2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

EXTRAPOLATION WITH MACHINE LEARNING BASED EARLY-STAGE ENERGY PREDICTION MODELS

Manav Mahan Singh[1] and Ian F. C. Smith[1]
[1] TUM Georg Nemetschek Institute Artificial Intelligence for the Built World, Garching near Munich, Germany

Abstract

Machine learning (ML) approaches are necessary to quickly predict energy performance at early stages. Although ML predictions are primarily useful for interpolation and many applications, early design stages require flexibility due to uncertain parameter values and vaguely defined building forms. We tested the ability of two ML approaches, artificial neural network (ANN) and convolutional neural network (CNN), to extrapolate for early-stage energy predictions. While ANN uses numerical parameters to represent building shape, CNN uses convoluted layers to learn geometrical representations. The generalisation of both networks deteriorates on extrapolated datasets with the CNN having a better performance than the ANN.

Introduction

A significant part of the global energy supply is required for building operation. Energy-efficient buildings reduce global energy consumption and contribute to sustainable development. Early stages of design give us ample opportunities to improve energy performance (Parasonis et al., 2012). However, it is a challenging task since any performance assessment at early stages require accommodating uncertain values of characteristics. This leads to simulating several hundreds of energy models to assess probabilistic performance (Singh et al., 2020; Van Gelder et al., 2014). Since high fidelity dynamic simulations for probabilistic energy predictions are time consuming at early stages, researchers proposed using surrogate models to reduce computation time (Roman et al., 2020; Westermann & Evins, 2019). This development has led to several performance evaluation tools for early design stages (Singh et al., 2022).

Surrogate modelling using machine learning (ML) are preferred due to their ability to capture complex interactions of building energy models (BEMs) (Singaravel et al., 2018). A generalisable function is trained that maps set of values of input variables to predict values of target variables. Researchers trained and tested ML networks for energy predictions at early stages using various approaches. Moreover, the accuracy of ML predictions is reported on test datasets that have values of characteristics within the same ranges as training datasets. The task of predicting values of target variables is referred to as interpolation since values of input variables are in the same range as the training data. When an ML network makes predictions on values of input variables that are beyond the ranges of training data, it is referred to as extrapolation. Prediction accuracy of ML models while extrapolating, can be enhanced when a model learns underlying functional relationship of the physical principle model (Sahoo et al., 2018) instead of memorising values of target variables. Researchers tested the ability of several ML algorithms to interpolate and extrapolate in other domains (Acharige & Johlin, 2022; McCartney et al., 2020). The work suggests that an ML algorithm having a good accuracy while interpolating may not have a good accuracy while extrapolating. Since the prediction accuracies for interpolation and extrapolation can be significantly different, it is necessary to test an ML network used for early stage for both interpolation and extrapolation. Singaravel et al., 2019 have tested ability of deep learning networks to predict energy performance of test dataset that has different number of floors than training dataset. There is no comprehensive study that tests ML networks for extrapolation to predict energy use. Different ranges of parameter values and building forms have not been tested for extrapolation.

BEMs use complex physical principles of heat and mass transfer to estimate energy use during design. While most characteristics can be represented numerically, shape is a complex feature. Traditional ML approaches represents shape using parameters such as relative compactness (Catalina et al., 2008; Li et al., 2019). Even with different test shapes than training shapes, it interpolates within parametric ranges of numeric characteristics. However, such simplified representations cannot capture the effect of solar radiation, self-shading, and orientation. Further, convolutional neural networks (CNN) are used to learn shape representations of buildings to capture its effects on energy predictions. The CNN approach is based on the argument that building shape that have a significant influence on the energy use are not well represented in artificial neural networks (ANN) (Mahan Singh & Geyer, 2021). Since building shape is a high-dimensional characteristic, it requires extrapolating even though values of numerical characteristics are within the range of training dataset (Balestriero et al., 2021). However, a true extrapolation test requires different ranges of values for numerical characteristics and different shapes than training dataset.

This, in case of ML for energy predictions, extrapolation refers to: i) extending beyond parametric ranges and ii) including new building shapes. For example, we trained an ML network on buildings with a floor area between 400-500 sq. m. and try to predict energy use for buildings with the floor area of 375 sq. m. This example refers to extrapolating beyond parametric ranges of training dataset. However, when we train ML network on box-shaped buildings and predict energy use for a complex form, it is extrapolating to a new building form that is not present in training dataset. In case of ANN, even a new form can have relative compactness in the range of training dataset.

Since building shapes in a test dataset can be significantly different from the shapes of the training dataset, we need to test the ability of ML networks to extrapolate for both scenarios – beyond parametric ranges and new shapes. We tested two approaches of training ML networks for early stage energy predictions in situation of interpolation and extrapolation. The first approach is a conventional ANN. It only uses numerical inputs to predict energy use intensity (EUI). The second approach is CNN that uses convolutional layers to capture shape information from the three dimensional building mass model. The next section describes both the approaches in detail.

ML networks to predict energy use intensity

The first two subsections describe two approaches to train ML based energy prediction model for the early stages – ANN and CNN. The third subsection describes the building geometry and parametric ranges of a training dataset and six test datasets.

Artificial neural network (ANN)

A simple approach to train an ML network for energy predictions is to feed all required information – numeric values of characteristics as input and values of target variables as output. This paper uses an ANN as shown in Figure 1.

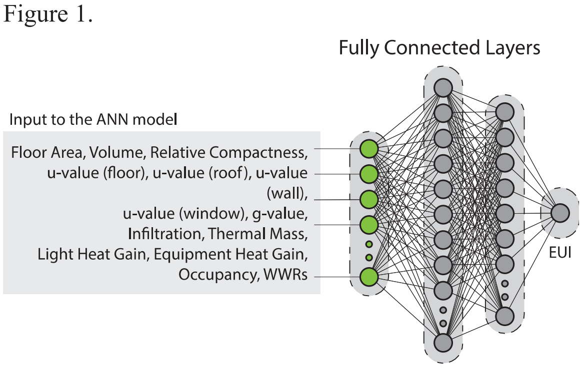

Figure 1: Configuration of the ANN

Figure 1 shows configuration of the ANN (one input layer, two hidden layers, and one output layer) to predict EUI. All inputs to the model are only numeric values of characteristics, including geometry which is represented by floor area, volume, and relative compactness. Relative compactness is the ratio of surface area over volume. The model has either one or two hidden layers with rectified linear unit (ReLU) activation function and L2 regularisation. L2 regularisation and early stopping is used to prevent overfitting (Ng, 2004; Prechelt, 1998). Values of input and output variables are scaled between 0 and 1 before feeding them into the network. We used a well-defined approach of leave-one-out cross validation to find a suitable value for hyperparameters (Wong, 2015; Yadav & Shukla, 2016). Training data is split in to two mutually exclusive sets of 80-20. 20% samples are used to calculate the validation loss. The model with the least validation loss is retained to report the accuracy of the ANN on test datasets. As mentioned in Table 1, we experimented with the number of hidden layers, number of neurons in each layer, regularisation coefficient, and learning rate. We tested eight random combinations of hyperparameter values. The best combination of hyperparameters that has the least validation loss is highlighted in bold (Table 1).

Table 1: Hyperparameters of the ANN

Hyperparameter	Values
Number of hidden layers	**1**, 2
Neurons (1st layer)	100, **200**, 500
Neurons (2nd layer)	0 (no 2nd layer), 20, 50
Regularisation coefficient	0.0003, **0.0001**, 0.00003
Learning rate	0.003, **0.001**, 0.0003

Convolution neural network (CNN)

The ability of CNN to learn relevant information from a structured array of data make them useful to approximate effects of building geometry on energy predictions. We used three dimensional (3D) mass model as an input to the CNN. This input along with other numerical characteristics are used to predict EUI.

Figure 2 shows the configuration of the CNN. The input to the ML model consists of two components. The first component is the 3D mass model. It is similar to 2D images with an additional dimension representing the 3D mass along z-axis. We structured the 3D mass model into an array of 24×100×100, 24 along the Z-axis and 100×100 on the XY-plane. One pixel is equivalent to 0.6 metres and have a value of either 0 or 1, representing empty space and the building, respectively. The CNN has a number of convolutional layers, followed by average pooling layer, dropout layer, and a flattening layer to process this data. The second component of the input is an array of numerical parameters which is similar to the input for ANN. The output of the convolution network is concatenated with this input and fed to a fully connected neural network to predict the EUI. There are a number of hyperparameters in the CNN. See (TensorFlow, 2022) for detailed definitions of convolutional layers, kernel size, kernel stride, pooling size, pooling stride, and dropout layer.

We experimented with the hyperparameters, mentioned in Table 2. Similar to the approach of finding the best combination of hyperparameters for the ANN, we tested eight random combinations. The CNN with the least validation loss is retained to report the accuracy on test datasets. The value of hyperparameters with the least validation loss are highlighted in bold.

Table 2: Hyperparameters of the CNN

Hyperparameter	Values
Convoluted layers	3, 5, **10**
Kernel size	(2, 3, 3), (2, 5, 5), **(3, 8, 8)**
Kernel stride	(1, 2, 2), **(2, 3, 3)**, (3, 5, 5)
Pooling size	(3, 5, 5), **(3, 8, 8)**, (4, 10, 10)
Pooling stride	(1, 2, 2), **(2, 3, 3)**, (3, 5, 5)
Neurons (1st layer)	100, 200, **500**
Neurons (2nd layer)	0 (no 2nd layer), **20**, 50
Regularisation coefficient	1e-4, 3e-5, **1e-5**
Learning rate	3e-3, **1e-3**, 3e-4

Training and test datasets

We tested this approach on a medium-sized office building in Germany. We defined the ranges of values of building characteristics based on the typical requirements of office buildings and relevant German standards. Table 3 describes the characteristics and ranges for their values, used in this paper for generating training, test, and extrapolated datasets. Figures 3 to 6 shows shapes from various datasets.

Training dataset: we sampled building characteristics in the range of (min, max) values, mentioned in Table 3, using Sobol sampling scheme. The training dataset contains 5000 samples. A sample, in the context of this paper, refers to one set of values of the building characteristics. The training dataset contains building shapes such as box, T-shape, U-shape, and L-shape, as shown in Figure 3. The shape for a sample is randomly selected from a set of 17 configurations which is resized and rotated according to the values of relevant characteristics. While these shapes are simple, they capture effects of self-shading and orientation. The relative compactness values for the training dataset range from 0.31 to 0.64. Relative compactness of a building are not controlled directly, and they depend on the shape, floor-to-floor height, and floor area. The values are only relevant for the ANN while analysing parametric ranges of characteristics. It does not suggest that two samples with the same value of relative compactness have similar complexity of forms.

Test dataset A and test dataset B: we sampled the parameters in the same range as training samples, using Latin hypercube sampling scheme to collect 50 training samples for each A and B test datasets. The samples in test dataset A have the same shapes as the training dataset. Test dataset B have a few additional shapes (from a total of 35 configurations), as shown in Figure 4. The relative compactness values for test dataset B are in the range of 0.32 to 0.62. Since test dataset B only have additional shapes than training and test dataset A, the difference in the prediction accuracy will show how well ML networks learn geometrical representation while the numerical characteristics are in the same parametric range.

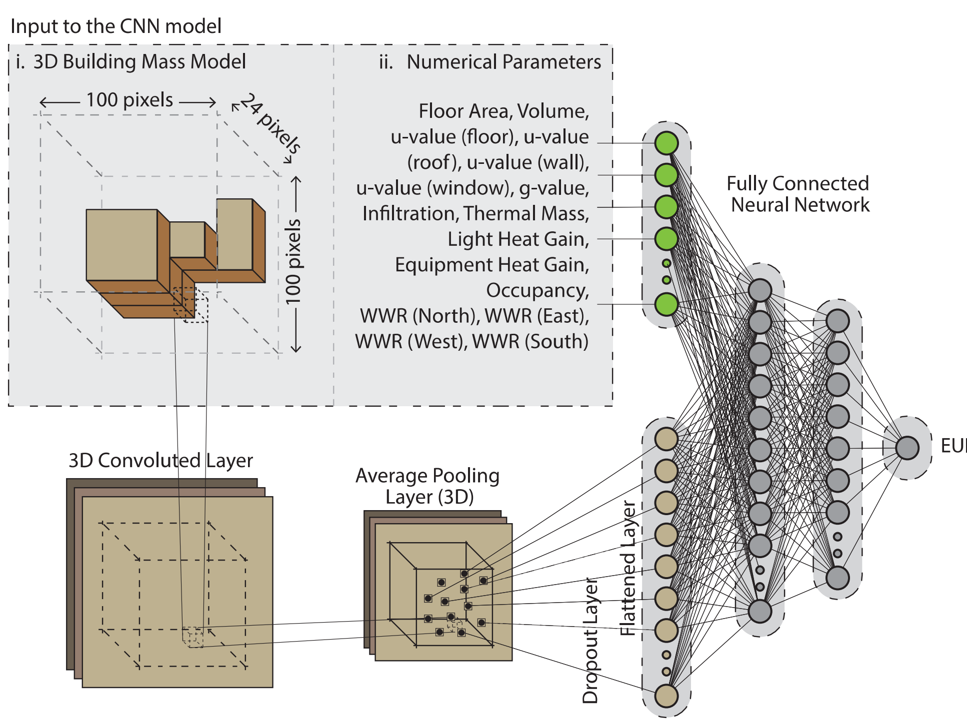

Figure 2: Configuration of the CNN

Extrapolated (Ext.) datasets 1A, 1B, 2A, and 2B: we sampled the parameters from Table 3 between the range (1. Left, 1. Right) using Latin hypercube sampling scheme to collect 50 samples for each ext. datasets, 1A and 1B. Similarly, the range of (2. Left, 2. Right) is used for test datasets, 2A and 2B. All extrapolated datasets do not contain any sample with parameters in the range of (min, max). Same as test dataset B, the extrapolated datasets 1B and 2B contain additional shapes, as shown in Figure 5 and Figure 6. Shapes for test dataset B are selected from a set of 70 configurations. The relative compactness values for ext. dataset 1B are in the range of 0.30 to 0.66 and for test dataset 2B are in the range of 0.32 to 0.65. Extrapolated datasets 1A and 2A have the same shapes as training and test dataset A, as shown in Figure 3.

The objective of creating multiple test datasets is to assess the ability of ML networks to generalize while extrapolating beyond ranges of numerical characteristics as well as new building shapes. Comparing prediction accuracy on test datasets A, 1A, and 2A will show how the prediction accuracy changes while increasing the range of values of characteristics. Similarly, comparisons of prediction accuracy on test datasets A against B, 1A against 1B, and 2A against 2B shows the changes in prediction accuracy when we include additional shapes. Finally, we can compare prediction on test datasets A against 1B or 2B to analyze the effect of changing both ranges of values for numerical characteristics and shapes.

Data collection

We created an EnergyPlus model and calibrated it using actual energy consumption of an office building near Munich, Germany. We calibrated an energy model using real energy consumption values and developed a tool that creates energy model of selected building shapes of required dimensions and parameters. We used this tool to generate BEMs and simulated these models to collect data for ML models. All training and test datasets are available at Kaggle (Singh, 2023). The ranges for values of building characteristics are based on previous studies and general requirements for new buildings in Germany. We considered a reasonable variation of around ±20% in parameter values for extrapolated datasets.

Table 3: Building characteristics and ranges for their values in training, test, and extrapolated datasets

Parameters	Unit	2. Left	1. Left	Min	Max	1. Right	2. Right
Floor Area Per Floor	m^2	400	425	450	550	575	600
Floor Height	m	3.0	3.2	3.3	3.6	3.7	3.8
No. of Floors	-	2	2	2	5	5	5
Orientation	°	0	0	0	90	90	90
u-value (Wall)	W/m^2K	0.15	0.17	0.18	0.22	0.23	0.25
u-value (Ground Floor)		0.15	0.17	0.18	0.22	0.23	0.25
u-value (Roof)		0.13	0.14	0.15	0.19	0.20	0.21
u-value (Window)		0.68	0.7	0.72	0.88	0.9	0.92
g-value	-	0.36	0.38	0.4	0.6	0.62	0.64
WWR[1]	-	0.1	0.15	0.2	0.8	0.85	0.9
Internal Mass	kJ/m^2K	20	22.5	25	35	37.5	40
Air Permeability	$m^3/h{\cdot}m^2$	5.8	5.85	5.9	6.1	6.15	6.2
Occupant Load	m^2/Person	20	21	22	24	25	26
Light Heat Load	W/m^2	5	5.5	6	8	8.5	9
Heating COP[2]	-	2.6	2.7	2.8	3.2	3.3	3.4
Cooling COP[2]		2.6	2.7	2.8	3.2	3.3	3.4
Boiler Efficiency		0.9	0.91	0.92	0.95	0.96	0.97

[1] Window-to-wall ratios (WWRs) vary independently in each direction.

[2] Coefficient-of-performance (COP)

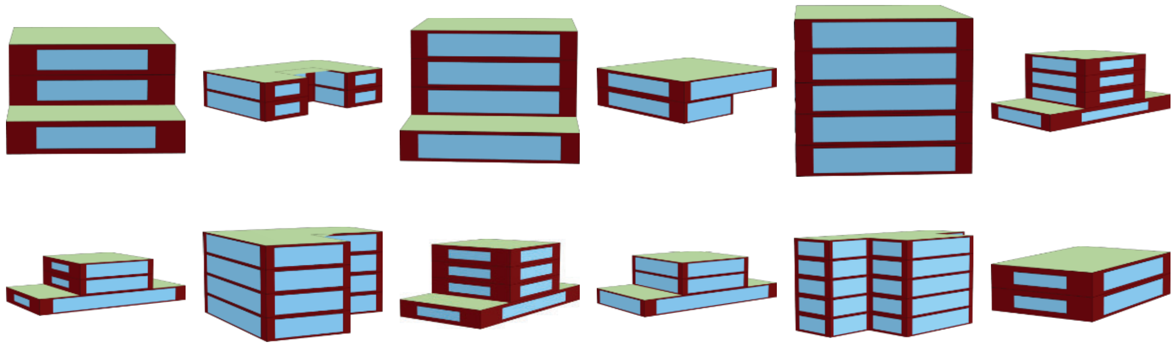

Figure 3: Building shapes of training dataset, test dataset A, and extrapolated dataset 1A and 2A

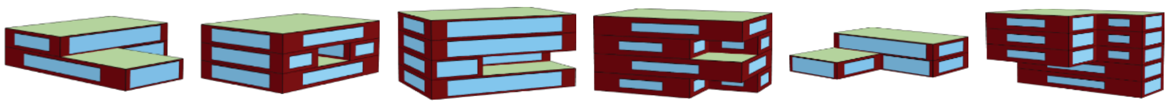

Figure 4: Building shapes of test dataset B

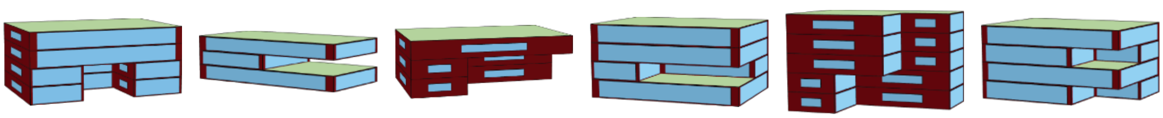

Figure 5: Building shapes of extrapolated samples 1B

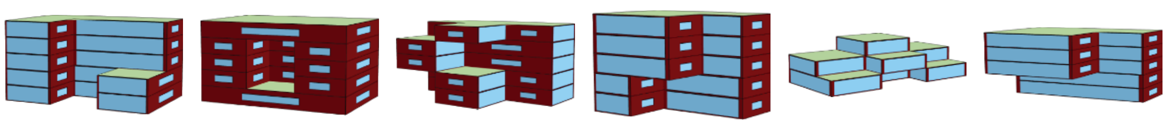

Figure 6: Building shapes of extrapolated dataset 2B

Results

The prediction errors of the ANN and the CNN are shown in Figures 7 and 8, respectively. The figures show the prediction error as residuals on the y-axis and the simulation values of EUI on the x-axis. As described earlier, we tested the prediction accuracy of the ML networks on six datasets, containing 50 samples each. Unfilled blue circles show samples from test dataset A and filled blue circles show samples from test dataset B. Similarly, red unfilled diamonds show samples from dataset 1A and green unfilled squares show samples from dataset 2A. Their filled counterparts show samples from datasets 1B and 2B, respectively. The prediction errors (residuals) increase as the samples move away from 0, above or below, with samples above showing over predictions and samples below showing under predictions.

Test dataset A contains the same building shapes as the training dataset while test dataset B contains a few additional shapes. Both, test datasets A and B have the same ranges for values of numerical characteristics as the training dataset. Hence, the prediction accuracy for test dataset A shows the model generalisation while interpolating within the same range of values of numerical characteristics and shapes. The difference between the prediction accuracies of test datasets A and B shows the effect of including new shapes on model generalisation.

The results show that the ANN has generally higher errors than the CNN for both test dataset A and B. The CNN has a root-mean-square-error (RMSE) of 0.38 kWh/a.m^2 for test dataset A; however, it increases quickly to 1.1 kWh/a.m^2 for test dataset B. In comparison, the ANN has more increase in the errors than the CNN.

Comparing the performances of the ANN and the CNN on extrapolated datasets 1A and 1B show that both the networks have lesser prediction accuracies while extrapolating than interpolating. However, the increase in the errors for the ANN are higher than the CNN. While the ANN has a mean-absolute-percentage-error (MAPE) of 3.6%, the CNN has MAPE of only 1.5%. Figure 6 shows that the ANN has both over and under predictions for the extrapolated dataset 1A. The CNN has most of the samples from extrapolated dataset 1A in the range of ±2 kWh/a.m^2. Only 3 samples have an underprediction of more than 2 kWh/a.m^2. The situation is no different for the ANN for extrapolated dataset 1B as many samples have an absolute error of more than 2 kWh/a.m^2 with a high RMSE of 3.3 kWh/a/m^2. The performance of the CNN on extrapolated dataset 1B is also worse than extrapolated dataset 1A, as the model has an RMSE of 2.1 kWh/a/m^2 and around 10 samples have a residual error of more than 2 kWh/a.m^2.

The performance of the ANN on both extrapolated dataset 2A and 2B is much worse than the 1A and 1B. Many

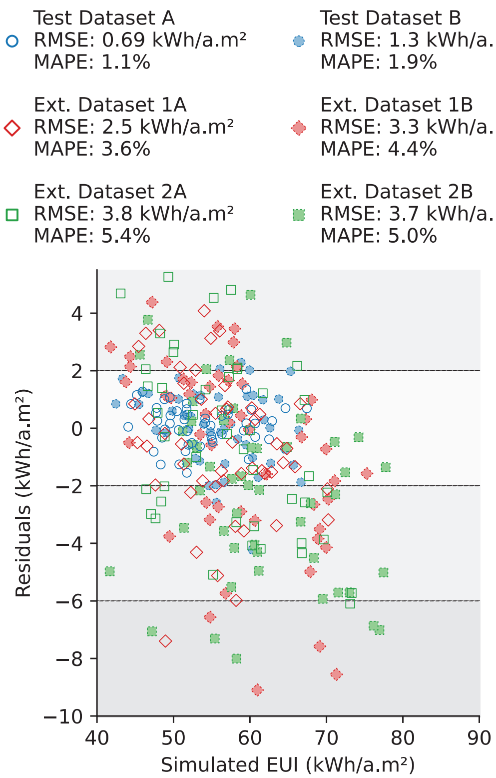

Figure 7: Residuals for the ANN

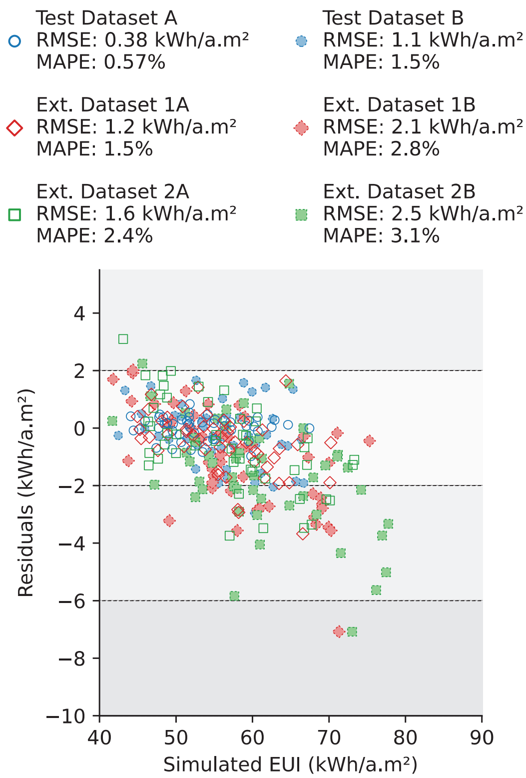

Figure 8: Residuals for the CNN

samples show under prediction of more than 2 kWh/a.m^2, with some samples having high percentage error of around 20%. Surprisingly, extrapolated dataset 2B have overall lower prediction errors (RMSE) than 2A that suggests including new shapes do not significantly influence the prediction errors. However, the metric can be misleading as we can see many filled green squares below -4 kWh/a.m2 that suggests there are many samples from dataset 2B that have high prediction errors, individually.

The performance of the CNN also deteriorates for extrapolated datasets 2A and 2B, compared to 1A and 1B; however, this loss of prediction accuracy is smaller than the ANN. The network has an RMSE of 2.5 kWh/a.m^2 for the dataset 2B. The change in the prediction errors from dataset 2A to 2B follows the same trend as from dataset 1A to 1B. The maximum error in an individual sample is around 10%. The increase is prediction errors from 1A to 2A and from 1B to 2B is also smaller than the ANN.

Discussion

We presented two ML approaches ANN and CNN to train a surrogate model for early-stage energy prediction and test their abilities to interpolate and extrapolate. The motivation of training these ML networks is to achieve computational efficiency over dynamic simulations. Since such ML networks are trained over a limited set of data, their use is also limited to that dataset. However, at the early stage design stage, extrapolation of solution is often carried out for parameter values that are outside the values of training dataset, i.e., extrapolation. We trained the ML networks on a training dataset and tested their prediction accuracies on several datasets. Multiple test datasets are created to distinguish between the prediction accuracy when the ranges of numerical characteristics are increased, and new shapes are included. The comparison of prediction accuracies shows that CNN has a better prediction accuracy than ANN. A possible reason is that CNN has more parameters and with convolutional layers, it can capture the effects of the shape on energy predictions. However, with simple network configuration, ANN also has a good prediction accuracy for test samples within the range of training datasets.

Aside from fewer parameters, the ANN relies on a small set of numerical inputs to represent building characteristics for energy predictions. The network represents building geometry using parameters such as floor area, height, number of floors, and relative compactness. Such model has high prediction errors while extrapolating beyond the parametric ranges and including new building shapes. In comparison, the CNN learns the effect of the building geometry from a three-dimensional model and thus, shows a better prediction accuracy while extrapolating. It shows that the CNN has an ability to learn about interactions well and can extrapolate better than the ANN.

Moreover, we extrapolated numerical parameters by around ±10% for test samples 1A and 1B, and ±20% for test samples 2A and 2B. This increase in the ranges is an arbitrary and requires further research in to design practices. It is possible that some building characteristics require more flexibility than others. Further, a surrogate model reduces computational time at the cost of accuracy. Since early-stage energy predictions are performed to compare and classify design options, a small error is allowed. An error is considered small if it does not affect the decision-making. This topic is discussed further in Singh, 2020.

Conclusions

We explored an important aspect of machine learning applications, extrapolating. In the application of machine learning for early stage energy predictions, extrapolation has been tested for both, beyond parametric ranges of training data as well as new building shapes. The building shape is a complicated feature to be represented by a few numerical parameters; hence, an approach of convolutional neural network to learn the effects of building shape on its energy performance is also tested.

We found that a simple artificial neural network which work extremely well for interpolation loses its prediction accuracy while extrapolating beyond parametric ranges. The accuracy reduces further when we extrapolate for a complex feature like building shape. The convolutional neural network is more reliable while extrapolating, either extending parametric ranges or including new shapes. Hence, a complex machine learning algorithm such as convolutional neural network performs well for extrapolation. It will be interesting to understand how three-dimensional data is processed through convolutional layers and what features does it extract for energy predictions. The characteristics of samples with high prediction errors can be studied to understand the causes and improve the performance of the model while extrapolating.

Acknowledgments

The research presented is supported by the TUM Georg Nemetschek Institute Artificial Intelligence for the Built World.

References

Acharige, D., & Johlin, E. (2022). Machine Learning in Interpolation and Extrapolation for Nanophotonic Inverse Design. *ACS Omega*, *7*(37), 33537–33547. https://doi.org/10.1021/acsomega.2c04526

Balestriero, R., Pesenti, J., & LeCun, Y. (2021). *Learning in High Dimension Always Amounts to Extrapolation*. http://arxiv.org/abs/2110.09485

Catalina, T., Virgone, J., & Blanco, E. (2008). Development and validation of regression models to predict monthly heating demand for residential buildings. *Energy and Buildings*, *40*(10), 1825–1832. https://doi.org/10.1016/j.enbuild.2008.04.001

Li, Z., Dai, J., Chen, H., & Lin, B. (2019). An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. *Building Simulation*, *12*(4), 665–681. https://doi.org/10.1007/s12273-019-0538-0

McCartney, M., Haeringer, M., & Polifke, W. (2020). Comparison of Machine Learning Algorithms in the Interpolation and Extrapolation of Flame Describing Functions. *Journal of Engineering for Gas Turbines and Power*, *142*(6). https://doi.org/10.1115/1.4045516

Ng, A. Y. (2004). Feature selection, L_1 vs. L_2 regularization, and rotational invariance. *Twenty-First International Conference on Machine Learning - ICML '04*, 78. https://doi.org/10.1145/1015330.1015435

Parasonis, J., Keizikas, A., Endriukaitytė, A., & Kalibatienė, D. (2012). Architectural Solutions to Increase the Energy Efficiency of Buildings. *Journal of Civil Engineering and Management*, *18*(1), 71–80. https://doi.org/10.3846/13923730.2011.652983

Prechelt, L. (1998). *Early Stopping - But When?* (pp. 55–69). https://doi.org/10.1007/3-540-49430-8_3

Roman, N. D., Bre, F., Fachinotti, V. D., & Lamberts, R. (2020). Application and characterization of metamodels based on artificial neural networks for building performance simulation: A systematic review. *Energy and Buildings*, *217*, 109972. https://doi.org/10.1016/j.enbuild.2020.109972

Sahoo, S. S., Lampert, C. H., & Martius, G. (2018, October 10). Learning Equations for Extrapolation and Control. *35th International Conference on Machine Learning*. http://proceedings.mlr.press/v80/sahoo18a/sahoo18a.pdf

Singaravel, S., Suykens, J., & Geyer, P. (2018). Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. *Advanced Engineering Informatics*, *38*, 81–90. https://doi.org/10.1016/j.aei.2018.06.004

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

Singh, M. M., Singaravel, S., Klein, R., & Geyer, P. (2020). Quick energy prediction and comparison of options at the early design stage. *Advanced Engineering Informatics*, *46*, 101185. https://doi.org/10.1016/j.aei.2020.101185

Singh, M. M., & Geyer, P. (2021, September 1). CNN-based quick energy prediction model using image analysis for shape information. https://doi.org/10.26868/25222708.2021.30250

Singh, M. M. (2023). extrapolation-3d-cnn-building-energy. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/5478752

Singh, M. M., Deb, C., & Geyer, P. (2022). Early-stage design support combining machine learning and building information modelling. *Automation in Construction*, *136*, 104147. https://doi.org/10.1016/j.autcon.2022.104147

TensorFlow. (2022). Convolutional Neural Network (CNN) | TensorFlow Core. https://www.tensorflow.org/tutorials/images/cnn

Vollmer, M., Schneider-Marin, P., Harter, H., & Lang, W. (2019). Innovation Bayerischer Bauindustrieverband e.V. Technische Universität München. https://www.tausendpfund.group/dateien/Innovation_2020.pdf

Van Gelder, L., Janssen, H., & Roels, S. (2014). Probabilistic design and analysis of building performances: Methodology and application example. *Energy and Buildings*, *79*, 202–211. https://doi.org/10.1016/j.enbuild.2014.04.042

Westermann, P., & Evins, R. (2019). Surrogate modelling for sustainable building design – A review. *Energy and Buildings*, *198*, 170–186. https://doi.org/10.1016/j.enbuild.2019.05.057

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839–2846. https://doi.org/10.1016/j.patcog.2015.03.009

Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78–83. https://doi.org/10.1109/IACC.2016.25