://keras.io (Accessed: 6 May 2019).

Fisher, A., Rudin, C. and Dominici, F., (2018). 'All Models are Wrong but many are Useful: Variable Importance for Black-Box, Proprietary, or Misspecified Prediction Models, using Model Class Reliance'. Available at: http://arxiv.org/abs/1801.01489.

Geyer, P. and Schlüter, A., (2014). 'Automated metamodel generation for Design Space Exploration and decision-making – A novel method supporting performance-oriented building design and retrofitting', *Applied Energy*, 119, pp. 537–556. doi: 10.1016/j.apenergy.2013.12.064.

Geyer, P. and Singaravel, S., (2018). 'Component-based machine learning for performance prediction in building design', *Applied Energy*, 228, pp. 1439–1453. doi: 10.1016/j.apenergy.2018.07.011.

Hooker, G. and Mentch, L., (2019). 'Please Stop Permuting Features: An Explanation and Alternatives'. doi: arXiv:1905.03151.

Ng, A. Y., (2020). *Ceiling Analysis: What Part of the Pipeline to Work on Next*, *Coursera*. Available at: https://www.coursera.org/lecture/machine-learning/ceiling-analysis-what-part-of-the-pipeline-to-work-on-next-LrJbq (Accessed: 8 May 2020).

Ng, R., (2019). *https://www.ritchieng.com/machine-learning-photo-ocr/*. Available at: https://www.ritchieng.com/machine-learning-photo-ocr/ (Accessed: 2 March 2020).

Roncancio, H., Hernandes, A. C. and Becker, M., (2013). 'Ceiling analysis of pedestrian recognition pipeline for an autonomous car application', in *2013 IEEE Workshop on Robot Vision (WORV)*. IEEE, pp. 215–220. doi: 10.1109/WORV.2013.6521941.

Sefaira, (2019). *Zoning Strategy - Perimeter / Core*. Available at: https://support.sefaira.com/hc/en-us/articles/115000093632-Zoning-Strategy-Perimeter-Core (Accessed: 16 December 2019).

Singaravel, S., Suykens, J. and Geyer, P., (2019). 'Deep convolutional learning for general early design stage prediction models', *Advanced Engineering Informatics*, 42, p. 100982. doi: 10.1016/j.aei.2019.100982.

Tian, W. *et al.,* (2018). 'Uncertainty and sensitivity analysis of energy assessment for office buildings based on Dempster-Shafer theory', *Energy Conversion and Management*, 174, pp.705–718. doi: 10.1016/j.enconman.2018.08.086.

US Department of Energy, (2018). 'EnergyPlus 8.9.0'. Available at: https://energyplus.net/documentation.