Ensemble learning for energy performance prediction of residential buildings

Razak Olu-Ajayi
Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, UK
r.olu-ajayi@herts.ac.uk

Hafiz Alaka
Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, UK
h.alaka@herts.ac.uk

Ismail Sulaimon
Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, UK
i.sulaimon@herts.ac.uk

Ketty Grishikashvili
Hertfordshire Business School, University of Hertfordshire, Hatfield, AL10 9AB, UK
K.grishikashvili@herts.ac.uk

Funlade Sunmola
School of Engineering and Technology, University of Hertfordshire, Hatfield, AL10 9AB, UK
f.sunmola@herts.ac.uk

Raphael Oseghale
Hertfordshire Business School, University of Hertfordshire, Hatfield, AL10 9AB, UK
r.oseghale@herts.ac.uk

Saheed Ajayi
School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds LS28AG, UK
S.Ajayi@leedsbeckett.ac.uk

ABSTRACT

In the past decades, the demand for energy in buildings has considerably amplified due to the increase in population and prompt urbanization. The high proportion of energy consumed by buildings engender major environmental problems causing climate change, air pollution and thermal pollution, which is detrimental to human existence. Therefore, the demand for understanding building energy efficiency and how it can be managed effectively is high within academics and society. This elevating concern has increasingly received attention and has been investigated from different perspectives using diverse machine learning techniques such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), among others. There have been applications of ML regression models for the prediction of energy consumption of operational buildings. However, the expedition to develop a reliable and accurate model remains elusive. Machine learning classification models can also contribute to bringing more insight to study this issue. In this research, the ensemble learning classification-based method was applied to predict the energy performance of residential buildings. Based on the United Kingdom (UK) Energy Performance Certificate (EPC) standard rating scale, this paper developed and compared six machine learning classification models, namely Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GB), Decision Tree (DT), K Nearest Neighbour (KNN) and Extra Trees (ET) for the prediction of building energy performance in terms of performance, feature importance, parameter optimization and computational efficiency. This result shows that ensemble learning produces good results for predicting the energy performance of buildings.

Keywords: Building Energy Performance, Energy Prediction, Ensemble Learning, Energy Rating, Machine Learning, Energy Efficiency, Classifier.

1. INTRODUCTION

The consumption of energy in buildings has triggered many environmental problems such as air pollution, thermal pollution, climate change, among others, which deploys a severe impact on the existence of mankind (Dandotiya, 2020). The United Nations Environment Program (UNEP), reports that residential and commercial buildings employ about 60% of worldwide electricity, 40% of global energy, 40 % of global resources, and emit approximately 1/3 of Green House Gas (GHG) (United Nations Environment Programme, 2017). Yet, the population of the world continues to grow exponentially and the relative demand for housing is equally bringing on the potential for more energy demand.

According to the United Kingdom (UK) Building Energy Efficiency Survey (BEES), the energy consumption of buildings in the UK accounted for 70 per cent of the total consumption and the four largest consumers of energy in the buildings were internal lighting, heating, catering and fridges (Building Energy Efficiency Survey, 2016). The higher the proportion of building energy consumption suggests the increase in environmental issues. However, buildings also proffer the most significant potential of alleviating these issues (Ibraheem *et al.*, 2017). Due to the increase in population and the need to construct more buildings thus causing more energy demand. There lies a necessity to understand building energy efficiency and how it can be improved, so new buildings can be constructed to be more energy efficient (TOPRAK *et al.*, 2017).

The demand for stable energy consumption cannot be over emphasized, and this has become evident as governments around the world are continuously implementing regulations, principles and occasionally incentives that are bespoke to enriching building energy savings initiatives (Amasyali and El-Gohary, 2021; Himeur *et al.*, 2020; Qiao, Yunusa-Kaltungo and Edwards, 2021). For example, the UK implemented a long term goal in 2008, to decrease CO2 emission by 26% in 2020 and 80% by 2050 (Department of Energy and Climate Change, 2009). In 2002, the EU directive on building energy performance, implemented a systematic framework for deducing energy performance. Member states have since generated certification systems for rating building energy performance (European Parliament, 2002). Consequently, The UK utilizes a standard rating scale to inform building owners of current energy consumed and effective policy measures towards enhancing building energy efficiency (Curtis, Devitt and Whelan, 2014).

In some cases, building energy rating is beneficial to property marketers as it serves as a premise to increase selling costs for buyers searching for energy efficient properties. Several studies have shown that buildings with excellent energy ratings can command premium pricing (Brounen and Kok, 2011; Cajias and Piazolo, 2012; Curtis, Devitt and Whelan, 2014). Hence, this may persuade property developers and building consumers to take energy performance into account before development or purchase. At the predevelopment stage, there is the availability of building features pertinent to the prediction of energy performance. There is adequate evidence showing that building properties (e.g wall properties, Roof properties, floor properties, wall thickness etc.) have high effect on the level of energy consumption (Brown, Southworth and Sarzynski, 2009; Hankey and Marshall, 2010; Guhathakurta and Williams, 2015; Robinson *et al.*, 2017; Li *et al.*, 2018).

The elevating concern of building energy consumption and its unfavourable impacts on the environment has increasingly received more attention from researchers worldwide (Li *et al.*, 2018). Researchers are utilising machine learning algorithms such as Support Vector Machine (SVM), Decision Tree (DT), Artificial Neural Network (ANN), among others, to better understand building energy efficiency (Dong, Cao and Lee, 2005; Li *et al.*, 2009; Tardioli *et al.*, 2015; Aversa *et al.*, 2016; Pham *et al.*, 2020). The development of an efficient energy demand management system requires a reliable and accurate model for predicting energy use (Aversa *et al.*, 2016). However, the expedition to develop a reliable and accurate model remains elusive due to several features that affect energy consumption in buildings, such as weather conditions, physical properties and occupant behaviour (Li *et al.*, 2018; Hamed and Nada, 2019). Furthermore, UNEP stipulated that the rate of energy consumed in a building and the physical structure of the building are often closely inter-connected (United Nations Environment

Programme, 2017). This study will explore the classification approach for predicting the energy performance of buildings specifically based on variables associated with weather and building properties.

1.1 Aim of this research

The primary aim of this research is to develop a model that enables property developers to predict the potential energy performance of a building based on its physical structure.

1.2 Research Objectives

Deriving from the research aim stated above and the theoretical rationale that is expounded in section 2 (Literature Review), the key objectives of the research are outlined below:

- To conduct a comparative analysis of six algorithms for predicting building energy performance.
- To examine the effect of feature selection on the performance of the model.
- To examine the effect of hyper parameter tuning on the performance of the model.
- To investigate the best performing model based on computational efficiency.

In this study, machine learning algorithms namely Support Vector Machine (SVM), Gradient Boosting (GB), Random Forest (RF), K Nearest Neighbour (KNN), Decision Tree (DT) and Extra Trees (ET) are applied in forecasting energy performance consumption in buildings. The rest of the paper is structured as follows: Section 2 delivers a brief background on the utilized machine learning techniques for predicting building energy performance. Section 3 describes the research methodology, feature engineering, data pre-processing, model framework and evaluation measures. Section 4 presents and discusses the findings and results, while Section 5 delivers the conclusions and the proposed future research.

2. LITERATURE REVIEW

Machine learning (ML) algorithm have been introduced by a number of researchers in the field of forecasting building energy consumption (Ekici and Aksoy, 2009; Qiong Li, Peng Ren, and Qinglin Meng, 2010; Robinson *et al.*, 2017; Li *et al.*, 2018; Amasyali and El-Gohary, 2021). However, the are no known exploration towards predicting energy performance based on the energy ratings. One ML recurrent model is called the Support Vector Machine (SVM). It has been introduced by a number of researchers in this field and recognized due to its generation of good results in small datasets (Li *et al.*, 2009; Qiong Li, Peng Ren, and Qinglin Meng, 2010; Aversa *et al.*, 2016). It is employed for classification which is also referred to as Support Vector Classifier (SVC). SVM was first proposed by Dong et al, through the utilization for prediction monthly electricity consumption of buildings. Dong et al stated that SVM produces better results than using neural networks (Dong, Cao and Lee, 2005). Ahmad et al reviewed the applications of ANN and SVM for forecasting building electrical energy consumption. They concluded that each method has their singular advantages and disadvantages thus proposing a combination of both models for better estimation performance (Ahmad *et al.*, 2014).
The training of a stable model that excels in all aspects is the main goal of Machine learning (Khan *et al.*, 2020). Support Vector Regression (SVR) is utilised in predicting building electricity consumption and cooling load for Heating, ventilation, and air conditioning (HVAC) system (Niu, Wang and Wu, 2010; Aversa *et al.*, 2016). Dong et al also implemented data driven method using customary machine learning algorithms namely Artificial Neural Network (ANN) and Support Vector regression (SVR) and hybrid techniques such as Least-square support vector machine (LS-SVM), Gaussian process regression (GPR) and Gaussian Mixture Model (GMM) for forecasting future electricity consumption and further established that the hybrid modelling approach performs relatively better for hourly energy predictions (Dong *et al.*, 2016).

Ensemble learning is a more sophisticated data-driven method focused on achieving a more comprehensive model. It is defined as a method of utilizing multiple algorithms to achieve better predictive performance (Wang and Srinivasan, 2017). Ensemble learning can split into three categories namely bagging, boosting, and stacking. Bagging focuses on achieving a model with lower difference than its components. For example Random Forest (RF), boosting will generally get a good model with less bias than the basic model such as Gradient Boosting (GB), while stacking mainly focuses on enhancing the prediction regardless of the variance or bias (Dong et al., 2021). In a study by Chae et al, ensemble learning and other types of ANN for forecasting building energy use. Results presented the ensemble model as the more effective model (Chae et al., 2016). In Hong Kong, Tso and Yau 2007 compared decision tree, neural network, and other regression methods for predicting weekly electricity consumption. The results concluded that decision tree and neural networks slightly outperforms other regression method (Tso and Yau, 2007). KNN also produced good results in predicting building energy consumption (Wang, Lu and Feng, 2020).

3. RESEARCH METHODS

In this research, the classification model is proposed to predict building energy performance solely based on weather data and variables associated with the building's physical structure. Residential buildings in cities around the UK are utilized in this study. Ensemble learning and other ML algorithms such as K Nearest Neighbour (KNN), Decision Tree (DT) and Support Vector Machine (SVM) are proposed for prediction, and model development is performed using python programming language on jupyter Integrated Development Environment (IDE). The data collected was examined and pre-processed to evade problems during model training. Subsequently, each model is trained and tested, after which, evaluation of the performance is conducted. These processes are divided into four steps- data description, data pre-processing, model development and evaluation.

3.1 Data description

Building properties and energy ratings data were collected from the UK Ministry of Housing Communities and Local Government (MHCLG) Repository. The meteorological data for the location of each building was collected from Meteostat repository. The building dataset consists of data for 1000 residential buildings within the United Kingdom (UK) each containing their respective current energy rating. The building properties in the dataset include floor area, wall type, roof type and so on. The meteorological data comprised of variables such as temperature, pressure, and wind speed. The monthly data containing monthly conditions from January 2020 to December 2020. After considering the available data, the variables were divided into two categories: building metadata and meteorological data as illustrated in Table 1 below:

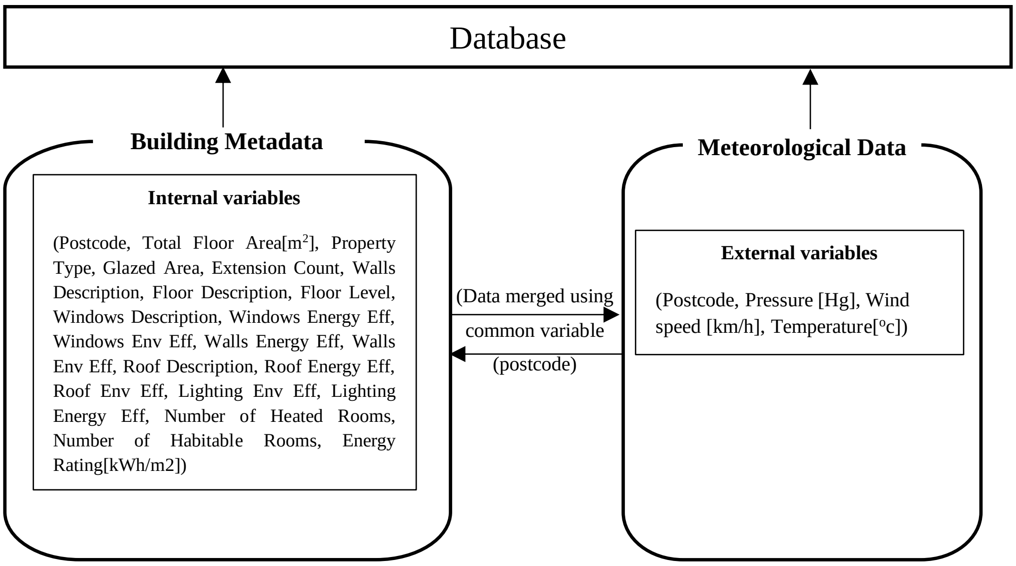

Figure 1. Building and Meteorological variables collected

3.2 Data pre-processing

The pre-processing of data stage is important for quality assurance of the database. This is a vital process in machine learning, although it is time consuming and computationally expensive (Shapi, Ramli and Awalin, 2021). Data comprising of missing or abnormal values must be processed before model training to avoid complexities. The building dataset was cleaned by removing buildings with multiple missing values which summed up to 114 buildings while the missing values of the meteorological data were managed by adopting the mean value imputation as specified by Newgard and Lewis, 2015. The mean value imputation could not be adopted for the building dataset as some of its variables are categorical namely windows, walls, roof, and lighting energy and environmental efficiency. These variables were assigned data using python. The independent variables include all the internal and external variables listed in figure 1 above, except energy rating which is the dependent variable. The building and meteorological datasets were merged using the common variable named postcode.

Data normalization is a method of data pre-processing that eliminates the influence of dimensions, as features often have different dimensions (Liu *et al.*, 2020). As presented in Figure 1, the merged data excluding the energy rating was normalized using friedman of the sklearn python package. The energy rating data was encoded using sklearn pre-processing python package. In the research by Zhang et al 2019, it was concluded that proper selection of features is closely associated with the accuracy of the prediction model (Zhang and Wen, 2019).

3.2.1 Feature selection (FS)

Building energy consumption prediction studies often select features only based on domain knowledge, however the utilization of feature selection generates better accuracy (Zhang and Wen, 2019). Feature selection (FS) is a major process in the implementation of machine learning because not all features selected are impactful on the predicted output. FS is regularly used to pinpoint unrelated and irrelevant features (Zhao and Magoulès, 2012). It is required to be employed for optimum model performance (Alaka *et al.*, 2018). In this research, Random Forest (RF) was employed to identify the most preferred input variables. The feature importance of random forest can be used to examine the role of each variable in the model. Random forest assesses the influence of adding or deleting a variable on the model and yields the relative dependence value for all variables as displayed in Figure 2 (Dong et al., 2021). The study evaluates the influence of feature selection on the model by adding and subtracting variables based on ranking. The model is developed with and without the lower ranked variables. Therefore, ten variables with the highest ranking were chosen, and all variables were selected to develop a model independently. The selected list of variables used are as follows: total floor area, wall energy efficiency, number habitable rooms, number heated rooms, floor description, walls description, roof description, temp, wind speed and pressure.

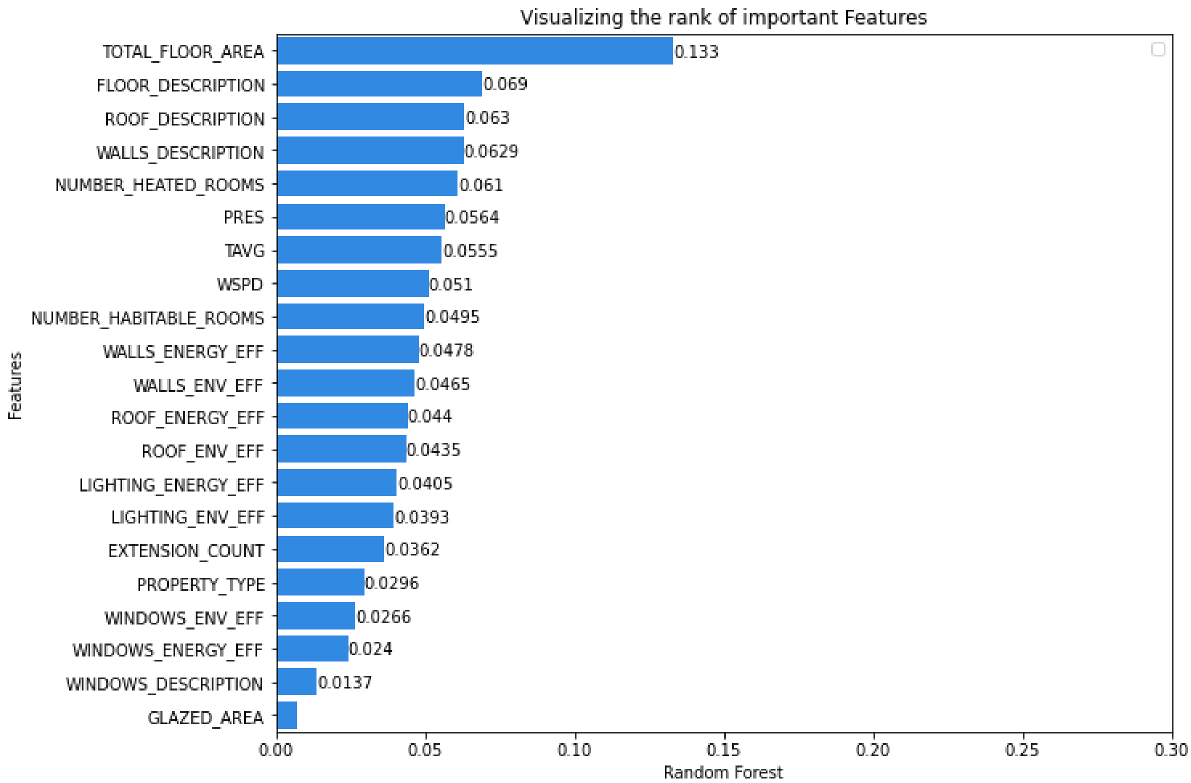

Figure 2. Visualization of feature importance using RF

3.3 Model development

After data pre-processing, the data was divided on two groups named train and test at a ratio of 7:3. The prediction approach is based on classification model for forecasting energy performance. The development of each model was conducted twice, with and without feature selection. The 70% train data was used to train each model while 30% test data are used to test the model based on specific input variables. Support Vector Machine (SVM), Gradient Boosting (GB), Random Forest (RF), K Nearest Neighbour (KNN), Decision Tree (DT) and Extra Trees (ET) are the classification algorithms utilized to develop a model for predicting energy performance.

3.4 Model evaluation

The evaluation of the prediction performance was measured using the following measures: Accuracy and Balanced accuracy. Accuracy is the most used metric for the evaluation of classifiers (Gonzalez-Abril *et al.*, 2014). It is the metric based on the exact match of the predicted output and actual values. It is also regarded as the probability that the model developed is correct. Balanced accuracy is also a known metric often used for multiclass classification problems to evaluate imbalanced datasets (Grandini, Bagli and Visani, 2020).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Balanced\ Accuracy = \frac{\frac{TP}{Total_{row1}} + \frac{TN}{Total_{row2}}}{2}$$

Where: TP is True Positive
TN is True Negative

4. RESULT AND DISCUSSION

The six machine learning models result is displayed in table 1. The result of this experiment shows that RF and GB model have similar results and are the highest performing model. The result also shows the models that perform best in terms of computational efficiency. All models produced relatively good results on accuracy. Decision Tree, however, trained in the shortest time in comparison with random forest and other models. The best model for accuracy, balance accuracy, and computational efficiency are presented in bold font in Table 1 below. This asserts RF has a potentially good model for predicting energy performance.

Table 1: Performance result for each developed model

Model	Training Time	Accuracy	Balanced Accuracy
Random Forest	0.23	**0.59**	0.27
Gradient Boosting	0.25	0.59	0.26
Extra Trees	0.19	0.58	**0.32**
Decision Tree	**0.02**	0.51	0.24
K Nearest Neighbors	0.05	0.50	0.20
Support Vector Machines	0.50	0.50	0.17

The box plot method was implemented to visualize the model’s performance with and without future selection as presented in Figure 3a-3d. Based on the boxplot, the accuracy plot presents Random Forest as the best performing model among the other models using feature selection while the accuracy plot without feature selection shows Gradient Boosting (GB) as the dominant model. In Figure 3c and 3d, The GB presents good results for accuracy and balance accuracy without feature selection. This suggests that the implementation of feature selection influences the model. However, the performance is predicated on the type of model selected. The boxplot without feature selection demonstrates that GB, DT and RF produce good results in both evaluation measures while SVC, DT and KNN produce lower performance.

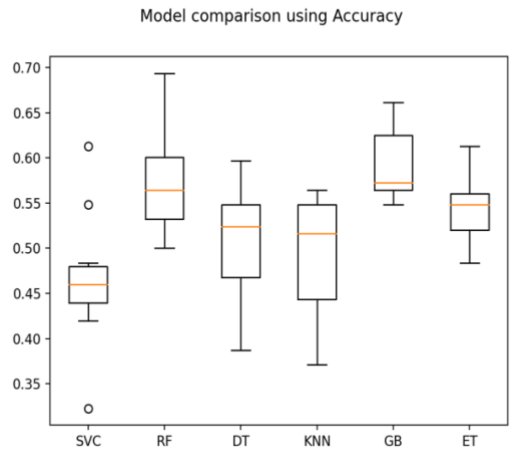

Figure 3a. Visualization of Boxplot with feature selection

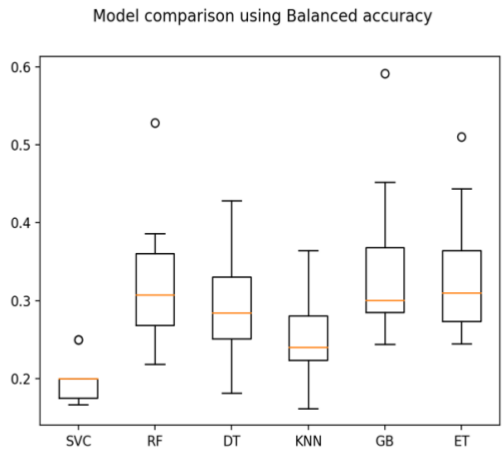

Figure 3b. Visualization of Boxplot with feature selection

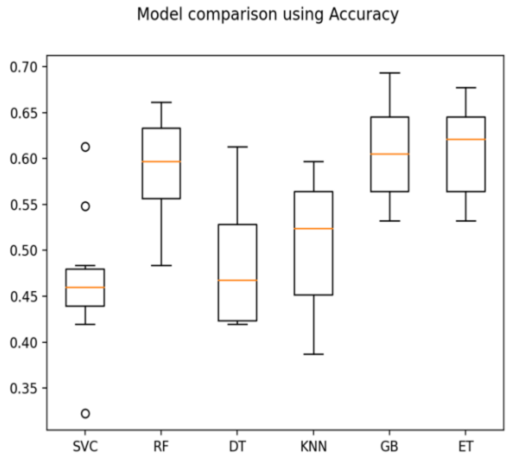

Figure 3c. Visualization of Boxplot without feature selection

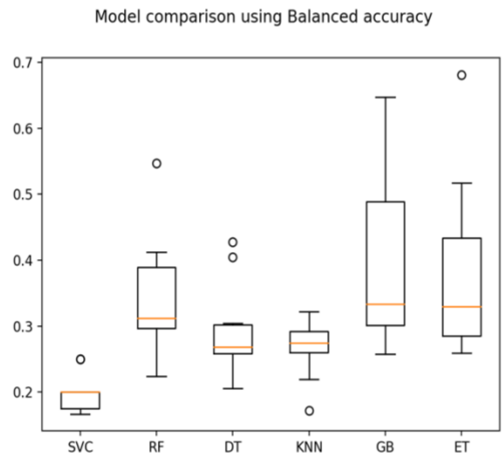

Figure 3d. Visualization of Boxplot without feature selection

To improve the accuracy of the performance model, hyper parameter tunning was applied to achieve better accuracy. The 10-fold cross validation was implemented, and the parameters are optimized based on validation results. The grid search method of the sklearn python package was utilised to identify the optimal parameter before training as presented in Table 2. Support Vector Machine (SVM), Gradient Boosting (GB), Random Forest (RF), K Nearest Neighbour (KNN), Decision Tree (DT) and Extra Trees (ET) were trained using the best parameter determined by grid search. Extra Trees (ET) and Random Forest (RF) produce good accuracy results after parameter optimization; However, it consumes more time to train the model. K Nearest Neighbour (KNN), Decision Tree (DT) and Support Vector Machine (SVM) achieve equal accuracy results, but DT performed best in terms of computational efficiency.

Table 2: Performance result after hyper parameter tunning

Model	Training Time	Accuracy	Balanced Accuracy
Random Forest	13s	0.62	0.24
Gradient Boosting	1m 38s	0.59	0.41
Extra Trees	10s	**0.64**	0.25
Decision Tree	2s	0.56	0.22
K Nearest Neighbors	**1s**	0.56	0.22
Support Vector Machines	16s	0.56	**0.51**

There are no known literatures that have explored the classification approach for predicting building energy performance using the UK standard rating scale. The prediction performance of Random Forest presents the potential of using it to predict the energy performance of a building before construction. However, it should be applied on a larger dataset with the possibility of achieving better accuracy. In previous studies, ensemble learning regression models have been applied for forecasting energy consumption and produce good results compared to recurrent model such as ANN. (Chae et al., 2016). Although, this is not directly comparable, Ensemble learning provides good results using the regression or classification model for forecasting building energy consumption or performance.

5. CONCLUSSION

In this research, ensemble learning was applied for the prediction of energy performance in residential buildings. Six models were developed and compared in terms of performance, feature importance, parameter optimization and computational efficiency. The input features were subject to feature selection that reduced 24 variables to 10 variables based on ranking or importance. The performance of models was compared with and without feature selection. The result shows good performance with feature importance however it is concluded that the achievement of a good performance model using feature selection is predicated on the type of algorithm selected. The decision tree performed best in terms of computational efficiency before parameter optimization. After hyper parameter tuning, the accuracy of the models increased significantly and in terms of computational efficiency, K Nearest Neighbor (KNN) performed best. In general, Extra Trees (ET) outperformed the other models utilized in this study to predict building energy performance of. The proposed futurological research should focus on applying machine learning classification algorithms to predicting building energy performance especially on a larger dataset. Although it can be computational and cost inefficient, future research should also optimize parameters on a wider range for models with promising potential for predicting building energy performance.

6. REFERNCES

Ahmad, A. S. *et al.* (2014) 'A review on applications of ANN and SVM for building electrical energy consumption forecasting', *Renewable and Sustainable Energy Reviews*, 33, pp. 102–109. doi: 10.1016/j.rser.2014.01.069.

Alaka, H. A. *et al.* (2018) 'Systematic review of bankruptcy prediction models: Towards a framework for tool selection', *Expert Systems with Applications*, 94, pp. 164–184. doi: 10.1016/j.eswa.2017.10.040.

Amasyali, K. and El-Gohary, N. (2021) 'Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings', *Renewable and Sustainable Energy Reviews*, 142, p. 110714. doi: 10.1016/j.rser.2021.110714.

Aversa, P. *et al.* (2016) 'Improved Thermal Transmittance Measurement with HFM Technique on Building Envelopes in the Mediterranean Area', *Selected Scientific Papers - Journal of Civil Engineering*, 11. doi: 10.1515/sspjce-2016-0017.

Brounen, D. and Kok, N. (2011) 'On the economics of energy labels in the housing market', *Journal of Environmental Economics and Management*, 62(2), pp. 166–179. doi: 10.1016/j.jeem.2010.11.006.

Brown, M. A., Southworth, F. and Sarzynski, A. (2009) 'The geography of metropolitan carbon footprints', *Policy and Society*, 27(4), pp. 285–304. doi: 10.1016/j.polsoc.2009.01.001.

Building Energy Efficiency Survey (2016) *Building Energy Efficiency Survey (BEES)*, *GOV.UK*. Available at: https://www.gov.uk/government/collections/non-domestic-buildings-energy-use-project (Accessed: 29 March 2021).

Cajias, M. and Piazolo, D. (2012) 'Green Performs Better: Energy Efficiency and Financial Return on Buildings', *Journal of Corporate Real Estate*, 15, pp. 53–72. doi: 10.1108/JCRE-12-2012-0031.

Chae, Y. T. *et al.* (2016) 'Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings', *Energy and Buildings*, 111, pp. 184–194. doi: 10.1016/j.enbuild.2015.11.045.

Curtis, J., Devitt, N. and Whelan, A. (2014) *Estimating Building Energy Ratings for the Residential Building Stock: Location and Occupancy*, *Papers*. WP489. Economic and Social Research Institute (ESRI). Available at: https://ideas.repec.org/p/esr/wpaper/wp489.html (Accessed: 15 June 2021).

Dandotiya, B. (2020) 'Climate-Change-and-Its-Impact-on-Terrestrial-Ecosystems', in. doi: 10.4018/978-1-7998-3343-7.ch007.

Department of Energy and Climate Change (2009) *Impact Assessment of the Climate Change Act*. Available at: https://www.legislation.gov.uk/ukia/2009/70/pdfs/ukia_20090070_en.pdf.

Dong, B. *et al.* (2016) ‘A hybrid model approach for forecasting future residential electricity consumption’, *Energy and Buildings*, 117, pp. 341–351. doi: 10.1016/j.enbuild.2015.09.033.

Dong, B., Cao, C. and Lee, S. E. (2005) ‘Applying support vector machines to predict building energy consumption in tropical region’, *Energy and Buildings*, 37(5), pp. 545–553. doi: 10.1016/j.enbuild.2004.09.009.

Dong, Z. *et al.* (2021) ‘Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification’, *Energy and Buildings*, 241, p. 110929. doi: 10.1016/j.enbuild.2021.110929.

Ekici, B. B. and Aksoy, U. T. (2009) ‘Prediction of building energy consumption by using artificial neural networks’, *Advances in Engineering Software*, 40(5), pp. 356–362. doi: 10.1016/j.advengsoft.2008.05.003.

European Parliament (2002) *Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings (repealed), https://webarchive.nationalarchives.gov.uk/eu-exit/https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02002L0091-20081211*. Queen’s Printer of Acts of Parliament. Available at: https://www.legislation.gov.uk/eudr/2002/91/2008-12-11 (Accessed: 15 June 2021).

Gonzalez-Abril, L. *et al.* (2014) ‘GSVM: An SVM for handling imbalanced accuracy between classes inbi-classification problems’, *Applied Soft Computing*, 17, pp. 23–31. doi: 10.1016/j.asoc.2013.12.013.

Grandini, M., Bagli, E. and Visani, G. (2020) ‘Metrics for Multi-Class Classification: an Overview’, *arXiv:2008.05756 [cs, stat]*. Available at: http://arxiv.org/abs/2008.05756 (Accessed: 17 June 2021).

Guhathakurta, S. and Williams, E. (2015) ‘Impact of Urban Form on Energy Use in Central City and Suburban Neighborhoods: Lessons from the Phoenix Metropolitan Region’, *Energy Procedia*, 75, pp. 2928–2933. doi: 10.1016/j.egypro.2015.07.594.

Hamed, M. and Nada, S. (2019) ‘Statistical Analysis for Economics of the Energy Development in North Zone of Cairo’, *International Journal of Finance & Economics*, 5, pp. 140–160.

Hankey, S. and Marshall, J. D. (2010) ‘Impacts of urban form on future US passenger-vehicle greenhouse gas emissions’, *Energy Policy*, 38(9), pp. 4880–4887. doi: 10.1016/j.enpol.2009.07.005.

Himeur, Y. *et al.* (2020) ‘Building power consumption datasets: Survey, taxonomy and future directions’, *Energy and Buildings*, 227, p. 110404. doi: 10.1016/j.enbuild.2020.110404.

Ibraheem, T. B. *et al.* (2017) ‘Renewable Energy Integration in African Buildings: Criteria and Prospects.’, *American Journal of Engineering Research (AJER)*, Volume-6(Issue-11), p. pp-39-43.

Khan, P. W. *et al.* (2020) ‘Machine Learning-Based Approach to Predict Energy Consumption of Renewable and Nonrenewable Power Sources’, *Energies*, 13(18), p. 4870. doi: 10.3390/en13184870.

Li, K. *et al.* (2018) ‘A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction’, *Energy and Buildings*, 174, pp. 323–334. doi: 10.1016/j.enbuild.2018.06.017.

Li, Q. *et al.* (2009) ‘Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks’, *Energy Conversion and Management*, 50(1), pp. 90–96. doi: 10.1016/j.enconman.2008.08.033.

Liu, Y. *et al.* (2020) ‘Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China’, *Journal of Cleaner Production*, 272, p. 122542. doi: 10.1016/j.jclepro.2020.122542.

Newgard, C. D. and Lewis, R. J. (2015) ‘Missing Data: How to Best Account for What Is Not Known’, *JAMA*, 314(9), pp. 940–941. doi: 10.1001/jama.2015.10516.

Niu, D., Wang, Y. and Wu, D. D. (2010) ‘Power load forecasting using support vector machine and ant colony optimization’, *Expert Systems with Applications*, 37(3), pp. 2531–2539. doi: 10.1016/j.eswa.2009.08.019.

Pham, A.-D. *et al.* (2020) ‘Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability’, *Journal of Cleaner Production*, 260, p. 121082. doi: 10.1016/j.jclepro.2020.121082.

Qiao, Q., Yunusa-Kaltungo, A. and Edwards, R. E. (2021) ‘Towards developing a systematic knowledge trend for building energy consumption prediction’, *Journal of Building Engineering*, 35, p. 101967. doi: 10.1016/j.jobe.2020.101967.

Qiong Li, Peng Ren, and Qinglin Meng (2010) ‘Prediction model of annual energy consumption of residential buildings’, in *2010 International Conference on Advances in Energy Engineering*. *2010 International Conference on Advances in Energy Engineering*, pp. 223–226. doi: 10.1109/ICAEE.2010.5557576.

Robinson, C. *et al.* (2017) ‘Machine learning approaches for estimating commercial building energy consumption’, *Applied Energy*, 208, pp. 889–904. doi: 10.1016/j.apenergy.2017.09.060.

Shapi, M. K. M., Ramli, N. A. and Awalin, L. J. (2021) ‘Energy consumption prediction by using machine learning for smart building: Case study in Malaysia’, *Developments in the Built Environment*, 5, p. 100037. doi: 10.1016/j.dibe.2020.100037.

Tardioli, G. *et al.* (2015) ‘Data Driven Approaches for Prediction of Building Energy Consumption at Urban Level’, *Energy Procedia*, 78, pp. 3378–3383. doi: 10.1016/j.egypro.2015.11.754.

TOPRAK, Ahmet *et al.* (2017) ‘International Journal of Intelligent Systems and Applications in Engineering’.

Tso, G. K. F. and Yau, K. K. W. (2007) ‘Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks’, *Energy*, 32(9), pp. 1761–1768. doi: 10.1016/j.energy.2006.11.010.

United Nations Environment Programme, U. N. (2017) *Sustainable buildings, UNEP - UN Environment Programme*. Available at: http://www.unep.org/explore-topics/resource-efficiency/what-we-do/cities/sustainable-buildings (Accessed: 16 March 2021).

Wang, R., Lu, S. and Feng, W. (2020) ‘A novel improved model for building energy consumption prediction based on model integration’, *Applied Energy*, 262, p. 114561. doi: 10.1016/j.apenergy.2020.114561.

Wang, Z. and Srinivasan, R. S. (2017) ‘A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models’, *Renewable and Sustainable Energy Reviews*, 75, pp. 796–808. doi: 10.1016/j.rser.2016.10.079.

Zhang, L. and Wen, J. (2019) ‘A systematic feature selection procedure for short-term data-driven building energy forecasting model development’, *Energy and Buildings*, 183, pp. 428–442. doi: 10.1016/j.enbuild.2018.11.010.

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