Random Forests and Artificial Neural Network for Predicting Daylight Illuminance and Energy Consumption

Muhammad Waseem Ahmad[1], Jean-Laurent Hippolyte[2],Monjur Mourshed[3], Yacine Rezgui[4]
BRE Centre for Sustainable Engineering, School of Engineering,
Cardiff University, Cardiff, CF24 3AA, UK
[1]AhmadM3@Cardiff.ac.uk; [2]HippolyteJ@Cardiff.ac.uk;
[3]MourshedM@Cardiff.ac.uk; [4]RezguiY@Cardiff.ac.uk

Abstract

Predicting energy consumption and daylight illuminance plays an important part in building lighting control strategies. The use of simplified or data-driven methods is often preferred where a fast response is needed e.g. as a performance evaluation engine for advanced real-time control and optimization applications. In this paper, we developed and then compared the performance of the widely-used artificial neural network (ANN) with random forest (RF), a recently developed ensemble-based algorithm. The target application was predicting the hourly energy consumption and daylight illuminance values of a classroom in Cardiff, UK. Overall, RF performed better than ANN for predicting daylight illuminance; with coefficients of determination (R^2) of 0.9881 and 0.9799 respectively. On the energy consumption testing dataset, ANN performed marginally better than RF with R^2 values of 0.9973 and 0.9966 respectively. RF performs internal cross-validation and is relatively easy to tune as it has few tuning parameters. The paper also highlighted possible future research directions.

Introduction

Buildings are responsible for 40% of the total global energy use and account for 30% of the total emission of CO_2, one of the greenhouse gases responsible for anthropogenic climate change (Ahmad et al., 2016). To mitigate this, building regulations have been developed or updated to reduce the impact of climate change and enhance the performance of buildings. With sustained reductions in a building's heating and cooling demands, the energy used by artificial lighting increases in relative terms (Ahmad et al., 2015). Daylight is an essential part of our life, and building occupants tend to prefer daylight over artificial lighting. It also provides a comfortable and effective learning environment in schools. An appropriate lighting level is necessary to satisfy both psychological and visual comfort conditions. Typically Venetian blinds have been used in buildings to control daylighting. In the literature, various automatic control strategies have been developed to enhance the thermal and daylighting performance of occupied spaces. One example of this research work is Ahmad et al. (2015), the authors proposed a genetic algorithm based method to adjust the window blind position. On the other hand, energy prediction strategies are one of the core components of building energy control and operational strategies (Li and Wen, 2014).

In recent years, a number of prediction approaches, either detailed or simplified, have been proposed and applied for predicting building energy consumption and daylight illuminance. These approaches can be broadly classified into three categories i.e. numerical, analytical and predictive. Numerical approaches (e.g. EnergyPlus, DAYSIM, RADIANCE, TRNSYS, etc.) often enable the user to evaluate designs with reduced uncertainties. However, these methods do not perform well in predicting the energy use and daylighting illuminance of occupied buildings as it is difficult to model occupants' behavior and how they interact with their buildings. On the other hand, predictive models (e.g. artificial neural networks, support vector machines, etc.) have been successfully applied to predict the energy consumption of occupied buildings. These models quickly perform predictions and thus are more suitable for real–time control purposes.

In a recent review, Ahmad et al. (2016) discussed several computational intelligence (CI) techniques for HVAC systems. It was mentioned that significant advances has been made in the past decades on the application of CI techniques for building energy applications. Most of these techniques use historical data to train a model or develop expert rules. With the evolution towards Internet of Things (IoT), there is an abundance of data available from buildings, and therefore these CI techniques can easily be applied to enhance building energy performance. Random Forest (RF) has been less explored for building energy and daylight illuminance predictions. RF does not require much fine-tuning of their hyper-parameters, and default parameters often give better results than being fine-tuned. On the other hand, artificial neural networks (ANNs) have been extensively used for energy and daylight predictions because of their fault–tolerant and robust nature. The objective of this study is to develop ANN and RF based models to predict the daylight illuminance and energy consumption of a classroom. This paper offers an alternative methodology to the existing energy consumption and illuminance prediction techniques.

Related work

In the literature, a large number of studies have focussed on using machine learning techniques to predict energy consumption. For building energy prediction, ANNs are the most popular choice among other computational intelligence techniques (Ahmad et al., 2016). In two different studies, both González and Zamarreño (2005) and Nizami and Al-Garni (1995) used a simple neural network to predict hourly values of building energy consumption by using weather and time stamp information as inputs to the models. Nizami and Al-Garni (1995) compared the results with a regression model and it was found that ANN performed better.

ANNs were also used by Kalogirou and Bojic (2000) to predict the energy use of a passive solar building. The authors developed different modules to predict outdoor and indoor air temperatures at next time step, as well as solar radiation and electrical heaters' state. Kreider et al. (1995) reported on the use of recurrent neural networks to predict cooling and heating energy consumption. ANNs are also being used to predict energy consumption for different climate zones by using envelope performance parameters, and heating and cooling degree days (Cheng-wen and Jian, 2010). Manufacturing industries show high fluctuations in their energy use and modeling energy consumption for these buildings could be a challenging task. Azadeh et al. (2008) predicted the annual electricity consumption of this type of building by using an ANN, where the results demonstrated their suitability for this purpose.

To the best of our knowledge, there are only a few studies that focussed on the application of decision trees for energy prediction.Tso and Yau (2007) and Yu et al. (2010) studied the use of decision trees for predicting energy demand and residential building energy performance respectively. Tso and Yau (2007) compared the results from neural network, decision trees, and regression analysis; and found that decision trees could be viable alternatives to understand energy patterns. One key advantage of decision trees is that the user can generate accurate models without having any computational knowledge. Yu et al. (2010) found that decision trees can produce accurate models for predicting building energy use intensity levels. Ahmad et al. (2017) compared the results of neural network and random forest (an ensemble-based method) for predicting hourly energy consumption, and found that ANN performed marginally better than random forest.

ANNs are also being used for predicting daylight illuminance in buildings. Hu and Olbina (2011) proposed an illuminance-based Venetian blind control method and used ANNs to predict illuminance values at two set-points. The author concluded that the proposed method has advantages for real-time blind control applications. Kazanasmaz et al. (2009) also used an ANN to determine daylight illuminance for an office building. The authors used building and weather parameters to predict illuminance values and found that the prediction accuracy of the model was approximately 98%. Ahmad et al. (2015) proposed a method for controlling Venetian blinds by using EnergyPlus as an evaluation engine. The authors stressed the need for a surrogate model to reduce the computational time required to run 1000s of simulations. Mourshed et al. (2011) studied the optimum design of artificial lighting and found that the search for an optimum design in a rugged solution space is a time consuming process, so there is a need to develop surrogate models.

Machine learning techniques

Random Forest

Random forests (RFs) are ensemble-based decision trees and were developed to overcome the shortcomings of traditional decision trees. In RF, like other ensemble learning techniques, the performance of a number of weak learners is boosted via a voting scheme. The main hallmarks of random forest include; 1) bootstrap sampling – randomly selecting number of samples with replacement, 2) random feature selection – randomly selecting only a small number of m features in the split of each node, 3) full depth decision tree growing, and 4) Out-of-bag error estimation – calculating error on the samples which were not selected during bootstrap sampling (Jiang et al., 2009).

In RF, a M number of decision trees are generated from a N number of training samples. For each tree, bootstrap sampling is performed to create a new training set. The new training dataset is then used to create a fully grown decision tree without pruning by using the 'classification and regression trees' (CART) technique (Duda et al., 2012). Instead of using all available features at each split of the node, only a small number of m features are randomly selected. This procedure is then repeated until M decision trees are created to form a randomly generated "Forest". For RF models, we used 1000 trees in the forest. One the hyper-parameters for RF is the number of randomly selected variables $mtry$ at each split node; according to Breiman (2001), the recommended value of $mtry$ is equal to $\sqrt{p}$ for classification problems (where p being the total number of predictors). The author also mentioned that $mtry << p$ should improve the performance of the model. For regression problems, the proposed default value of $mtry$ is $p/3$.

Artificial Neural Network

Artificial Neural Networks (ANNs) learn the relationships between inputs and outputs by using a training dataset and do not need any information about the system as they are black-box models. The inspiration for ANNs comes from the functioning of biological neurons. ANNs consist of a number of layers of neuron-like processing units, which are interconnected with each other. Different neural network strategies have been developed in the literature e.g. feed-forward, Hopfield, Elman, self-organising maps, and radial basis networks (Krenker et al., 2011). Feed-forward ANNs are the most widely used and generic neural network types and have been used for solving problems in various research fields.

This paper uses a feed-forward neural network trained by the Broyden-Flatcher-Goldfarb-Shano (BFGS) algorithm. In this study, only one hidden layer used, as more hidden layers did not improve the prediction power of the neural network. This concurs with Principe et al. (1999), who state that for most applications, one hidden layer should be adequate. We also varied the number of neurons between 10-20 for both the energy consumption and daylight illuminance models, by using the stepwise searching method. We found that for both models, using more than 10 neurons in the hidden layer did not significantly improve the accuracy of the models, so we used 10 hidden layer neurons. As the paper is focussed on random forest and its comparison with ANNs, the details about searching the neurons in the hidden layer, number of hidden layers and training algorithms are not discussed here.

In order to develop machine learning models; we used the implementation of RF included in the scikit-learn module of the Python programming language, and the neurolab module for the development of artificial neural network.

Methodology

Energy model

A model of a classroom of a school building was created in EnergyPlus as a use case to develop and validate prediction models. The school building is located in Cardiff, UK and is a BREEAM excellent (≈ LEED platinum) rated building. The dimensions of the classroom were 9.0 m (width)× 9.5 m (depth) × 3.6 m (height). It was assumed that the classroom has a 30% window-to-wall ratio on its southern façade, which consists of a double glazing (3 mm Generic PYR B Clear + 13 mm air gap + 3 mm Generic Clear). An interior Venetian blind was modeled on the window, which has a thermal conductivity of 0.9 W/mK and slat beam solar reflectance for both front and back sides were set to 0.8. The slat width was modeled as 2.5 cm and the slat angle was fixed at 4°. The illuminance levels were calculated at one sensor point (SP), which was located in the middle of the classroom at a height of 0.8 m.

Energy consumption and hourly illuminance values at the sensor point were obtained by running EnergyPlus simulations. The weather file for Cardiff, Wales, UK was used for these simulations. The cooling and heating demand of the model was met by using a purchased air system (ideal load air system). The system meets the energy demand by providing the required supply air capacity at the specified temperature. The heating and cooling set points are defined according to CIBSE (2006) Guide A i.e. 22°C and 24°C for heating and cooling respectively. 9 people were modeled with an activity level of 60 W/m^2. A continuous dimming control was modeled to control artificial lighting based on the lux level calculated at the SP. The sensor point has an illuminance set-point value of 300 lx (CIBSE, 2006). The outputs from the models were energy consumption (sum of heating, cooling and lighting energy consumption) and daylight illuminance at the SP. We also considered different input variables to improved the prediction accuracy of the machine learning models. The considered inputs variables were; solar altitude angle, solar azimuth angle, direct normal radiation, diffuse horizontal radiation, day of the week, hour of the day, month of the year, outdoor dry-bulb air temperature, wind speed, outdoor air relative humidity, window blind position, occupancy and previous hour values of daylight illuminance and energy consumption.

Evaluation metrics

To evaluate the performance of the models, several metrics were calculated the testing data samples: the root mean squared error (RMSE), the coefficient of variation (CV), the mean absolute deviation and the coefficient of determination (R^2), as formalised in Equations (1) to (4).

$$CV = \frac{\sqrt{\frac{\sum_{i=1}^{N}(\mathrm{y}_i-\hat{\mathrm{y}}_i)^2}{N}}}{\bar{\mathrm{y}}} \times 100 \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{y}_i-\hat{\mathrm{y}}_i)^2}{N}} \tag{2}$$

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|\hat{\mathrm{y}}_i-\mathrm{y}_i| \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\mathrm{y}_i-\hat{\mathrm{y}}_i)^2}{\sum_{i=1}^{N}(\mathrm{y}_i-\bar{\mathrm{y}})^2} \tag{4}$$

where $\hat{\mathrm{y}}_\mathrm{i}$ is the predicted output value, y_i is the actual output variable for the ith sample in the testing subset, $\bar{\mathrm{y}}$ is the mean of the observed values and N is the number of samples in the testing subset. The MAD metric calculates the average distance between each data value and its mean for a given dataset. CV is a measure of variation in error with respect to the actual consumption means, R^2 is used to evaluate the closeness of fit, where perfectly fitted model will have an R^2 of 1. We have used RMSE as our primary metric, and other performance metrics were used as tie breakers; they were considered only when RMSE did not provide any statistical difference between two prediction models.

Results and discussion

The variables used to train the prediction models, along with their maximum and minimum values, are listed in Table 1. The variable importance plots for energy consumption and daylight illuminance are shown in Figure 1. These plots were produced by replacing each input variable in turn by random noise in the RF and then analyzing the deterioration of the performance of the random forest models. The resulting deterioration is a measure of predictor's importance, for regression problems the most widely used score is the increase in the mean of the error of a tree (mean squared error) (Vincenzi et al., 2011). As shown in Figure 1, it is evident that for both energy consumption and daylight illuminance, the previous hour's values are the most important variables. For energy consumption, other important variables include: occupancy schedule, outdoor dry-bulb temperature, indoor air temperature, blind schedule, altitude angle, month of the year, diffuse solar radiation, hour of the day, azimuth angle, total transmitted solar radiation and direct solar radiation. For daylight illuminance prediction; total transmitted solar radiation, direct solar radiation, altitude angle, diffuse solar radiation, azimuth angle, blind schedule, month of the year and hour of the day were the most influential variables. We only used these input variables for further analysis (i.e tuning hyper-parameters, training and testing of models).

Table 1: Summary of the input variables considered in the model construction.

Variable	Minimum	Maximum	Unit
Outdoor air temperature	-3.7	26	°C
Outdoor relative humidity	30.63	100	%
Wind speed	0	22.5	m/sec
Diffuse solar radiation	0	383.17	W/m^2
Direct solar radiation	0	613.03	W/m^2
Solar azimuth angle	4.96	352.06	deg.
Solar altitude angle	-61.50	61.54	deg.
Total trans. solar rad.	0	2355.84	W
Previous hour's D.I.	0	26330.62	lx
Indoor air temperature	4.61	30.91	°C
Occupancy	0	1	–
*Window blind	0	1	–
Previous hour's E.C.	0	7.22	kWh
Hour of the day	0	23	–
Day of the week	0	6	–
Month of the year	1	12	–

Note: *Dichotomous variable

The observed effect of tree depth on the performance

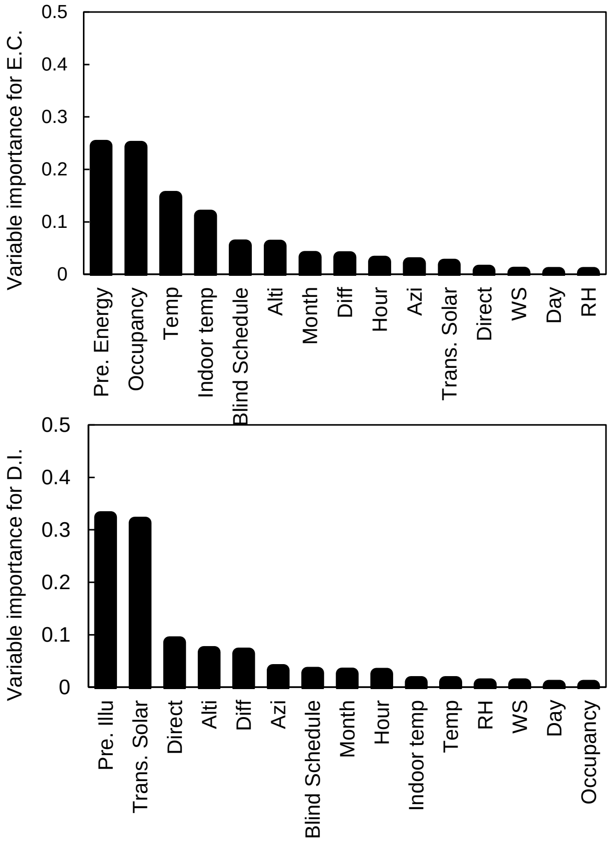

Figure 1: Variable importance for energy consumption daylight illuminance prediction models.

of a Random Forest models is shown in Figure 2 and Table 2. For both cases, it was found that a forest constructed with deeper trees resulted in better accuracy. A maximum depth of 1 resulted in poor performance and led to under-fitting. From results, it is shown that using trees deeper than 10 levels did not enhance the performance significantly and therefore, we used a maximum depth of 10 levels for both energy consumption and daylight illuminance prediction models. The forest with trees depth of 10 levels resulted in the lowest values of RMSE (0.0540), CV (14.8952%), MAD (0.019) and a higher R^2 (0.9959) value for energy consumption. For daylight illuminance, using a tree depth of 10 levels resulted in RMSE, CV and R^2 values of 238.108, 36.429% and 0.9867 respectively.

Table 3 shows the effects of varying the number of features on the accuracy of the RF models.

the performance of RF models while varying the number of features. Increasing *mtry* can improve the predictive performance of the RF model as there is a higher number of predictors being available each node of the tree. Contrary to the norm, we observed a deterioration in the performance of the RF model when using more than 7 features. As shown in Table 3, the resulting CV values for the model with max features equal to 7 was 13.101% and 36.067% for energy

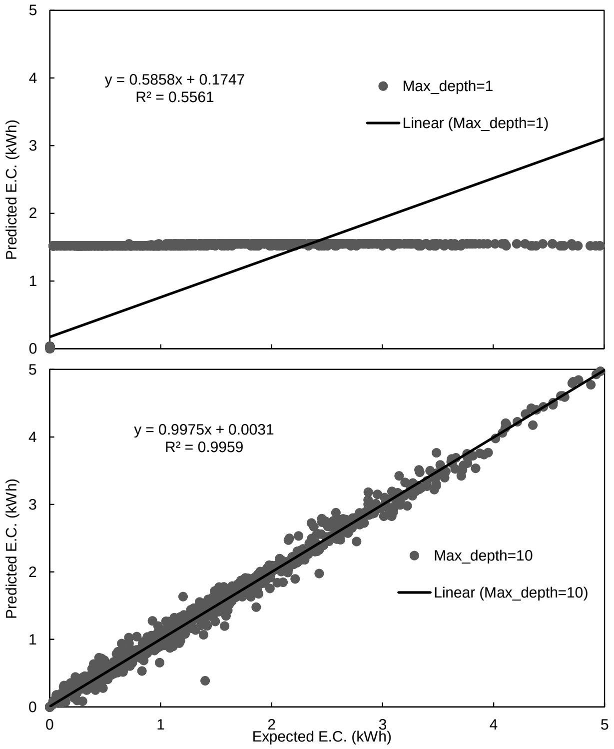

Figure 2: RF models for energy consumption with different maximum depths.

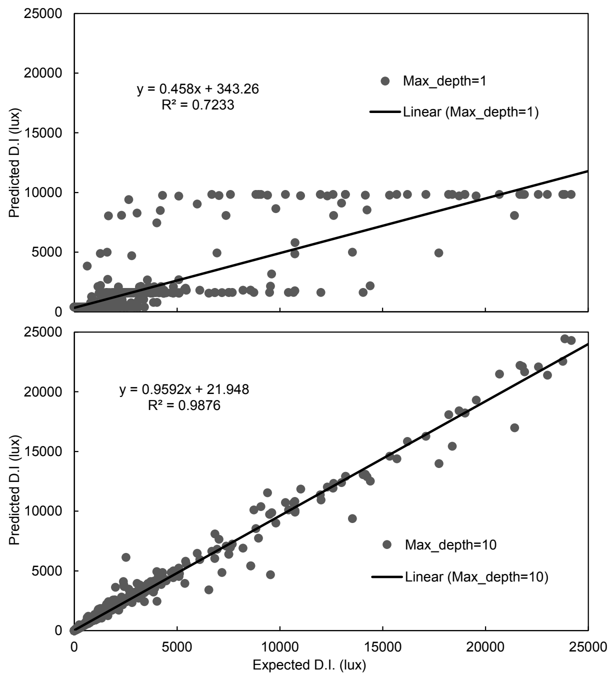

Figure 3: RF models for daylight illuminance with different maximum depths.

Table 2: Maximum depth for RF

Max depth	**RMSE**	**CV**	**MAD**	$\mathbf{R^2}$
Energy consumption				
1	0.5644	155.6710	0.238	0.5537
3	0.2166	59.7446	0.083	0.9342
5	0.1135	31.3014	0.043	0.9820
7	0.0741	20.4428	0.027	0.9923
10	0.0540	14.8952	0.019	0.9959
13	0.0496	13.6816	0.017	0.9966
15	0.0492	13.5786	0.016	0.9966
20	0.0491	13.5438	0.016	0.9966
Daylight Illuminance				
1	1264.396	193.446	603.779	0.6260
3	650.598	99.538	194.066	0.9010
5	373.824	57.193	97.200	0.9673
7	271.102	41.477	60.927	0.9828
10	238.108	36.429	44.915	0.9867
13	235.500	36.030	41.058	0.9870
15	234.368	35.857	40.419	0.9871
20	234.936	35.944	40.368	0.9870

consumption and daylight illuminance respectively, which was higher than considering minimum feature ($mtry = 1$, $\mathrm{CV}_{E.C} = 37.818\%$ and $\mathrm{CV}_{D.I} = 61.205\%$) and $mtry_{E.C} = 12$ ($\mathrm{CV}_{E.C} = 14.895\%$), $mtry_{D.I} = 9$ ($\mathrm{CV}_{D.I} = 36.429\%$). The results showed the same behaviour for other performance metrics. Increasing $mtry$ also reduced the diversity of the individual tree in the forest and therefore the performance of the RF deteriorated when considering more than 7 features. It is also worth mentioning that the construction of an RF with more features is computationally intensive and hence slower.

To evaluate the accuracy and generalization capabilities of the developed models, both models were used to predict energy consumption and daylight illuminance on an unseen dataset (i.e. the dataset was not used during the training or validation stages). The results from both models are shown in Figure 4. Moreover, RMSE and R^2 of the testing and validation samples using RF and ANN models were compared, and are shown in Table 4. From Table 4 and Figure 4,

Table 3: Maximum Features for RF

Max features	**RMSE**	**CV**	**MAD**	$\mathbf{R^2}$
Energy consumption				
1	0.137	37.818	0.060	0.9736
3	0.064	17.833	0.024	0.9942
5	0.051	14.199	0.019	0.9963
7	0.048	13.101	0.017	0.9968
10	0.049	13.544	0.017	0.9966
12	0.054	14.895	0.019	0.9959
Daylight Illuminance				
1	400.046	61.205	86.806	0.9626
3	270.917	41.449	52.871	0.9828
5	242.970	37.173	47.027	0.9862
7	235.738	36.067	45.335	0.9870
8	236.302	36.153	44.645	0.9870
9	238.108	36.429	44.915	0.9867

Table 4: Comparison of ANN and RF for validation and testing datasets

	Validation		Testing	
Model	**RMSE**	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$
RF (E.C)	0.048	0.9968	0.0559	0.9966
ANN (E.C)	0.043	0.9975	0.0493	0.9973
RF (D.I)	235.738	0.9870	227.87	0.9881
ANN (D.I)	282.354	0.9814	278.498	0.9799

it is evident that the Random Forest model performed better at predicting daylight illuminance, had lower RMSE values and higher R^2 values. On the other hand, ANN performed slightly better while predicting energy consumption. Predicting daylight illuminance was a challenging task as the illuminance at SP has drastic fluctuations. RF models performed well at predicting both higher and lower values, whereas ANN struggled to predict particularly lower values. There were even two high values of daylight illuminance which the ANN model predicted as values close to zero. One of the advantages of RF, as an ensemble algorithm, is that it can efficiently deal with any missing values in the input values. Both studied models did capture the relationship between input and output variables and can be used as evaluation engines.

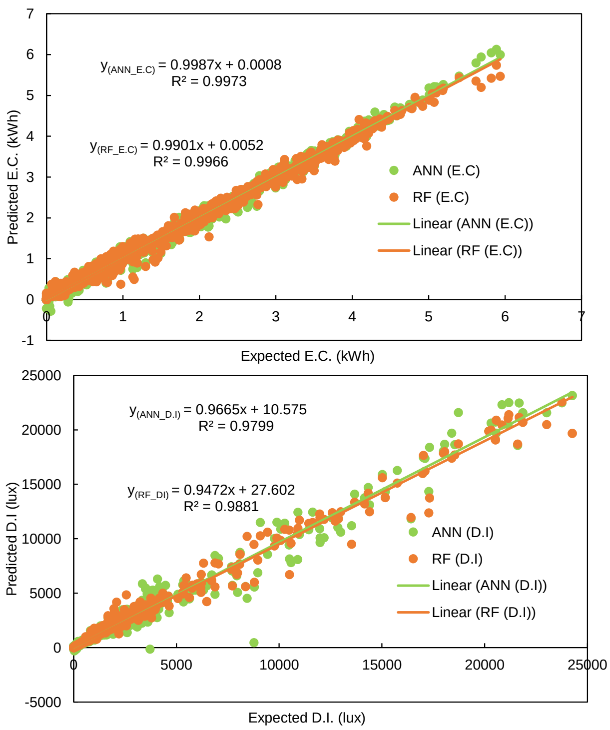

Figure 4: Testing results for RF and ANN models

The results showed that both models can be valuable computational intelligence tools to predict energy consumption and daylight illuminance. We also found that the training time of the RF model was much less than the ANN model (a few seconds versus a few minutes). This time may vary from problem to problem and also depends on other factors (e.g. number of trees in the forest, tree depth, number of random features selected at each split). Our results concur with Siroky et al. (2009), who state that Random Forests are faster to train and tune than other ML techniques.

Conclusions

The paper presented two machine learning algorithms to predict energy consumption and daylight illuminance, based on a simulation model of a classroom. Based on the evaluation metrics of RMSE, CV, MAD and R^2, it was found that the developed models can be feasible and effective for predicting the hourly daylight illuminance at a set-point and energy consumption. On the testing dataset, ANN performed marginally better than RF for predicting energy consumption with a RMSE value of 0.0559 as compared to 0.0493. Whereas, on daylight illuminance prediction's validation dataset, RF model provided better results than the ANN model. The paper also used RF as a method to calculate variable importance score, which is a useful method for dimensionality reduction in order to improve model's performance on high-dimensional datasets.

Whilst the paper was focussed on developing ML models for a specific classroom, the study will be extended in the future to include a diverse range of building types. In future, the performance of the proposed models will be compared with other ensemble based algorithms e.g. Gradient Boosted Regression Trees (Friedman, 2002) and Extremely Randomised Trees (Geurts et al., 2006). In the paper, the models were developed by using a Test Reference year (TRY weather file), however, in future actual weather conditions will be used to replicate the proposed work.

Acknowledgement

The authors acknowledge financial support from the European Commission Seventh Framework Programme (FP7), grant reference – 609154.

References

Ahmad, M., M. Mourshed, J.-L. Hippolyte, Y. Rezgui, and H. Li (2015). Optimising the scheduled operation of window blinds to enhance occupant comfort. In *Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India*, pp. 2393–2400.

Ahmad, M. W., M. Mourshed, D. Mundow, M. Sisinni, and Y. Rezgui (2016). Building energy metering and environmental monitoring – a state-of-the-art review and directions for future research. *Energy and Buildings 120*, 85 – 102.

Ahmad, M. W., M. Mourshed, and Y. Rezgui (2017). Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings 147*, 77 – 89.

Ahmad, M. W., M. Mourshed, B. Yuce, and Y. Rezgui (2016). Computational intelligence techniques for hvac systems: A review. *Building Simulation 9*(4), 359–398.

Azadeh, A., S. Ghaderi, and S. Sohrabkhani (2008). Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management 49*(8), 2272 – 2278.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Cheng-wen, Y. and Y. Jian (2010, May). Application of ann for the prediction of building energy consumption at different climate zones with hdd and cdd. In *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, Volume 3, pp. 286–289.

CIBSE, G. A. (2006). Environmental design. *The Chartered Institution of Building Services Engineers, London*.

Duda, R. O., P. E. Hart, and D. G. Stork (2012). *Pattern classification*. John Wiley & Sons.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis 38*(4), 367–378.

Geurts, P., D. Ernst, and L. Wehenkel (2006). Extremely randomized trees. *Machine learning 63*(1), 3–42.

González, P. A. and J. M. Zamarreño (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings 37*(6), 595 – 601.

Hu, J. and S. Olbina (2011). Illuminance-based slat angle selection model for automated control of split blinds. *Building and Environment 46*(3), 786 – 796.

Jiang, R., W. Tang, X. Wu, and W. Fu (2009). A random forest approach to the detection of epistatic interactions in case-control studies. *BMC bioinformatics 10*(1), 1.

Kalogirou, S. A. and M. Bojic (2000). Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy 25*(5), 479 – 491.

Kazanasmaz, T., M. Gnaydin, and S. Binol (2009). Artificial neural networks to predict daylight illuminance in office buildings. *Building and Environment 44*(8), 1751 – 1757.

Kreider, J., D. Claridge, P. Curtiss, R. Dodier, J. Haberl, and M. Krarti (1995). Building energy use prediction and system identification using recurrent neural networks. *Journal of solar energy engineering 117*(3), 161–166.

Krenker, A., J. Bester, and A. Kos (2011). Introduction to the artificial neural networks. *Artificial neural networks: methodological advances and biomedical applications*, 1–18.

Li, X. and J. Wen (2014). Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews 37*, 517–537.

Mourshed, M., S. Shikder, and A. D. Price (2011). Phi-array: A novel method for fitness visualization and decision making in evolutionary design optimization. *Advanced Engineering Informatics 25*(4), 676 – 687. Special Section: Advances and Challenges in Computing in Civil and Building Engineering.

Nizami, S. J. and A. Z. Al-Garni (1995). Forecasting electric energy consumption using neural networks. *Energy Policy 23*(12), 1097 – 1104.

Principe, J. C., N. R. Euliano, and W. C. Lefebvre (1999). *Neural and adaptive systems: fundamentals through simulations with CD-ROM*. John Wiley & Sons, Inc.

Siroky, D. S. et al. (2009). Navigating random forests and related advances in algorithmic modeling. *Statistics Surveys 3*, 147–163.

Tso, G. K. and K. K. Yau (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy 32*(9), 1761 – 1768.

Vincenzi, S., M. Zucchetta, P. Franzoi, M. Pellizzato, F. Pranovi, G. A. D. Leo, and P. Torricelli (2011). Application of a random forest algorithm to predict spatial distribution of the potential yield of ruditapes philippinarum in the venice lagoon, italy. *Ecological Modelling 222*(8), 1471 – 1478.

Yu, Z., F. Haghighat, B. C. Fung, and H. Yoshino (2010). A decision tree method for building energy demand modeling. *Energy and Buildings 42*(10), 1637 – 1646.